ARTICLES
PUBLISHED ONLINE: 28 APRIL 2014 | DOI: 10.1038/NGEO2144
nature geoscience

Reduction in local ozone levels in urban São Paulo due to a shift from ethanol to gasoline use

Alberto Salvo[1]* and Franz M. Geiger[2]

Ethanol-based vehicles are thought to generate less pollution than gasoline-based vehicles, because ethanol emissions contain lower concentrations of mono-nitrogen oxides than those from gasoline emissions. However, the predicted effect of various gasoline/ethanol blends on the concentration of atmospheric pollutants such as ozone varies between model and laboratory studies, including those that seek to simulate the same environmental conditions. Here, we report the consequences of a real-world shift in fuel use in the subtropical megacity of São Paulo, Brazil, brought on by large-scale fluctuations in the price of ethanol relative to gasoline between 2009 and 2011. We use highly spatially and temporally resolved observations of road traffic levels, meteorology and pollutant concentrations, together with a consumer demand model, to show that ambient ozone concentrations fell by about 20% as the share of bi-fuel vehicles burning gasoline rose from 14 to 76%. In contrast, nitric oxide and carbon monoxide concentrations increased. We caution that although gasoline use seems to lower ozone levels in the São Paulo metropolitan area relative to ethanol use, strategies to reduce ozone pollution require knowledge of the local chemistry and consideration of other pollutants, particularly fine particles.

Ozone levels are relatively high in São Paulo, with hourly concentrations above 75 and 125 $\mu g\,m^{-3}$, respectively, being 2.7 and 5.3 times more likely than for PM10 in our sample. Light transportation is a key contributor to air pollution in this gridlocked metropolis[1–3], with large public health implications[4–7]. In 2011, 40% of the city's six million active light-duty vehicles—probably accounting for over one-half of all light-vehicle distance travelled—possessed bi-fuel capability. This capability allowed consumers to choose between gasoline (an E25 or E20 blend) and ethanol E100 at the pump[8], as both fuels were ubiquitous among São Paulo's retailers[9,10]. In recent years, government-controlled gasoline prices held steady whereas market-set sugarcane ethanol prices tracked the significant swings in the world price of sugar[8,10]. Large fluctuations in the relative price of ethanol between 2009 and 2011 led to large-scale switching out of ethanol and into gasoline as ethanol prices soared, and back to ethanol when prices dropped, as evidenced by aggregate shipments reported by wholesalers for the state of São Paulo (Fig. 1), as well as revealed-choice surveys of consumers[11,12]. For perspective, wholesaler reports suggest that the unblended (pure) gasoline component shifted between 42% and 68% of total gasoline-plus-ethanol light-vehicle distance travelled (Supplementary Information Part A).

This empirical setting provides a rarely observed opportunity to examine whether urban air pollution was impacted by emissions that transitioned between gasoline and ethanol—both combustion and evaporation. São Paulo city currently features clogged roads but limited industrial activity and residential heating. Electricity generation is mostly hydroelectric. The shifts in the fuel mix occurred over relatively short time windows during which meteorological conditions and vehicle usage, including ridership of public transport, were broadly similar. These fuel mix shifts were a response to exogenously varying relative prices, and to a temporary change in the gasoline blend mandate, not to concerns over air quality; furthermore, evidence established herein indicates that the relative price variation did not significantly impact road traffic. Such characteristics, together with the existence of extended air quality, weather and vehicle traffic monitoring networks, make São Paulo a unique natural laboratory for studying the impact of gasoline versus ethanol fuel combustion on urban air pollution.

To date, work relating fuel mix with air quality has largely focused on the chemical analysis of vehicle exhaust[13–23], on how varying emissions affect air chemistry via smog chamber models[24], or on computer simulations of atmospheric science[25–27]. As described in Supplementary Information Part B, tailpipe emissions tests tend to show that less NO and NO_2 but significantly more aldehydes are produced from ethanol-dominant versus gasoline-dominant fuel, with the differences in emissions depending on vehicle characteristics and fuel composition. One chamber study suggests health benefits when switching from straight gasoline to ethanol blends in certain vehicles[28]. Reductions in ozone concentrations ranging from 14 to 55% were simulated specifically for air monitoring stations in the São Paulo metropolis in September 2004 on assessing a hypothetical increase in the ethanol share of total gasoline-plus-ethanol consumption from 34% of distance travelled in the base case to 97% in the simulated case[25]. In contrast, computer simulations with explicit chemical mechanisms applied to the Los Angeles metropolitan area showed some public health risks associated with ethanol in terms of increased ozone, formaldehyde and acetaldehyde concentrations, especially in colder temperatures[26]. The current state of knowledge regarding urban air chemistry predicts that all other things being equal, fuel/engine combinations that reduce NO_X emissions from tailpipes should lead to decreases in ambient O_3 concentrations in the NO_X-limited regime but increases in O_3 levels in the hydrocarbon-limited regime, with recent inventories highlighting the importance of biogenic sources of hydrocarbon emissions (for example, isoprene) in addition to anthropogenic sources[29–31]. Review articles underscore the need for data-based studies examining the impact on air quality due to consumer adoption of alternative fuels and vehicles[32,33].

Beyond science but no less important to society owing to its influence on human behaviour, conventional wisdom seems

[1]Department of Economics, National University of Singapore, Singapore 117570, Singapore, [2]Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA. *e-mail: albertosalvo@nus.edu.sg

