PRE-PRINT DATE: 18 February, 2014

Reduction in Local Ozone Levels in Urban São Paulo Due to a Shift from Ethanol to Gasoline Use

Alberto Salvo* and Franz M. Geiger

Department of Economics, National University of Singapore and Department of Chemistry, Northwestern University

*Corresponding Author

It has been proposed that lower NO_X emission fuels such as ethanol can mitigate air pollution from vehicles burning oil-based hydrocarbons. Yet, existing modeling and laboratory studies, even those seeking to simulate the same environment, vary in their predictions of how gasoline/ethanol blends affect atmospheric pollutant concentrations, including ozone. Importantly, ambient concentrations have not been evaluated during an actual – as opposed to hypothetical – shift in fuel mix in a real-world environment. Here, we report the first such study, for the subtropical megacity of São Paulo, Brazil. We combine detailed street-hour level data on regulated pollutant concentrations, meteorology, and traffic with fuel shares from a consumer demand model to compare concentrations across subsamples that differ only in the fuel mix but are otherwise similar in meteorology, anthropogenic activity, and biogenic emissions. As the gasoline share of the bi-fuel light-duty vehicle fleet rose by 62 percentage points, we estimate a robust and statistically significant reduction of about 20% in ozone concentrations, and less precise increases in NO and CO concentrations. We propose that our "model-free" analysis potentially accounts for the interaction between anthropogenic and biogenic emissions and caution that successful strategies against ozone pollution require knowledge of the

DOI: 10.1038/NGEO2144
FOR FINAL PUBLISHED VERSION, SUPPLEMENTARY INFORMATION, AND DATA ARCHIVE, ACCESS:
http://www.nature.com/ngeo/journal/vaop/ncurrent/full/ngeo2144.html

local chemistry and analysis beyond the presently monitored pollutants, most notably fine particles.

The World's Largest Bi-Fuel Vehicle Fleet, Fuel Choice, and Air Quality

Ozone levels are relatively high in São Paulo, with hourly concentrations above 75 and 125 μg/m^3, respectively, 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 6 million active light-duty vehicles – likely accounting for over one-half of all light vehicle distance traveled – 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 traveled (see Supplementary Materials 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; further, evidence established herein indicates that 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 the Supplementary Materials 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 traveled 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 at 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 (e.g., isoprene) on top of anthropogenic ones[29-31]. Review articles underscore the need for data-based studies examining the air quality impacts of consumer adoption of alternative fuels and vehicles[32,33].

Beyond science but no less important to society due to its influence on human behavior, conventional wisdom appears to associate ethanol with improved environmental outcomes, including air quality (e.g., surveys of Brazilian ethanol consumers[12], comments by the ethanol industry at energy hearings in the US Senate[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 center.

Analysis of Concentrations, Traffic, and Meteorology at Street-Hour Level

Our study cross-examines a large amount of measured data, detailed at the street-hour level (Fig. 2), from several sources: (i) concentrations of regulated “priority” pollutants, namely O_3, NO, NO_2, and CO (including SO_2 and PM10 in the Supplementary Materials), measured by spatially differentiated air monitoring stations maintained by the

environmental authority of the state of São Paulo (CETESB)[2]; (ii) meteorological conditions measured at these same CETESB stations as well as recorded by the Institute for Meteorology (INMET)[37]; and (iii) controls for vehicle traffic congestion and speed obtained from the city traffic authority (CET)[38]. We combine the extensive pollutant-meteorology-traffic data with: (iv) weekly gasoline and ethanol prices at the pump, obtained from the National Agency for Oil, Biofuels and Natural Gas (ANP)[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 I. The table reports regression estimates for hourly O_3 concentrations measured in the early afternoon (13:00 to 16:00) on non-holiday (regular) weekdays. Since 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 (shaded) reports mean effects and precision across all 12 ozone monitor-level regressions. The fuel mix variables – the main variables of interest, see Methods below – are s_t^{gas}, the share of bi-fuel vehicles fueled 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 Materials 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. 3, 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. Fig. 4 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, etc. – are increasing in radiation and temperature and decreasing in 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 additionally controls for the total extension of traffic congestion (i.e., 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 hours preceding an observation. In addition to road congestion at the citywide level, traffic covariates in column VI now include two local measures, namely: (i) the sum of congestion only in the region of the city where the monitoring station is located (e.g., North in Fig. 2); and (ii) a weighted sum 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 wind is blowing (i.e., we include $f(W_t, T_t)$ in regression equation (1) below).

Ozone Reduction and NO and CO Increase with Shift into Gasoline

Across specifications IV to VII of Table I, the average fuel mix effect λ_1 (see (1)) on the ozone concentration is estimated at -20.7 to -30.2 μg/m^3, with standard errors (s.e.) of under 9 μg/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 μg/m^3 ($\hat{\lambda}_1$ averaged across the columns times 0.62). This 15 μg/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 μg/m^3. Fig. 5 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 5 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), likely due to the smaller magnitude of the ethanol-gasoline shift over this episode (see Supplementary Materials 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 (2) in Methods) 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 II presents the same analysis for NO, NO_2 and CO, based on concentrations measured in the morning rush hours (07:00 to 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 Materials).

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 to VII, a 62-percentage-point rise in the gasoline share is associated with increases of 17.3 μg/m^3 (s.e. 7.6 μg/m^3) and 0.22 ppm (s.e. 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. 5 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.

Fig. 6 offers an intuitive illustration of our method and of our result for ozone. Panel a 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 panel b 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 μg/m^3 – this is similar to the mean coefficient across station-specific regressions in column VI, Table 1.

In Supplementary Materials Part F we perform a placebo test and subject our baseline results to a number of additional 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 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 and Ozone Abatement

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 – though larger than the one we observe in the data – is of comparable magnitude. Our joint data 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 μg/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 modeling 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 while that of ozone falls upon 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:

$$concentrat\,ion_t = \lambda_1 s_t^{gas} + \lambda_2 e20_t^{gas} + W_t'\Delta^W + T_t'\Delta^T + fixedeffec\,ts_t + trend_t + \varepsilon_t. \qquad (1)$$

An observation *t* is an hour-date pair, e.g., for the Diadema station, early afternoon (13:00 to 16:00), non-holiday weekday regression, an observation is 14:00 on Monday, March 14, 2011 (this was not a public holiday). The dependent variable $concentration_t$ corresponds to a pollutant that is measured at the station, e.g., O_3, in the measured units (μg/m^3) or a logarithmic transform thereof. Both fuel mix variables, s_t^{gas} and $e20_t^{gas}$, increase in the proportion of gasoline, though the shift from ethanol to gasoline as $e20_t^{gas}$

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 Materials, 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 (March 14, 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:

$$E[s_t^{gas}\varepsilon_t \mid X_t] = 0, \text{ where } X_t := (W_t, T_t, fixedeffects_t, trend_t). \qquad (2)$$

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, while quite predictable, did not significantly vary with fuel prices during the sample period. This finding can be rationalized on different counts, namely: (i) 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 (see Supplementary Materials Part E); (ii) the existence of "repressed

demand" for vehicle usage that has been argued in the face of widespread gridlock[43,44]; and (iii) 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 fueling their bi-fuel vehicles with the fuel that yields the lowest $ per distance traveled. By this assumption, consumers would switch from ethanol to gasoline, $s_t^{gas} = 1$, whenever the per-liter price of ethanol surpassed around 70% of the per-liter 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 behavior 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. Fig. 3 reports how the gasoline share, s_t^{gas}, varies in the sample: (i) (panel a) with the per-liter ethanol-to-gasoline price ratio – notice that there is no kink at the approximate 70% "parity" threshold, at which $/mile traveled on either fuel is about the same; and (ii) (panel b) over time (see Supplementary Materials Part A for demand modeling and estimation).

Figure Captions

Fig. 1. Shifting fuel quantities and prices between November 2008 and July 2011. a, The monthly share of blended gasoline purchased at retail (E25 or E20) of total estimated light-vehicle distance traveled, prepared from wholesale shipment reports. **b,** The weekly per-liter price of regular ethanol (E100), denoted p_e, divided by the per-liter price of regular blended gasoline, denoted p_g, at pumps in the city of São Paulo (left ordinate). Shades of grey indicate the ranges of the 5th-25th, 25th-50th (indicated by the black curve), 50th-75th, and 75th-95th percentile of the distribution of p_e/p_g across retailers, as well as the 70% "parity" threshold widely reported by the media, at which $/kilometer equalizes (dashed blue horizontal line). Right ordinate 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, authors' calculations.

Fig. 2. Street-hour level data on pollutant concentrations, meteorology and traffic congestion. a, Map of the environmental authority's air monitoring stations, which often double as weather stations, in the São Paulo metropolitan area, superimposed on the road network, monitored every 30 minutes 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 January 31 to February 6, 2011. Sources: CETESB, INMET, CET, ANP.

Fig. 3. Gradual transitions between ethanol and gasoline in bi-fuel vehicles. a, In-sample variation in the median per-liter 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 (ref. 12, specification III, Table 2) 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. Sources: ANP, ref. 12, DETRAN-SP, Fenabrave, Sindipeças, authors' estimates.

Fig. 4. Variation in the gasoline share around a linear or quadratic trend. a, s_t^{gas} plotted against date in the sample (November 1, 2008 to May 31, 2011, excluding the colder winter months from June to September). **b**, Residuals of a regression of s_t^{gas} on a linear trend plotted against date. The black line in each panel denotes the best linear predictor (the best quadratic predictor lies on top of its linear counterpart).

Fig. 5. 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 the different times of 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 (Tables I and II). See the footnote to Table I on how the error bars were obtained.

Fig. 6. Intuitive illustration of the method. Partialling out the effect of meteorological and traffic conditions, seasonality, trending omitted factors, etc. on O_3 concentrations to identify the effect of the fuel mix. The panels plot, for all 12 O_3-monitoring stations in the early afternoon hours on non-holiday weekdays: **a**, measured O_3 concentrations in μg/m^3 against the gasoline share s_t^{gas}; and **b**, 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. Color intensity indicates local density in each of 128×128 bins. For panel b, 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. Source: Specification VI (with station fixed effects included).

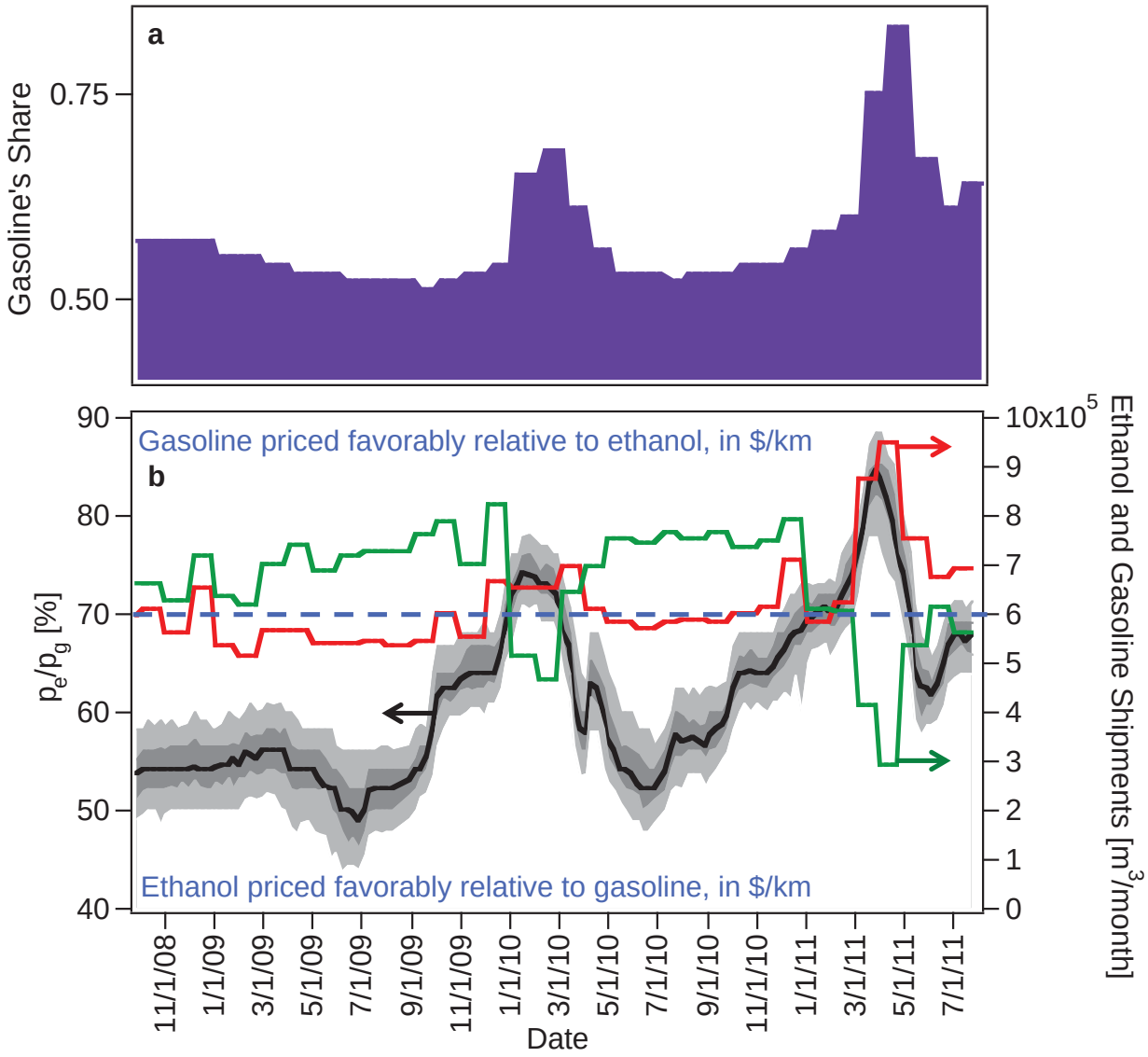
a
Gasoline's Share
0.75
0.50
b
Gasoline priced favorably relative to ethanol, in $/km
Ethanol priced favorably relative to gasoline, in $/km
p_e/p_g [%]
90
80
70
60
50
40
10x10^5
9
8
7
6
5
4
3
2
1
0
Ethanol and Gasoline Shipments [m^3/month]
11/1/08
1/1/09
3/1/09
5/1/09
7/1/09
9/1/09
11/1/09
1/1/10
3/1/10
5/1/10
7/1/10
9/1/10
11/1/10
1/1/11
3/1/11
5/1/11
7/1/11
Date

Fig. 1

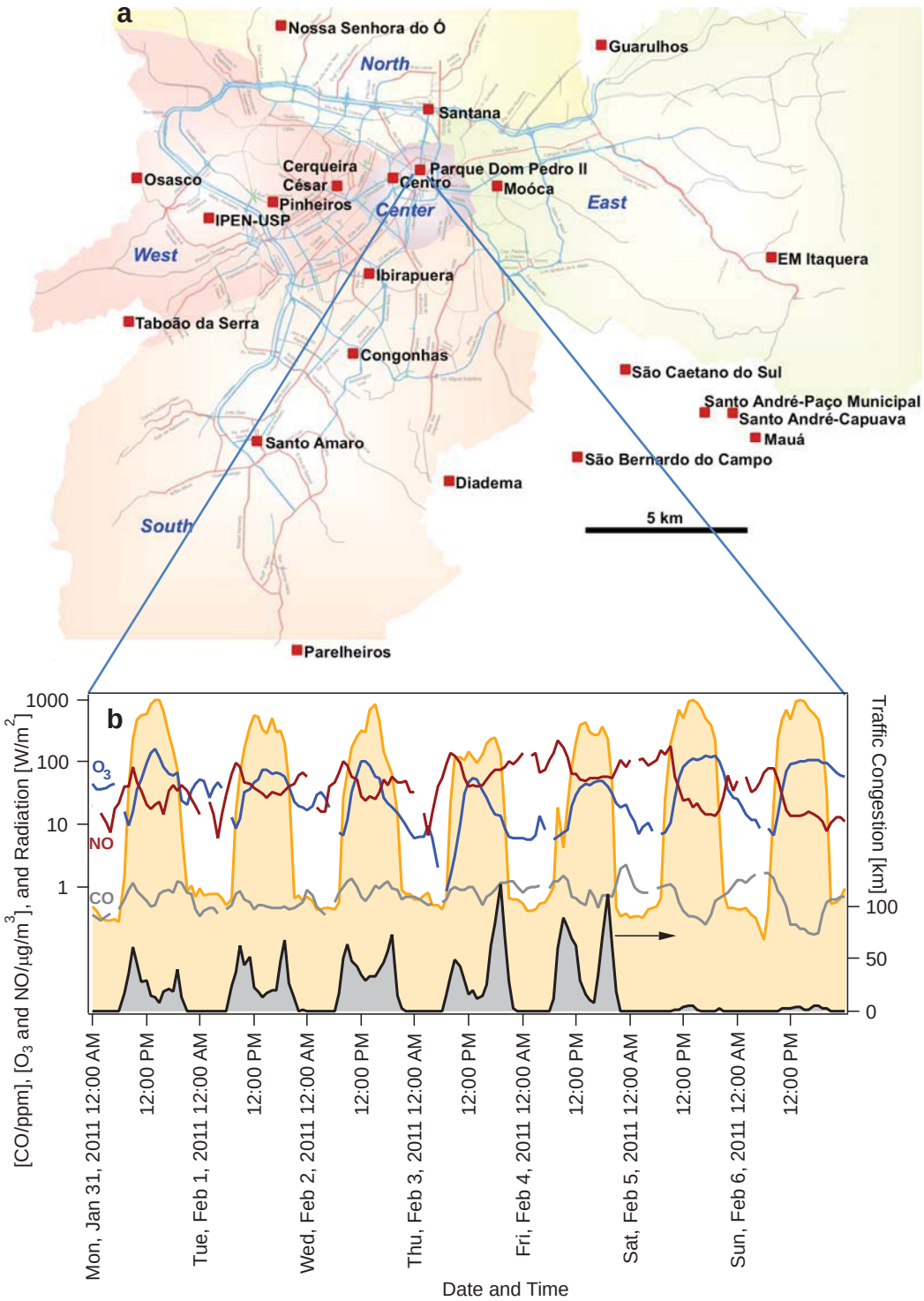
a
Nossa Senhora do Ó
Guarulhos
North
Santana
Cerqueira César
Parque Dom Pedro II
Centro
Moóca
Osasco
Pinheiros
IPEN-USP
Center
East
West
EM Itaquera
Ibirapuera
Taboão da Serra
Congonhas
São Caetano do Sul
Santo André-Paço Municipal
Santo André-Capuava
Mauá
Santo Amaro
São Bernardo do Campo
Diadema
5 km
South
Parelheiros
b
1000
100
10
1
O3
NO
CO
[CO/ppm], [O3 and NO/μg/m3], and Radiation [W/m2]
Traffic Congestion [km]
100
50
0
Mon, Jan 31, 2011 12:00 AM
12:00 PM
Tue, Feb 1, 2011 12:00 AM
12:00 PM
Wed, Feb 2, 2011 12:00 AM
12:00 PM
Thu, Feb 3, 2011 12:00 AM
12:00 PM
Fri, Feb 4, 2011 12:00 AM
12:00 PM
Sat, Feb 5, 2011 12:00 AM
12:00 PM
Sun, Feb 6, 2011 12:00 AM
12:00 PM
Date and Time

Fig. 2

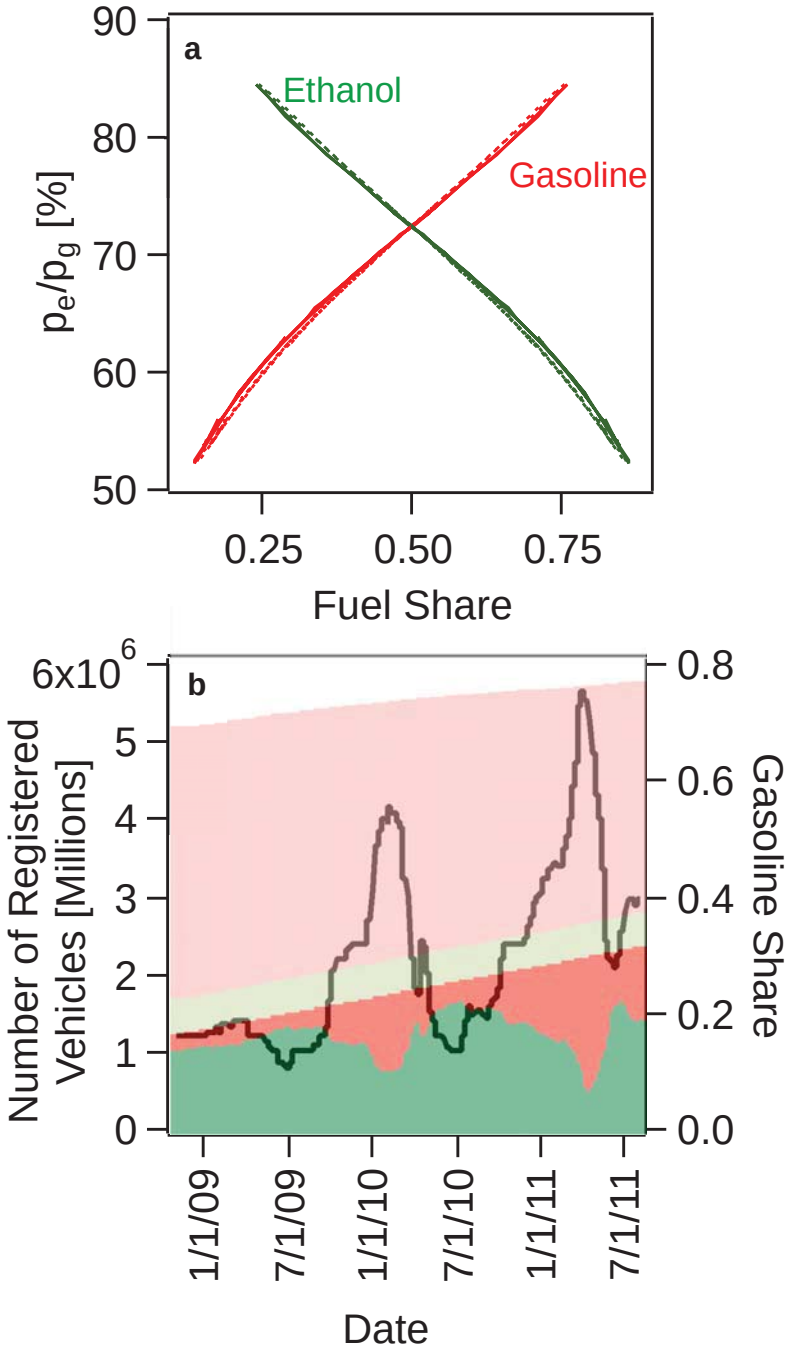
a
Ethanol
Gasoline
p_e/p_g [%]
90
80
70
60
50
0.25
0.50
0.75
Fuel Share
b
$6x10^6$
5
4
3
2
1
0
Number of Registered Vehicles [Millions]
0.8
0.6
0.4
0.2
0.0
Gasoline Share
1/1/09
7/1/09
1/1/10
7/1/10
1/1/11
7/1/11
Date

Fig. 3

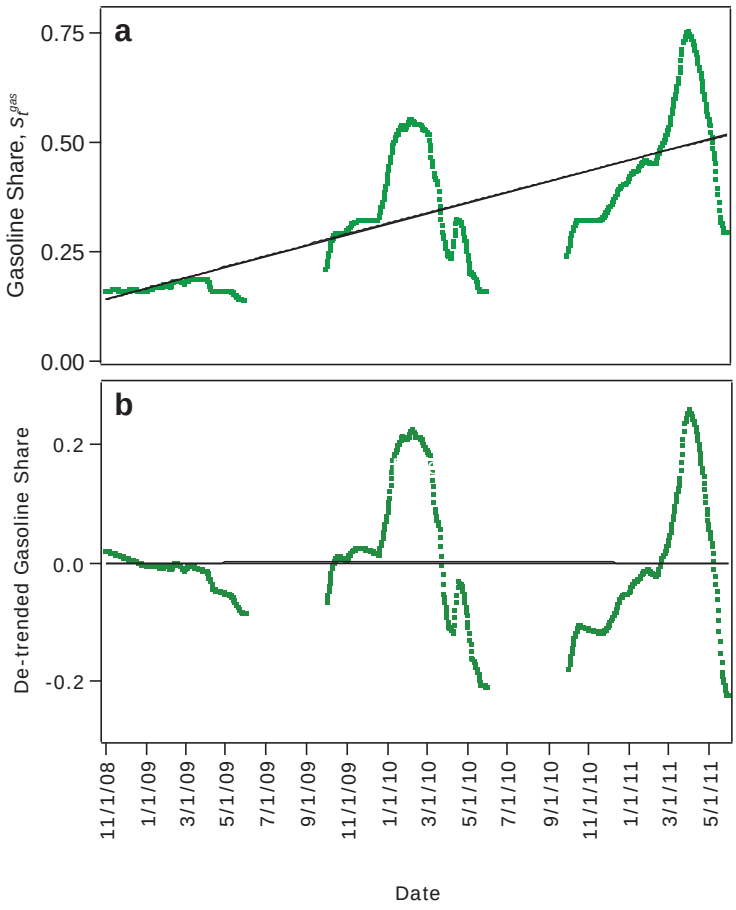
a
b
0.75
0.50
0.25
0.00
Gasoline Share, s_t^{gas}
0.2
0.0
-0.2
De-trended Gasoline Share
11/1/08
1/1/09
3/1/09
5/1/09
7/1/09
9/1/09
11/1/09
1/1/10
3/1/10
5/1/10
7/1/10
9/1/10
11/1/10
1/1/11
3/1/11
5/1/11
Date