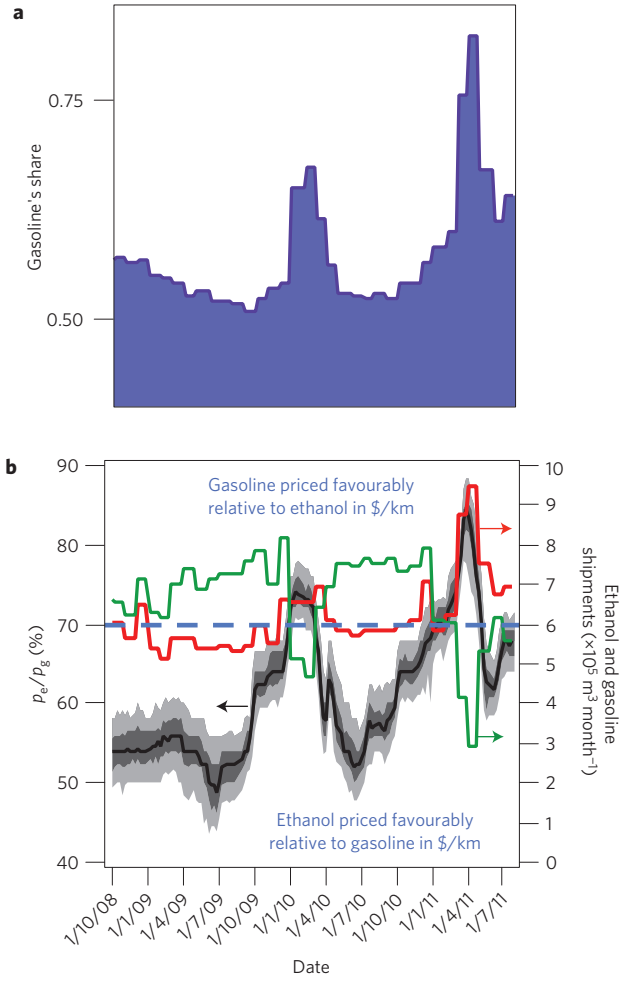

Figure 1 | Shifting fuel quantities and prices between October 2008 and July 2011. a, The monthly share of blended gasoline purchased at retail (E25 or E20) of total estimated light-vehicle distance travelled, prepared from wholesale shipment reports. **b**, The weekly per-litre price of regular ethanol (E100), denoted p_e, divided by the per-litre price of regular blended gasoline, denoted p_g, at pumps in the city of São Paulo (left *y* axis). The black curve indicates the median, and grey shading represents the ranges of the 5th–25th (lower light grey), 25th–75th (dark grey), and 75th–95th (upper light grey) percentile of the distribution of p_e/p_g across retailers. The 70% 'parity' threshold widely reported by the media, at which \$ km^{-1} equalizes, is marked by the dashed blue horizontal line. The right *y* axis shows monthly reported shipments of all grades of blended gasoline (red) versus ethanol (green) from wholesalers to retailers located in the state of São Paulo. Sources: ANP, Inmetro and authors' calculations.

to associate ethanol with improved environmental outcomes, including air quality (for example, surveys of Brazilian ethanol consumers[12], comments by the ethanol industry at US Senate hearings[34,35], and an interview with a former Secretary of the Environment in Brazil)[9,36]. Despite their importance, the above studies and claims have not yet been benchmarked against the chemical composition of air measured before, during, and after an actual rather than hypothetical large-scale switch from a fossil fuel over to a biofuel in a large urban centre.

Street-hour analysis of concentrations and controls

Our study cross-examines a large amount of measured data, detailed at the street-hour level (Fig. 2) from several sources: concentrations of regulated 'priority' pollutants, namely O_3, NO, NO_2, and CO (including SO_2 and PM10 in the Supplementary Information), measured by spatially differentiated air monitoring stations maintained by the environmental authority of the state of São Paulo (CETESB; ref. 2); meteorological conditions measured at these same CETESB stations as well as recorded by the Institute for Meteorology (INMET; ref. 37); and controls for vehicle traffic congestion and speed obtained from the city traffic authority (CET; ref. 38). We combine the extensive pollutant-meteorology-traffic data with: weekly gasoline and ethanol prices at the pump, obtained from the National Agency for Oil, Biofuels and Natural Gas (ANP; ref. 10), which in turn feeds a consumer demand system estimated from survey data[12] (or, to check the robustness of our findings, fuel shares based on available monthly wholesaler reports)[11].

Our main result, for ozone, is summarized in Table 1. The table reports regression estimates for hourly O_3 concentrations measured in the early afternoon (13:00–16:00) on non-holiday (regular) weekdays. As our baseline regressions avoid pooling observations for different locations, the top panel reports estimated coefficients and standard errors for one of the ozone monitors, by way of example, whereas the bottom panel reports mean effects and precision across all 12 ozone monitor-level regressions. The fuel mix variables—the main variables of interest, see Methods—are s_t^{gas}, the share of bi-fuel vehicles fuelled with blended gasoline over ethanol, and $e20_t^{gas}$, an indicator variable for the three-month period during which the government mandated the distribution of gasoline as E20 rather than the usual E25. Standard errors account for the fact that in these particular regressions the gasoline share is estimated rather than measured (more rigorously, $\hat{s}_t^{gas}$ rather than s_t^{gas}), and standard errors for the means allow for correlation across stations. Estimates for other times of day and types of day, for each individual station, are provided in Supplementary Information Part F. Columns I through VII indicate how coefficients on the fuel mix are impacted by progressively adding controls to soak up residual variation. We focus the discussion on the gasoline share, and subsequently comment on the temporary gasoline blend requirement change. As seen in Fig. 3a–c, the gasoline share varied 62 percentage points over the sample period, thus an in-sample effect is obtained by multiplying the estimated coefficient on s_t^{gas} by 0.62.

In the absence of controls, ozone concentrations and the gasoline share are (on average) not statistically significantly associated (column I). The inclusion of a linear trend, in column II, shifts this association, as one would expect given any underlying trend in ozone concentrations and the fact that the gasoline share between November 2008 and May 2011 also trends. Our purpose is to exploit the significant fuel mix variation around this trend. Supplementary Information Fig. 13 illustrates the variation remaining in the gasoline share once a linear trend (or a quadratic one) has been partialled out.

In column III, the mean relationship between O_3 concentrations and the gasoline share becomes negative—but not significantly so—on adding fixed effects for the week of the year, day of the week, and hour of the day. Week-of-year dummies, in particular, raise explanatory power considerably, as these capture seasonal variation in pollutant concentrations. Intuitively, this specification compares ozone pollution on a given year's week when the de-trended gasoline share was high with pollution on the same week in another year when the share was low—within location, time of day, and day type.

Columns IV to VII report specifications that control for different functions of contemporaneous and lagged measures of meteorological and traffic conditions (denoted W_t and T_t, respectively, these enter as logarithmic transforms of their units of observation). In column IV, the addition of five contemporaneous meteorological covariates boosts the power of these location-time-day-type specific regressions to predict O_3 concentrations, with R^2 growing on average from 22 to 69%. Ozone concentrations—already conditioning on early afternoon, week of the year, and so on—are increasing in radiation and temperature and decreasing in

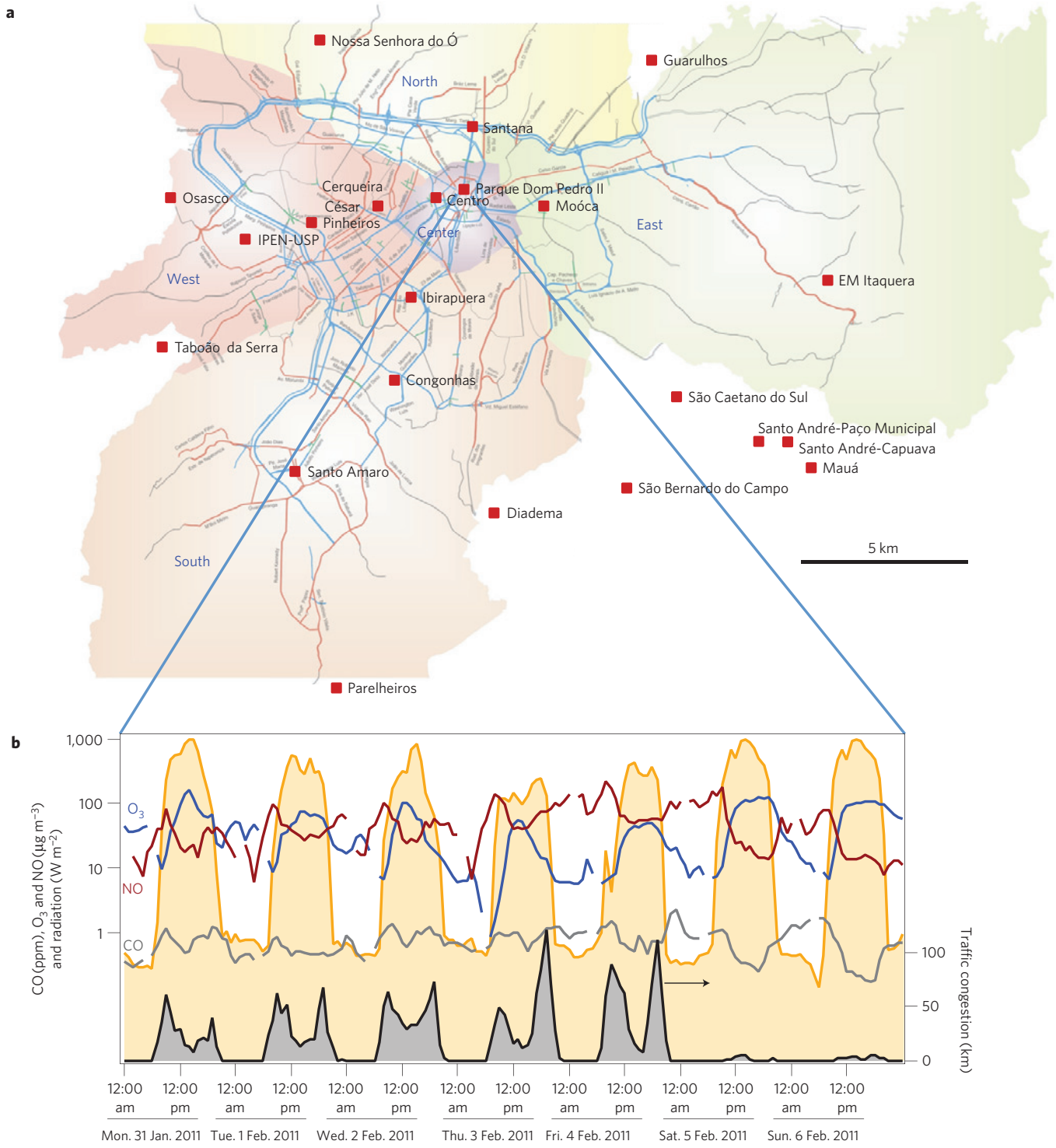

Figure 2 | Street-hour analysis of concentrations and controls. a, Map of the environmental authority's air monitoring stations (red squares), which often double as weather stations, in the São Paulo metropolitan area, superimposed on the road network monitored every 30 min by the traffic authority for traffic congestion, in São Paulo city. **b**, Measured concentrations of O_3 (blue), NO (brown), and CO (grey), and radiation (yellow), at generic stations and citywide extension of traffic congestion (black), by hour from 31 January to 6 February 2011. Sources: CETESB, INMET, CET and ANP.

humidity (a correlate of precipitation) and wind speed. Importantly, the coefficient on the gasoline share, averaged across the 12 O_3-monitoring stations, becomes more negative and is more precisely estimated. Column V also controls for the total extension of traffic congestion (that is, idling vehicles) reported contemporaneously over a monitored 840-km road network across the city. Column VI adds lagged meteorological and traffic covariates to account for variation in conditions up to 18 h preceding an observation. In addition to road congestion at the citywide level, traffic covariates in column VI now include two local measures, namely: the sum of congestion only in the region of the city where the monitoring station is located (for example, North in Fig. 2); and a weighted sum