Fig. 4

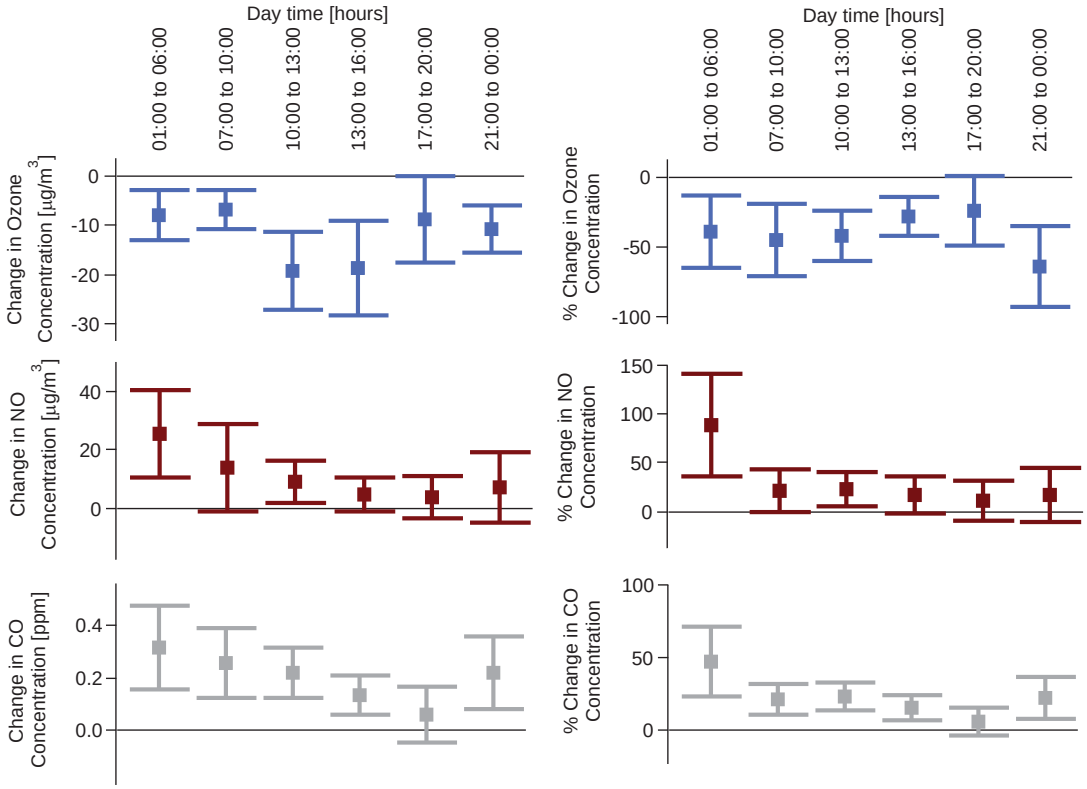
Day time [hours]
01:00 to 06:00
07:00 to 10:00
10:00 to 13:00
13:00 to 16:00
17:00 to 20:00
21:00 to 00:00
Change in Ozone Concentration [µg/m3]
0
-10
-20
-30
Change in NO Concentration [µg/m3]
40
20
0
Change in CO Concentration [ppm]
0.4
0.2
0.0
Day time [hours]
01:00 to 06:00
07:00 to 10:00
10:00 to 13:00
13:00 to 16:00
17:00 to 20:00
21:00 to 00:00
% Change in Ozone Concentration
0
-50
-100
% Change in NO Concentration
150
100
50
0
% Change in CO Concentration
100
50
0

Fig. 5

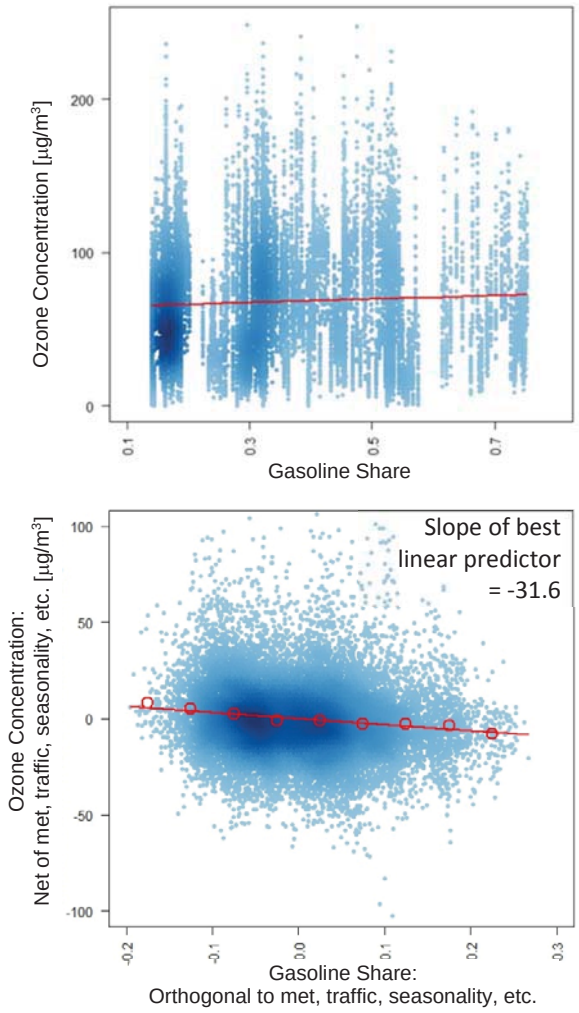
Ozone Concentration [µg/m³]
200
100
0
0.1
0.3
0.5
0.7
Gasoline Share
Ozone Concentration:
Net of met, traffic, seasonality, etc. [µg/m³]
100
50
0
-50
-100
-0.2
-0.1
0.0
0.1
0.2
0.3
Slope of best
linear predictor
= -31.6
Gasoline Share:
Orthogonal to met, traffic, seasonality, etc.

Fig. 6

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 Pollution* **195**, 201-213, doi:10.1007/s11270-008-9740-x (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] Portuguese. (Companhia Ambiental do Estado de São Paulo, Sao 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] Portuguese. (Centro de Estudos Integrados sobre Meio Ambiente e Mudanças Climáticas, Universidade Federal do Rio de Janeiro, São Paulo, 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, doi:DOI 10.1111/j.1539-6924.2011.01630.x (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, doi:Doi 10.1016/S0140-6736(07)60037-3 (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. *American journal of epidemiology* **162**, 140-148, doi:10.1093/aje/kwi173 (2005).

7 Currie, J. & Walker, R. Traffic Congestion and Infant Health: Evidence from E-ZPass. *Am Econ J-Appl Econ* **3**, 65-90, doi:Doi 10.1257/App.3.1.65 (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, doi:DOI 10.1126/science.1137013 (2007).

10 *Please see http://www.anp.gov.br/preco/prc/Resumo_Por_Municipio_Index.asp*.

11 *Please see 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 Conversion and Management* **44**, 1547-1561, doi:10.1016/s0196-8904(02)00166-8 (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. *Atmospheric Environment* **42**, 4498-4516, doi:10.1016/j.atmosenv.2008.01.061 (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. *Atmospheric Environment* **37**, 949-957, doi:10.1016/s1352-2310(02)00973-1 (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.

Atmospheric Environment **36**, 403-410, doi:10.1016/s1352-2310(01)00508-8 (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. *Journal of Hazardous Materials* **123**, 29-34, doi:10.1016/j.jhazmat.2005.03.046 (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. *Atmospheric Environment* **36**, 3495-3503, doi:10.1016/s1352-2310(02)00288-1 (2002).

19 Lynd, L. R. Overview and Evaluation of Fuel Ethanol from Cellulosic Biomass: Technology, Economics, the Environment, and Policy. *Annual Review of Energy and the Environment* **21**, 403-465, doi:doi:10.1146/annurev.energy.21.1.403 (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. *Environmental Science and Technology* **31**, 1302-1307, doi:10.1021/es960514r (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. *Atmospheric Environment* **35**, 4399-4406, doi:10.1016/s1352-2310(01)00248-5 (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. *Renewable Energy* **31**, 2534-2542, doi:10.1016/j.renene.2006.01.004 (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. *Journal of Automobile Engineering* **223**, 941 (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. *Journal of the Brazilian Chemical Society* **15**, 646-651 (2004).

25 Martins, L. D. A., M.F.;. Emission Scenario Assessment of Gasohol Reformulation Proposals and Ethanol Use in the Metropolitan Area of São Paulo. *The Open Atmospheric Science Journal* **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. *Atmospheric Environment* **44**, 1192-1199, doi:10.1016/j.atmosenv.2009.12.024 (2010).

27 Jacobson, M. Z. Effects of Ethanol (E85) versus Gasoline Vehicles on Cancer and Mortality in the United States. *Environmental Science and Technology* **41**, 4150-4157, doi:10.1021/es062085v (2007).

28 Beer, T. *et al.* The Health Impacts of Ethanol Blend Petrol. *Energies* **4**, 352-367, doi:10.3390/en4020352 (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 University Press, 1999).
32 Anderson, L. G. Ethanol fuel use in Brazil: air quality impacts. *Energ Environ Sci* **2**, 1015-1037, doi:Doi 10.1039/B906057j (2009).
33 Romieu, I., Weitzenfeld, H. & Finkelman, J. Urban Air-Pollution in Latin-America and the Caribbean. *J Air Waste Manage* **41**, 1166-1171 (1991).
34 *Please see http://www.gpo.gov/fdsys/pkg/CHRG-106shrg71515/html/CHRG-106shrg71515.htm*.
35 *Please see http://lugar.senate.gov/energy/hearings/pdf/060622/Carvalho_slides2.pdf*.
36 *Please see http://cbn.globoradio.globo.com/home/HOME.htm*.
37 *Please see http://www.inmet.gov.br/*.
38 *Please see 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. *Science of the Total Environment* **408**, 1612-1620 (2010).
40 Cohen, R. C. *et al.* Quantitative constraints on the atmospheric chemistry of nitrogen oxides: An analysis along chemical coordinates. *Journal of Geophysical Research: Atmospheres* **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, doi:DOI 10.1257/aer.101.6.2687 (2011).
42 Hughes, J. E., Knittel, C. R. & Sperling, D. Evidence of a shift in the short-run price elasticity of gasoline demand. *Energ 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, doi:DOI 10.1257/aer.101.6.2616 (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, doi:Doi 10.1257/Pol.1.1.106 (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*, doi:Doi 10.1257/Pol.1.1.106 (Forthcoming).
47 Goolsbee, A. & Petrin, A. The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* **72**, 351-381, doi:DOI 10.1111/j.1468-0262.2004.00494.x (2004).

Acknowledgments.
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 fueling stations. In particular, we thank Wagner Baptista, Cristina Costa, Alaor Dall'Antonia Jr, Fábio Henkes, Masayuki Kuromoto, Carlos Lacava, Dario Garcia Medeiros, Rui César Melo, Roseni dos Santos, and Telma Paulino Senaubar. We thank Thomas Aguirre for summarizing descriptive CETESB and CPTEC weather reports, Mark Peterson for preparing a literature review as well as an initial written and graphical description of the pollutant and meteorological data, and Sam Ritchey for listing the GPS coordinates of road segments. We thank Sofya Budanova, Elizabeth Lehman, and Charles Maalouf for research assistance, as well as Joel Brito, Meghan Busse, Jiaxiu He, Erin Mansur, and Shaun McRae for helpful comments. AS 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. FMG 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 contribution statement
AS conceived of the research; AS and FMG analyzed the data and wrote the paper.

Table I. Predicting Ozone (μg/m^3), non-holiday weekday, 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**	**26.1**	**-14.9**	**-20.7**	**-20.1**	**-23.2**	**-18.8**
	(11.5)	**(15.8)**	**(20.1)**	**(12.5)**	**(12.4)**	**(12.9)**	**(12.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.1**	**-1.6**	**-3.0**	**1.3**	**1.4**	**5.0**	**2.3**
	(4.4)	**(4.4)**	**(4.9)**	**(2.6)**	**(2.7)**	**(2.7)**	**(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	0.0		0.0
				(0.4)	(0.5)		(0.4)
Humidity				-36.5	-36.6		-37.6
				(6.7)	(6.7)		(6.9)
Radiation				10.6	10.6		9.8
				(1.9)	(1.9)		(1.9)
Temperature				89.6	89.4		93.8
				(12.9)	(12.9)		(13.0)
Wind speed				-11.9	-11.8		-10.2
				(3.6)	(3.6)		(3.5)
Citywide traffic congestion: contemporaneous conditions	No	No	No	No	Yes	Yes	Yes
Total extension of congestion across city					-0.7		1.2
					(1.8)		(2.1)
Meteorology: conditions lagged up to 18 hours	No	No	No	No	No	Yes	No
Meteorology: Pairwise interactions of contemporan.cond.	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 hours	No	No	No	No	No	Yes	No
Interactions of wind direction & 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	1414	1414	1414	1401	1401	1371	1397
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**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Selected meterology and traffic (mean estimates across 12 station-specific regressions)							
Precipitation				0.2	0.2		0.2
				(0.4)	(0.4)		(0.4)
Humidity				-46.5	-45.9		-45.3
				(4.7)	(4.7)		(4.7)
Radiation				7.3	7.5		6.8
				(1.2)	(1.2)		(1.2)
Temperature				85.2	86.0		91.7
				(9.0)	(9.0)		(9.4)
Wind speed				-12.7	-12.7		-12.4
				(2.4)	(2.4)		(2.3)
Total extension of congestion across city					2.1		1.8
					(1.2)		(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	1574	1574	1574	1560	1560	1531	1515
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

Notes: The top panel reports coefficients and standard errors (in parentheses) for one of the station-specific regressions. The bottom panel (shaded) reports means for selected effects across regressions for each of the 12 ozone-monitoring stations. Standard errors are calculated by bootstrapping (200 samples each): (i) the consumer-level fuel choice data, to account for sampling variation in the gasoline share, and (ii) 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 November 1, 2008 to May 31, 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.

Table II. Predicting NO (μg/m^3), NO_2 (μg/m3), and CO (ppm), non-holiday weekday, 07:00 to 10:00 readings only

Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Dependent variable: NO concentration (μg/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**	**3.6**	**45.6**	**29.2**	**28.2**	**22.7**	**31.7**
	(12.6)	**(19.7)**	**(21.1)**	**(12.2)**	**(12.3)**	**(12.4)**	**(11.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**6.7**	**6.1**	**8.3**	**11.8**	**11.8**	**9.7**	**11.0**
	(4.5)	**(4.5)**	**(5.5)**	**(3.8)**	**(3.8)**	**(4.0)**	**(3.9)**
Selected meterology and traffic (mean estimates across 9 station-specific stations)							
Precipitation				-0.4	-0.4		-0.5
				(0.4)	(0.5)		(0.5)
Humidity				-19.6	-19.7		-18.1
				(11.6)	(11.7)		(11.0)
Radiation				3.5	3.6		3.5
				(1.1)	(1.1)		(1.1)
Temperature				-17.0	-17.0		-25.7
				(13.8)	(13.8)		(13.8)
Wind speed				-52.3	-52.3		-51.2
				(3.2)	(3.2)		(3.2)
Total extension of congestion across city					1.4		1.8
					(2.9)		(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	1529	1529	1529	1514	1514	1489	1464
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 (μg/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**	**-2.6**	**5.0**	**-5.3**	**-6.2**	**5.1**	**-5.8**
	(3.6)	**(5.0)**	**(6.4)**	**(4.7)**	**(4.7)**	**(4.4)**	**(4.8)**
Three month period with gasoline E20, $e20_t^{gas}$	**3.3**	**3.3**	**3.7**	**4.6**	**4.5**	**-0.7**	**4.5**
	(1.7)	**(1.7)**	**(1.9)**	**(1.3)**	**(1.3)**	**(1.2)**	**(1.4)**
Selected meterology and traffic (mean estimates across 9 station-specific stations)							
Precipitation				0.3	0.3		0.2
				(0.2)	(0.2)		(0.2)
Humidity				-28.2	-28.3		-26.7
				(4.7)	(4.7)		(4.7)
Radiation				1.3	1.3		1.4
				(0.5)	(0.5)		(0.5)
Temperature				33.6	33.6		31.2
				(4.7)	(4.8)		(4.8)
Wind speed				-10.4	-10.4		-9.9
				(0.8)	(0.8)		(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	1529	1529	1529	1514	1514	1489	1464
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**	**0.35**	**0.66**	**0.34**	**0.33**	**0.41**	**0.34**
	(0.13)	**(0.21)**	**(0.22)**	**(0.12)**	**(0.12)**	**(0.11)**	**(0.12)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.08**	**0.07**	**0.10**	**0.17**	**0.17**	**0.11**	**0.16**
	(0.05)	**(0.05)**	**(0.06)**	**(0.04)**	**(0.04)**	**(0.04)**	**(0.04)**
Selected meterology and traffic (mean estimates across 11 station-specific stations)							
Precipitation				0.00	0.00		0.0
				(0.01)	(0.01)		(0.01)
Humidity				0.16	0.16		0.20
				(0.12)	(0.12)		(0.12)
Radiation				0.06	0.06		0.06
				(0.01)	(0.01)		(0.01)
Temperature				0.70	0.70		0.63
				(0.14)	(0.14)		(0.14)
Wind speed				-0.49	-0.49		-0.48
				(0.03)	(0.03)		(0.03)
Total extension of congestion across city					0.01		0.01
					(0.03)		(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	1565	1565	1565	1548	1548	1523	1505
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

Notes: See notes to Table I. 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.

Supplementary Materials (SM) for

Reduction in Local Ozone Levels in Urban São Paulo Due to a Shift from Ethanol to Gasoline Use

Alberto Salvo* and Franz M. Geiger

Department of Economics, National University of Singapore and Department of Chemistry, Northwestern University

This file includes:

Materials and Methods

A. Fuel prices at the pump and consumers' choice of fuel.
B. Studies involving tailpipe emissions, smog chambers, and computer models.
C. Description of air monitoring stations.
D. Meteorology.
E. Vehicle traffic.
F. Further details on methods, results, and robustness.

Figs. S1 to S12

Tables SI to SXVII

References to the Supplementary Materials

(NOT FOR PUBLICATION) Other Supplementary Materials, to be made available on the authors' webpages once the embargo is lifted, include:

Appendix, containing estimates detailed by pollutant-station pair, along with further supplementary figures

A. Fuel prices at the pump and consumers' choice of fuel.

1. Fuel prices and blends at the pump in the São Paulo metropolitan area. In recent years, supply-side economic shocks – such as a poor sugarcane harvest in India in late 2009 and another hike in the world sugar price in early 2011[1-3] – have led to large fluctuations in the consumer price of sugarcane ethanol (E100), a motor fuel that is widely retailed across Brazil. Fig. S1 shows the price of *etanol hidratado comum*, "regular hydrated ethanol," at São Paulo's pumps over some weeks in our sample.[4] We adjust the time series for inflation so that prices in 2010 and in 2011 are comparable in terms of consumer purchasing power.[5]

By contrast to ethanol prices, which are deregulated, prices for the substitute fuel, gasoline, are in effect controlled by the government via wholesale prices at the refinery. Fig. S1 also shows prices of *gasolina C comum*, "regular gasoline C," over the same weeks. Over the sample period, the central government largely held gasoline prices constant in nominal terms (Fig. S2 below).[6] To the extent that gasoline prices at the pump varied a little, this reflects the fact that gasoline retailed in Brazil contains a 20–25% "anhydrous" ethanol component by volume (E20 or E25). For perspective, between November 1, 2009 and February 7, 2010, ethanol prices rose 23%, from 1.66 to 2.04 R$/liter, compared with gasoline prices rising only 2% (median, inflation-adjusted prices). Ethanol price variation was even more pronounced the following year. From November 7, 2010 to April 17, 2011, ethanol prices rose 33%, from 1.67 to 2.22 R$/liter, with gasoline prices rising 6%.

These prices are based on large weekly surveys that were representative of the population of retailers in the city of São Paulo, with a median of 349 retail outlets ("gas stations") sampled per week and a minimum of 261. The surveys also indicate that the availability of both ethanol and gasoline was ubiquitous. Across the weekly samples, the distribution of the number of retail outlets which: (i) did not carry regular ethanol has a median of 0 (a maximum of 2 out of around 350 outlets had run out of ethanol in one weekly survey), and (ii) did not carry regular gasoline has a median of 0 (at most 4 outlets were observed without gasoline). In 2010, most retailers also sold "midgrade gasoline," *gasolina C aditivada*, at about a 5% markup to regular gasoline, and only a few further sold "premium gasoline," *gasolina C premium*.[7] "Midgrade ethanol," *etanol hidratado aditivado*, was introduced in 2011 by selected retail outlets in upmarket neighborhoods.[8]

The federal blending mandate that dictates the ethanol content in the gasoline ("gasohol") fuel that is available to consumers has applied equally to regular gasoline and to midgrade and premium gasoline varieties. This blending requirement varied temporarily during our sample period, a shift we also exploit in our analysis. Specifically, starting on February 1, 2010, distributors were required to shift to E20, from E25 earlier, on gasoline shipments to retailers, but shifted back to gasoline E25 from May 1, 2010 on.[9] To be clear, consumers did not face a choice between E25 and E20 at the pump: vehicles running on any *gasolina C* purchased in January 2010 were burning E25, whereas those operating on the fuel purchased under the same name a month later were burning E20.

Fig. S2 reports almost identical price variation, from another source and covering the entire sample period, for gasoline and ethanol in the São Paulo metropolitan area.[10] The figure plots the monthly price index (base price index in October 2008 = 100). We include a price index for diesel oil, a fuel that was almost exclusively used in heavy-duty vehicles such as trucks and buses (historically, Brazil's government severely limited the penetration of diesel in the light-duty vehicle fleet). After a 5% price adjustment in mid 2009, diesel prices stayed constant in nominal (inflation-unadjusted) terms and gradually declined in real (inflation-adjusted) terms. This situation suggests that potentially confounding effects on air quality through variation in heavy-duty vehicle traffic – driven by variation in diesel prices – around the time of each ethanol price hike are of lesser concern. Our baseline pollutant regressions allow for a linear or quadratic trend in the date, and in robustness tests we add the price of diesel as well as ridership on the public transport system.

It is important to emphasize that neither changes to the prices consumers paid for fuels nor changes to blending requirements were driven by concerns over air quality. We model both the ethanol price increases, and the political reaction by which blended gasoline was changed from E25 to E20 and then back to E25, as being exogenous to air pollution in the São Paulo metropolis. As we show in SM Part E, there is also little evidence that fuel price variation impacted vehicle usage, as proxied by measured road congestion and speeds in the city of São Paulo. Similarly, public transport ridership data, also reported in SM Part E, suggest that motorists did not noticeably substitute into public transport as ethanol prices rose.

As for channels other than transportation that might have influenced pollution as ethanol prices varied, we note that the sugar industry accounts for a small fraction of the diversified

economy of the state of São Paulo, let alone that of its capital city. One estimate put the country's entire sugar-ethanol sectoral Gross Domestic Product (GDP) at US$ 48 billion,[11] compared to a GDP for Brazil of US$ 2.5 trillion,[12] with the state of São Paulo accounting for one-third of national GDP, i.e., US$ 840 billion.[13] Thus, stronger ethanol (and sugar) prices are unlikely to have spilled over, via an income effect, to increased spending and emissions in other sectors, which could otherwise potentially confound our research design. In a robustness test, we include an index for industrial activity. Moreover, ethanol is not an input to energy-consuming sectors other than personal transportation, so stronger ethanol prices would not have dampened non-transportation activity.

2. São Paulo's active vehicle stock: Size and composition by fuel type. The number and composition of vehicles circulating in the São Paulo metropolitan area are estimated with some degree of uncertainty. While registration data for new vehicles are fairly reliable, vehicle usage and scrappage rates by vintage are only rough estimates. The state of São Paulo's Department of Motor Vehicles (DETRAN-SP) estimated the active fleet in the state's capital city in July 2011 at 5.9 million light-duty vehicles (passenger vehicles including sport utility vehicles, minivans and light pickup trucks), 0.9 million two-wheelers and 0.3 million heavy-duty vehicles (trucks and buses).[14] These figures may be inflated to the extent that the assumed rates of vehicle scrappage (through ageing, collision, and theft followed by dismantling) are understated, as suggested by observed cohort-specific sales of auto-parts.[15] This overstating of the active vehicle stock is likely to be more severe for the older single-fuel vehicles than for newer bi-fuel vehicles (more below).

An engineer at São Paulo's traffic authority (CET) estimated that only one-third of the active fleet circulated on any one day.[16] It has been argued that widespread gridlock leads to "repressed demand," such that expansion to road capacity – or higher fuel prices – would not necessarily relieve traffic congestion.[17] This argument is consistent with the "Fundamental Law of Road Congestion."[18] Indeed, as we show in SM Part E, we do not find evidence that higher ethanol prices relieved traffic congestion.

An estimate of the composition of the light-duty vehicle stock for the state of São Paulo at the end of 2008 was: (i) 30% bi-fuel gasoline-ethanol, such vehicles having been introduced with much success by automakers in 2003; (ii) 57% gasoline-captive, such vehicles were sold

primarily prior to 2005; and (iii) 12% ethanol-captive, also of pre-2005 vintage. The penetration of natural gas in São Paulo was minimal.[6] With bi-fuel engines – known as "flexible fuel" – accounting for nine-tenths of new sales,[6] and considering vehicle scrappage rates estimated by the auto-parts industry trade association,[15] our estimate for the penetration of bi-fuel engines in the light vehicle stock in 2011 is about 40%. The sugar industry trade association UNICA put this penetration as high as 51% across the country.[11] Importantly, relative to single-fuel vehicles, bi-fuel vehicles were on average newer, owned by wealthier households, and likely to be used more intensively, thus pushing their probable share of total light-vehicle distance traveled within the São Paulo metropolitan area to over (if not well over) 50% by 2011. For perspective, a government study assumes the usage of new vehicles to be 10 times that of 30-year-old vehicles remaining in operation (20,000 kilometers against 2,000 km per year).[19] Available data for the United States also indicate that average usage (vehicle miles traveled) declines significantly with vintage.[20,21] Since the relative usage of single-fuel vehicles is unknown, in our regression analysis we effectively interpret their largely gasoline consumption as a "background level" of emissions; recall that gasoline prices varied substantially less and that we account for trends.

Two-wheelers with bi-fuel gasoline-ethanol capability were introduced in the market only in 2009. Though the share of these motorcycles was growing, over the sample period the majority of motorcycles were single-fuel, operating on gasoline. Our conjecture that by 2011 bi-fuel vehicles may have accounted for half of overall distance traveled by light-duty vehicles (four-wheelers) is supported by back-of-the-envelope calculations that exclude gasoline-only two-wheelers from the gasoline versus ethanol shares of wholesale shipments that we report below. Finally, the vast majority of trucks and buses were single-fuel, predominantly operating on diesel.

3. Fuel economy, effective fuel prices in $ per distance traveled, and "price parity." As shown in Fig. 1 of the main text, the ethanol-to-gasoline (per-volume) price ratio peaked at 71%-78% in January 2010 (the first and ninth deciles across the sample of retailers) and at a higher 78%-88% in March 2011. On each occasion, ethanol prices rose to a level at which they stood at a substantial premium relative to gasoline, in terms of $ per km traveled, starting from a level only a few months earlier at which they were substantially discounted relative to gasoline.

Similarly, on both occasions, ethanol prices returned to levels at which they were again substantially discounted relative to gasoline, only a few months after peaking.