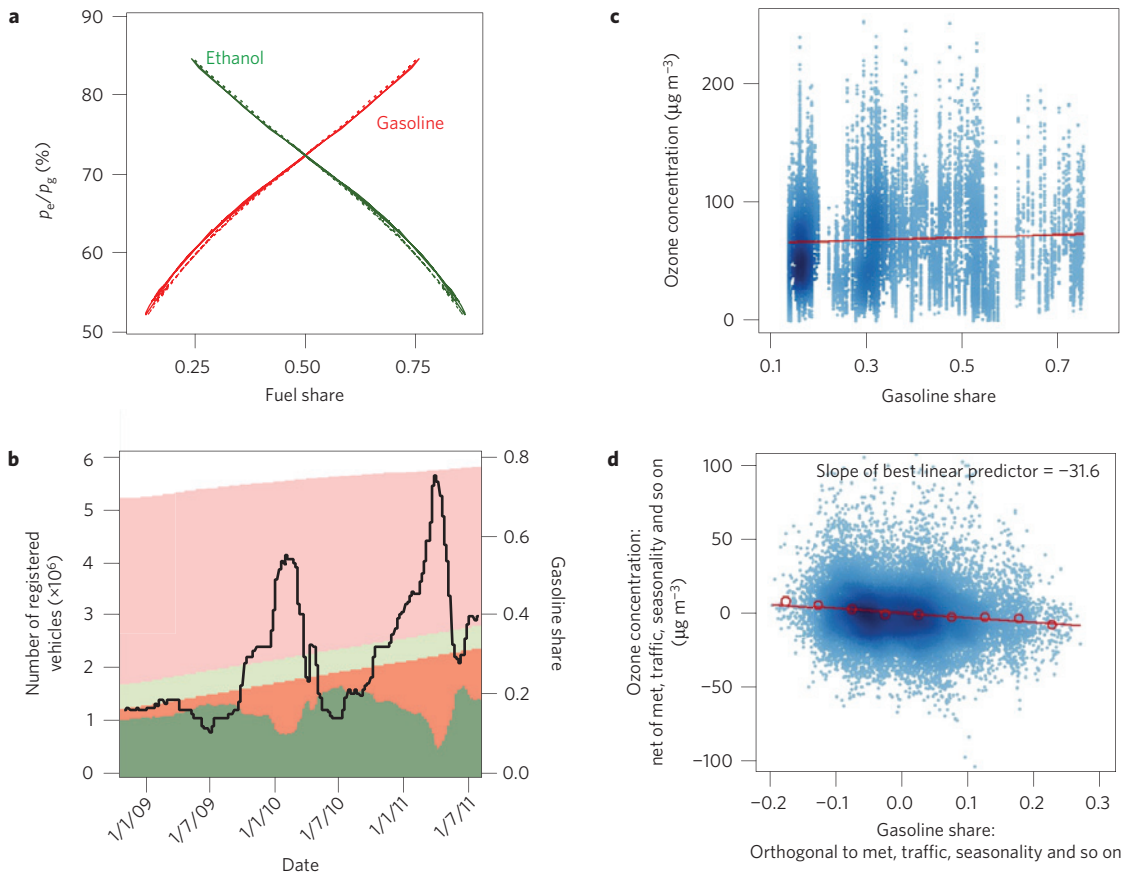

Figure 3 | Gradual transitions between ethanol and gasoline in bi-fuel vehicles. a, In-sample variation in the median per-litre regular ethanol-to-gasoline price ratio, p_e/p_g, against the corresponding shares of blended gasoline (E25, regular and midgrade) and ethanol (E100, regular) chosen by bi-fuel vehicle consumers at the pump. Solid lines indicate shares predicted by a multinomial probit specification[12] (Table 2, specification III) and dashed lines indicate shares predicted by an alternative multinomial logit specification (using data in ref. 12). **b**, Variation in the predicted gasoline share s_t^{gas} (multinomial probit specification), and counts of bi-fuel and single-fuel vehicles burning gasoline or ethanol, between November 2008 and July 2011. Bi-fuel vehicles on gasoline or ethanol are marked by the dark red and dark green areas, respectively, and single-fuel vehicles on gasoline or ethanol are marked by the light red and light green areas, respectively. Official single-fuel vehicle counts are neither adjusted for less usage nor for being overstated relative to bi-fuel vehicles. **c**,**d**, Partialling out the effect of meteorological and traffic conditions, seasonality, trending omitted factors, and so on, on O_3 concentrations to identify the effect of the fuel mix. These two panels plot, for all 12 O_3-monitoring stations in the early afternoon hours on non-holiday weekdays: measured O_3 concentrations in $\mu g\,m^{-3}$ against the gasoline share s_t^{gas} (**c**); and residuals of a regression of O_3 concentrations on all explanatory variables other than s_t^{gas} against the residuals of a regression of s_t^{gas} on these same explanatory variables (**d**). Colour intensity indicates local density in each of 128 × 128 bins. For **d** the best linear predictor over all points (no binning) is marked with a red line, and mean ozone residuals in bins of width 0.05 along the horizontal axis (−0.2 to −0.15, −0.15 to −0.10, ..., 0.20 to 0.25) are marked by red circles at the horizontal midpoint. Sources: **a**,**b**, ANP, ref. 12, DETRAN-SP, Fenabrave, Sindipeças, authors' estimates; **c**,**d**, specification VI (with station fixed effects included).

of congestion recorded along traffic corridors that are in proximity to the station, where the weights are given by the inverse distance from each corridor to the station. In column VI, coefficients on W_t and T_t are too numerous to report. Finally, relative to column V, column VII adds interactions of traffic congestion in the regions of the city that surround a station and the direction from which the wind is blowing (that is, we include $f(W_t, T_t)$ in regression equation (1) below).

O_3 reduction and NO and CO increase with shift to gasoline

Across specifications IV–VII of Table 1, the average fuel mix effect λ_1 (see equation (1)) on the ozone concentration is estimated at −20.7 to −30.2 $\mu g\,m^{-3}$, with standard errors on mean (s.e.m.) of under 9 $\mu g\,m^{-3}$. This range of estimates corresponds to a statistically significant reduction in ambient ozone levels, as the gasoline share rose by 62 percentage points, of about 15 $\mu g\,m^{-3}$ ($\hat{\lambda}_1$ averaged across the columns times 0.62). This 15 $\mu g\,m^{-3}$ drop amounts to 22% of the mean value of the dependent variable—observed ozone concentrations in the early afternoon on non-holiday weekdays average 68 $\mu g\,m^{-3}$. Figure 4 plots mean changes in ozone concentrations, as the gasoline share rose 62 percentage points, estimated not only for the early afternoon but also for other times of the day (using specification VI).

We also estimate a mean negative coefficient λ_2 on the gasoline E20 blend dummy variable, $e20_t^{gas}$, suggesting that O_3 concentrations were similarly lower in the three-month period from February to April 2010, during which the gasoline fuel dispensed to consumers contained five percentage points less ethanol (more gasoline) by volume. This estimated negative effect $\hat{\lambda}_2$, however, is only marginally significant (columns IV and V) to insignificantly different from zero (column VII), probably as a result of the smaller magnitude of the ethanol–gasoline shift over this episode (Supplementary Information Part F). Nonetheless, that we estimate λ_1 and λ_2 to be of the same sign—based on continuously valued and discretely valued variables, respectively—increases our confidence that our identifying assumption holds (see equation (2) in Methods)

Table 1 | Predicting ozone ($\mu g\,m^{-3}$) on non-holiday weekdays; 13:00 to 16:00 readings only.

Specification	I	II	III	IV	V	VI	VII
Example of one station-level regression: Station ID 1							
Main variables of interest							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	3.0 (11.5)	26.1 (15.8)	−14.9 (20.1)	−20.7 (12.5)	−20.1 (12.4)	−23.2 (12.9)	−18.8 (12.0)
Three-month period with gasoline E20, $e20_t^{gas}$	−0.1 (4.4)	−1.6 (4.4)	−3.0 (4.9)	1.3 (2.6)	1.4 (2.7)	5.0 (2.7)	2.3 (2.6)
Control variables							
Trend (linear)	No	Yes	Yes	Yes	Yes	Yes	Yes
Week-of-year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Day-of-week fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Hour-of-day fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Meteorology: contemporaneous conditions	No	No	No	Yes	Yes	Yes	Yes
Precipitation				0.0 (0.4)	0.0 (0.5)		0.0 (0.4)
Humidity				−36.5 (6.7)	−36.6 (6.7)		−37.6 (6.9)
Radiation				10.6 (1.9)	10.6 (1.9)		9.8 (1.9)
Temperature				89.6 (12.9)	89.4 (12.9)		93.8 (13.0)
Wind speed				−11.9 (3.6)	−11.8 (3.6)		−10.2 (3.5)
Citywide traffic congestion: contemporaneous conditions	No	No	No	No	Yes	Yes	Yes
Total extension of congestion across city					−0.7 (1.8)		1.2 (2.1)
Meteorology: conditions lagged up to 18 h	No	No	No	No	No	Yes	No
Meteorology: Pairwise interactions of contemporaneous conditions	No	No	No	No	No	Yes	No
Local traffic congestion: contemporaneous conditions	No	No	No	No	No	Yes	No
Traffic congestion, citywide and local: lagged up to 18 h	No	No	No	No	No	Yes	No
Interactions of wind direction and traffic in other regions	No	No	No	No	No	No	Yes
R2	0.0%	0.8%	25.2%	67.3%	67.4%	74.0%	68.6%
Number of observations	1,414	1,414	1,414	1,401	1,401	1,371	1,397
Number of regressors	3	4	43	48	49	95	54
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	9.6 (9.0)	19.3 (12.9)	−5.9 (16.2)	−21.9 (8.3)	−24.7 (8.3)	−30.2 (7.9)	−20.7 (8.6)
Three-month period with gasoline E20, $e20_t^{gas}$	−5.6 (4.0)	−6.1 (4.0)	−5.3 (4.2)	−3.9 (2.3)	−4.5 (2.2)	−2.8 (2.0)	−1.9 (2.4)
Selected meterology and traffic (mean estimates across 12 station-specific regressions)							
Precipitation				0.2 (0.4)	0.2 (0.4)		0.2 (0.4)
Humidity				−46.5 (4.7)	−45.9 (4.7)		−45.3 (4.7)
Radiation				7.3 (1.2)	7.5 (1.2)		6.8 (1.2)
Temperature				85.2 (9.0)	86.0 (9.0)		91.7 (9.4)
Wind speed				−12.7 (2.4)	−12.7 (2.4)		−12.4 (2.3)
Total extension of congestion across city					2.1 (1.2)		1.8 (1.1)
Mean across 12 station-specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1,574	1,574	1,574	1,560	1,560	1,531	1,515
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7

The top panel reports coefficients and standard errors (in parentheses) for one of the station-specific regressions. The bottom panel reports means and standard errors (in parentheses) for selected effects across regressions for each of the 12 ozone-monitoring stations. Standard errors are calculated by bootstrapping (200 samples each): the consumer-level fuel choice data, to account for sampling variation in the gasoline share, and the pollutant-meterology-traffic data, clustering by date. Standard errors on station-level estimates are the standard deviations of coefficients over the 200 replications. Standard errors on means across stations are calculated by averaging, for each replication, coefficients across stations, and computing the standard deviation, over replications, of these means. An observation is an hour-date pair falling within the specified time of day and type of day. The sample period is 1 November 2008–31 May 2011, excluding the colder months of June to September. Ordinary Least Squares estimates. Local traffic conditions entering in specification VI are the extension of congestion in the region of the city where the station is located and the inverse-distance weighted sum of congestion in nearby roads. In specification VII, interactions of wind direction and traffic congestion in other regions is for contemporaneous conditions.