Following U.S. EPA guidelines, dynamometer-based laboratory measures of fuel economy (energy efficiency) available from the National Institute for Metrology indicate that locally sold bi-fuel vehicles when operated under either "urban" or "highway" cycles averaged a distance around 30% less on a liter of ethanol than on a liter of gasoline.[22,23] For example, the popular "Fiat Palio ELX 1.0 2010 Flex," when new and driven in the city, reportedly produced 6.9 km/liter running on ethanol E100 against 9.9 km/liter of gasoline E22. That is, taking k to denote km per liter and its subscript to denote the fuel, $k_e/k_g \approx 70\%$ for this vehicle model. The mean k_e/k_g across 67 tested vehicles, of varying segments, makes, models and versions, was 67.7% under the urban cycle, with a standard deviation of 1.9%. Using the same lab data, the mean predicted value for k_e/k_g across a market-share-weighted sample of 2160 bi-fuel vehicles, adjusting for the exact composition of gasoline (mostly E25) on different dates in early 2010, was estimated at 68.7%, with a standard deviation of 1.6%.[7]

Consistent with fuel economy measurements, São Paulo's media, including the radio which the city's motorists typically tuned into for traffic updates, routinely informed them that ethanol and gasoline prices were effectively equalized, in R$ per km, when the ethanol price per liter divided by the gasoline price per liter reached 70%, i.e., $p_e/p_g \approx 70\%$. By way of illustration, audio files from the leading *Rádio CBN Notícias*, with content aired during the sample period, are available from the authors upon request.[24] In addition to the media, fuel retailer attendants – who typically fueled vehicles in Brazil – were available to provide advice to motorists with regard to the competitively priced fuel, gasoline or ethanol, at the pump on the day. Phone-based interview evidence suggests that this media-reported 70% "price parity" threshold was recalled by a substantial proportion of the relevant consumer population.[7]

4. Consumer choice between gasoline and ethanol. In light of the above information, it is not surprising that most drivers of bi-fuel vehicles, who had the ability to switch between gasoline and ethanol as relative prices fluctuated, indeed *did* switch. This switch-over from ethanol to gasoline and back to ethanol over the course of 2009/10, and again during 2010/11, can be seen in the aggregate fuel shipments reported by wholesalers, shown in Fig. 1 of the main text. By aggregate, we mean monthly shipments throughout the wider state of São Paulo. Less aggregated

weekly quantity data for retailers in the city of São Paulo are not available – in contrast to price data. We return to wholesale quantities in the next subsection, as robustness tests of our results use an alternative measure of fuel mix variation calculated from such aggregate reports.

Our baseline pollutant regressions use consumer shares predicted by Salvo and Huse's multinomial probit choice model.[7,25] The gasoline share among bi-fuel consumers, s_t^{gas}, is based on Salvo and Huse's surveyed distribution of consumer and vehicle characteristics for the city of São Paulo, as well as variation in fuel prices reported over the sample period in weekly surveys of the city's retailers, as described above. Since retailers were surveyed on a weekly rather than daily basis, with surveys largely taking place early in the week (and centered around Tuesday), we predict s_t^{gas} using fuel prices that vary daily based on linear interpolation of prices that are observed to vary weekly about the reference weekday, Tuesday. Specifically, Salvo and Huse (p.259) model consumer i, observed at fuel retailer l, as choosing fuel $f \in \{regular\ gasoline,\ midgrade\ gasoline,\ ethanol\}$ that maximizes utility:

$$u_{fi} = \alpha\left(p_{fl}/k_{fi}\right) + x_l'\beta_{1f} + x_i'\beta_{2f} + \varepsilon_{fi}$$

where p_{fl}/k_{fi} are retailer-vehicle specific fuel prices in R$ per km driven, x_l and x_i are vectors containing other observed retailer and consumer/vehicle characteristics that shift choice probabilities, and unobserved idiosyncratic tastes ε_{fi} follow a multivariate Normal distribution with mean zero and covariance matrix Ω, i.e., $\varepsilon \sim MVN(0, \Omega)$. The choice set for those consumers purchasing fuel at the low proportion of retailers that do not carry midgrade gasoline, as surveyed by Salvo and Huse, includes only regular gasoline and ethanol. Thus, for example, the probability that a consumer, facing the full choice set, purchases ethanol is given by:

$$\begin{aligned}\Pr(consumer\ i\ chooses\ e\ over\ g\ and\ midg) &= \Pr\left(u_{gi} - u_{ei} \leq 0 \cap u_{midgi} - u_{ei} \leq 0\right) \\ &= \Phi\left(\left(\alpha(p_{el}/k_{ei}) + x_{li}'\beta_e\right) - \left(\alpha\left(p_{fl}/k_{fi}\right) + x_{li}'\beta_f\right), \Omega_{-e}\right),\ \ f = g, midg\end{aligned}$$

where Φ is the CDF of the bivariate normal random variable $\left(\varepsilon_g - \varepsilon_e\ , \varepsilon_{midg} - \varepsilon_e\ \right)$ with mean zero vector and covariance matrix Ω_{-e}. We base our main variable of interest, the gasoline share s_t^{gas}, on Salvo and Huse's specification III (Table 2, p. 39). For convenience, we reproduce their estimates for parameters (α, β, Ω) in Table S1. Estimation is by Maximum Likelihood.

To account for fuel stored in vehicles' tanks, following consumer purchase but prior to combustion, s_t^{gas} is predicted using four-day lagged prices at the pump. Salvo and Huse (p.258)

report that the median consumer purchased fuel once a week. Thus, the gasoline share of combustion on day t is predicted from fuel prices at the pump 7/2 ≈ 4 days earlier (and in a robustness test we increase consumer stocks to 7 days). To account for sampling variation in generating a prediction for s_t^{gas} (a variable that is estimated rather than observed), we bootstrap Salvo and Huse's original sample of consumers observed making choices at the pump. Thus, for every one of 200 bootstrap samples, b = 1, …, 200, we obtain a different gasoline choice probability $s_t^{gas,b}$ (for each combination of fuel prices). We subsequently use these "first-step" bootstrap samples to make inference from our "second-step" pollutant regressions (see SM Part F).

An alternative consumer-level choice model to that of Salvo and Huse is the multinomial logit, where consumer i at fuel retailer l chooses fuel f to maximize utility:

$$u_{fi} = \alpha_i(p_{fl}/k_{fi}) + x_l'\beta_{1f} + x_i'\beta_{2f} + \varepsilon_{fi}$$

Here, unobserved idiosyncratic tastes are assumed to be distributed Extreme Value Type 1 (EVT1), and the price sensitivity parameter is allowed to vary by consumer type. Parameter estimates for this alternative specification are reported in Table S1. As Fig. 3 in the main text shows, predicted shares for this alternative specification are very similar to those predicted by Salvo and Huse's specification.

5. Aggregate reports of gasoline and ethanol shipments by wholesalers. As stated, shifts in the fuel mix were captured in reported wholesale shipments. For example, among the São Paulo state distributors that submitted reports, 950,000 m^3 of gasoline and 290,000 m^3 of ethanol were shipped in April 2011, as relative ethanol prices peaked, compared with 750,000 m^3 of gasoline and 530,000 m^3 of ethanol in May 2011, as ethanol prices dropped (see Fig. 1 of the main text).

We calculate an alternative measure of the gasoline share as follows. We convert the aggregate fuel quantity series, separately for gasoline and ethanol, from cubic meters to vehicle distance traveled, using light-vehicle fuel economy rates of: (i) 7.05 km/liter on ethanol E100; (ii) 10.23 km/liter on gasoline E25; and (iii) 10.42 km/liter on gasoline E20 (temporarily mandated between February and April 2010). Fig. S3 depicts the sum of these gasoline and ethanol "vehicle kilometers traveled," as well as the gasoline share of this total. Our conversion ignores the fact that a small but unknown fraction of gasoline shipments is used to power motorcycles, with higher km/liter. It also does not account for the lower fuel efficiency of older

single-fuel (mostly gasoline-powered) vehicles that still circulate, for which there is little data. One can view the converted series as fuel quantities expressed in a common unit of energy, considering the same equipment and the same driving cycle. As Fig. S3 shows, this alternative gasoline share grew from 51% of total distance traveled in September 2009 to 68% in February 2010, and from 53% in September 2010 to 83% in April 2011. Notice that this alternative share varies less than the baseline gasoline share, s_t^{gas}, in part because the baseline share relates to choices in the subpopulation of bi-fuel vehicles.

We caution that variation in the alternative share may not be accurate because it is based on reported data that is too aggregated, both temporally (monthly) and spatially (shipments throughout the state). Further, as a proxy for fuel consumption in the São Paulo metropolis, the monthly state-level wholesale shipment data on which the alternative share is based may not be comprehensive, may not accurately capture interstate over in-state shipments, and does not account for variation in downstream inventories.[7] Nevertheless, the aggregate data does indicate that fuel switching occurred at a large scale. Excluding the ethanol component from the (blended) gasoline series suggests that the share of pure gasoline varied from 42% to 58% of distance traveled over 2009/2010, and from 43% to 68% over 2010/2011, i.e., it grew by roughly one half at the expense of ethanol.

Despite the limitations of such aggregate, likely non-comprehensive fuel shipment data, we can also use it to look for any suggestive evidence of an "intensive margin" of fuel consumption, that is, whether consumers cut back on vehicle usage as ethanol prices rose (this is a question we address with detailed traffic data in SM Part E). Here we take the sum of gasoline and ethanol "vehicle kilometers traveled" depicted in Fig. S3 – a time series with 61 monthly observations – and regress this distance driven on month-of-year fixed effects (12 dummy variables less one, as we include an intercept), a quadratic trend, and fuel prices (either the ethanol-to-gasoline price ratio, or the ethanol price and the gasoline price). The month-of-year dummies capture seasonality in consumer driving behavior as well as in fuel retailers' purchasing behavior; for example, the state's retailers tend to stock up in December in advance of the yearend school vacation into January and the Carnival month of February (see below). These month-of-year fixed effects play the role of week-of-year fixed effects in our pollutant regressions. The trend captures any underlying variation in, for example, economic activity, such

as growth between 2006 and mid 2008 (prior to the sample period for our pollutant regressions, where we similarly allow for a trend).

We find no evidence in these exploratory regressions that aggregate fuel consumption patterns were significantly associated with fuel prices. In the total distance traveled regression that includes the price of ethanol relative to gasoline (as in Fig. 1 in the main text) as an explanatory variable, we obtain that: (i) the trend is estimated to be significantly increasing and concave, with growth in billion km traveled slowing by late 2008; (ii) there is evidence of seasonality, with December exhibiting higher shipments ahead of January and February; and (iii) the coefficient on the ethanol-to-gasoline price ratio is positive and not significant (a point estimate of 0.005, per percentage point in p_e/p_g, with robust standard error of 0.009). In the regression that includes the (inflation-adjusted) prices of ethanol and gasoline as separate regressors, we obtain the same trend and seasonal patterns as well as insignificantly positive estimated coefficients on the gasoline price and the ethanol price.

B. Studies involving tailpipe emissions, smog chambers, and computer models.

Compared with gasoline, whose chemical composition is complex as it contains olefins, aromatics, paraffins, additives, nitrogen- and sulfur-containing organic species,[26] ethanol fuel is often viewed as a "cleaner" alternative.[27,28] This assumption is based on some tailpipe emissions studies which suggest that increasing ethanol content in gasoline fuel is associated with: (i) reductions in CO and hydrocarbon emissions;[29-35] (ii) reductions in particulate concentrations in the coarse modes (2.5 to 10 microns);[32,36] (iii) reductions in NO_X emissions,[37-40] though measured NO_X emission trends appear inconsistent;[29,31,32,34,41-44] (iv) reductions in SO_2 emissions, though significant amounts of SO_2 may be emitted depending on the sulfur additives and lubricants in ethanol-burning engines;[45] (v) reductions in 1,3-butadiene and benzene concentrations;[40] and (vi) significant increases in aldehyde emissions.[33,40,45,46] Increases in CO_2 concentrations as a result of more complete combustion of ethanol have also been reported.[29,30] These results are often rationalized in terms of the differences in the fuels' heats of combustion,[47,48] which are 28-30 MJ/kg for ethanol[37-39] and 40-47 MJ/kg for gasoline E25.[30,38,39,49-51] Emissions tests show some variance to the model, condition, and engine setup of the vehicles tested, as well as the local formulation of fuels.

The differences in NO_X emissions in particular have been attributed to: (i) oxidation of nitrogen-containing compounds in the fuel with oxygen in the combustion chamber (fuel NO_X); (ii) high temperature oxidation of nitrogen molecules in the chamber (thermal NO_X) via the Zeldovich cycle;[47,48] and (iii) reactions of nitrogen with hydrocarbon radicals formed in the fuel at high temperature.[48,52] One thus may expect that more NO_X will be produced during gasoline relative to ethanol combustion, and that the effect is compounded for gasoline with high organic nitrogen content. Not accounted for is the possible introduction of nitrogen species into ethanol from ethanol production, storage, and transport as well as lubricant additives used for ethanol combustion, which can be substantial and coincide with other additional elements not typically associated with ethanol combustion, such as Zn, Cu, Cr, Pb and even Pt.[53]

Pollutant concentrations in the natural system are not merely reflections of tailpipe emissions, but also may depend significantly on meteorological conditions and concentrations of intermediates and products resulting from reactions of tailpipe and evaporative emissions with other natural and anthropogenic atmospheric constituents.[54] Studies for the São Paulo metropolitan area tend to report high atmospheric concentrations of acetaldehyde and ethanol, and significant levels of photochemical smog.[55-58] However, compared to the number of tailpipe emissions studies, there exist unfortunately far fewer experimental studies examining the influence of vehicle transportation fuels on atmospheric chemical composition.[33,59-61] One such study, which adopts a smog chamber approach, finds that ethanol resulted in around 30% more ozone than gasoline E22-E24.[62]

This study is supported by a recent detailed atmospheric chemistry modeling study,[60] which concludes that powering vehicles with ethanol E85 versus gasoline would increase ozone concentrations from 7 to 40 ppb for the conditions studied, and also increase ambient levels of formaldehyde, acetaldehyde, and peroxyacetyl nitrate. Such modeling studies[60,61] seem to be contradicted by other experimental observations that show decreases in ozone levels and the number of smog days when ethanol was added to fuels.[63]

Specifically, a modeling study by Martins and Andrade for the São Paulo metropolitan area suggests that running the entire light-duty vehicle fleet on pure ethanol would reduce exceedance frequencies of ozone.[57] The study calibrates the California Institute of Technology photochemical Eulerian model[64] to the local environments of ozone-monitoring stations that our study also considers, including Diadema, Ibirapuera, Moóca, Pinheiros, and São Caetano do Sul

(see SM Part C). Martins and Andrade consider a scenario in which the São Paulo vehicle fleet prevailing in 2004 – assumed to comprise 70% passenger vehicles running on gasoline E25, 15% passenger vehicles on ethanol E100, 9% motorcycles on gasoline E25, and 6% heavy-duty vehicles on diesel – were to *hypothetically* experience all passenger vehicles shifting to E100. Their "results suggest that implementing (such a) scenario would improve air quality in the metropolitan area of São Paulo," (p.166) with simulated average and peak ozone concentrations for the week of September 6, 2004 predicted to fall by 16-50% and 14-55% against the gasoline-dominant base case under actual fleet emissions. The fuel mix changes considered in the Martins and Andrade simulation are of fairly comparable order of magnitude (though larger) to those of our study. Our calculations based on aggregate reported shipments for 2004 (SM Part A), and an assumption regarding the relative usage of motorcycles, suggest that Martins and Andrade assess a hypothetical increase in the "pure" ethanol share (including the ethanol component in blended gasoline) of total gasoline-plus-ethanol combustion from 34% of distance traveled in the base case to 97% in the simulated case (less than 100% as gasoline still powers motorcycles). By way of comparison, we observe a pure ethanol share at two different points in time in our sample varying between 32% in April 2011 and 57% in September 2010. Despite the fairly comparable in-sample fuel mix variation and the common São Paulo setting, our findings based on field data are at variance with those of Martins and Andrade.

Review articles underscore the need for data-based studies that examine the air quality impacts of consumer adoption of alternative fuels and vehicles. One article reviews work analyzing variation in the São Paulo light-duty fleet between 1981 and 1985, during which the number of single-fuel ethanol vehicles rose from 84,000 to 500,000, and the number of single-fuel gasoline vehicles fell by 18%.[65] This change in the fuel mix coincided with reported reductions in (average or maximum) ambient SO_2 and CO levels, but persistent PM10 pollution, which was attributed to fixed sources and the heavy-duty fleet. The authors discuss possible changes in O_3 production rates due to the presence of aldehydes in the air, which they list in Table 3 to be largely invariant with time for various intervals studied. Another review article summarizes studies of the air quality impacts of ethanol use in Brazil and concludes that "(f)or the most part, we ignore the hundreds of individual compounds that are actually emitted from the vehicle in that broad range of VOC compounds. As fuel composition changes, it is necessary to look at the details of the VOCs and how they change with changing fuel composition. If we do

not, there can be dramatic effects on air quality, as we have seen in Brazil" (p.1034).[66] The author sums up the outlook on bi-fuel gasoline-ethanol vehicles by stating that "(t)here are very little data in the literature dealing with identifying, quantifying and reducing the emissions of the unregulated pollutants from these vehicles" (p.1035).

C. Description of air monitoring stations.

1. Data source and measurement. The pollutant concentration data we use, as well as a large part of the meteorological data (see SM Part D), were collected by the Companhia Ambiental do Estado de São Paulo (CETESB) at twenty-two air monitoring stations located throughout the São Paulo metropolitan area, as shown in Fig. 2 in the main text (for a larger area see Fig. A-S1 in the Appendix). Data were recorded at hourly intervals,[67] with hourly observations based on at least 720 automatic readings within the hour. The calibration of instruments measuring concentration of O_3, NO_X and CO occurred during the 60 minutes immediately preceding 06:00, 01:00 and 05:00, respectively, thus reliable measures were not available for one early hour every morning. With regard to the other measured pollutants, calibration for SO_2 took place during the hour leading up to 04:00, and no downtime was needed for PM10, for which 24 hourly measures were provided every day.

2. Descriptive statistics. Table A-S1, in the Appendix to the Supplementary Materials, displays station images and GPS coordinates for the twenty-two individual stations, along with a list of the pollutants measured at each facility and the availability of measures during the sample period.[68] For cost reasons, most stations measured only a subset of the regulated pollutants. Occasionally, instruments were out of order and measurements may be missing for a given parameter.

Table SII summarizes pollutant concentration levels and ratios for specified times of the day at stations monitoring O_3, NO_X and CO with overall data availability of at least 70% during the sample period (as noted in Table A-S1). A strong sign of immediate anthropogenic activity is the concentration of NO.[69] For example, three NO_X-monitoring stations that were somewhat removed from large roadways and busy intersections – namely, Ibirapuera, IPEN-USP and Mauá – recorded lower NO concentrations (denoted [NO]) on average (less than 30 $\mu g/m^3$ in the early morning), consistent with their relative distance from vehicle traffic. These relatively removed

locations also featured lower [NO] to [NO_2] ratios (less than 0.9) compared to stations located right by roads, consistent with NO being produced preferentially over NO_2 during combustion. Consistently, O_3 concentrations tended to be somewhat higher (around 80 μg/m^3 in the early afternoon) at these relatively removed stations compared to stations right by roads, as O_3 reacts away with NO.[70,71] Further, as one would expect, CO concentrations tended to be higher at stations located by busy intersections (e.g., Cerqueira César and Congonhas exceeding 1 ppm).
We find the consistency between satellite images of stations' surroundings and recorded pollutant concentrations to be reassuring.

D. Meteorology.

1. Data sources. Table SIII reports the sources – in terms of institution and location – of the hourly meteorological data that we use in our pollutant regressions. Also reported is the very high availability of hourly measurements during the sample period. For example, the lowest availability, at 94% of the maximum number of possible hourly measurements, is for atmospheric pressure, which was measured only at the Ibirapuera station.

a. Imputing of missing hourly precipitation values. In a very small number of instances, we used descriptive daily weather reports from CETESB (*Boletim de Qualidade do Ar – Condições Meteorológicas*) to impute some missing hourly precipitation values, since precipitation was measured every hour at only one station, maintained by the Institute for Meteorology (INMET), in Santana in the northern region of the city. Namely, we imputed zero for some hourly observations around which non-missing hourly precipitation values were zero and which fell on days that were reported as dry in CETESB's daily weather summaries. From INMET we obtained further precipitation data collected manually (also in Santana) but over longer intervals than one hour: 00:00-12:00 GMT (Greenwich Meridian Time), 12:00-18:00 GMT and 18:00-24:00 GMT. We then imputed zero for some missing hourly observations that fell on intervals that INMET's lower-frequency data indicated as completely dry. We did not impute other missing hourly precipitation observations.

b. Imputing of missing hourly atmospheric pressure values. Similarly, we imputed some missing values for atmospheric pressure, which was measured at a single station (Table SIII). We linearly interpolated atmospheric pressure that was measured at hours immediately preceding and immediately subsequent to an hour for which pressure was unrecorded. Remaining missing

values were imputed based on the linear prediction of atmospheric pressure regressed on a full set of date fixed effects (November 1, 2008 through July 31, 2011) and hour-of-day fixed effects (01:00 through 23:00); the R^2 for this regression was 85%. We include atmospheric pressure among weather controls only in the robustness tests of our baseline pollutant regressions.

2. Overview of meteorology in the São Paulo metropolitan area. Fig. S4 plots times series, at the hourly interval, for some meteorological variables over the period November 1, 2008 to July 31, 2011. Other figures accompanying this overview, highlighting specific variables, can be found in Figs. A-S2 to A-S5 in the Appendix. Temperatures remain moderate throughout the year, ranging from 15-30°C between the months of October to May, which include the summer, and dropping to 10-20°C in the colder months between June and September. Relative humidity can fall below 50% during night-time. The prevailing winds blow from the Southeast and the Northwest, with speeds below 5 m/s. Precipitation follows the annual seasonality that is typical for this sub-tropical region. Though less common than in the dry winter months of July and August, completely dry days during the summer months of January and February are still observed, and moderately dry days are common. Radiation levels follow the seasonal variation.

3. Spatial correlation of variation in meteorological conditions. Since meteorological conditions were measured at only a reduced number of locations in the São Paulo metropolis, ranging between one and five stations for each variable (Table SIII), we specify meteorological controls W_t in our pollutant regressions (equation (1) in the main text) using mean measurements across the available weather stations that monitored each variable. We find this approach reasonable after ensuring that meteorological conditions measured at different locations in the metropolis do indeed exhibit strong spatial correlation. Specifically, figures indicating such high spatial correlation are available upon request for: (i) temperature measured in Pinheiros, São Caetano do Sul and Taboão da Serra stations; (ii) relative humidity at these same three stations; (iii) radiation at Ibirapuera and Paulínia stations, the latter station being located 100 km northwest of the city of São Paulo; (iv) wind speed at Ibirapuera, Moóca, Osasco, Pinheiros and Santana stations; (v) wind direction at four of these stations; and (vi) precipitation, for INMET's lower-frequency data collected in Santana and additionally in Guarulhos, also in the São Paulo metropolis, and Sorocaba, located 100 km west of the metropolis.

On predicting pollutant concentrations at each air monitoring station, what is important is not that the level of each meteorological variable remain invariant across space, rather, it is that *temporal variation* in meteorology correlate across space. For example, (i) when we observe a relatively sunny afternoon in radiation-measuring Ibirapuera station, it is the case that other air monitoring stations throughout the metropolis tend to experience relatively sunny conditions too; (ii) rising temperatures as detected by the Pinheiros station on the western side tend to occur alongside rising temperatures at ozone-measuring Móoca station on the eastern side; and (iii) when it rains in Santana, where our single source of hourly precipitation data is located, it tends to rain elsewhere in the metropolis.

4. Pollutant concentrations and meteorology. The following paragraphs briefly discuss scatterplots of pollutant concentrations against several meteorological parameters. These scatterplots, prepared from our sample, can be found in Figs. A-S6 to A-S9 in the Appendix. For each particular pollutant, we plot hourly observations at each CETESB station that monitors that pollutant's concentration.

a. Correlation of pollutant concentrations with wind speed and direction. Appendix Fig. A-S6 plots the concentrations of O_3, NO, CO, PM10 and PM2.5 against wind speed and indicates that the air is generally well mixed. The upper bounds to NO and CO concentrations tend to decrease with wind speed (see panel A). Logarithmic ordinate representations show that the lower bounds to measured concentrations tend to increase with wind speed (panel B). Scatterplots of these concentrations with respect to wind direction (Appendix Fig. A-S7) seem to suggest that the concentrations of O_3, NO, CO, and PM10 do not fall below a certain threshold when wind blows from the Southwest, but inspection indicates that this is due to low data density for wind directions other than those corresponding to the prevailing winds.[53] Scatterplots of concentration against wind direction that restrict plotted observations to hours experiencing wind speeds in excess of 0.5 m/s are very similar (not shown for brevity).

b. Correlation of pollutant concentrations with precipitation. Appendix Fig. A-S8, panel A shows a weak clearing effect that precipitation has on the concentrations of O_3, NO, CO, PM10 and PM2.5 for varying amounts of precipitation, highlighting why precipitation needs to be controlled for in our pollutant regressions. A logarithmic ordinate representation in panel B shows that this clearing effect is not akin to reducing all concentrations to zero.

c. Correlation of pollutant concentrations with radiation. Appendix Fig. A-S9, panel A shows that the concentrations of O_3 are generally high when radiation is high, which is expected from the well-known gas-phase photochemistry of urban air.[72] This relationship is robust to restricting O_3 measurements to the 13:00 through 16:00 hours of the afternoon, to illustrate sunny versus cloudy weather. There is a negative association between radiation and the concentrations of NO and CO, and no clear association with the concentrations of PM10 and PM2.5. A logarithmic ordinate representation in panel B shows that such relationships with radiation are also observed for lower concentration levels of O_3, NO, and CO, but are indeed minor or even negligible for PM2.5 and PM10.

E. Vehicle traffic.

1. Data sources (including the holiday calendar). Data on vehicle traffic throughout the city of São Paulo between November 2008 and July 2011 was obtained from the city's traffic authority, Companhia de Engenharia de Tráfego (CET).[73] We use two datasets, "traffic congestion" and "traffic speed." The first dataset records, at 30-minute intervals each day of the year, including weekends and public holidays, which road segments out of an extensive fixed grid were "congested," i.e., it informs which street-half-hour pairs are associated with idling vehicles. The monitored network comprises the city's main roads and corridors, with a total extension of 840 kilometers including opposite directions and express lanes for the few existing urban highways (e.g., the *Marginais*). The monitored grid was last expanded in 2007,[74] prior to the start of our sample period, and can be view in the backdrop to Fig. 2A in the main text. To illustrate one data point, on November 30, 2010 at 19:30, 0.75 km of congestion was reported on the segment "de Rocha Azevedo até acesso da Rebouças" in the "Consolação/Paraíso" direction of the "Avenida Paulista" corridor. This corridor is assigned by CET to the city's Center region, shaded in purple in Fig. 2A of the main text. The traffic congestion data can be aggregated spatially to different levels, such as: (i) the street level, e.g., km of congestion along the Avenida Paulista between Rocha Azevedo and Rebouças; (ii) the region level, e.g., Center; and (iii) the citywide level, i.e., total congestion across all five regions of the city – Center, North, East, South, and West. The media, particularly the local radio, widely reports on such measures of *lentidão* (slowness), in particular, the citywide km of congestion (spatial aggregation (iii) just noted). Because the traffic

congestion data are so comprehensive, the traffic controls T_t in our baseline pollutant regressions (equation (1) in the main text) are based on this first dataset (see SM Part F).