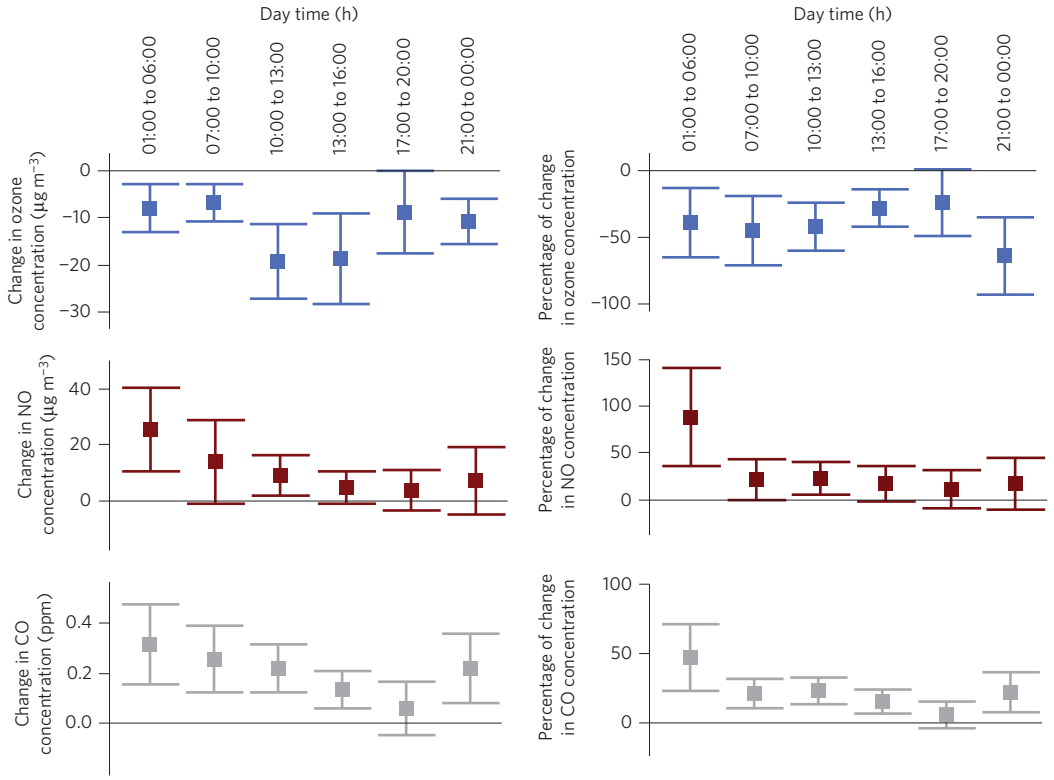

Figure 4 | Estimated changes in O_3, NO, and CO concentrations as the gasoline share, s_t^{gas}, rose by 62 percentage points. Mean effects across regressions for the stations monitoring each given pollutant, for different times on a non-holiday weekday. The left panels plot the 95% confidence intervals for (the mean across monitors of) $0.62\hat{\lambda}_1$ and the right panels express these confidence intervals as proportions of mean recorded concentrations at the different times of a non-holiday weekday. Source: specification VI estimates (Table 1 and Table 2). See the footnote to Table 1 on how the error bars were obtained.

and that the negative coefficient on s_t^{gas} is not being driven by some time-varying omitted variable that, after controlling for a linear trend, still happens to be spuriously correlated with the gasoline share. We note that our results are very robust to replacing the linear trend by a quadratic one.

Table 2 presents the same analysis for NO, NO_2 and CO, based on concentrations measured in the morning rush hours (07:00–10:00) of non-holiday weekdays at 9 NO_X-monitoring stations and 11 CO-monitoring stations. For brevity, the table reports estimated effects averaged across station-specific regressions (individual estimates are provided in the Supplementary Information).

Point estimates of the effect λ_1 of raising the gasoline share tend to be positive for NO and for CO, but these effects are less precisely estimated than for O_3. As with O_3, the estimated effect $\hat{\lambda}_2$ of the step change in the gasoline blend is of the same sign as $\hat{\lambda}_1$. Averaging across specifications IV–VII, a 62-percentage-point rise in the gasoline share is associated with increases of 17.3 $\mu g\,m^{-3}$ (s.e.m. 7.6 $\mu g\,m^{-3}$) and 0.22 ppm (s.e.m. 0.07 ppm) in ambient NO and CO concentrations, respectively—amounting to 26 and 18% of the mean readings during morning rush hours (also see Fig. 4 for other times of the day). Finally, the estimated effect λ_1 for NO_2 is not significantly different from zero—point estimates are smaller and noisier than those for NO, whose ambient concentrations are around 40% higher compared with NO_2.

Figure 3c,d offers an intuitive illustration of our method and of our result for ozone. Figure 3c plots O_3 concentrations measured in the early afternoon hours on non-holiday weekdays against the gasoline share s_t^{gas}. There happens to be a positive relationship in the raw data (only to illustrate, here we pool observations at all O_3 monitors). The vertical axis in Fig. 3d shows fitted residuals of a regression of O_3 concentrations on all independent variables except s_t^{gas} (we use specification VI plus monitor fixed effects for this pooled regression). These residual concentrations are the variation in ozone that is left unexplained once variation in meteorology, traffic, seasonality, trending omitted factors, and the gasoline blend change are accounted for. The horizontal axis plots residuals from a regression of s_t^{gas} on the same vector of independent variables: these share residuals capture the component of variation in the gasoline-over-ethanol consumer choice that is orthogonal to the other regressors. The relationship between the residual O_3 concentrations and the residual gasoline share is negative, with a $\hat{\lambda}_1$ slope of $-31.6\,\mu g\,m^{-3}$; this is similar to the mean coefficient across station-specific regressions in column VI, Table 1.

In Supplementary Information Part F we perform a placebo test and subject our baseline results to a number of extra robustness checks, including: specifying dependent variables as logarithmic transforms of the units of measurement; keeping the colder months of June to September in the sample; controlling for recorded traffic speeds on top of congestion; controlling for the real price of diesel, monthly ridership on the public transport system, monthly physical industrial production for the state of São Paulo, and employment or wages in the metropolis. We also note that gone are the days in which the city of São Paulo was an industrial hub, and that now the electricity that serves southeastern Brazil is predominantly generated by hydropower. In sum, factors that might otherwise confound identification of the effect of the fuel mix on air quality are less of a concern in the present study.

Towards quantitative benchmarks for model studies

Our results stand at variance to those of the recent computer simulation that was calibrated to the São Paulo system[25] which predicted large reductions in ozone concentrations from a hypothetical switch to ethanol that—although larger than the one we observe in the data—is of comparable magnitude. Our joint data

Table 2 | Predicting NO ($\mu g\,m^{-3}$), NO_2 ($\mu g\,m^{-3}$), and CO (ppm) on non-holiday weekdays; 07:00–10:00 readings only.

Specification	I	II	III	IV	V	VI	VII
Dependent variable: NO concentration ($\mu g\,m^{-3}$)							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	−17.4 (12.6)	3.6 (19.7)	45.6 (21.1)	29.2 (12.2)	28.2 (12.3)	22.7 (12.4)	31.7 (11.9)
Three-month period with gasoline E20, $e20_t^{gas}$	6.7 (4.5)	6.1 (4.5)	8.3 (5.5)	11.8 (3.8)	11.8 (3.8)	9.7 (4.0)	11.0 (3.9)
Selected meterology and traffic (mean estimates across 9 station-specific stations)							
Precipitation				−0.4 (0.4)	−0.4 (0.5)		−0.5 (0.5)
Humidity				−19.6 (11.6)	−19.7 (11.7)		−18.1 (11.0)
Radiation				3.5 (1.1)	3.6 (1.1)		3.5 (1.1)
Temperature				−17.0 (13.8)	−17.0 (13.8)		−25.7 (13.8)
Wind speed				−52.3 (3.2)	−52.3 (3.2)		−51.2 (3.2)
Total extension of congestion across city					1.4 (2.9)		1.8 (2.9)
Mean across 9 station-specific regressions							
R2	0.9%	1.5%	21.0%	43.2%	43.3%	53.2%	45.9%
Number of observations	1,529	1,529	1,529	1,514	1,514	1,489	1,464
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	66.5	66.5	66.5	66.5	66.5	67.0	66.8
Dependent variable: NO_2 concentration ($\mu g\,m^{-3}$)							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	−4.4 (3.6)	−2.6 (5.0)	5.0 (6.4)	−5.3 (4.7)	−6.2 (4.7)	5.1 (4.4)	−5.8 (4.8)
Three-month period with gasoline E20, $e20_t^{gas}$	3.3 (1.7)	3.3 (1.7)	3.7 (1.9)	4.6 (1.3)	4.5 (1.3)	−0.7 (1.2)	4.5 (1.4)
Selected meterology and traffic (mean estimates across 9 station-specific stations)							
Precipitation				0.3 (0.2)	0.3 (0.2)		0.2 (0.2)
Humidity				−28.2 (4.7)	−28.3 (4.7)		−26.7 (4.7)
Radiation				1.3 (0.5)	1.3 (0.5)		1.4 (0.5)
Temperature				33.6 (4.7)	33.6 (4.8)		31.2 (4.8)
Wind speed				−10.4 (0.8)	−10.4 (0.8)		−9.9 (0.8)
Total extension of congestion across city					1.1 (1.0)		1.0 (1.0)
Mean across station specific regressions							
R2	2.9%	4.7%	21.9%	40.3%	40.4%	58.4%	42.9%
Number of observations	1,529	1,529	1,529	1,514	1,514	1,489	1,464
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	48.1	48.1	48.1	48.0	48.0	48.1	47.9
Dependent variable: CO concentration (ppm)							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	0.15 (0.13)	0.35 (0.21)	0.66 (0.22)	0.34 (0.12)	0.33 (0.12)	0.41 (0.11)	0.34 (0.12)
Three-month period with gasoline E20, $e20_t^{gas}$	0.08 (0.05)	0.07 (0.05)	0.10 (0.06)	0.17 (0.04)	0.17 (0.04)	0.11 (0.04)	0.16 (0.04)
Selected meterology and traffic (mean estimates across 11 station-specific stations)							
Precipitation				0.00 (0.01)	0.00 (0.01)		0.0 (0.01)
Humidity				0.16 (0.12)	0.16 (0.12)		0.20 (0.12)
Radiation				0.06 (0.01)	0.06 (0.01)		0.06 (0.01)
Temperature				0.70 (0.14)	0.70 (0.14)		0.63 (0.14)
Wind speed				−0.49 (0.03)	−0.49 (0.03)		−0.48 (0.03)
Total extension of congestion across city					0.01 (0.03)		0.01 (0.03)
Mean across 11 station-specific regressions							
R2	1.2%	2.0%	24.6%	48.9%	49.0%	59.8%	51.5%
Number of observations	1,565	1,565	1,565	1,548	1,548	1,523	1,505
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	1.20	1.20	1.20	1.20	1.20	1.20	1.20

See footnote to Table 1. Estimated mean coefficients and standard errors on means (in parentheses) across station-specific regressions (9 stations monitoring nitrogen oxides and 11 stations monitoring CO). Standard errors account for estimation of the gasoline share.

analysis of pollutant concentrations, meteorological and road traffic conditions, and consumer fuel choice indicates that early-afternoon O_3 concentrations declined by an average 15 $\mu g\,m^{-3}$ (22% of the sample mean) as the share of bi-fuel vehicles burning *gasoline* grew from 14 to 76%. Such empirical findings are consistent with the modelling hypothesis that O_3 production over the São Paulo metropolis may be hydrocarbon-limited[39], whereby higher NO_X emissions (from gasoline) would result in reductions in ambient ozone. Hydrocarbon-limited O_3 production would also rationalize why O_3 levels tend to increase, and NO_X and CO levels tend to decrease, on weekends, when road traffic congestion falls. Such an interpretation for our São Paulo result should be contrasted with the claim that '(m)easurements and model calculations now show that O_3 production over most of the United States is primarily NO_X-limited, not hydrocarbon-limited'[31]. Clearly, successful strategies against ozone pollution require knowledge of the local regime. Moreover, with access to the relevant air (and other) monitoring data for the area outside of the heavily urbanized São Paulo metropolis, our approach is potentially applicable for the estimation of ozone and NO_X concentrations in suburban or rural areas downwind.

Our present study has shown that under atmospheric conditions observed in São Paulo, concentrations of two air pollutants, specifically NO and CO, may increase whereas that of ozone falls on raising the gasoline fuel share. We caution that the concentration of particles, specifically fine particulate matter, may also increase under that situation. Given that the method presented here allows, in principle, for the evaluation of how different fuel mixes impact pollutants other than ozone and NO_X, such as particulate matter, it is our view that studies such as ours may help inform scientists and policymakers alike on the benefits and disadvantages that certain fuel mixes may have on ambient levels of pollutants, be they in the gas or condensed phase.

Methods

Not unlike the 'chemical coordinates' approach put forth by Cohen *et al.*[40], we apply a multivariate regression analysis of a real-world dataset exhibiting, in the present case, rich and exogenous time variation in fuel mix[41]. The idea is to compare pollutant concentrations across subsamples which differ only in the fuel mix—gasoline versus ethanol—but are otherwise similar with regard to other determinants of air quality, including meteorology, anthropogenic activity, and biogenic activity. We directly control for variation in local meteorological and vehicle traffic conditions, contemporaneously and in the several hours that precede an observation. Our regressions flexibly predict a pollutant's concentration specific to the location of the air monitor and time and type of day, using a relatively short sample period during which the fuel mix varied, namely late 2008 to mid 2011. We drop the colder months from June to September from the baseline sample. We thus control for unobserved variation that might potentially confound our inference of the effect of the fuel mix on air quality.