The second dataset is available only for weekdays, excluding public holidays, and provides the traffic speeds (or, equivalently, travel times) recorded at 08:30 – the morning traffic "peak time," as defined by the traffic authority – and again at 18:00 – the afternoon peak time – along each of 36 monitored road segments located in all five regions of the city. To illustrate one data point, the time it took vehicles to move along the *Avenida Rebouças* from the *Rua Joaquim Antunes* crossing to the *Rua Lisboa* crossing, extending over a distance of 440 meters, was recorded to be 106 seconds and 239 seconds in the morning and afternoon peak times, respectively, of Wednesday, November 25, 2009; these recorded times correspond to speeds of 14.9 km/hour (kph) in the morning and 6.6 kph in the afternoon. The next morning, on Thursday, November 26, 2009, traffic along this same road segment was flowing more freely, reaching a speed of 27.8 kph (and the afternoon speed was similar to the day before). Unlike traffic congestion data, which was collected without interruption and with complete availability between November 2008 and July 2011, traffic speeds were not recorded in November 2008, over the first half of December 2008 and at the end of January 2010. There are further missing data on specific date-time-segment combinations. Considering that there were 663 weekdays excluding public holidays between November 2008 and July 2011, and 36 monitored road segments, speed data availability is 15327/(663×36) ≈ 64% for morning observations (i.e., there are 15,327 recorded morning speed observations in the data) and 14356/(663×36) ≈ 60% for afternoons. We additionally control for traffic speeds in robustness tests of our baseline pollutant regressions.

Because the traffic authority administers, by force of municipal law, certain vehicle circulation restrictions which apply only to regular workdays, it maintains a calendar of public holidays (*feriados*) and yearend school vacations (*férias escolares*) during which such restrictions did not apply.[75] The light-vehicle *rodízio* ("rotation") restricts a rotating one-fifth of the lift-duty fleet from circulating in most of the city during morning and evening rush hours on any given non-holiday weekday, Monday through Friday, based on the last digit of the vehicle's number plate. Importantly, these driving restrictions, as well as similar restrictions that apply to the circulation of certain heavy-duty vehicles, were in effect throughout our sample period.[76]

Even though we control for observed traffic conditions via vector T_t, our pollutant regressions are estimated separately for non-holiday weekdays and other types of day, based on the traffic authority's calendar (as well as separately by location of measurement and time of day). To be clear, our definition of non-holiday weekdays excludes the yearend school vacation, typically starting on December 24 and lasting two weeks, during which traffic might flow a bit more freely. We test for robustness by adding back weekdays during the school vacation fortnight to the non-holiday weekday sample. Further, weekdays that fall between an official public holiday and the weekend – known as *ponte* ("bridge") days – are effectively considered public holidays by employers, schools, and the traffic authority. Thus, for example, we accordingly specify – and the traffic congestion data bears this out – the Monday prior to *Finados* ("Day of the Dead") holiday on Tuesday November 2, 2010 also as a public holiday.

2. Pollutant concentrations and traffic. To illustrate the relationship between ambient pollution and road traffic in the sample, Fig. S5 shows that the hourly concentrations of NO and CO at given locations increase in the amount of traffic congestion that was contemporaneously recorded in the surrounding regions of the city. The scatterplots report concentrations during morning and evening rush hours at two air monitoring stations, Congonhas and Taboão da Serra, both located in close proximity to roads. Upon request, similar scatterplots are available where, instead of only contemporaneous traffic congestion, we plot the sum of contemporaneous congestion and congestion recorded 30 minutes earlier, 60 minutes earlier, etc., to capture the build-up and diffusion of emissions.

Further illustrating the ability of traffic to predict local air pollution is the point we made in SM Part C that, in the cross-section of stations, stations that were somewhat removed from large roadways, such as Ibirapuera and IPEN-USP, recorded lower NO and CO concentrations. Similarly, conditioning on location and time of day, lower NO and CO concentrations observed on Sundays and public holidays, compared to non-holiday weekdays, is consistent with there being less vehicles on the road in the former than in the latter – though still a considerable number of vehicles circulate even on Sundays and public holidays.

3. Did consumers cut back on vehicle usage, and possibly switch to public transportation, as ethanol prices rose? The analysis that follows provides evidence that traffic conditions

observed between November 2008 and July 2011 did not significantly vary as relative ethanol prices varied. Rising ethanol prices, beginning in mid 2009 and again in mid 2010, did not ease traffic congestion, raise traffic speeds, or increase ridership in the public transportation system. Similarly, when ethanol prices began falling in March 2010 and again in April 2011, motorists did not take to their vehicles more often. We show that traffic in the city of São Paulo, while often congested, is *predictably* congested based primarily on time of day, type of day and precipitation shocks – and that fuel prices do not provide additional power to predict observed traffic. This is important, as one might otherwise be concerned that potential variation in vehicle usage caused by variation in fuel prices – the "intensive margin" – may confound any causal effect of variation in the fuel mix – the "extensive margin" – on air quality, which is the object of this study.

There are several ways to interpret this finding. In general, studies of consumer demand for private vehicle use tend to find that fuel consumption, or vehicle kilometers traveled, is not sensitive to fuel prices in the short run (say, over the space of a few years), in part due to the poor availability of substitutes.[77] One should further note that in our specific setting, there was much less variation in the pump price of gasoline – a close substitute to ethanol fuel, if not the only fuel, for over nine-tenths of São Paulo's light-duty vehicle fleet (Fig. S2).[6] Owners of gasoline-dedicated vehicles were less exposed to fluctuations in the price of ethanol. Among the population of bi-fuel vehicle motorists, while some chose to stay with ethanol as prices rose, the majority switched to gasoline, so the welfare effect from a given ethanol price increase on these consumers was similarly muted, limiting the likely effect on vehicle kilometers traveled. Finally, the existence of "repressed demand" for vehicle usage that has been argued in the face of widespread gridlock, as noted in SM Part A,[17,18] suggests that reduced vehicle usage by any price-sensitive gasoline-averse consumer in the population, were traffic congestion to have eased on this margin, might have been offset by increased vehicle usage by other motorists.

While we do control for traffic, and thus vehicle usage, directly in our pollutant regressions (equation (1) in the main text), we are reassured by the exogeneity of traffic to fuel prices that we verify empirically, and are able to interpret in light of the setting.

a. Evidence from traffic congestion. We estimate the following regression equation:

$$traffic_t = f(daytype_t * time_t, W_t * daytype_t * time_t, event_t, fuelprices_t * daytype_t * time_t, \varepsilon_t)$$

where the dependent variable $traffic_t$ is the recorded traffic congestion in km at some level of spatial aggregation (i.e., across the city, within a specific region of the city, or along a specific road corridor of the monitored grid) at 30-minute intervals between November 1, 2008 at 00:00 and July 31, 2011 at 23:30. There are 48,144 observations (1003 days × 48 half-hourly observations per day). The explanatory variables (i) $daytype_t * time_t$ are a full set of type-of-day fixed effects ("regular Monday,"…, "regular Friday," "yearend weekday," "regular Saturday," "yearend Saturday," "regular Sunday," "yearend Sunday," and "public holiday") interacted with 47 half-hour fixed effects (00:30, 01:00, …, 23:30); (ii) $W_t * daytype_t * time_t$ are interactions of contemporaneous (last 60 minutes) and lagged ($t-2$ hours to $t-1$ hour, $t-3$ to $t-2$, …, $t-13$ to $t-9$) precipitation covariates with a (reduced) set of type-of-day-and-time-of-day fixed effects (e.g., "regular weekday 07:00 or 07:30"); and (iii) $event_t$ are fixed effects that control for: (iii-a) the time window between 17:00 and 22:30 that precedes a long weekend, when many residents head out of town and congestion tends to rise further, and (iii-b) the time window between 17:00 and 22:30 that closes a long weekend, as residents return to town. To control for extreme (but predictable) traffic patterns around 2010 World Cup soccer matches that fell on the middle of weekdays, when most commuters returned home to watch the game and subsequently stayed at home, vector $event_t$ also includes fixed effects for: (iii-c) the two-hour window prior to a match with a 15:30 kick-off time, (iii-d) the four-hour window prior to a match with an 11:00 kick-off, and (iii-e) the hours between 17:00 and 20:30 following a match. We also include (for brevity, subsumed in $daytype_t * time_t$ in the regression equation) week-of-year fixed effects interacted with the reduced set of type-of-day-and-time-of-day fixed effects. We flexibly include a quadratic function of the contemporaneous and lagged precipitation covariates, i.e., squares and cross-terms of all precipitation lags, each term interacted with type-of-day-and-time-of-day fixed effects.

The main covariates of interest in these traffic regressions are $fuelprices_t$, a vector that contains: (i) the per-liter prices of gasoline and ethanol, p_g and p_e, respectively, inflation-adjusted, as in Fig. S1; (ii) the "energy-adjusted" price (or "price heuristic") for the competitively priced fuel among substitute fuels gasoline and ethanol, $min(0.7p_g, p_e)$, also shown in Fig. S1; and (iii) the real price index for diesel, as in Fig. S2. To illustrate, expressing prices as differences relative to their values at the start of the sample period on November 1, 2008, $(p_g, p_e, 0.7p_g, diesel\ index)$ equaled $(-0.21, +0.17, +0.17, -0.17)$ on November 7,

2010, compared to $(-0.05, +0.72, +0.41, -0.20)$ on April 18, 2011, as ethanol prices peaked. That is, ethanol in April 2011 was priced R$ 0.72 higher than in November 2008 and R$ $0.72 - 0.17 = 0.55$ higher than in November 2010. We interact each of the four fuel price covariates with type-of-day-and-time-of-day fixed effects.

Finally, ε_t is the econometric error, accounting for unobserved drivers of traffic such as vehicle accidents or breakdowns (e.g., a truck breaking down along the *Marginal Tietê* during rush hours can wreak havoc), signal failures, public works, strikes, protests, etc. Notice that, consistent with fossil-fuel prices being controlled at the national level by the federal government and ethanol prices tracking developments in the world market for sugar (SM Part A), we take fuel prices to be exogenous to traffic conditions in the city of São Paulo.

Fig. S6 illustrates predictions of a linear version of the above traffic regression equation estimated by Ordinary Least Squares (OLS). The dependent variable is citywide traffic congestion in km. For brevity, we show results graphically for specific dates only, Monday, April 18, 2011 to Wednesday, April 20, 2011 in the left panel, and Wednesday, February 16, 2011 in the right panel. A full report of estimated coefficients and standard errors is available upon request, as are model predictions for other dates in the sample. Here, we simply state that the joint predictive power of the 3,010 explanatory variables (on 47,969 observations) is very high, with R^2 of 89.9%, and we illustrate by means of two counterfactual exercises the separate effects of fuel price variation (left panel) and precipitation (right panel) on traffic congestion.

The goodness-of-fit of the model is very high, with model predictions closely approximating observed citywide traffic congestion. Traffic conditions across São Paulo city are quite predictable, even on the evening before *Tiradentes* holiday on Thursday, April 21, 2011 when many city dwellers were leaving town and the extension of congestion exceeded 150 km. The left panel shows that traffic conditions would not have changed significantly had fuel prices on the week of April 18, 2011 been equal to fuel prices on the week of November 7, 2010, when ethanol prices were substantially (0.55 R$/liter) lower. In particular, the estimated model allows us to reject the hypothesis that traffic congestion would have been higher had ethanol prices been lower. The right panel shows that traffic congestion during the evening rush hours of Wednesday, February 16, 2011, in the aftermath of 30 mm of cumulative precipitation between 17:00 and 19:00, was approximately double relative to a hypothetical scenario in which the city were not to have suffered a precipitation shock.

Results are robust to including radiation covariates (to proxy for visibility) and to adding type-of-day-and-time-of-day specific linear time trends. Also available upon request are results for variations around the specification presented above, including fitting the logarithm of traffic congestion in km (plus 1 km), and regressing traffic congestion at other levels of spatial aggregation (e.g., traffic congestion in the South region of the city).

b. Evidence from traffic speed. We now consider the second traffic dataset, on traffic speeds recorded for a panel of road segments at 08:30 ("AM peak time") and at 18:00 ("PM peak time") on weekdays between December 15, 2008 and July 29, 2011. We again find that the evidence weighs in favor of the exogeneity of traffic conditions to fuel prices in the sample.

We base the dependent variable in these traffic regressions on the inverse of traffic speed, expressed in hours per kilometer (hpk). In the example we provided on describing the data, in which a particular 440-meter distance was covered in 106 and 239 seconds at 08:30 and 18:00, respectively, on a given day, inverse speeds correspond to 0.067 and 0.15 hpk. That is, it took vehicles over twice as long to cover the given extension in the afternoon rush than in the morning one. We estimate a similar regression equation to that for traffic congestion:

$$traffic_t = f(daytype_t * time_t, W_t * daytype_t * time_t, event_t, fuelprices_t * daytype_t * time_t, \varepsilon_t)$$

The dependent variable $traffic_t$ is now the 75th percentile, median or 25th percentile of the distribution of inverse traffic speeds (travel times) recorded across the 36 monitored road segments in the city (alternatively, similar regressions can be estimated at the region or road segment levels—see below). Time subscript t corresponds to morning and afternoon peak hours (08:30 or 18:00) of weekdays in the sample. The type-of-day fixed effects (e.g., "regular Monday") are now interacted with two time-of-day fixed effects, $time_t$, corresponding to the morning (08:30) or afternoon (18:00) measurement. Precipitation covariates W_t are the accumulation of rain over the 12 hours that precede an observation, separately by three-hour windows. The vector $event_t$ includes similar controls for long-weekend traffic and World Cup soccer. The vector of fuel prices $fuelprices_t$ again contains the covariates of main interest. We flexibly include a quadratic function of precipitation covariates and a quadratic function of fuel price covariates, where each term is interacted with type-of-day and time-of-day fixed effects.

Fig. S7 indicates the good fit of a linear version of the model, estimated by OLS, at explaining the 75th, median, and 25th percentiles of the empirical distribution of inverse traffic speed – the figure illustrates model predictions for weekdays between April 4, 2011 and April

20, 2011. The R^2 for the median inverse speed regression is 81.7% (1262 observations, 362 regressors). It tended to take less time to travel a given distance in the morning than in the afternoon, and this is the case for different quantiles of the cross-sectional distribution, i.e., for more congested (as shown by the 75th percentile of inverse traffic speeds) and less congested (25th percentile) road segments alike. As reported in the analysis of traffic congestion, feeding the model counterfactual fuel prices from November 8, 2010, when ethanol prices were substantially lower, does not yield inverse traffic speed predictions that are significantly different compared to predictions using actual April 2011 fuel prices. This conclusion is robust to employing quantile instead of OLS regression.

Further, Table SIV reports results for *separate* regressions for each speed-monitored time of day and road segment pair, e.g., the first regression considers travel times at 08:30 along the *Avenida Aricanduva* from *Rua Julio Colaço* to *acesso da Radial Leste para o Elevado*. The table reports regression estimates for time-road pairs with overall data availability of at least 70% during the sample period, namely 22 roads with travel times recorded for the morning peak time and 21 roads for the afternoon peak time, totaling 43 regressions. An observation is a non-holiday weekday during the sample period, namely November 1, 2008 to May 31, 2011, excluding the colder months of June to September. The dependent variable is the recorded travel time (inverse speed), expressed in minutes per kilometer. Thus, for example, morning travel times were on average highest – and traffic speeds lowest – along the *Radial Leste* from *R. Piratininga* to *R. Alm. Brasil*, averaging 7.0 min/km (8.6 km/h). As regressors, we include week-of-year fixed effects, day-of-week fixed effects, a linear trend, contemporaneous precipitation conditions, traffic events (traffic surrounding long weekends or World Cup soccer matches as explained above), and – as the main covariate of interest – the gasoline share among bi-fuel consumers, s_t^{gas} (noting that results are robust to variations around this specification, such as specifying fuel prices rather than the fuel mix, a quadratic rather than linear trend, and including lagged precipitation).

The exercise confirms the finding above that there is no significant association between the gasoline share (or the relative price of ethanol) and observed travel times during the sample period. The mean estimated coefficient across 43 time-road regressions is only 0.02 (to be interpreted in min/km if higher ethanol prices were to induce s_t^{gas} to shift from 0 to 1), with a

mean standard error of 0.99 min/km. The estimated coefficient on s_t^{gas} is significantly different from zero (at the 5% significance level) in only 9 of the 43 regressions. Travel times are positively associated with s_t^{gas} in 4 regressions and negatively associated with s_t^{gas} in 5 regressions. (Here, we did not adjust standard errors for the fact that the gasoline share is a predicted rather than measured variable, as doing this would not change our conclusion that there is no significant association between the fuel mix and travel times.)

c. Evidence from public transportation ridership records. About 15,000 predominantly diesel-powered buses, of varying passenger capacity levels (maximum ranging from 21 to 190 persons), regularly carried passengers on 1,300 routes throughout the São Paulo metropolitan area.[78,79] Fig. S8 reports monthly ridership on the public bus transport system, alongside the ethanol-to-gasoline price ratio, from November 2008 to July 2011. Ridership was quite stable over the period, tending to fall in the month of January due to the yearend vacation period, and similarly in the winter month of July in which schools also break (in both cases, these days are excluded from the sample period in our pollutant regressions). Importantly, there is no indication that bi-fuel vehicle motorists might have taken to public transport as ethanol prices rose. This is consistent with our finding above that traffic congestion and traffic speeds (of all vehicle types on the road, light-duty, heavy-duty, and motorcycles) were not significantly affected by variation in fuel prices over the study period.

F. Further details on methods, results, and robustness.

1. Sources of emissions other than traffic. In SM Part A, we argued that higher ethanol prices were unlikely to have impacted broader industrial and commercial activity in the São Paulo metropolis. Fig. S9 plots indices of industrial (manufacturing) activity for the state of São Paulo and across twelve other relatively industrialized Brazilian states, alongside the ethanol-to-gasoline price ratio, from November 2008 to July 2011.[80] The figure indicates that manufacturing, and thus emissions associated with manufacturing, evolved similarly in the state of São Paulo as in other states over the period of study. There is no indication of income spillovers from higher sugar and ethanol prices inducing, for example, household purchases of durable goods, which might then stimulate steelmaking and thus industrial emissions in São Paulo state. As noted earlier, Brazil's sugarcane growing and processing activity – a large part of

which is based in northwestern São Paulo state, 400 km from its capital city – accounts for a small share of the state's vast economy.

On top of that, it is important to note that gone are the days in which the state's capital city was an industrial hub. A greenhouse gas emissions inventory prepared for the Office of the Mayor of the municipality of São Paulo (*Prefeitura do Município de São Paulo*) estimated that in 2003 industrial activity, commercial activity and electricity generation accounted, respectively, for 7%, 3% and 2% of municipal emissions among energy users, with transportation accounting for a full 79% of the municipality's emissions (and residences accounting for the remaining 10%).[81] Similarly, in 2010 industrial users accounted for 15% of electricity consumption in the São Paulo municipality against an average 51% across São Paulo state's 644 other municipalities.[82] Vehicle-related emissions (including evaporative) are the major source of local air pollution in the city of São Paulo,[46] whose economy nowadays specializes in services. We further note that the electricity that serves southeastern Brazil is predominantly generated by hydropower. In 2009, hydroelectric plants accounted for 73% of the electricity generating capacity that was installed in the state of São Paulo, namely 14,226 MW out of a total 19,555 MW.[83] In sum, factors that might otherwise confound the identification of the effect of the fuel mix on local air quality are less of a concern in our study.

2. Meteorological and traffic controls.

a. Meteorology. Meteorological conditions – precipitation, relative humidity, radiation, temperature, and wind speed – enter specifications IV to VII of our baseline model (equation (1) and Tables I and II in the main text) as logarithmic transforms of their units of observation (plus 0.001, a normalization to deal with 0 mm of precipitation and 0 m/s of wind speed). Results are robust to controlling for meteorological covariates directly in their reported units. In addition to conditions recorded contemporaneously to pollution, specification VI controls for meteorological conditions over the 18 hours that precede each observation of pollution (indexed by t, an hour-date pair). Specifically, we add the following lagged readings of meteorological covariates: (i) for precipitation, radiation, and wind speed we include $t-1$ hour, $t-2$, the mean across $t-3$ hour and $t-4$ hour, the mean from $t-5$ to $t-8$, and the mean from $t-9$ to $t-18$; and (ii) for relative humidity and temperature, which tend to be more stable over adjacent hours, we include $t-6$ and $t-18$. Specification VI additionally includes ten possible pairwise interactions of the

five contemporaneous meteorological conditions. We have further tested for robustness of our results by additionally specifying squares and cross-terms of contemporaneous and lagged meteorological readings.

b. Traffic. With regard to traffic controls, specifications vary not only depending on whether or not lagged (on top of contemporaneous) traffic conditions are included, but also on the level of spatial aggregation. Compared to meteorological conditions (SM Part D), traffic conditions (SM Part E) can exhibit lower spatial correlation, and we can exploit the spatial dimension of our traffic data. Traffic congestion covariates enter our baseline pollutant regressions in logarithmic transforms of measures in km (plus a minor 1 km normalization to deal with the occasional zero congestion, say at 03:00 on a given night). Results are again robust to not applying this logarithmic transformation.

Specification V controls for total extension of traffic congestion reported contemporaneously across the citywide 840-km monitored road network. To capture the build-up of vehicle emissions over the 18 hours preceding an air quality reading, specification VI adds the same lag structure for traffic congestion as described above for, e.g., precipitation, namely, $t-1$ hour, $t-2$, the mean across $t-3$ hour and $t-4$ hour, the mean from $t-5$ to $t-8$, and the mean from $t-9$ to $t-18$.

On top of citywide congestion, specification VI adds two measures of traffic congestion that are local to the air monitoring station (specifically, it adds contemporaneous and lagged values for each local measure). The first local measure takes the arithmetic sum of the extension of congestion recorded only on road corridors inside the region of the city where the air monitoring station is located. For example, for Pinheiros station this considers roads in the West region of the city. The second local measure takes the weighted sum of congested extensions recorded along local roads, where weights are given by the inverse of the "Haversine" distance between the air monitoring station and the point on the given road corridor that is closest to the station. For example, on Saturday, March 19, 2011 at 09:00, 0.541 km and 0.390 km of congestion were recorded, respectively, along the "Fernando Vieira de Mello Túnel (Rebouças)" and "Rebouças/ Eusébio Matoso, Av" corridors, among a total of 46 western corridors (a fixed panel for the region). Given respective (minimum) distances to the Pinheiros station of 1.368 km and 1.213 km, and inverse distances to Pinheiros for the 46 corridors summing to 25.938 km^{-1}, this second local measure of the congestion surrounding Pinheiros at this time is computed to be

(0.541/1.368+0.390/1.213+0)/25.938 = 0.0276 km (on this Saturday morning, vehicles were not idling on the other 44 western corridors).

c. Interaction of meteorology and traffic. Specification VII adds contemporaneous interactions between the direction from which the wind was blowing[84] and traffic congestion in each of four "other" regions of the city – the city's five regions excluding the air monitoring station's "own" region, for which local traffic is separately controlled. Specifically, to the station-specific pollutant regression equation (1) in the main text, we add four covariates of the following type, one for each region other than a station's own:

$$1\left(\theta_{endpoint\_1}^{station,other\_region} < \theta_t^{wind} < \theta_{endpoint\_2}^{station,other\_region}\right)$$
$$\times\, traffic_t^{other\_region} \times \cos(\theta_t^{wind} - \theta_{midpoint}^{station,other\_region})$$

For each air monitoring station-other region pair, consider a circular sector with center at the station and exactly enclosing the other region, thus defining two radii by the endpoints of the arc only just encompassing the region as viewed from the station. We explain the notation by example: for the Pinheiros station-South region pair (Fig. 2 in the main text), and measuring angles in degrees from North clockwise (Table SIII), the two radii intersecting at Pinheiros station would lie at angles $\theta_{endpoint\_1}^{Pinheiros,South} = 115°$ and $\theta_{endpoint\_2}^{Pinheiros,South} = 189°$. Thus pollution observed in Pinheiros is allowed to covary with traffic congestion in the South region, $traffic_t^{South}$, when wind, from direction θ_t^{wind}, blows over the South region on its way to Pinheiros, such that the indicator function $1(115° < \theta_t^{wind} < 189°)$ turns on, multiplied by a weight. This weight is the cosine of the angle between the direction of the wind over Pinheiros, θ_t^{wind}, and the radius that splits the Pinheiros-South circular sector in half, $\theta_{midpoint}^{Pinheiros,South} = \frac{115+189}{2} = 152°$. For the Pinheiros-South pair, this weight is maximal (one) when the direction from which wind blows is $152°$. For all station-other region pairs, the spatial configuration ensures that this weight is never negative when the indicator is 1. For air monitoring stations where wind direction is not measured, we take wind direction at the closest among the four stations measuring wind direction (Table SIII).

3. Detailed results for O_3, NO_X and CO. Following Tables I and II in the main text, Tables SV to SXII report, for specifications I to VII, mean fuel mix effects on:

- Table SV: **O_3** averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for **non-holiday weekdays**, at different times of the day;
- Table SVI: **O_3** averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for **Sundays/Public Holidays**, at different times of the day;
- Table SVII: **NO** averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for **non-holiday weekdays**, at different times of the day;
- Table SVIII: **NO** averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for **Sundays/Public Holidays**, at different times of the day;
- Table SIX: **NO_2** averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for **non-holiday weekdays**, at different times of the day;
- Table SX: **NO_2** averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for **Sundays/Public Holidays**, at different times of the day;
- Table SXI: **CO** averaged over 11 regressions, one for each of the 11 CO-monitoring stations, for non-holiday weekdays, at different times of the day;
- Table SXII: **CO** averaged over 11 regressions, one for each of the 11 CO-monitoring stations, for **Sundays/Public Holidays**, at different times of the day.

Each table consists of 6 panels, one panel for each time of the day:

- 01:00 to 06:00;
- 07:00 to 10:00;
- 10:00 to 13:00;
- 13:00 to 16:00;
- 17:00 to 20:00;
- 21:00 to 00:00.

Tables SV (for O_3), SVII (NO), SIX (NO_2), and SXI (CO) subsume the results reported in Tables I (bottom panel) and II in the main text.

Estimates for each individual regression, i.e., by pollutant, by air monitoring station, by type of day, and by time of day, are provided in Tables A-S2 to A-S49 in the Appendix to the Supplementary Materials. For example, Table A-S2 reports regression estimates for hourly O_3 concentrations measured at Parque Dom Pedro II (station 1) on non-holiday weekdays between 01:00 and 06:00, as controls are progressively added under specifications I to VII.

Standard errors on monitor-level pollutant regression coefficients and on cross-monitor mean coefficients are estimated as follows. We draw: (i) 200 bootstrap samples on the pollutant-meteorology-traffic data, where each replication is a bootstrap sample of dates; and (ii) 200 bootstrap samples on the original fuel choice dataset, where each replication is a bootstrap sample of consumers choosing fuel at the pump, from which we predict gasoline choice probabilities for every price vector (using a multinomial probit model, as explained in SM Part A). Bootstrap sampling (ii) generates a different vector of choice probabilities $s_t^{gas,b}$ for every one of 200 bootstrap samples, b = 1, …, 200. We then pair a vector $s_t^{gas,b}$ with a different bootstrap sample on the pollutant-meteorology-traffic data (from (i)) and, for each replication, estimate the station-specific pollutant regression equation. The standard error on an estimated coefficient, for a given station, type and time of day, is then the standard deviation of point estimates across 200 replications. To obtain the standard error on the cross-station mean, for every replication we average point estimates across stations (for the given pollutant, type and time of day) and then take the standard deviation of these cross-monitor means across the 200 replications. In a robustness test reported below, we raise the number of bootstrap samples – on both sets of data, the "first-step" consumer-level choice dataset and the "second-step" date-level pollutant-meteorology-traffic data – from 200 to 500 replications.

Fig. S10 plots estimated changes (under specification VI) in O_3, NO, and CO concentrations for Sundays and Public Holidays at different times of the day, as the gasoline share, s_t^{gas}, rose by 62 percentage points. Fig. S10 is the counterpart to Fig. 5 in the main text, which plots results for non-holiday weekdays. Like Fig. 5, Fig. S10 shows the 95% confidence interval for the "in-sample" effect $0.62\lambda_1$.

The following comments complement the discussion provided in the main text. First, the signs of the estimated effects are robust across non-holiday weekdays and Sundays/Public Holidays – in particular, ozone concentrations decrease in the gasoline share – though confidence intervals are wider for Sundays/Public Holidays, in part because these are based on only one-third as many observations compared to the non-holiday weekday subsample. The estimated coefficient on the gasoline blend change, λ_2, tends to be of the same sign but of lesser magnitude than the estimated coefficient on the share of bi-fuel consumers choosing gasoline over ethanol at the pump, λ_1. This is consistent with the relative magnitudes of the associated fuel mix

variation: our calculations suggest that the mandated gasoline shift from E25 to E20 increased the "pure" gasoline share (excluding anhydrous ethanol) of total gasoline-plus-ethanol combustion from 55% (hypothetically) to 58% (base February 2010). Similarly, the blend change back to E25 in May 2010 lowered the pure gasoline share from 46% (hypothetically) to 43% (base May 2010). This 3 percentage-point change in the fuel mix is moderate compared to the extent of consumer switching at the pump induced by fluctuating relative prices during our sample period.

Second, the "diurnal cycle" of the estimated effects is similar across non-holiday weekdays and Sundays/Public Holidays. Third, across types of day, times of day, and stations, we generally find that meteorology, commuting patterns and fuel mix jointly exhibit high power to predict measured pollutant concentrations. This is indicated by the mean R^2 that are reported in Tables SV to SXII (and R^2 by individual regression in Tables A-S2 to A-S49). The R^2 tend to come in higher for O_3 regressions, e.g., (for specification VI) peaking at 74-80% in the late morning and early afternoon hours; by comparison, R^2 average 57-64% for NO regressions and 62-74% for CO regressions in the late afternoon hours. "Unobserved heterogeneity" in pollutant concentrations in the early afternoon hours relative to midnight tends to be lower for O_3 but higher for NO_X and CO. Importantly, this diurnal cycle in the predictive power of the regression model underscores the importance of specifying each time of day as a unique local environment, thus keeping the model flexible rather than pooling very distinct time observations together into the same regression. Similarly, the predictive power of the regression model tends to be higher – i.e., unobserved heterogeneity tends to be lower – for Sundays/Public Holidays than for non-holiday weekdays. The high but varying R^2 illustrate the relative joint explanatory power across pollutants of the observed meteorological, road traffic, week-of-year (among other) fixed effects, and fuel mix controls, already conditioning on the type of day, time of the day, and location. By construction, specifying regressions at the day type – time – station level *already explained away a large part of the total variation* in pollutant concentrations (see SM Parts C, D and E).

A final comment relates not to regression estimates but to differences in recorded pollutant concentrations on non-holiday weekdays and Sundays/Public Holidays, and our interpretation of these differences. Tables SV to SXII provide mean concentrations (values for the dependent variables in the regressions) detailed by type of day and time of day. Ozone concentrations tend to be higher on Sundays/Public Holidays, when there are less vehicles on the

road as evidenced by traffic congestion data, compared to non-holiday weekdays. For example, in our sample mean O_3 readings between 13:00 and 16:00 are 80 μg/m^3 on Sundays/Public Holidays and 68 μg/m^3 on non-holiday weekdays. By contrast, NO_X and CO concentrations tend to be lower on Sundays/Public Holidays compared to non-holiday weekdays. For example, between 07:00 and 10:00, NO and CO concentrations respectively average 24 μg/m^3 and 0.73 ppm on Sundays/Public Holidays, but 67 μg/m^3 and 1.20 ppm on non-holiday weekdays. We take this descriptive evidence that on weekdays NO_X and CO concentrations are higher, but ozone pollution is less severe, than on weekends as being consistent with the overall interpretation of our findings, that O_3 production over the São Paulo metropolis may be hydrocarbon-limited. Higher NO_X emissions from weekday road traffic is associated with lower ambient O_3, much as air monitoring stations located right by roads tend to record higher NO (and CO) and lower O_3 concentrations compared to stations that are somewhat removed from roads.

4. Other measured pollutants: SO_2 and PM10. Similar to Fig. 5 in the main text and Fig. S10, Figs. S11 and S12 report mean estimated effects on SO_2 and PM10 concentrations, recorded at 4 and 15 stations respectively, from in-sample gains in the gasoline share, $0.62\lambda_1$, by type of day and time of day. Raising the gasoline share at the expense of ethanol is associated with increases in ambient levels of SO_2 but reductions in PM10, though the estimated effects are mostly statistically insignificant. Compared with O_3, NO_X, and CO, the regression model's explanatory power for SO_2 and PM10 tends to be lower, with maximal R^2 of around 40-45% for SO_2 regressions and 50-55% for PM10 (again higher on Sundays/Public Holidays than for non-holiday weekdays).

5. Robustness tests. Tables SXIII to SXVII report several robustness tests that were motivated in the preceding sections and in the main text. For brevity, for most robustness tests we report cross-monitor mean fuel mix estimates, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, for O_3 regressions in the early afternoon on non-holiday weekdays.

- Table SXIII: The dependent variable is the **natural logarithm** of the sum **of measured concentration** and 0.001, i.e., ln([O_3]+0.001), rather than concentration [O_3] in μg/m^3 (we shift measures by a low 0.001 to deal with very low measured concentrations).

- Table SXIV: The top panel of Table SXIV reproduces mean fuel mix estimates for our baseline specifications (bottom panel of Table I in the main text). Robustness tests reported in the table are:
    - Robustness (a): Specify a **quadratic trend** in the date rather than a linear one.
    - Robustness (b): Specify **meteorological covariates** directly **in their reported units** rather than the logarithmic transform.
    - Robustness (c): Include **atmospheric pressure** in the set of meteorological covariates.
    - Robustness (d): Specify **traffic congestion** covariates directly **in their reported units** rather than the logarithmic transform.
    - Robustness (e): Additionally control for the **median inverse traffic speed** (travel time) recorded at 08:30 (morning rush time when traffic speeds are measured) across 36 monitored road segments, interacted with day-of-week dummies for non-holiday weekdays.
- Table SXV: Robustness tests reported in Table SXV are:
    - Robustness (f): Expand sample to include **weekdays during the school vacation fortnight** (typically starting December 24).
    - Robustness (g): Expand sample to include **the colder months of June to September**.
    - Robustness (h): Control for the real **price of diesel** (Fig. S2).
    - Robustness (i): Control for **ridership on the public transport system** in the São Paulo metropolis (Fig. S8, a monthly series of total ridership divided by the number of non-holiday weekdays each month).
    - Robustness (j): Control for **physical industrial production** in the state of São Paulo (Fig. S9, a monthly series).
- Table SXVI: Robustness tests reported in Table SXVI are:
    - Robustness (k): Control for the São Paulo metropolitan area's **"economically active" population** in millions, according to IBGE's Monthly Employment Survey (*Pessoas economicamente ativas*, *Pesquisa Mensal de Emprego*).[85]
    - Robustness (l): Control for the São Paulo metropolitan area's **work force** in millions, according to IBGE's Monthly Employment Survey (*Pessoas ocupadas*,

Pesquisa Mensal de Emprego).[86] The correlation coefficient between this series and the economically active population from November 2008 to July 2011 is 0.73.

- Robustness (m): Control for the São Paulo metropolitan area's **mean real earnings**, according to IBGE's Monthly Employment Survey (*Rendimento médio real do trabalho principal*, *Pesquisa Mensal de Emprego*).[87]
- Robustness (n): Base the gasoline share on predictions of a **multinomial logit choice model** rather than a multinomial probit model (Fig. 3 in the main text).
- Robustness (o): Base the gasoline share on **aggregate monthly reports of gasoline and ethanol shipments by wholesalers** for the state of São Paulo (Fig. 1, panel a in the main text). The correlation coefficient between this series and the baseline gasoline share, based on consumer demand predicted from weekly prices, is a tight 0.84 (November 2008 to July 2011), which explains the robustness of our baseline estimates. As stated in the notes to Table SXVI, this robustness test does not require a first-step correction for the gasoline share as the variable is now based on data rather than on an estimate.

- Table SXVII: The following robustness tests are reported:
  - Robustness (p): Increase the **number of bootstrap samples** on both consumer choice and pollutant-meteorology-traffic datasets from 200 to 500 replications.
  - Robustness (q): Predict s_t^{gas} using seven-day, rather than four-day, lagged prices at the pump, thus assuming a **large volume of fuel stored in vehicles' tanks**.
  - Robustness (r): **Instrument** for s_t^{gas} using the median ethanol-to-gasoline price ratio p_e/p_g.

6. A placebo test. We briefly perform an (admittedly coarse) "placebo test" of the effect of the fuel mix on ambient ozone levels. We implement this test on the non-holiday weekday, early afternoon subsample, for which we have more observations (relative to Sundays/Public Holidays) and when ozone readings tend to be higher (relative to other times of the day); further, as reported above, our regression model (specification VI) has higher explanatory power in the early afternoon.

Consider the null hypothesis that the gasoline versus ethanol fuel mix had no effect on ozone concentrations. In particular, under the null, $\lambda_1 = \lambda_2 = 0$ in regression equation (1) of the main text. We now re-estimate this constrained model, again separately for each of the 12 O_3-monitoring stations, but to the constrained specification ($\lambda_1 = \lambda_2 = 0$) we include six indicator variables defined around the weeks in which ethanol prices peaked in January/February 2010 and in March/April 2011 but were, by contrast, stable in early 2009:

$$\begin{aligned} concentrat\ ion_t &= 0 + \delta_1 gas\_2010_t + \delta_2 etoh\_2010_t + \delta_3 gas\_2011_t + \delta_4 etoh\_2011_t \\ &+ \delta_5 i\_gas\_2009_t + \delta_6 i\_etoh\_2009_t + W_t'\Delta^W + T_t'\Delta^T + fixedeffec\ ts_t + trend_t + \varepsilon_t. \end{aligned} \tag{1'}$$

Specifically, the indicator variables we introduce to the constrained version of (1) are such that:

1. gas_2010_t takes the value 1 if observation *t* falls in the period January 16, 2010 to March 5, 2010, when the gasoline share of bi-fuel vehicles is predicted to have exceeded 0.5, and 0 otherwise (see Fig. 3, panel b in the main text);
2. $etoh\_2010_t$ takes the value 1 if observation *t* falls in the dates immediately *surrounding* the period indicated by gas_2010_t, namely November 15, 2009 to January 15, 2010 (prior to the first ethanol price peak) or March 6, 2010 to May 6, 2010 (after the peak), and 0 otherwise;
3. gas_2011_t takes the value 1 if observation *t* falls in the period February 13, 2011 to May 14, 2011, when the predicted gasoline share again exceeded 0.5, and 0 otherwise;
4. $etoh\_2011_t$ takes on the value 1 if observation *t* falls in the dates immediately surrounding the period indicated by gas_2011_t, namely November 12, 2010 to February 12, 2011 or May 15, 2011 to May 31, 2011, and 0 otherwise (recalling that our regression sample comprises the months of October through May, thus the cut-off on May 31).

Under the null, ozone concentrations recorded in January/February 2010 should not have been different, all else equal, to concentrations observed in the preceding or subsequent weeks when the gasoline share was lower, i.e., $\delta_1 - \delta_2 = 0$. Similarly, under the null, ozone concentrations in March/April 2011 should have been similar to concentrations recorded in the surrounding weeks, i.e., $\delta_3 - \delta_4 = 0$. In contrast to the two years that followed, ethanol prices at the turn of 2008/09 were relatively stable, which suggests the following placebo. Define an additional pair of dummy variables representing "*imaginary* phases" for gasoline and ethanol at that time:

5. $i\_gas\_2009_t$ based on the same cut-off dates used to define gas_2010_t but turned back exactly one year; so, for example, $i\_gas\_2009_t$ takes the value 1 for February 2009 observations despite gasoline *not* being favorably priced relative to ethanol, and thus the gasoline share being a low 0.2, at that point in time (gasoline was favorably priced a year later, in February 2010, with the gasoline share reaching 0.6);
6. $i\_etoh\_2009_t$ based similarly on the cut-off dates used to define $etoh\_2010_t$ though turned back exactly one year.

Under both the null hypothesis that the fuel mix had no effect on ozone concentrations, as well as the alternative hypothesis that it did have an effect, we expect $\delta_5 - \delta_6 \approx 0$. This is because relative ethanol prices were stable between November 2008 and May 2009, with p_e/p_g at about 55% (see Fig. 1, panel b in the main text), compared to the same calendar months a year later, and another year after that, when they fluctuated sharply.

We obtain that ozone concentrations in the dates indicated by gas_2010_t were statistically significantly lower than in the dates indicated by $etoh\_2010_t$. Specifically, the mean estimate for $\delta_1 - \delta_2$ across 12 station-specific regressions (each estimated on the non-holiday weekday, 13:00 to 16:00 subsample) is -8.6 μg/m^3, with standard error of 3.6 μg/m^3. Similarly, we find that ozone concentrations in the dates indicated by gas_2011_t were statistically significantly lower than in the dates indicated by $etoh\_2011_t$, with mean $\delta_3 - \delta_4$ estimated at -8.3 μg/m^3 with standard error of 3.1 μg/m^3. By contrast, ozone concentrations in early 2009 did not vary as they did in the subsequent years: mean $\delta_5 - \delta_6$ is estimated at +0.1 μg/m^3 with standard error 3.8 μg/m^3. We compute standard errors on cross-monitor mean estimates similarly to above, noting that there is no correction for a first step model, since (1') does not include an imputed variable. In conclusion, we reject the null hypothesis that a varying fuel mix did affect on ozone concentrations, against the alternative that recorded ozone levels were lower as ethanol prices peaked and bi-fuel vehicle motorists shifted to gasoline. (We note, however, that these indicator variables do not best capture the feature that consumer fuel switching occurs gradually as relative prices vary, as their values jump discretely at the cut-off dates.)

Fig. S1. The median per-liter price of regular ethanol (E100, thick green line) and the median per-liter price of regular gasoline (E25 or E20, thin red line) in large weekly surveys of retail fueling stations in the city of São Paulo, during specific weeks in our sample: Ethanol prices peaked in January/February 2010 (**A**) and again in March/April 2011 (**B**). Prices are in Brazilian Reais (denoted R$) per liter, at constant July 2011 Brazil CPI (*IPCA Brasil*) terms, adjusting for inflation of about 6% per year. We also plot the price of gasoline multiplied by the 70% "parity" ratio, the widely reported ratio at which $/mile equalizes across the two substitute fuels (dashed red line). Sources: ANP, IBGE.

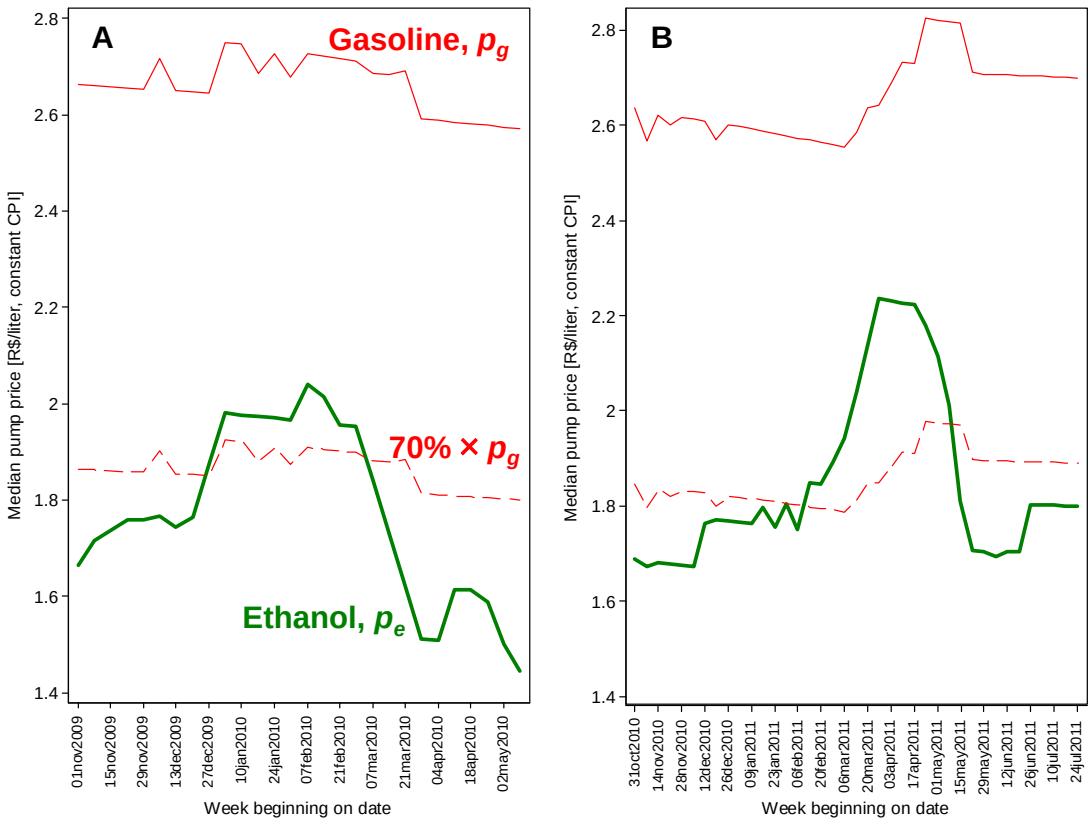

Fig. S2. Monthly price indices for regular ethanol (E100, thick green line), regular gasoline (E25 or E20, crossed red line), and diesel oil (dashed black line) at the pump in the São Paulo metropolitan area, from October 2008 to July 2011. In the top panel, price indices are deflated to account for variation in the economy-wide price level (*IPCA Brasil*). In the bottom panel, price indices are not adjusted for inflation. Base October 2008 = 100. Source: IBGE.

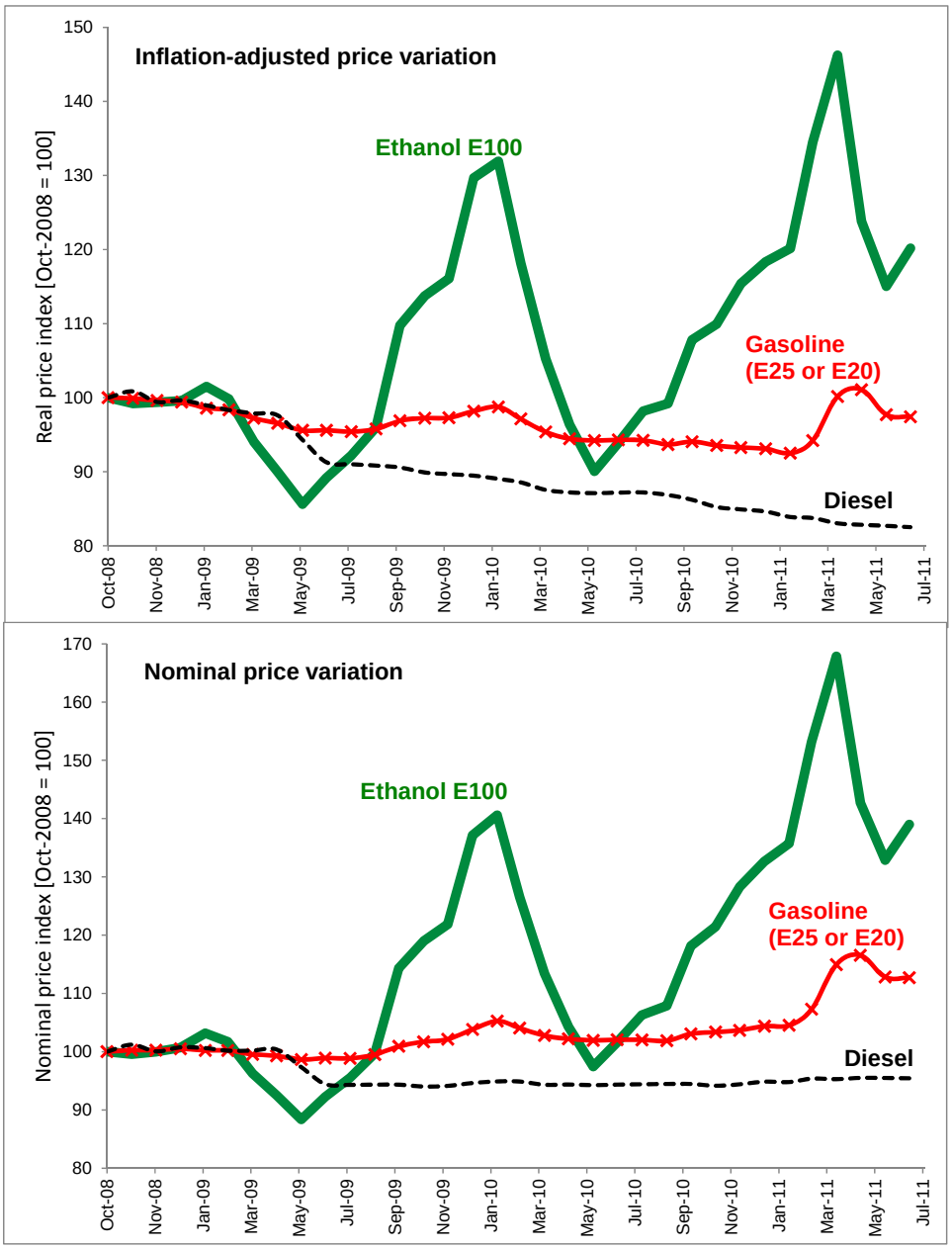

Fig. S3. Monthly shipments of gasoline (E25 or E20) and ethanol (E100), reported by wholesalers for São Paulo state, converted from cubic meters to "vehicle kilometers traveled." On preparing the figure, we ignore that gasoline is also used by motorcycles, thus overstating gasoline's share of light-duty distance traveled, particularly at lower values. The solid blue line (right vertical axis) indicates estimated billion vehicle kilometers driven on either gasoline or ethanol. The marked red line (left vertical axis) indicates gasoline's share of total distance traveled. Sources: ANP, Salvo and Huse (2013), authors' calculations.

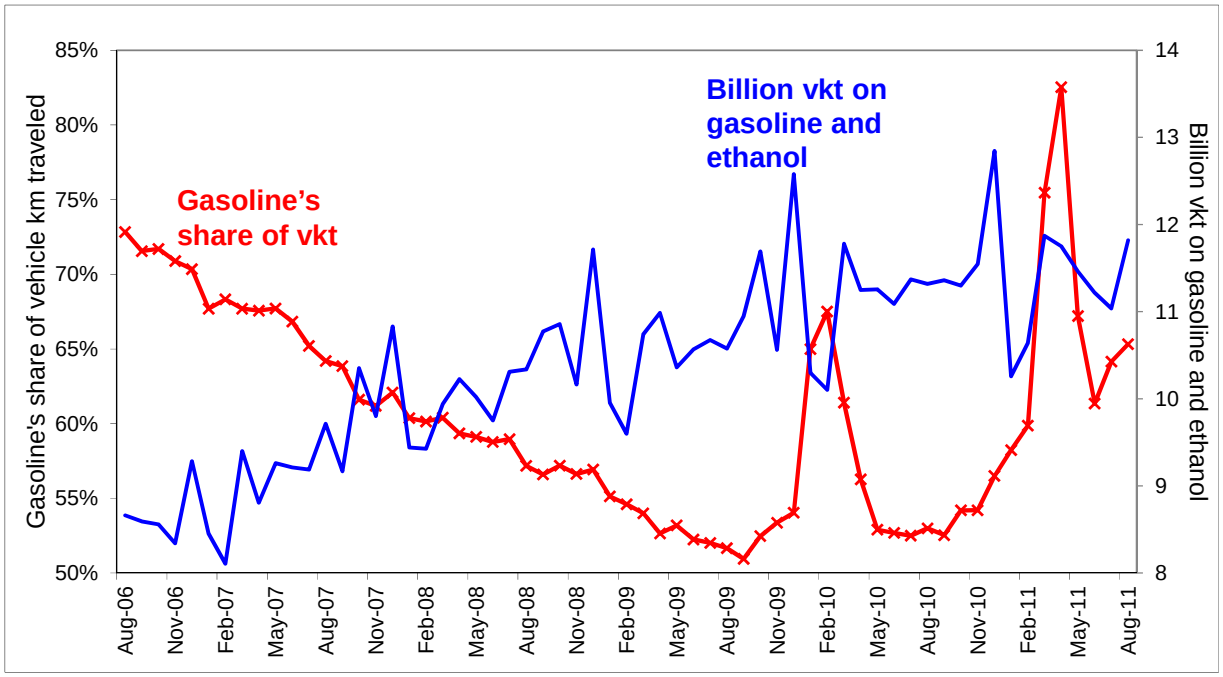

Fig. S4. (Top to bottom, on the left vertical axis) Radiation, precipitation, relative humidity, and wind speed. (Top to bottom, on the right vertical axis) Temperature and wind direction. Hourly observations from November 1, 2008 to July 31, 2011; see SM Part D for location of measurement. Sources: CETESB, INMET.