Our baseline regression equation, which we estimate separately by location of measurement and time and type of day, takes the following form:

$$\text{concentration}_t = \lambda_1 s_t^{\text{gas}} + \lambda_2 e20_t^{\text{gas}} + W_t'\Delta^W + T_t'\Delta^T + \text{fixedeffects}_t + \text{trend}_t + \varepsilon_t \quad (1)$$

An observation t is an hour–date pair, for example, for the Diadema station, early afternoon (13:00–16:00), non-holiday weekday regression, an observation is 14:00 on Monday 14 March 2011 (this was not a public holiday). The dependent variable concentration_t corresponds to a pollutant that is measured at the station, for example, O_3, in the measured units ($\mu g\,m^{-3}$) or a logarithmic transform thereof. Both fuel mix variables, s_t^{gas} and $e20_t^{\text{gas}}$, increase in the proportion of gasoline, although the shift from ethanol to gasoline as $e20_t^{\text{gas}}$ changes from 0 to 1 is of lesser magnitude—we thus expect the effect λ_2 to have a lower magnitude than, but exhibit the same sign as, λ_1. W_t and T_t are vectors of contemporaneous and lagged meteorological and traffic controls that are local to the particular station of measurement, as detailed in the Supplementary Information, and Δ^W and Δ^T are coefficients. To account for seasonal variation, we include full sets of week-of-year, day-of-week, and hour-of-day fixed effects; for the Diadema station observation in the example, the week 11 (14 March 2011), Monday, and 14:00 indicators would be on. We also allow for a linear or quadratic trend in the date to control for potentially confounding omitted time-varying factors. The identifying assumption is that, conditional on controls, the residual is uncorrelated with the fuel mix, in particular:

$$E[s_t^{\text{gas}}\varepsilon_t | X_t] = 0, \quad \text{where } X_t := (W_t, T_t, \text{fixedeffects}_t, \text{trend}_t) \quad (2)$$

A concern that might arise in a real-world—as opposed to lab or synthetic—setting such as ours is the possibility that consumers may have cut back on vehicle usage when faced with rising ethanol prices. If this were the case, not controlling for vehicle usage would confound our estimation of the effect of varying the fuel mix on air quality, as the corresponding orthogonality condition (without covariates T_t) would not hold. Two points should be noted. First, we do add detailed controls for local and citywide road traffic congestion and speed recorded at the hourly level. Second, we show that traffic conditions and thus vehicle usage, although predictable, did not significantly vary with fuel prices during the sample period. This finding can be rationalized on different counts, namely: the typically price-inelastic short-run demand for vehicle usage due to the poor availability of substitutes[42], including public transportation, as evidenced by ridership records (Supplementary Information Part E); the existence of 'repressed demand' for vehicle usage that has been argued in the face of widespread gridlock[43,44]; and the relatively subdued variation in the price of gasoline—which can fuel nine-tenths of São Paulo's light-duty fleet of bi-fuel and single-fuel vehicles.

With regard to the gasoline share among bi-fuel consumers s_t^{gas}, one approach[45] would be to assume that consumers perceive gasoline and ethanol to be 'perfect substitutes', thus fuelling their bi-fuel vehicles with the fuel that yields the lowest $ per distance travelled. By this assumption, consumers would switch from ethanol to gasoline, $s_t^{\text{gas}} = 1$, whenever the per-litre price of ethanol surpassed around 70% of the per-litre price of gasoline, and s_t^{gas} would be 0 otherwise. The analysis could then follow a regression discontinuity design[46]. However, surveys of Brazilian motorists making choices at the pump have shown that there is substantial heterogeneity in consumer behaviour and that, rather than discontinuously, fuel switching occurs gradually over a wide range of relative price variation[12]. Our measure of s_t^{gas}, which ranges from 14 to 76% in-sample, is obtained from an estimated consumer demand system, based on the multinomial probit model[47]. For robustness, we obtain a similar gasoline share on estimating an alternative consumer-level choice model based on the multinomial logit. Figure 3 reports how the gasoline share, s_t^{gas}, varies in the sample: (Fig. 3a) with the per-litre ethanol-to-gasoline price ratio—notice that there is no kink at the approximate 70% 'parity' threshold, at which $/mile travelled on either fuel is about the same; and (Fig. 3b) over time (see Supplementary Information Part A for demand modelling and estimation). The data archive can be accessed at http://bit.do/salvo_geiger_data.

Received 2 October 2013; accepted 18 March 2014; published online 28 April 2014

References

1. Martins, L. & Andrade, M. Ozone formation potentials of volatile organic compounds and ozone sensitivity to their emission in the megacity of São Paulo, Brazil. *Water Air Soil Poll.* **195,** 201–213 (2008).
2. CETESB, *Relatório Anual sobre a Qualidade do Ar no Estado de São Paulo [Annual Report on Air Quality in the State of São Paulo]* (Companhia Ambiental do Estado de São Paulo, 2010).
3. La Rovere, E. L. *Inventário de Emissões de Gases de Efeito Estufa do Município de São Paulo* [*Greenhouse Gas Emissions Inventory for the Municipality of São Paulo*] (Centro de Estudos Integrados sobre Meio Ambiente e Mudanças Climáticas, Universidade Federal do Rio de Janeiro, 2005).
4. Fann, N. *et al.* Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Anal.* **32,** 81–95 (2012).
5. Gauderman, W. J. *et al.* Effect of exposure to traffic on lung development from 10 to 18 years of age: A cohort study. *Lancet* **369,** 571–577 (2007).
6. Ponce, N. A., Hoggatt, K. J., Wilhelm, M. & Ritz, B. Preterm birth: The interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods. *Am. J. Epid.* **162,** 140–148 (2005).
7. Currie, J. & Walker, R. Traffic congestion and infant health: Evidence from E-ZPass. *Am. Econ. J.-Appl. Econ.* **3,** 65–90 (2011).
8. Salvo, A. & Huse, C. Is arbitrage tying the price of ethanol to that of gasoline? Evidence from the uptake of flexible-fuel technology. *Energy J.* **32,** 119–148 (2010).
9. Goldemberg, J. Ethanol for a sustainable energy future. *Science* **315,** 808–810 (2007).
10. http://www.anp.gov.br/preco/prc/Resumo_Por_Municipio_Index.asp
11. http://www.anp.gov.br/?dw=11031
12. Salvo, A. & Huse, C. Build it, but will they come? Evidence from consumer choice between gasoline and sugarcane ethanol. *J. Environ. Econ. Manag.* 251–279 (2013).