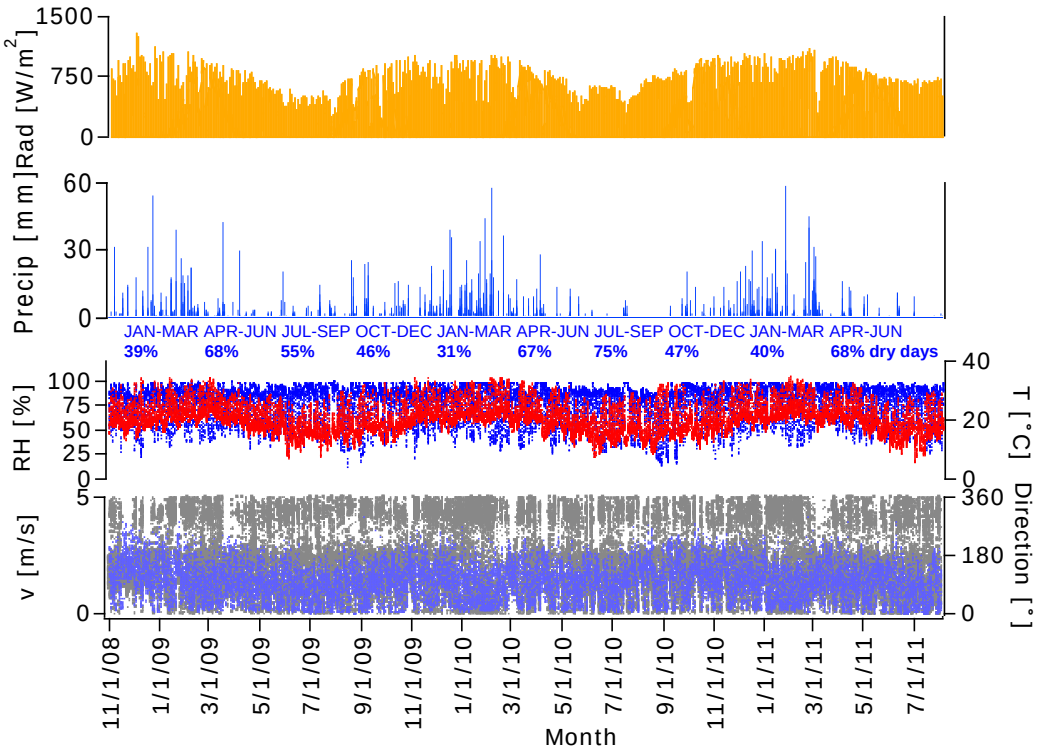

Fig. S5. Scatterplots of hourly concentrations of NO (brown) and CO (gray) versus contemporaneous traffic congestion, for the Congonhas (**A**) and Taboão da Serra (**B**) air monitoring stations. An observation is a station-hour pair between 07:00 and 09:00 or between 18:00 and 21:00. The sample period is November 1, 2008 to May 31, 2011, excluding the colder months of June to September. Traffic congestion for Congonhas is the total km recorded in the South region of the city; congestion for Taboão da Serra is the total km recorded in the West and South regions of the city. To amplify the plots over the relevant (linear) range we cut them off above 40 km of congestion in the South region (Congonhas) and 70 km of congestion in the South and West regions (Taboão da Serra). Source: CETESB, CET.

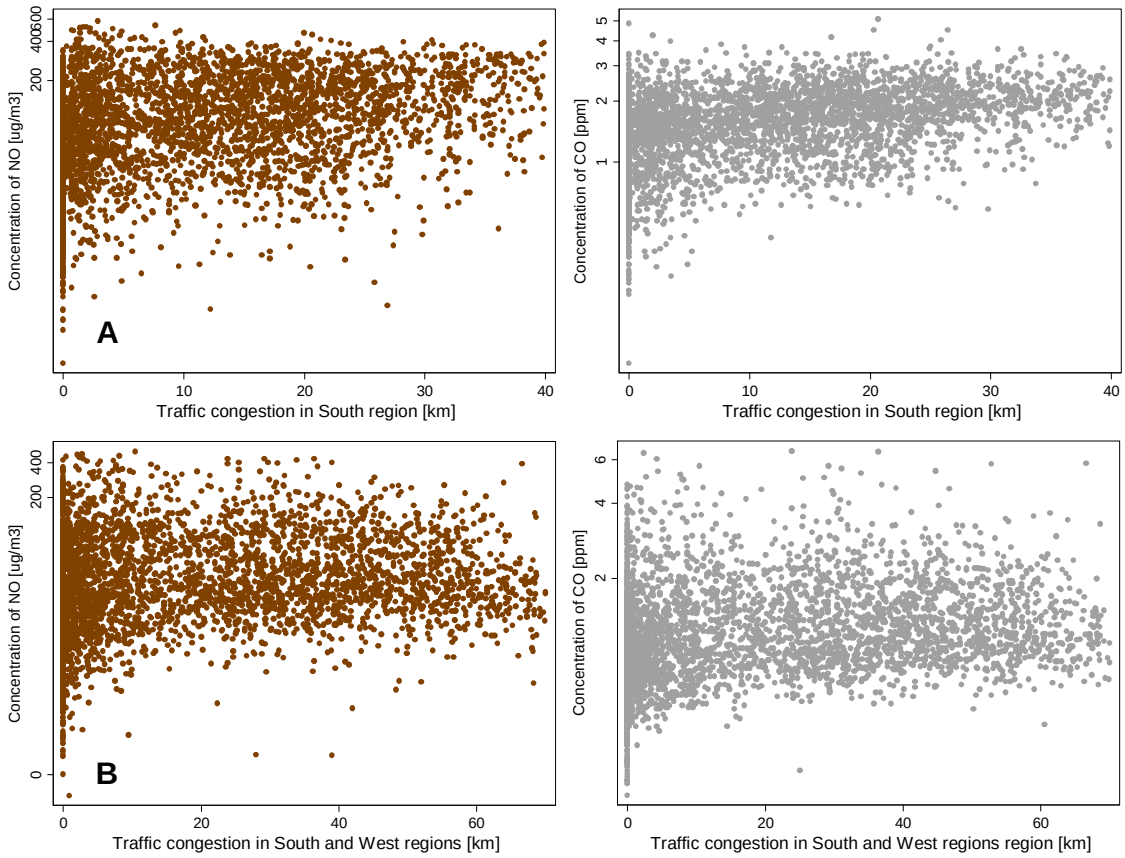

Fig. S6. Citywide traffic congestion in km on Monday, April 18, 2011 to Wednesday, April 20, 2011 (**left**) and Wednesday, February 16, 2011 (**right**): Actual recorded congestion (black), predicted congestion under actual fuel prices and actual precipitation (dotted blue, 95% C.I. shown), and predicted congestion under counterfactual fuel prices (**left**, dashed red, 95% C.I. shown) or counterfactual precipitation (**right**, dashed red, 95% C.I. shown). Predictions (actuals and counterfactuals) computed from OLS regression estimates as discussed in the text. Counterfactual fuel prices (**left** only) are those observed in the week of November 7, 2010, in particular, the hypothetical pump price of ethanol is R$ 1.67 compared to an actual price of R$ 2.22 in the week of April 18, 2011. Counterfactual precipitation (**right** only) is 0 mm compared to an actual 30 mm between 17:00 and 19:00 on February 16, 2011.

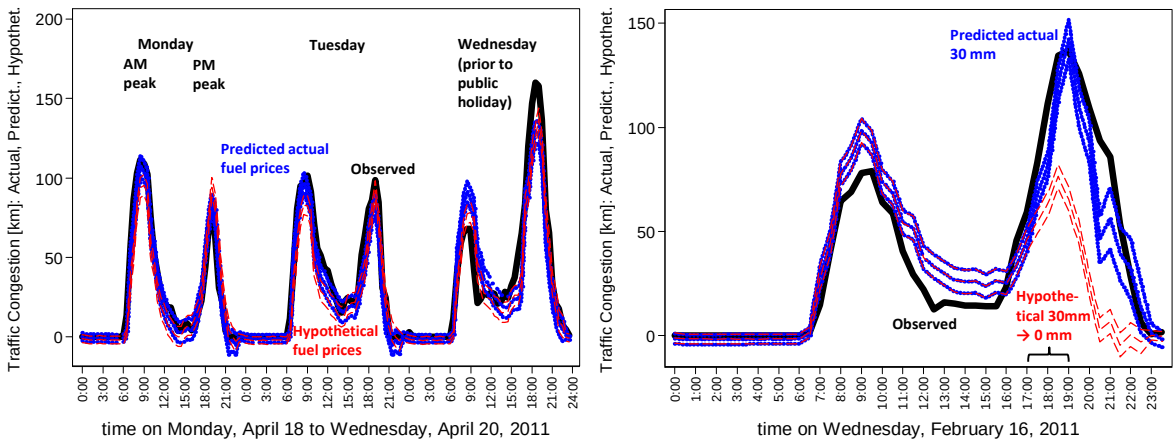

Fig. S7. 75th percentile (**A**), median (**B**), and 25th percentile (**C**) of the distribution of inverse traffic speed, in hours per kilometer (hpk), recorded along 36 road segments located across the city, at 08:30 ("AM") and again at 18:00 ("PM"), on weekdays between Monday, April 4, 2011 and Wednesday, April 2011: Actual recorded inverse speed (black), predicted inverse speed under actual fuel prices (dotted blue, 95% C.I. shown), and predicted inverse speed under counterfactual fuel prices (dashed red, 95% C.I. shown). Predictions (actuals and counterfactuals) computed from OLS regression estimates as discussed in the text. Counterfactual fuel prices are those observed on November 8, 2010.

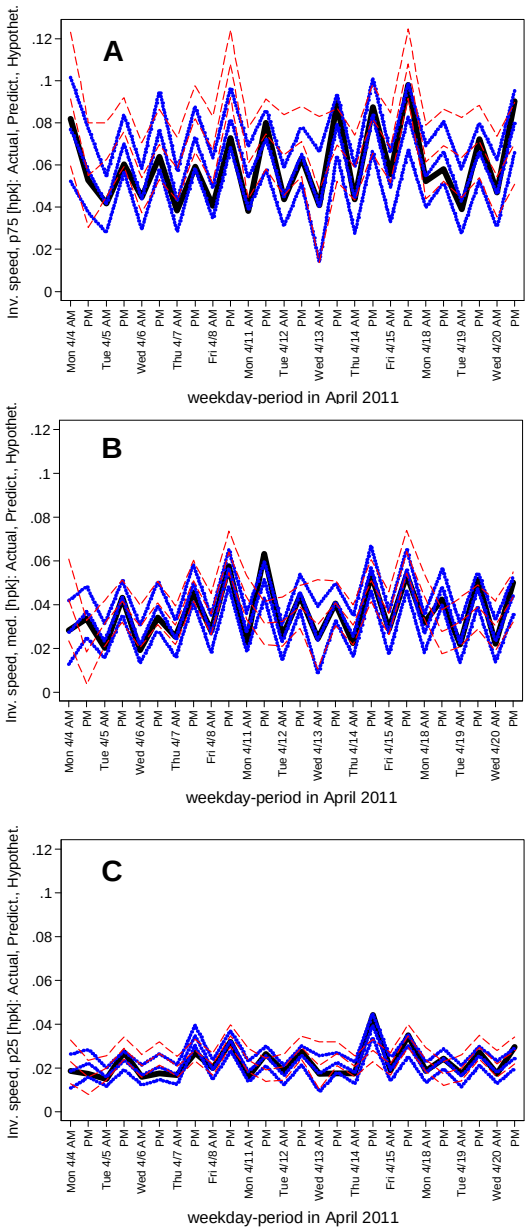

Fig. S8. Ridership on the public transport system in the São Paulo metropolitan area (left vertical axis) and average ethanol-to-gasoline price ratio across the city's pumps (dashed green line, right vertical axis), by month from November 2008 to July 2011. Average daily rates of ridership in millions of passengers per day are reported by dividing the month's total ridership: (i) by the number of calendar days in the month (lower black line), and (ii) by the number of non-holiday weekdays (upper black line). Sources: SPTrans (Prefeitura de São Paulo, Transportes), ANP.

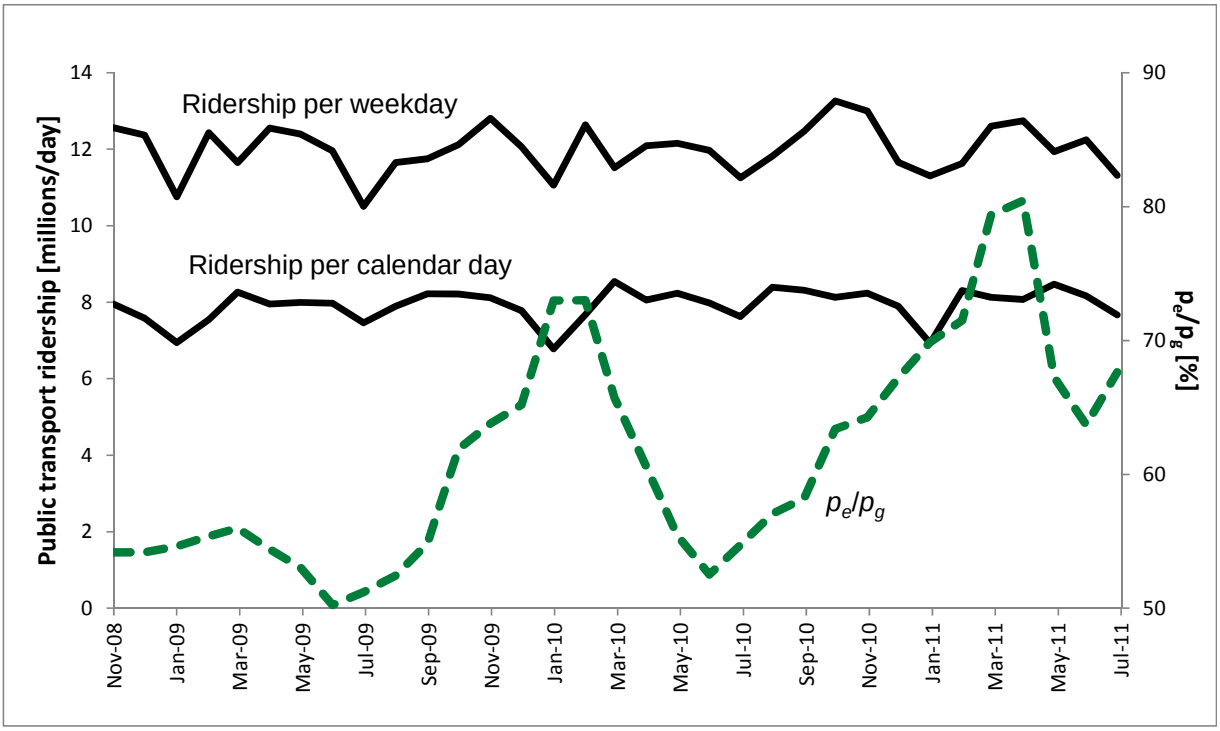

Fig. S9. Index of physical industrial production for the state of São Paulo (thick black line, left vertical axis) and (the mean across indices) for twelve other relatively industrialized states of Brazil (black line with markers), alongside the average ethanol-to-gasoline price ratio in the city of São Paulo (dashed green line, right vertical axis), by month from November 2008 to July 2011. Industrial production indices are normalized at 100 in mid 2002. Sources: IBGE (Pesquisa Industrial Mensal, Produção Física Regional), ANP

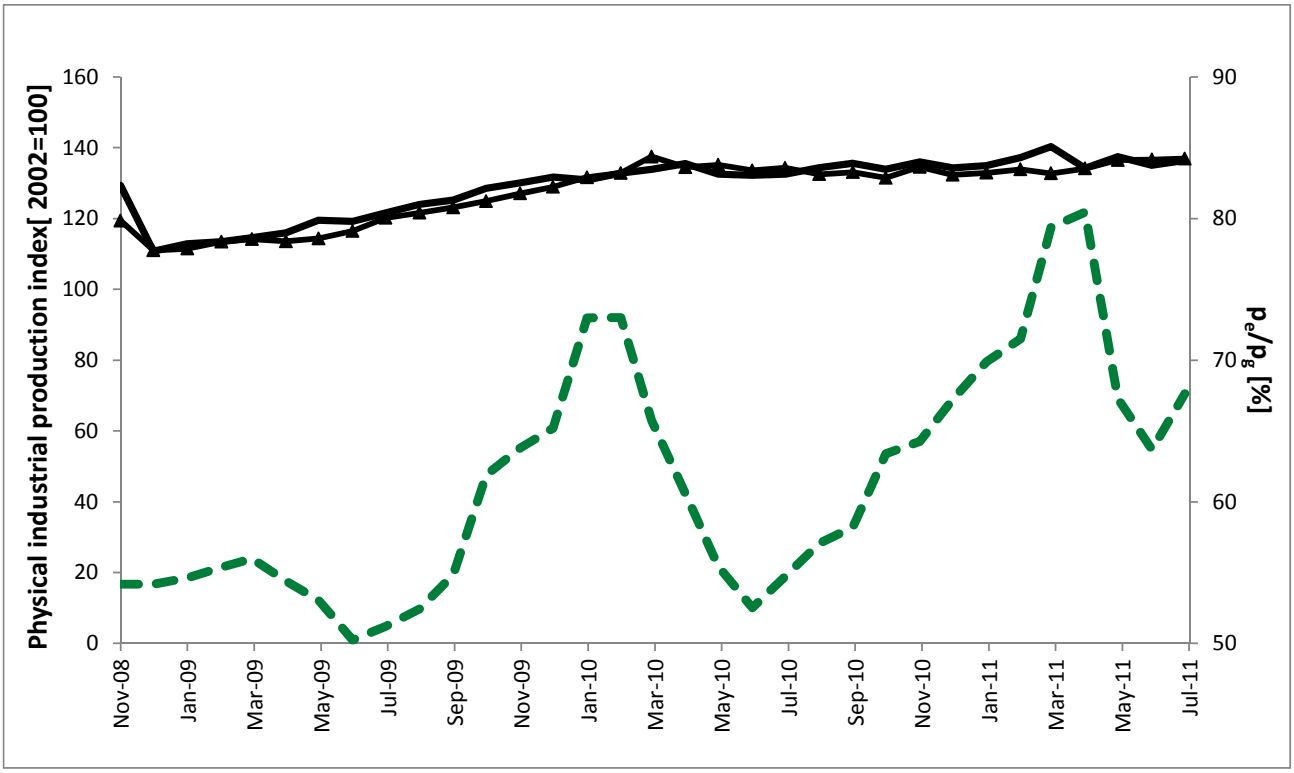

Fig. S10. Estimated changes in O_3, NO, and CO concentrations as the gasoline share, s_t^{gas}, rose by 62 percentage points. Mean effects across regressions for the stations monitoring each given pollutant, for the different times of a Sunday/Public Holiday. The left axis plots the 95% confidence intervals for (the mean across monitors of) $0.62\hat{\lambda}_1$ (bars) and the right axis expresses these confidence intervals as proportions of mean recorded concentrations (dashed lines) at the different times of a Sunday/Public Holiday. Source: Specification VI estimates. See the footnote to Table I on how the error bars were obtained.

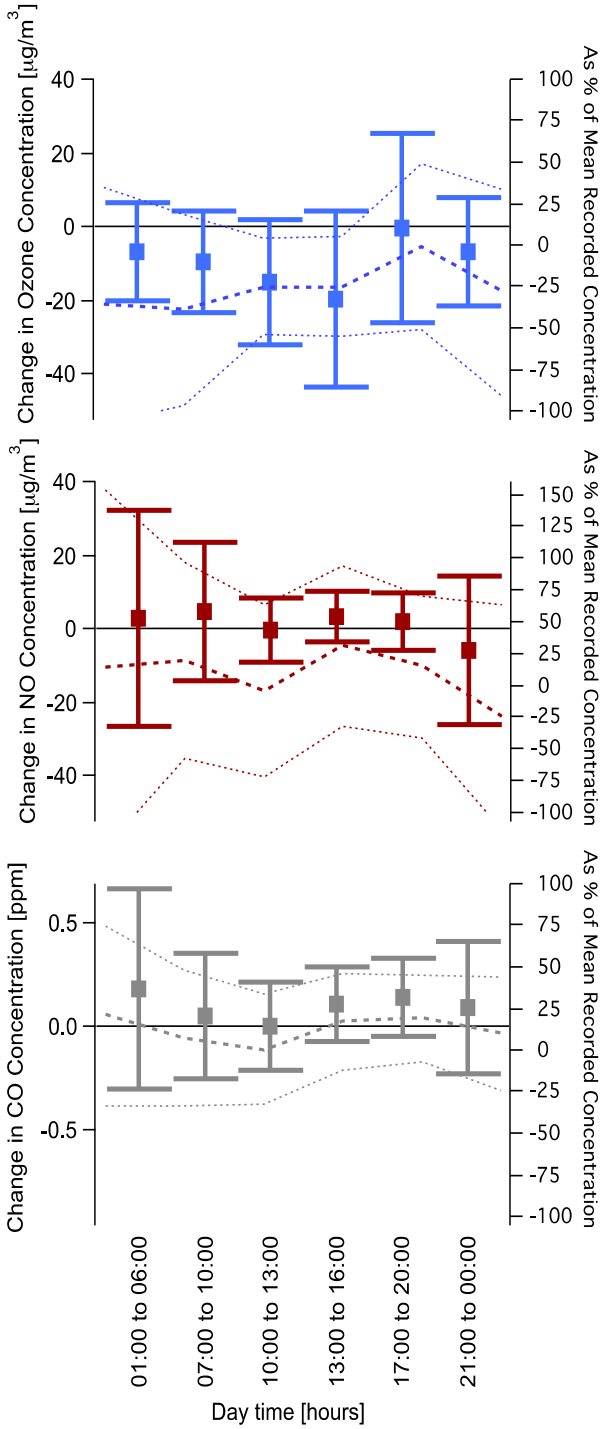

Fig. S11. Estimated changes in SO_2 and PM10 concentrations as the gasoline share, s_t^{gas}, rose by 62 percentage points. Mean effects across regressions for the stations monitoring each given pollutant, for the different times of a non-holiday weekday. The left axis plots the 95% confidence intervals for (the mean across monitors of) $0.62\hat{\lambda}_1$ (bars) and the right axis expresses these confidence intervals as proportions of mean recorded concentrations (dashed lines) at the different times of a non-holiday weekday. Source: Specification VI estimates. See the footnote to Table I on how the error bars were obtained.

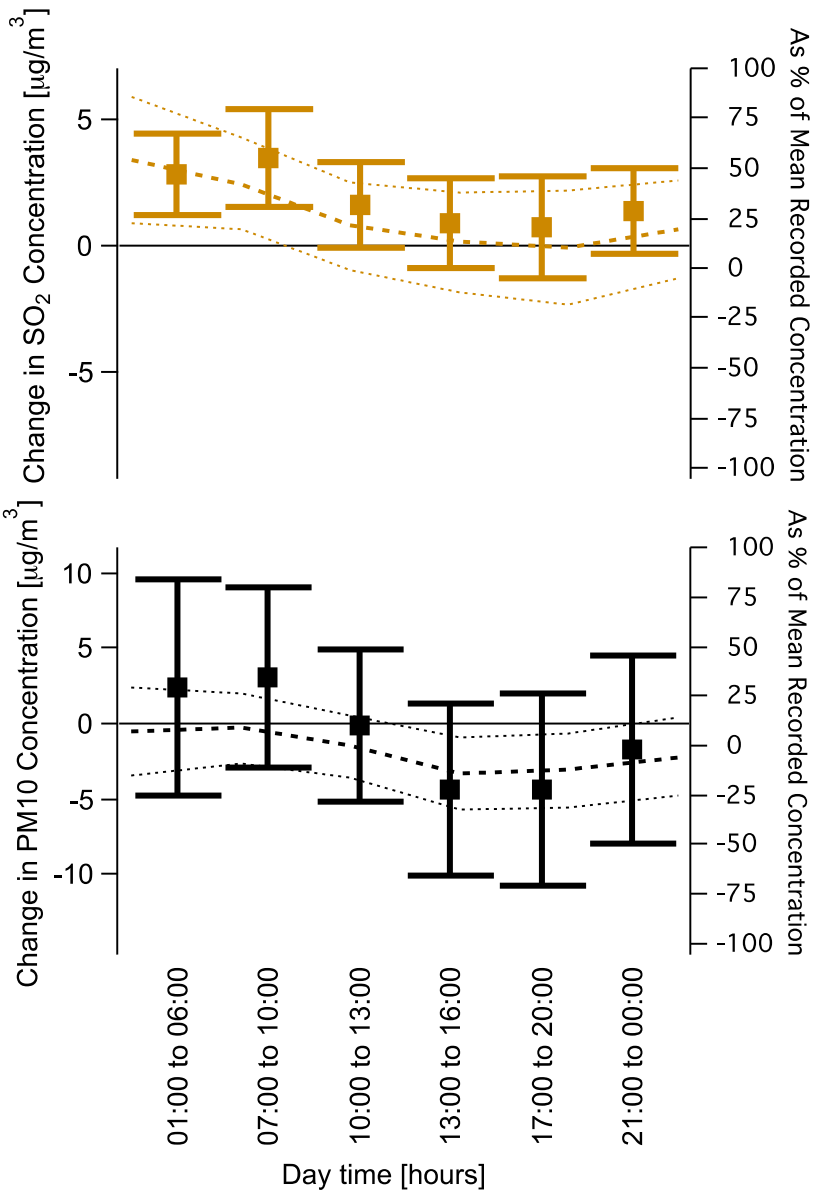

Fig. S12. Estimated changes in SO_2 and PM10 concentrations as the gasoline share, s_t^{gas}, rose by 62 percentage points. Mean effects across regressions for the stations monitoring each given pollutant, for the different times of a Sunday/Public Holiday. The left axis plots the 95% confidence intervals for (the mean across monitors of) $0.62\hat{\lambda}_1$ (bars) and the right axis expresses these confidence intervals as proportions of mean recorded concentrations (dashed lines) at the different times of a Sunday/Public Holiday. Source: Specification VI estimates. See the footnote to Table I on how the error bars were obtained.

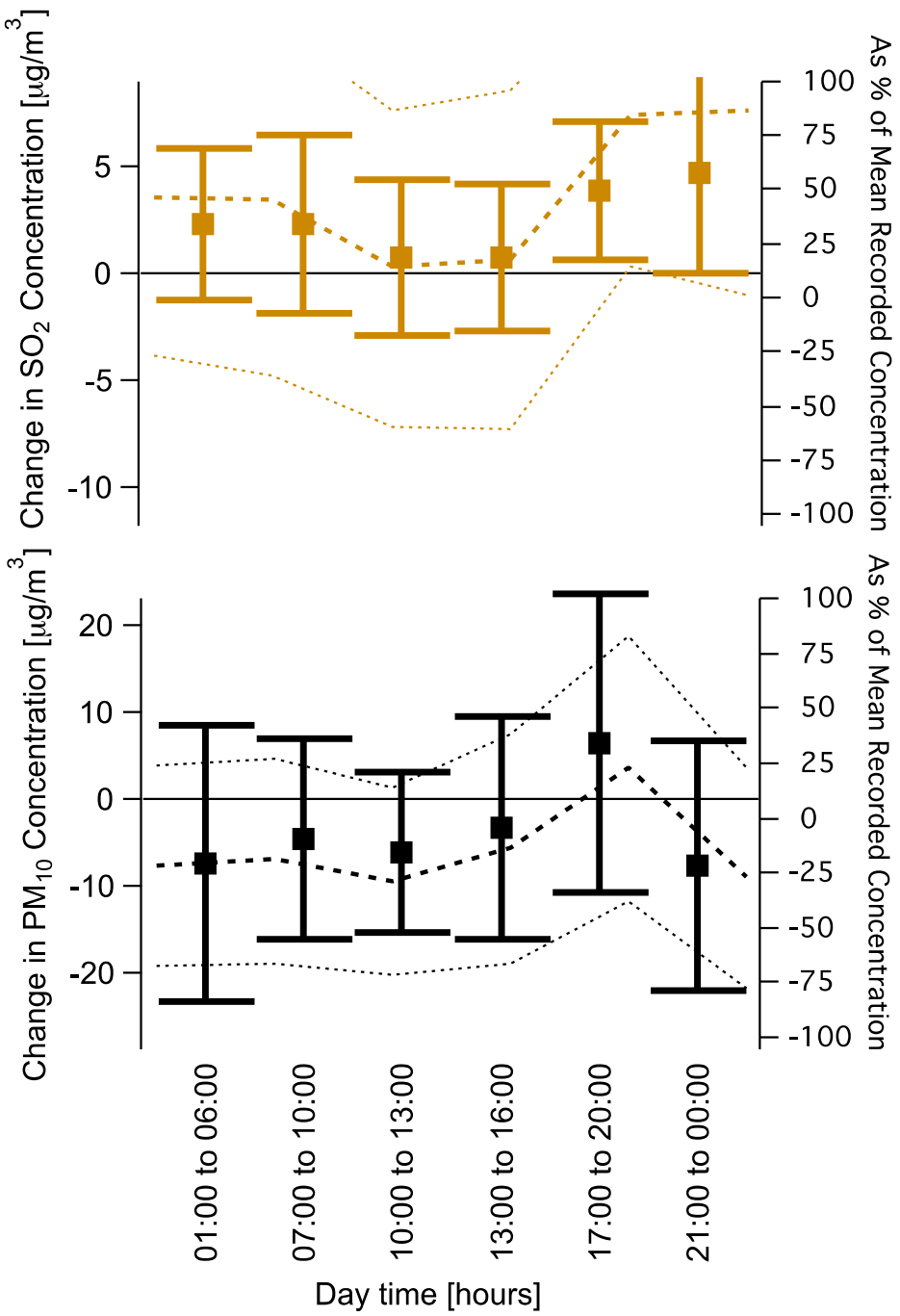

Table SI. Estimated coefficients and standard errors for two consumer-level fuel demand models. The first column reports multinomial probit estimates from Salvo and Huse (2013), Table 2, p. 261, specification III. The second column reports multinomial logit estimates on the same data to indicate robustness with respect to a specification with additional interactions.

Table SI. Estimated coefficients and standard errors for two consumer-level fuel demand models.

Specification:		Multinomial Probit Salvo-Huse (2013)		Multinomial Logit (Robustness)	
		coeff.	s.e.	coeff.	s.e.
Price of fuel (R$/km)		-19.81	(1.47)	-209.6	(51.6)
Price of fuel (R$/km) * Extensive user (DV)				-0.97	(3.52)
Price of fuel (R$/km) * log vehicle price (R$)				17.53	(4.93)
Gasoline fixed effect	*g*	-0.43	(0.47)	-5.30	(2.11)
Midgrade gasoline fixed effect	*midg*	-1.83	(1.01)	-11.12	(2.98)
Female consumer (DV)	× *g*	0.12	(0.09)	0.15	(0.11)
	× *midg*	-0.27	(0.18)	-0.39	(0.20)
Consumer aged 25 to 40 years (DV)	× *g*	0.15	(0.15)	0.22	(0.16)
	× *midg*	0.42	(0.22)	0.70	(0.36)
Consumer aged 40 to 65 years (DV)	× *g*	0.16	(0.14)	0.23	(0.17)
	× *midg*	0.50	(0.27)	0.86	(0.36)
Consumer aged more than 65 years (DV)	× *g*	0.87	(0.28)	1.17	(0.32)
	× *midg*	1.21	(0.38)	1.81	(0.53)
Consumer attained secondary school (and no more) (DV)	× *g*	-0.15	(0.18)	-0.21	(0.21)
	× *midg*	-0.10	(0.21)	-0.22	(0.35)
Consumer is college educated (DV)	× *g*	-0.12	(0.18)	-0.17	(0.20)
	× *midg*	0.03	(0.23)	0.00	(0.33)
Consumer is an extensive vehicle user (DV)	× *g*	0.26	(0.09)	0.33	(0.14)
	× *midg*	0.33	(0.14)	0.46	(0.19)
Consumer drives an expensive vehicle (DV)	× *g*	0.11	(0.09)		
	× *midg*	0.21	(0.12)		
Log vehicle price (R$)	× *g*			0.44	(0.21)
	× *midg*			0.86	(0.30)
Value of 12 vehicles sampled at retailer (R$m)	× *g*	-0.42	(1.07)	-0.06	(1.16)
	× *midg*	-0.03	(1.65)	-0.94	(1.91)
Number of consumers		2160		2160	
Total number of alternatives		6288		6288	
σ_{midg}		1.37	(1.47)		
$\rho_{e,midg}$		0.56	(0.62)		

Notes: An observation is an alternative that a bi-fuel motorist faces among regular gasoline (always available), ethanol (always available) and midgrade gasoline (when available at the pump). "DV" denotes a dummy variable. Robust standard errors (s.e.), clustered by fuel retailer-day, in parentheses. Maximum Likelihood estimates.

Table SII. Mean selected pollutant concentrations and concentration ratios across air monitoring stations. The sample period is November 1, 2008 to May 31, 2011, excluding the colder months of June to September. See Table A-S1 for data availability by pollutant-station pair. Source: Authors' calculations based on CETESB hourly measures.

Station name (Station ID)	**[O_3]** (μg/m^3) Mean over dates of mean within date (13:00-16:00)	**[NO]** (μg/m^3) Mean over dates of mean within date (07:00-09:00)	**[NO_X]** (ppb) Mean over dates of mean within date (07:00-09:00)	**[CO]** (ppm) Mean over dates of mean within date (07:00-09:00 & 18:00-21:00)	**[NO]:[NO_2]** Mean over dates of ratio of within-date means*	**[NO]:[O_3]** Mean over dates of ratio of within-date means**
Parque Dom Pedro II (1)	64.10	53.19	70.69	0.79	0.94	1.77
Santana (2)	71.03	-	-	-	-	-
Moóca (3)	67.60	-	-	0.62	-	-
Ibirapuera (5)	88.05	21.13	37.02	0.73	0.50	0.28
Nossa Senhora do Ó (6)	65.87	-	-	-	-	-
São Caetano do Sul (7)	84.53	47.68	61.55	0.96	1.04	0.67
Congonhas (8)	-	112.64	120.69	1.56	1.79	-
Cerqueira César (10)	-	81.80	91.84	1.00	1.61	-
Diadema (15)	65.94	-	-	-	-	-
Osasco (17)	-	-	-	1.87	-	-
Santo André-Capuava (18)	72.88	-	-	-	-	-
Taboão da Serra (20)	-	82.40	88.51	1.28	2.04	-
Mauá (22)	73.36	20.65	32.80	-	0.59	0.32
Parelheiros (29)	57.31	-	-	0.83	-	-
Pinheiros (27)	63.30	72.68	79.17	1.09	2.09	1.70
IPEN-USP (31)	84.82	26.67	37.50	0.50	0.83	0.35

* Mean of (same-date 07:00-09:00 mean [NO] / same-date 07:00-09:00 mean [NO_2]), both numerator and denominator measured in μg/m^3; ** Mean of (same-date 07:00-09:00 mean [NO] / same-date 13:00-16:00 mean [O_3]), both numerator and denominator measured in μg/m^3

Table SIII. Hourly data on meteorological conditions: Source, location, and availability in the raw data. The sample period is November 1, 2008 to May 31, 2011, excluding the colder months of June to September. Availability is computed as the proportion of the maximum number of possible hourly measurements, namely 698 days × 24 hourly values per day, equaling 16,752 values.

Meteorological variable	Unit	Institution	Location | availability (%)
Temperature	degrees Celsius	CETESB	Air monitoring stations (see Fig. 2A of the main text): - Pinheiros, entire sample | 97% - São Caetano do Sul, entire sample | 97% - Taboão da Serra, entire sample | 96%
Relative humidity	%	CETESB	Air monitoring stations: - Pinheiros, until May 2010 | 98% - São Caetano do Sul, entire sample | 97% - Taboão da Serra, entire sample | 96%
Radiation	W/m^2	CETESB	Air monitoring stations: - Ibirapuera, until May 2010 | 97% - Paulínia (outside São Paulo*), entire sample | 93%
Atmospheric pressure	hPa	CETESB	Air monitoring stations: - Ibirapuera, entire sample | 94%
Wind speed	m/s	CETESB	Air monitoring stations: - Ibirapuera, entire sample | 95% - Moóca, from Feb 2009 on except Nov 2010 | 97% - Osasco, entire sample | 96% - Pinheiros, entire sample | 98% - Santana, entire sample | 97%
Wind direction	degrees from North (clockwise, direction from which it originates)	CETESB	Air monitoring stations: - Ibirapuera, entire sample | 94% - Osasco, entire sample | 96% - Pinheiros, entire sample | 98% - Santana, entire sample | 97%
Precipitation	mm	INMET	A701 Santana (Lat. 23°30'S, Lon. 46°37'W), close to CETESB's Santana station, entire sample | 99%

* Because the only station measuring radiation in the São Paulo metropolitan area, Ibirapuera, does not do so over the entire sample period, we also consider radiation measurements in CETESB's Paulínia station, 100 km northwest of the city of São Paulo. See text.

Table SIV. Time-of-day-and-road-segment specific regressions of travel time (equivalently, inverse traffic speed), in minutes per kilometer, on covariates, including the gasoline share among bi-fuel consumers. A separate regression is estimated for each speed-monitored time-road pair.

Table SIV. Predicting Travel Times on Major Road Segments (minutes/km), 08:30 or 18:00, non-holiday weekday

Time of Day & Road (Segment)	Prop. BFVs burning gasoline Coefficient	(Std. Error)	R2	Number Observ.	Number Regress.	Mean Dep.Var.
08:30 Aricanduva (Julio Colaço to acess.Radial Leste p/Elev.)	2.068	(1.059)	35.2%	348	39	2.47
08:30 Bandeirantes (R.Arapuã to Rod. dos Imigrantes)	-3.931	(0.546)	33.9%	271	39	2.33
08:30 Consolação (Acesso para Amaral Gurgel to R.Piauí)	0.913	(0.371)	16.3%	370	39	1.68
08:30 Dr. Arnaldo (R.Teodoro Sampaio to R.Major Natanael)	-2.373	(0.860)	20.6%	377	39	3.33
08:30 Ibirapuera (Av.dos Eucaliptos to R.Lavandisca)	0.515	(0.965)	34.9%	357	39	2.80
08:30 João Dias (Inicio da Pte.João Dias to Rep.Centro Afric.)	0.438	(1.278)	27.8%	371	39	4.60
08:30 Ligação Leste-Oeste (Pça Pér.Byington to Vd.Liberdade)	0.389	(0.723)	14.1%	377	39	1.62
08:30 Marginal Pinheiros (Pte.Morumbi to Av.J.Rob.Marinho)	-0.016	(0.863)	23.6%	326	39	1.41
08:30 Marginal Tietê (R.Jacirendi to Pte.Tatuapé)	1.178	(1.088)	29.9%	284	39	2.95
08:30 Maria Paula (R.Sto Antonio to Brig.Luis Antonio)	-0.781	(0.325)	26.6%	370	39	0.96
08:30 Paulista (Al.Casa Branca to Brig.Luis Antonio)	-0.520	(0.338)	25.8%	374	39	2.62
08:30 Prof. Abraão de Moraes (Av.Bq.de Saúde to R.Rib.Lac.)	-0.571	(0.352)	12.0%	360	39	1.19
08:30 Radial Leste br (Av.Mgda.Maria Alves to Vdo.V.Matilde)	0.159	(0.321)	14.1%	357	39	1.18
08:30 Radial Leste br (R.Tuiuti to R.Apucarana)	2.282	(1.155)	19.7%	377	39	1.94
08:30 Radial Leste mo (R.Piratininga to R.Alm.Brasil)	2.620	(0.940)	37.9%	377	39	6.99
08:30 Rebouças (R.Joaquim Antunes to R.Lisboa)	0.447	(0.921)	22.7%	379	39	3.39
08:30 Vinte e Três de Maio cont (Vd.Ben.Port. to Vd.Sta Gen.)	-0.077	(0.133)	16.5%	378	39	0.94
08:30 Consolação (R.Antonio Carlos to Av.Paulista)	-0.625	(0.787)	40.1%	364	39	1.95
08:30 Marginal Tietê (Pte.Limão to Pte.Julio de Mesq. Neto)	0.345	(0.378)	15.5%	346	39	0.91
08:30 Nove de Julho (Av.São Gabriel to Av.Brasil)	-0.667	(0.988)	25.0%	348	39	2.15
08:30 Tiradentes cont (Pass.do Correio to Pass.da R.Mauá)	3.777	(1.622)	34.3%	354	39	6.12
08:30 Vinte e Três de Maio cont (Vd.Ped.Tol. to Av.J.Ma.Whit.)	-0.646	(0.431)	25.5%	377	39	2.52
18:00 Aricanduva (Julio Colaço to acess.Radial Leste p/Elev.)	0.026	(0.213)	22.0%	350	40	0.91
18:00 Bandeirantes (R.Arapuã to Rod. dos Imigrantes)	-0.780	(0.927)	38.7%	262	40	2.70
18:00 Consolação (Acesso para Amaral Gurgel to R.Piauí)	1.990	(1.092)	19.6%	337	40	2.82
18:00 Dr. Arnaldo (R.Teodoro Sampaio to R.Major Natanael)	1.457	(1.186)	23.5%	362	40	3.59
18:00 Ibirapuera (Av.dos Eucaliptos to R.Lavandisca)	0.031	(1.329)	21.5%	326	40	5.61
18:00 João Dias (Inicio da Pte.João Dias to Rep.Centro Afric.)	-2.313	(1.883)	16.3%	369	40	6.64
18:00 Ligação Leste-Oeste (Pça Pér.Byington to Vd.Liberdade)	-2.636	(0.873)	28.4%	370	40	2.73
18:00 Marginal Pinheiros (Pte.Morumbi to Av.J.Rob.Marinho)	-0.590	(0.700)	22.0%	315	40	2.01
18:00 Marginal Tietê (R.Jacirendi to Pte.Tatuapé)	0.343	(0.310)	40.5%	275	40	2.11
18:00 Maria Paula (R.Sto Antonio to Brig.Luis Antonio)	-6.007	(1.749)	21.8%	354	40	2.54
18:00 Paulista (Al.Casa Branca to Brig.Luis Antonio)	-0.791	(1.945)	15.2%	366	40	5.49
18:00 Prof. Abraão de Moraes (Av.Bq.de Saúde to R.Rib.Lac.)	-5.951	(0.941)	39.7%	348	40	2.37
18:00 Radial Leste br (Av.Mgda.Maria Alves to Vdo.V.Matilde)	-0.166	(0.393)	15.0%	361	40	1.66
18:00 Radial Leste br (R.Tuiuti to R.Apucarana)	1.412	(0.825)	29.5%	343	40	4.21
18:00 Radial Leste mo (R.Piratininga to R.Alm.Brasil)	-5.757	(1.374)	27.4%	374	40	6.36
18:00 Vinte e Três de Maio cont (Vd.Ben.Port. to Vd. Sta Gen.)	2.384	(1.543)	29.6%	351	40	4.00
18:00 Luís Inácio de Anhaia Melo (R.Ibitirama to R.Dianop.)	0.219	(0.714)	31.7%	357	40	2.41
18:00 Marginal Tietê (Pte.Limão to Pte.Julio de Mesq. Neto)	8.314	(3.251)	44.0%	334	40	5.73
18:00 Nove de Julho (Av.São Gabriel to Av.Brasil)	3.670	(2.509)	15.6%	339	40	4.59
18:00 Tiradentes cont (Pass.do Correio to Pass.da R.Mauá)	2.336	(1.583)	29.0%	315	40	4.24
18:00 Vinte e Três de Maio cont (Vd.Ped.Tol. to Av.J.Ma.Whit.)	-1.077	(0.878)	17.8%	364	40	2.28

Notes: Coefficients and standard errors (in parentheses) on the gasoline share among bi-fuel consumers for the indicated road-hour specific regression (standard errors shown in this table do not account for sampling variation in the prediction of the gasoline share; doing so would only strengthen our conclusion, as discussed in the text). An observation is a non-holiday weekday. The sample period is November 1, 2008 to May 31, 2011, excluding the colder months of June to September. Ordinary Least Squares estimates. Specification includes a trend, week-of-year fixed effects, day-of-week fixed effects, precipitation (logarithm of accumulation in the last hour plus 0.001), and traffic events surrounding long weekends and World Cup soccer matches.

Table SV. Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SVI. Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for Sundays/Public Holidays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SVII. Mean fuel mix effects and precision on NO averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for non-holiday weekdays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SVIII. Mean fuel mix effects and precision on NO averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for Sundays/Public Holidays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SIX. Mean fuel mix effects and precision on NO_2 averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for non-holiday weekdays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SX. Mean fuel mix effects and precision on NO_2 averaged over 9 regressions, one for each of the 9 NO_X-monitoring stations, for Sundays/Public Holidays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SXI. Mean fuel mix effects and precision on CO averaged over 11 regressions, one for each of the 11 CO-monitoring stations, for non-holiday weekdays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SXII. Mean fuel mix effects and precision on CO averaged over 11 regressions, one for each of the 11 CO-monitoring stations, for Sundays/Public Holidays, at different times of the day. Estimates for each individual regression are provided in the Appendix.

Table SV. Predicting Ozone (μg/m^3), non-holiday weekday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**1.0**	**-19.2**	**-24.3**	**-15.3**	**-15.1**	**-12.7**	**-15.8**
	(3.7)	**(5.8)**	**(6.8)**	**(4.3)**	**(4.3)**	**(4.3)**	**(4.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**-3.2**	**-2.4**	**-1.6**	**-2.0**	**-1.9**	**-1.4**	**-2.0**
	(1.5)	**(1.5)**	**(1.8)**	**(1.2)**	**(1.2)**	**(1.2)**	**(1.3)**
Mean across 12 station specific regressions							
R2	1.8%	6.8%	22.9%	48.4%	48.5%	57.5%	49.2%
Number of observations	1957	1957	1957	1943	1943	1890	1896
Number of regressors	3	4	44	49	50	97	54
Mean value of dependent variable	20.3	20.3	20.3	20.3	20.3	20.0	20.3
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**3.9**	**-4.7**	**-10.0**	**-10.0**	**-11.5**	**-10.9**	**-11.4**
	(2.0)	**(2.9)**	**(3.5)**	**(2.9)**	**(2.9)**	**(3.2)**	**(2.8)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.2**	**0.5**	**-0.6**	**-1.5**	**-1.6**	**-2.2**	**-1.3**
	(0.8)	**(0.8)**	**(0.9)**	**(0.8)**	**(0.8)**	**(0.8)**	**(0.8)**
Mean across 12 station specific regressions							
R2	2.3%	5.2%	45.0%	58.8%	59.0%	66.2%	60.2%
Number of observations	1562	1562	1562	1547	1547	1521	1493
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	14.9	14.9	14.9	14.8	14.8	14.8	14.9
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**7.7**	**10.0**	**-13.6**	**-24.8**	**-27.5**	**-30.8**	**-26.4**
	(5.6)	**(7.9)**	**(9.8)**	**(5.6)**	**(5.6)**	**(6.6)**	**(5.8)**
Three month period with gasoline E20, $e20_t^{gas}$	**1.9**	**1.8**	**-1.7**	**-2.3**	**-2.9**	**-2.0**	**-1.4**
	(2.7)	**(2.7)**	**(2.9)**	**(1.7)**	**(1.6)**	**(1.7)**	**(1.9)**
Mean across 12 station specific regressions							
R2	1.4%	2.3%	40.1%	70.1%	70.3%	76.3%	70.8%
Number of observations	1556	1556	1556	1538	1538	1514	1493
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	45.5	45.5	45.5	45.4	45.4	45.4	45.3
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**17.4**	**28.2**	**4.6**	**-10.8**	**-13.4**	**-14.2**	**-8.6**
	(7.3)	**(10.1)**	**(12.6)**	**(8.7)**	**(8.5)**	**(7.3)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-6.5**	**-7.0**	**-1.1**	**-2.9**	**-3.2**	**-2.0**	**-1.4**
	(2.5)	**(2.5)**	**(3.1)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.0)**
Mean across 12 station specific regressions							
R2	2.0%	2.7%	35.5%	63.3%	63.7%	73.7%	65.0%
Number of observations	1587	1587	1587	1578	1578	1544	1535
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	36.0	36.0	36.0	36.1	36.1	36.1	36.0
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.0**	**-3.1**	**-12.8**	**-14.5**	**-14.5**	**-17.1**	**-13.6**
	(2.9)	**(4.1)**	**(4.6)**	**(4.0)**	**(4.0)**	**(3.9)**	**(3.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**-4.8**	**-4.3**	**-2.7**	**-2.2**	**-2.2**	**-1.6**	**-1.7**
	(1.0)	**(1.0)**	**(1.2)**	**(1.2)**	**(1.2)**	**(1.2)**	**(1.1)**
Mean across 12 station specific regressions							
R2	3.6%	6.3%	20.5%	37.9%	37.9%	46.1%	39.0%
Number of observations	1585	1585	1585	1575	1575	1537	1531
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	16.6	16.6	16.6	16.6	16.6	16.5	16.6

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SVI. Predicting Ozone (μg/m^3), Sunday/Public Holiday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**10.6**	**3.1**	**-1.6**	**-9.9**	**-10.0**	**-10.9**	**-11.9**
	(6.0)	**(10.1)**	**(14.5)**	**(11.6)**	**(11.6)**	**(11.0)**	**(11.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.9**	**-5.4**	**-4.9**	**-1.9**	**-1.8**	**-2.0**	**-1.5**
	(2.0)	**(2.1)**	**(2.8)**	**(3.1)**	**(3.1)**	**(2.9)**	**(2.8)**
Mean across 12 station specific regressions							
R2	4.0%	5.5%	27.6%	50.3%	50.5%	68.4%	52.0%
Number of observations	661	661	661	651	651	642	626
Number of regressors	3	4	41	46	47	93	50
Mean value of dependent variable	18.9	18.9	18.9	18.7	18.7	18.7	18.6
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**5.2**	**-7.8**	**-7.3**	**-11.2**	**-10.9**	**-15.4**	**-10.4**
	(5.6)	**(8.7)**	**(11.0)**	**(11.0)**	**(10.8)**	**(11.5)**	**(11.1)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.1**	**-4.3**	**-3.7**	**-4.8**	**-4.9**	**-4.9**	**-4.9**
	(1.9)	**(1.8)**	**(2.6)**	**(2.5)**	**(2.5)**	**(2.8)**	**(2.7)**
Mean across 12 station specific regressions							
R2	2.6%	4.7%	53.8%	65.6%	65.8%	74.6%	66.5%
Number of observations	526	526	526	517	517	516	497
Number of regressors	3	4	40	45	46	92	48
Mean value of dependent variable	24.5	24.5	24.5	24.3	24.3	24.2	24.3
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**10.3**	**-0.3**	**-26.3**	**-22.5**	**-22.6**	**-24.1**	**-28.1**
	(11.4)	**(17.7)**	**(19.7)**	**(15.1)**	**(15.2)**	**(14.0)**	**(16.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**-1.5**	**-0.8**	**-0.4**	**-5.1**	**-5.1**	**-1.4**	**-7.5**
	(3.9)	**(3.9)**	**(5.7)**	**(4.7)**	**(4.7)**	**(4.7)**	**(5.0)**
Mean across 12 station specific regressions							
R2	1.7%	2.8%	48.8%	70.1%	70.3%	80.4%	71.1%
Number of observations	529	529	529	520	520	512	493
Number of regressors	3	4	40	45	46	91	49
Mean value of dependent variable	59.9	59.9	59.9	59.5	59.5	59.2	59.3
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.2**	**-9.5**	**-47.5**	**-26.8**	**-27.5**	**-32.1**	**-27.7**
	(17.5)	**(26.8)**	**(32.3)**	**(20.0)**	**(19.9)**	**(19.8)**	**(20.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.1**	**0.5**	**5.1**	**-7.1**	**-7.2**	**-0.8**	**-8.7**
	(7.0)	**(7.1)**	**(10.2)**	**(5.7)**	**(5.8)**	**(5.2)**	**(6.7)**
Mean across 12 station specific regressions							
R2	1.8%	2.5%	31.7%	65.1%	65.4%	77.5%	66.9%
Number of observations	531	531	531	522	522	510	496
Number of regressors	3	4	40	45	46	91	50
Mean value of dependent variable	80.7	80.7	80.7	79.7	79.7	79.5	79.1
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**16.7**	**23.1**	**-20.0**	**-9.8**	**-8.0**	**-0.7**	**-13.3**
	(14.9)	**(21.6)**	**(24.4)**	**(17.8)**	**(18.3)**	**(21.2)**	**(18.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**4.4**	**4.0**	**6.8**	**-2.0**	**-1.6**	**-1.2**	**-3.7**
	(7.1)	**(7.0)**	**(8.4)**	**(4.6)**	**(4.7)**	**(5.0)**	**(4.2)**
Mean across 12 station specific regressions							
R2	2.5%	3.2%	44.2%	68.2%	68.5%	78.2%	68.8%
Number of observations	530	530	530	527	527	505	500
Number of regressors	3	4	40	45	46	91	50
Mean value of dependent variable	52.4	52.4	52.4	52.2	52.2	51.9	51.2
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**15.0**	**6.3**	**-21.7**	**-17.6**	**-16.9**	**-10.8**	**-21.2**
	(6.8)	**(10.3)**	**(14.8)**	**(12.9)**	**(12.8)**	**(12.1)**	**(11.7)**
Three month period with gasoline E20, $e20_t^{gas}$	**-3.0**	**-2.4**	**-3.3**	**-2.5**	**-2.3**	**-3.9**	**-3.6**
	(3.5)	**(3.6)**	**(5.1)**	**(3.8)**	**(3.8)**	**(3.3)**	**(3.1)**
Mean across 12 station specific regressions							
R2	3.0%	4.4%	32.2%	48.7%	48.9%	63.0%	50.6%
Number of observations	533	533	533	529	529	511	505
Number of regressors	3	4	40	45	46	92	50
Mean value of dependent variable	23.8	23.8	23.8	23.8	23.8	23.5	23.4