13. Al-Hasan, M. Effect of ethanol-unleaded gasoline blends on engine performance and exhaust emission. *Energy Convers. Manag.* **44,** 1547–1561 (2003).
14. Graham, L. A., Belisle, S. L. & Baas, C-L. Emissions from light duty gasoline vehicles operating on low blend ethanol gasoline and E85. *Atmos. Environ.* **42,** 4498–4516 (2008).
15. He, B-Q., Jian-Xin, W., Hao, J-M., Yan, X-G. & Xiao, J-H. A study on emission characteristics of an EFI engine with ethanol blended gasoline fuels. *Atmos. Environ.* **37,** 949–957 (2003).
16. Hsieh, W-D., Chen, R-H., Wu, T-L. & Lin, T-H. Engine performance and pollutant emission of an SI engine using ethanol-gasoline blended fuels. *Atmos. Environ.* **36,** 403–410 (2002).
17. Jia, L-W., Shen, M-Q., Wang, J. & Lin, M-Q. Influence of ethanol–gasoline blended fuel on emission characteristics from a four-stroke motorcycle engine. *J. Haz. Mat.* **123,** 29–34 (2005).
18. Leong, S. T., Muttamara, S. & Laortanakul, P. Applicability of gasoline containing ethanol as Thailand's alternative fuel to curb toxic VOC pollutants from automobile emission. *Atmos. Environ.* **36,** 3495–3503 (2002).
19. Lynd, L. R. Overview and evaluation of fuel ethanol from cellulosic biomass: Technology, economics, the environment, and policy. *Ann. Rev. Energy Environ.* **21,** 403–465 (1996).
20. Mulawa, P. A. *et al.* Effect of ambient temperature and E-10 fuel on primary exhaust particulate matter emissions from light-duty vehicles. *Environ. Sci. Technol.* **31,** 1302–1307 (1997).
21. Poulopoulos, S. G., Samaras, D. P. & Philippopoulos, C. J. Regulated and unregulated emissions from an internal combustion engine operating on ethanol-containing fuels. *Atmos. Environ.* **35,** 4399–4406 (2001).
22. Topgül, T., Yücesu, H. S., Çinar, C. & Koca, A. The effects of ethanol–unleaded gasoline blends and ignition timing on engine performance and exhaust emissions. *Renew. Energy* **31,** 2534–2542 (2006).
23. Yoon, S. H., Ha, S. Y., Roh, H. G. & Lee, C. S. Effect of bioethanol as an alternative fuel on the emissions reduction characteristics and combustion stability in a spark ignition engine. *J. Auto Engin.* **223,** 941–951 (2009).
24. Pereira, P. A. d. P., Santos, L. M. B., Sousa, E. T. & Andrade, J. B. d. Alcohol- and gasohol-fuels: a comparative chamber study of photochemical ozone formation. *J. Braz. Chem. Soc.* **15,** 646–651 (2004).
25. Martins, L. D. & Andrade, M. F. Emission scenario assessment of gasohol reformulation proposals and ethanol use in the metropolitan area of São Paulo. *Open Atmos. Sci. J.* **2,** 166–175 (2008).
26. Ginnebaugh, D. L., Liang, J. & Jacobson, M. Z. Examining the temperature dependence of ethanol (E85) versus gasoline emissions on air pollution with a largely-explicit chemical mechanism. *Atmos. Environ.* **44,** 1192–1199 (2010).
27. Jacobson, M. Z. Effects of ethanol (E85 versus gasoline vehicles on cancer and mortality in the United States. *Environ. Sci. Technol.* **41,** 4150–4157 (2007).
28. Beer, T. *et al.* The health impacts of ethanol blend petrol. *Energies* **4,** 352–367 (2011).
29. Finlayson-Pitts, B. & Pitts, J. *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications* (Academic Press, 2000).
30. Seinfeld, J. H. & Pandis, S. N *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (Wiley, 1998).
31. Jacob, D. J. *Introduction to Atmospheric Chemistry* (Princeton Univ. Press, 1999).
32. Anderson, L. G. Ethanol fuel use in Brazil: Air quality impacts. *Energy Environ. Sci.* **2,** 1015–1037 (2009).
33. Romieu, I., Weitzenfeld, H. & Finkelman, J. Urban air pollution in Latin America and the Caribbean. *J. Air Waste Manag.* **41,** 1166–1171 (1991).
34. *US Senate Hearing 106-953* (US Government Printing Office, 2000).
35. *US Senate Hearing 109-857* (*Presentation by Mr Eduardo Pereira de Carvalho, president of UNICA, downloaded on November 17 2011 from website of Office of US Senator Richard G. Lugar*) (US Government Printing Office, 2006).
36. http://cbn.globoradio.globo.com/home/HOME.htm
37. http://www.inmet.gov.br/
38. http://cetsp1.cetsp.com.br/monitransmapa/agora/
39. Orlando, J. P., Alvim, D. S., Yamazaki, A., Corrêa, S. M. & Gatti, L. V. Ozone precursors for the São Paulo metropolitan area. *Sci. Tot. Environ.* **408,** 1612–1620 (2010).
40. Cohen, R. C. *et al.* Quantitative constraints on the atmospheric chemistry of nitrogen oxides: An analysis along chemical coordinates. *J. Geophys. Res.: Atmos.* **105,** 24283–24304 (2000).
41. Auffhammer, M. & Kellogg, R. Clearing the air? The effects of gasoline content regulation on air quality. *Am. Econ. Rev.* **101,** 2687–2722 (2011).
42. Hughes, J. E., Knittel, C. R. & Sperling, D. Evidence of a shift in the short-run price elasticity of gasoline demand. *Energy J.* **29,** 113–134 (2008).
43. Duranton, G. & Turner, M. A. The fundamental law of road congestion: Evidence from US cities. *Am. Econ. Rev.* **101,** 2616–2652 (2011).
44. Vickrey, W. S. Congestion theory and transport investment. *Am. Econ. Rev.* **59,** 251–260 (1969).
45. Holland, S. P., Hughes, J. E. & Knittel, C. R. Greenhouse gas reductions under low carbon fuel standards? *Am. Econ. J.-Econ. Polic.* **1,** 106–146 (2009).
46. Bento, A., Kaffine, D., Roth, K. & Zaragoza-Watkins, M. The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector. *Am. Econ. J.-Econ. Polic.* (in the press).
47. Goolsbee, A. & Petrin, A. The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* **72,** 351–381 (2004).

Acknowledgements

We gratefully acknowledge numerous people from CETESB, INMET, CET and ANP for generously sharing their data; CBN Notícias for sharing their newscasts; and Raízen for sharing access to their fuelling stations. In particular, we thank W. Baptista, C. Costa, A. Dall'Antonia Jr, F. Henkes, M. Kuromoto, C. Lacava, D.G. Medeiros, R.C. Melo, R. dos Santos and T. P. Senaubar. We thank T. Aguirre for summarizing descriptive CETESB and CPTEC weather reports, M. Peterson for preparing a literature review as well as an initial written and graphical description of the pollutant and meteorological data, and S. Ritchey for listing the GPS coordinates of road segments. We thank S. Budanova, E. Lehman and C. Maalouf for research assistance, as well as J. Brito, M. Busse, J. He, E. Mansur and S. McRae for helpful comments. A.S. acknowledges support from the Initiative for Sustainability and Energy at Northwestern University (ISEN) and from the Dean's Office at the Kellogg School of Management, Northwestern University. F.M.G. thanks the NSF Atmospheric and Geospace Science division for support under grant # NSF ATM-0533436 and gratefully acknowledges support from an Irving M. Klotz professorship in physical chemistry.

Author contributions

A.S. conceived the research; A.S. and F.M.G. analysed the data and wrote the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S.

Competing financial interests

The authors declare no competing financial interests.