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SVII. Predicting NO (μg/m^3), non-holiday weekday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**6.7**	**1.6**	**57.0**	**38.2**	**38.7**	**41.1**	**41.6**
	(12.9)	**(20.1)**	**(19.9)**	**(14.3)**	**(14.4)**	**(12.3)**	**(14.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**4.3**	**4.6**	**3.7**	**2.4**	**2.5**	**-1.0**	**3.4**
	(4.6)	**(4.6)**	**(5.3)**	**(4.1)**	**(4.1)**	**(4.1)**	**(4.4)**
Mean across 9 station specific regressions							
R2	0.4%	0.7%	19.1%	37.0%	37.1%	50.1%	37.5%
Number of observations	1926	1926	1926	1909	1909	1862	1859
Number of regressors	3	4	44	49	50	97	53
Mean value of dependent variable	28.1	28.1	28.1	28.1	28.1	28.5	28.5
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-17.4**	**3.6**	**45.6**	**29.2**	**28.2**	**22.7**	**31.7**
	(12.6)	**(19.7)**	**(21.1)**	**(12.2)**	**(12.3)**	**(12.4)**	**(11.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**6.7**	**6.1**	**8.3**	**11.8**	**11.8**	**9.7**	**11.0**
	(4.5)	**(4.5)**	**(5.5)**	**(3.8)**	**(3.8)**	**(4.0)**	**(3.9)**
Mean across 9 station specific regressions							
R2	0.9%	1.5%	21.0%	43.2%	43.3%	53.2%	45.9%
Number of observations	1529	1529	1529	1514	1514	1489	1464
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
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-30.3**	**-7.1**	**23.2**	**22.0**	**20.8**	**14.4**	**23.7**
	(4.4)	**(6.9)**	**(7.9)**	**(6.3)**	**(6.4)**	**(5.9)**	**(5.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.3**	**-0.4**	**3.6**	**5.9**	**5.7**	**5.2**	**6.4**
	(1.4)	**(1.3)**	**(1.6)**	**(1.5)**	**(1.6)**	**(1.5)**	**(1.8)**
Mean across 9 station specific regressions							
R2	2.4%	3.7%	30.5%	42.9%	43.1%	52.1%	49.6%
Number of observations	1521	1521	1521	1504	1504	1481	1456
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	38.8	38.8	38.8	38.8	38.8	38.9	38.8
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-32.6**	**-3.8**	**15.6**	**17.3**	**14.0**	**7.8**	**17.1**
	(3.3)	**(4.3)**	**(5.7)**	**(4.9)**	**(5.0)**	**(4.7)**	**(4.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**4.3**	**3.4**	**4.2**	**4.1**	**3.5**	**2.2**	**5.2**
	(1.5)	**(1.3)**	**(1.3)**	**(1.3)**	**(1.3)**	**(1.3)**	**(1.6)**
Mean across 9 station specific regressions							
R2	3.8%	6.1%	17.3%	35.5%	35.6%	48.4%	41.1%
Number of observations	1538	1538	1538	1525	1525	1497	1474
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	29.0	29.0	29.0	29.1	29.1	29.2	29.2
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-38.5**	**-18.6**	**8.7**	**15.3**	**14.3**	**6.2**	**17.4**
	(5.0)	**(7.0)**	**(7.5)**	**(6.0)**	**(5.8)**	**(5.8)**	**(6.1)**
Three month period with gasoline E20, $e20_t^{gas}$	**7.2**	**6.5**	**4.5**	**3.2**	**3.1**	**2.5**	**5.1**
	(2.0)	**(2.0)**	**(2.5)**	**(1.8)**	**(1.8)**	**(1.6)**	**(1.9)**
Mean across 9 station specific regressions							
R2	3.9%	5.5%	19.6%	39.8%	40.0%	56.6%	43.1%
Number of observations	1562	1562	1562	1553	1553	1519	1504
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	34.8	34.8	34.8	34.6	34.6	34.8	34.6
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-36.7**	**-41.7**	**17.3**	**8.2**	**8.3**	**11.6**	**10.7**
	(10.0)	**(17.8)**	**(12.7)**	**(8.8)**	**(8.8)**	**(9.9)**	**(9.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**13.0**	**13.3**	**9.8**	**7.3**	**7.2**	**4.2**	**7.7**
	(5.1)	**(5.1)**	**(4.8)**	**(3.0)**	**(3.0)**	**(3.0)**	**(3.1)**
Mean across 9 station specific regressions							
R2	2.0%	2.6%	19.7%	45.1%	45.2%	60.4%	46.3%
Number of observations	1567	1567	1567	1556	1556	1519	1509
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	42.9	42.9	42.9	42.9	42.9	43.3	43.1

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SVIII. Predicting NO (μg/m^3), Sunday/Public Holiday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-21.5**	**-28.1**	**9.9**	**15.3**	**14.1**	**4.7**	**13.2**
	(12.5)	**(19.6)**	**(31.7)**	**(25.2)**	**(24.3)**	**(24.3)**	**(25.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**8.5**	**8.9**	**7.2**	**2.1**	**1.6**	**0.5**	**2.0**
	(5.0)	**(5.0)**	**(7.4)**	**(7.0)**	**(7.0)**	**(6.4)**	**(7.0)**
Mean across 9 station specific regressions							
R2	2.4%	2.8%	24.7%	42.4%	42.5%	60.1%	43.6%
Number of observations	653	653	653	644	644	637	625
Number of regressors	3	4	41	46	46	93	50
Mean value of dependent variable	20.7	20.7	20.7	20.9	20.9	21.0	21.2
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-10.2**	**-5.7**	**2.8**	**5.3**	**4.7**	**7.4**	**4.5**
	(8.9)	**(16.1)**	**(19.7)**	**(16.0)**	**(15.2)**	**(15.5)**	**(16.1)**
Three month period with gasoline E20, $e20_t^{gas}$	**6.5**	**6.3**	**4.7**	**4.8**	**5.0**	**3.7**	**4.4**
	(4.3)	**(4.3)**	**(4.8)**	**(4.1)**	**(4.1)**	**(5.2)**	**(4.2)**
Mean across 9 station specific regressions							
R2	1.7%	2.6%	25.8%	38.8%	39.8%	55.4%	40.9%
Number of observations	522	522	522	513	513	512	497
Number of regressors	3	4	40	44	45	92	48
Mean value of dependent variable	24.2	24.2	24.2	24.3	24.3	24.4	24.4
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-6.2**	**1.8**	**1.7**	**3.0**	**2.7**	**-1.0**	**0.1**
	(3.2)	**(5.3)**	**(6.7)**	**(7.0)**	**(6.4)**	**(7.1)**	**(7.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**1.0**	**0.6**	**-0.4**	**0.8**	**0.8**	**0.7**	**-0.2**
	(1.8)	**(1.7)**	**(1.7)**	**(1.6)**	**(1.5)**	**(1.9)**	**(1.9)**
Mean across 9 station specific regressions							
R2	1.5%	2.7%	33.3%	40.0%	42.1%	56.1%	44.4%
Number of observations	521	521	521	512	512	504	491
Number of regressors	3	4	40	44	45	91	49
Mean value of dependent variable	12.7	12.7	12.7	12.8	12.8	12.8	12.8
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-7.9**	**1.2**	**4.6**	**6.0**	**5.1**	**5.4**	**4.5**
	(2.6)	**(4.0)**	**(5.6)**	**(5.2)**	**(4.6)**	**(5.6)**	**(5.7)**
Three month period with gasoline E20, $e20_t^{gas}$	**2.2**	**1.7**	**0.8**	**1.7**	**1.1**	**1.2**	**1.1**
	(1.6)	**(1.5)**	**(1.8)**	**(1.5)**	**(1.3)**	**(1.5)**	**(2.0)**
Mean across 9 station specific regressions							
R2	3.7%	5.2%	25.9%	40.3%	43.2%	56.6%	46.9%
Number of observations	521	521	521	513	513	501	491
Number of regressors	3	4	40	44	45	91	49
Mean value of dependent variable	10.7	10.7	10.7	10.8	10.8	10.8	10.8
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-13.6**	**-8.7**	**7.5**	**2.9**	**0.9**	**3.3**	**0.0**
	(3.8)	**(6.2)**	**(7.2)**	**(5.3)**	**(5.3)**	**(6.4)**	**(5.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**2.0**	**1.7**	**1.3**	**1.3**	**0.9**	**2.1**	**0.1**
	(1.6)	**(1.6)**	**(2.1)**	**(1.6)**	**(1.5)**	**(1.6)**	**(1.9)**
Mean across 9 station specific regressions							
R2	2.0%	2.7%	23.1%	40.7%	42.0%	63.7%	44.7%
Number of observations	521	521	521	517	517	498	494
Number of regressors	3	4	40	45	46	91	49
Mean value of dependent variable	13.7	13.7	13.7	13.8	13.8	13.9	13.8
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-25.9**	**-34.4**	**15.7**	**5.3**	**3.0**	**-9.5**	**3.6**
	(11.6)	**(17.9)**	**(23.2)**	**(16.4)**	**(16.7)**	**(16.7)**	**(17.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**4.9**	**5.4**	**5.5**	**-1.5**	**-2.2**	**-2.6**	**-1.8**
	(4.0)	**(4.0)**	**(6.1)**	**(5.5)**	**(5.4)**	**(5.4)**	**(5.9)**
Mean across 9 station specific regressions							
R2	1.9%	2.7%	29.3%	46.0%	46.5%	67.6%	48.1%
Number of observations	524	524	524	520	520	503	500
Number of regressors	3	4	40	45	46	92	50
Mean value of dependent variable	22.8	22.8	22.8	22.8	22.8	23.1	23.0

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SIX. Predicting NO_2 (μg/m^3), non-holiday weekday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.5**	**5.4**	**12.7**	**5.0**	**4.8**	**6.3**	**6.7**
	(3.8)	**(5.8)**	**(7.4)**	**(4.4)**	**(4.4)**	**(3.9)**	**(4.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**3.6**	**3.4**	**5.2**	**2.2**	**2.1**	**0.4**	**2.7**
	(2.1)	**(2.1)**	**(2.2)**	**(1.3)**	**(1.4)**	**(1.3)**	**(1.4)**
Mean across 9 station specific regressions							
R2	2.4%	3.4%	21.4%	55.2%	55.2%	67.0%	56.0%
Number of observations	1926	1926	1926	1909	1909	1862	1859
Number of regressors	3	4	44	49	50	96	53
Mean value of dependent variable	30.9	30.9	30.9	30.8	30.8	30.9	30.8
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-4.4**	**-2.6**	**5.0**	**-5.3**	**-6.2**	**5.1**	**-5.8**
	(3.6)	**(5.0)**	**(6.4)**	**(4.7)**	**(4.7)**	**(4.4)**	**(4.8)**
Three month period with gasoline E20, $e20_t^{gas}$	**3.3**	**3.3**	**3.7**	**4.6**	**4.5**	**-0.7**	**4.5**
	(1.7)	**(1.7)**	**(1.9)**	**(1.3)**	**(1.3)**	**(1.2)**	**(1.4)**
Mean across 9 station specific regressions							
R2	2.9%	4.7%	21.9%	40.3%	40.4%	58.4%	42.9%
Number of observations	1529	1529	1529	1514	1514	1489	1464
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
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-12.2**	**-6.7**	**13.1**	**0.9**	**0.9**	**3.8**	**2.0**
	(4.3)	**(6.7)**	**(8.0)**	**(4.9)**	**(4.8)**	**(4.9)**	**(5.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.2**	**-0.3**	**1.6**	**5.5**	**5.5**	**2.8**	**6.6**
	(2.0)	**(2.0)**	**(2.1)**	**(1.3)**	**(1.4)**	**(1.3)**	**(1.4)**
Mean across 9 station specific regressions							
R2	2.8%	4.7%	21.7%	48.8%	49.0%	59.6%	52.5%
Number of observations	1521	1521	1521	1504	1504	1481	1456
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	48.2	48.2	48.2	48.1	48.1	48.2	48.1
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-16.8**	**-9.5**	**6.4**	**-2.3**	**-4.5**	**-2.8**	**-4.5**
	(3.4)	**(5.2)**	**(6.4)**	**(4.4)**	**(4.5)**	**(4.5)**	**(4.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.4**	**0.2**	**0.9**	**3.5**	**3.1**	**3.6**	**4.2**
	(1.4)	**(1.4)**	**(1.4)**	**(1.1)**	**(1.2)**	**(1.2)**	**(1.3)**
Mean across 9 station specific regressions							
R2	4.2%	7.3%	19.7%	46.8%	47.0%	55.7%	50.1%
Number of observations	1538	1538	1538	1525	1525	1497	1474
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	43.5	43.5	43.5	43.6	43.6	43.6	43.6
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-16.5**	**-15.6**	**0.2**	**-10.1**	**-12.4**	**-8.5**	**-11.9**
	(5.3)	**(7.3)**	**(8.6)**	**(5.9)**	**(5.7)**	**(5.2)**	**(5.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**5.7**	**5.8**	**4.7**	**5.6**	**5.3**	**5.2**	**6.5**
	(2.5)	**(2.5)**	**(2.6)**	**(1.5)**	**(1.4)**	**(1.4)**	**(1.4)**
Mean across 9 station specific regressions							
R2	3.3%	5.3%	24.4%	55.6%	56.4%	66.8%	58.3%
Number of observations	1562	1562	1562	1553	1553	1519	1504
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	51.5	51.5	51.5	51.4	51.4	51.6	51.3
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-9.3**	**0.0**	**8.7**	**2.7**	**2.7**	**4.6**	**2.0**
	(5.5)	**(7.7)**	**(9.0)**	**(5.7)**	**(5.7)**	**(5.0)**	**(5.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**7.4**	**7.1**	**8.4**	**3.9**	**3.7**	**1.2**	**4.0**
	(2.8)	**(2.8)**	**(2.7)**	**(1.6)**	**(1.6)**	**(1.4)**	**(1.5)**
Mean across 9 station specific regressions							
R2	2.8%	3.9%	20.3%	59.0%	59.4%	68.9%	60.3%
Number of observations	1567	1567	1567	1556	1556	1519	1509
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	49.7	49.7	49.7	49.6	49.6	49.9	49.4

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SX. Predicting NO_2 (μg/m^3), Sunday/Public Holiday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-2.7**	**-0.2**	**0.0**	**-8.5**	**-8.7**	**-12.9**	**-9.8**
	(6.6)	**(12.0)**	**(16.1)**	**(11.4)**	**(11.4)**	**(11.0)**	**(10.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**7.1**	**7.0**	**6.0**	**4.3**	**4.1**	**4.4**	**3.6**
	(3.3)	**(3.3)**	**(3.5)**	**(3.5)**	**(3.5)**	**(3.6)**	**(3.5)**
Mean across 9 station specific regressions							
R2	3.4%	4.2%	36.3%	61.1%	61.2%	72.8%	63.6%
Number of observations	653	653	653	644	644	637	625
Number of regressors	3	4	41	45	46	93	49
Mean value of dependent variable	31.2	31.2	31.2	31.2	31.2	31.2	31.3
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-2.5**	**6.1**	**-4.6**	**-10.9**	**-11.2**	**-11.8**	**-11.6**
	(5.4)	**(10.0)**	**(12.0)**	**(10.1)**	**(10.1)**	**(9.5)**	**(10.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**5.6**	**5.1**	**2.3**	**3.1**	**3.2**	**0.9**	**3.3**
	(3.0)	**(2.9)**	**(2.9)**	**(2.4)**	**(2.4)**	**(2.8)**	**(2.5)**
Mean across 9 station specific regressions							
R2	3.9%	5.6%	36.8%	49.7%	50.6%	67.6%	51.4%
Number of observations	522	522	522	513	513	512	497
Number of regressors	3	4	40	44	45	92	48
Mean value of dependent variable	31.2	31.2	31.2	31.2	31.2	31.2	31.1
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-3.8**	**1.1**	**-10.8**	**-12.3**	**-12.7**	**-16.1**	**-14.3**
	(5.2)	**(10.0)**	**(12.5)**	**(11.2)**	**(10.7)**	**(10.5)**	**(11.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**1.3**	**1.0**	**-0.9**	**1.1**	**1.1**	**-0.2**	**0.4**
	(2.6)	**(2.6)**	**(2.7)**	**(2.0)**	**(1.9)**	**(2.8)**	**(2.0)**
Mean across 9 station specific regressions							
R2	2.3%	3.7%	33.0%	48.8%	51.7%	67.4%	54.4%
Number of observations	521	521	521	512	512	504	491
Number of regressors	3	4	40	44	45	91	49
Mean value of dependent variable	26.8	26.8	26.8	26.8	26.8	26.8	26.7
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-7.9**	**-1.0**	**0.1**	**-0.1**	**-0.8**	**1.4**	**0.6**
	(3.4)	**(5.9)**	**(8.4)**	**(8.2)**	**(7.5)**	**(7.5)**	**(8.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**1.1**	**0.7**	**0.4**	**0.9**	**0.3**	**1.6**	**0.4**
	(1.8)	**(1.8)**	**(2.3)**	**(1.9)**	**(1.9)**	**(1.9)**	**(2.2)**
Mean across 9 station specific regressions							
R2	4.6%	7.2%	29.7%	44.9%	48.7%	65.1%	50.9%
Number of observations	521	521	521	513	513	501	491
Number of regressors	3	4	40	44	45	91	49
Mean value of dependent variable	24.6	24.6	24.6	24.6	24.6	24.6	24.7
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-8.9**	**-6.1**	**8.8**	**-5.3**	**-7.2**	**3.2**	**-6.7**
	(6.3)	**(9.1)**	**(12.4)**	**(8.9)**	**(8.8)**	**(8.4)**	**(9.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**4.5**	**4.4**	**4.2**	**2.9**	**2.5**	**1.0**	**2.9**
	(2.4)	**(2.5)**	**(3.1)**	**(2.2)**	**(2.2)**	**(2.5)**	**(2.5)**
Mean across 9 station specific regressions							
R2	3.3%	4.6%	38.6%	59.9%	61.2%	75.0%	63.7%
Number of observations	521	521	521	517	517	498	494
Number of regressors	3	4	40	45	46	91	49
Mean value of dependent variable	30.5	30.5	30.5	30.4	30.4	30.6	30.4
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 9 station-specific regressions)							
Proportion FFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-8.8**	**0.5**	**8.4**	**-1.6**	**-3.3**	**-2.2**	**-1.2**
	(8.4)	**(12.5)**	**(18.0)**	**(11.7)**	**(11.7)**	**(11.1)**	**(11.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**7.8**	**7.3**	**6.7**	**0.9**	**0.5**	**0.6**	**1.5**
	(3.2)	**(3.3)**	**(3.8)**	**(3.0)**	**(2.9)**	**(3.1)**	**(2.9)**
Mean across 9 station specific regressions							
R2	3.7%	4.9%	29.1%	58.0%	58.6%	72.1%	60.5%
Number of observations	524	524	524	520	520	503	500
Number of regressors	3	4	40	45	46	92	50
Mean value of dependent variable	39.0	39.0	39.0	38.9	38.9	38.9	38.9

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXI. Predicting CO (ppm), non-holiday weekday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.20**	**0.23**	**0.78**	**0.40**	**0.40**	**0.51**	**0.41**
	(0.15)	**(0.23)**	**(0.24)**	**(0.15)**	**(0.15)**	**(0.13)**	**(0.16)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.04**	**0.04**	**0.08**	**0.05**	**0.05**	**0.02**	**0.05**
	(0.06)	**(0.06)**	**(0.07)**	**(0.04)**	**(0.04)**	**(0.03)**	**(0.04)**
Mean across 11 station specific regressions							
R2	1.1%	1.4%	21.3%	51.1%	51.2%	62.8%	51.6%
Number of observations	1923	1923	1923	1907	1907	1859	1865
Number of regressors	3	4	44	49	50	96	54
Mean value of dependent variable	0.66	0.66	0.66	0.66	0.66	0.66	0.66
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.15**	**0.35**	**0.66**	**0.34**	**0.33**	**0.41**	**0.34**
	(0.13)	**(0.21)**	**(0.22)**	**(0.12)**	**(0.12)**	**(0.11)**	**(0.12)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.08**	**0.07**	**0.10**	**0.17**	**0.17**	**0.11**	**0.16**
	(0.05)	**(0.05)**	**(0.06)**	**(0.04)**	**(0.04)**	**(0.04)**	**(0.04)**
Mean across 11 station specific regressions							
R2	1.2%	2.0%	24.6%	48.9%	49.0%	59.8%	51.5%
Number of observations	1565	1565	1565	1548	1548	1523	1505
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
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.03**	**0.20**	**0.51**	**0.29**	**0.30**	**0.36**	**0.30**
	(0.07)	**(0.11)**	**(0.13)**	**(0.07)**	**(0.07)**	**(0.08)**	**(0.07)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.03**	**-0.03**	**0.04**	**0.11**	**0.12**	**0.09**	**0.11**
	(0.03)	**(0.03)**	**(0.03)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**
Mean across 11 station specific regressions							
R2	1.1%	2.7%	24.3%	48.5%	48.8%	57.8%	53.5%
Number of observations	1554	1554	1554	1535	1535	1512	1498
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.04**	**0.15**	**0.33**	**0.17**	**0.14**	**0.21**	**0.14**
	(0.05)	**(0.09)**	**(0.10)**	**(0.06)**	**(0.06)**	**(0.06)**	**(0.06)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.02**	**-0.03**	**0.00**	**0.06**	**0.05**	**0.05**	**0.05**
	(0.02)	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**
Mean across 11 station specific regressions							
R2	1.8%	4.2%	17.2%	42.1%	42.4%	52.3%	47.1%
Number of observations	1561	1561	1561	1547	1547	1520	1507
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	0.82	0.82	0.82	0.82	0.82	0.82	0.82
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.02**	**-0.03**	**0.22**	**0.08**	**0.05**	**0.10**	**0.06**
	(0.09)	**(0.15)**	**(0.15)**	**(0.09)**	**(0.08)**	**(0.09)**	**(0.08)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.05**	**0.06**	**0.03**	**0.06**	**0.05**	**0.05**	**0.06**
	(0.04)	**(0.04)**	**(0.05)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**
Mean across 11 station specific regressions							
R2	1.2%	2.2%	21.9%	52.4%	52.8%	62.3%	55.2%
Number of observations	1582	1582	1582	1572	1572	1539	1534
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	1.02	1.02	1.02	1.02	1.02	1.02	1.02
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.00**	**0.04**	**0.53**	**0.26**	**0.26**	**0.36**	**0.25**
	(0.14)	**(0.23)**	**(0.20)**	**(0.11)**	**(0.11)**	**(0.11)**	**(0.11)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.14**	**0.14**	**0.14**	**0.12**	**0.12**	**0.07**	**0.11**
	(0.07)	**(0.07)**	**(0.07)**	**(0.04)**	**(0.04)**	**(0.04)**	**(0.04)**
Mean across 11 station specific regressions							
R2	1.1%	1.5%	18.5%	56.5%	56.7%	67.4%	57.5%
Number of observations	1583	1583	1583	1571	1571	1535	1534
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	1.01	1.01	1.01	1.01	1.01	1.01	1.01

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXII. Predicting CO (ppm), Sunday/Public Holiday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.06**	**0.08**	**0.73**	**0.57**	**0.56**	**0.30**	**0.55**
	(0.25)	**(0.45)**	**(0.63)**	**(0.42)**	**(0.41)**	**(0.40)**	**(0.43)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.11**	**0.10**	**0.15**	**0.10**	**0.09**	**0.08**	**0.09**
	(0.11)	**(0.11)**	**(0.13)**	**(0.11)**	**(0.10)**	**(0.09)**	**(0.11)**
Mean across 11 station specific regressions							
R2	0.9%	1.5%	29.0%	56.4%	56.7%	73.1%	57.6%
Number of observations	652	652	652	643	643	636	628
Number of regressors	3	4	41	46	47	93	51
Mean value of dependent variable	0.89	0.89	0.89	0.89	0.89	0.89	0.89
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.03**	**0.17**	**0.16**	**0.04**	**0.03**	**0.08**	**0.04**
	(0.15)	**(0.28)**	**(0.36)**	**(0.26)**	**(0.26)**	**(0.25)**	**(0.26)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.08**	**0.07**	**0.08**	**0.11**	**0.11**	**0.05**	**0.11**
	(0.06)	**(0.06)**	**(0.07)**	**(0.06)**	**(0.06)**	**(0.07)**	**(0.06)**
Mean across 11 station specific regressions							
R2	1.7%	3.0%	28.3%	47.9%	48.6%	64.4%	49.5%
Number of observations	529	529	529	520	520	519	507
Number of regressors	3	4	40	45	46	92	48
Mean value of dependent variable	0.73	0.73	0.73	0.73	0.73	0.73	0.73
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.02**	**0.21**	**0.18**	**0.00**	**0.00**	**0.00**	**0.03**
	(0.09)	**(0.15)**	**(0.19)**	**(0.16)**	**(0.15)**	**(0.17)**	**(0.17)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.01**	**0.00**	**0.01**	**0.05**	**0.05**	**0.02**	**0.06**
	(0.03)	**(0.03)**	**(0.04)**	**(0.03)**	**(0.03)**	**(0.05)**	**(0.03)**
Mean across 11 station specific regressions							
R2	2.6%	6.3%	29.8%	45.8%	48.6%	63.3%	50.5%
Number of observations	529	529	529	519	519	511	504
Number of regressors	3	4	40	45	46	91	50
Mean value of dependent variable	0.63	0.63	0.63	0.63	0.63	0.63	0.63
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.01**	**0.25**	**0.30**	**0.12**	**0.10**	**0.17**	**0.15**
	(0.06)	**(0.09)**	**(0.14)**	**(0.13)**	**(0.13)**	**(0.15)**	**(0.13)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.00**	**-0.02**	**0.00**	**0.04**	**0.03**	**0.04**	**0.04**
	(0.03)	**(0.03)**	**(0.04)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**
Mean across 11 station specific regressions							
R2	3.0%	8.2%	31.8%	44.1%	46.3%	61.3%	48.5%
Number of observations	530	530	530	522	522	509	505
Number of regressors	3	4	40	45	46	91	50
Mean value of dependent variable	0.62	0.62	0.62	0.62	0.62	0.62	0.62
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.04**	**0.21**	**0.51**	**0.11**	**0.08**	**0.22**	**0.09**
	(0.12)	**(0.19)**	**(0.21)**	**(0.12)**	**(0.13)**	**(0.15)**	**(0.13)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.04**	**0.03**	**0.04**	**0.04**	**0.03**	**0.01**	**0.03**
	(0.04)	**(0.05)**	**(0.06)**	**(0.03)**	**(0.03)**	**(0.04)**	**(0.03)**
Mean across 11 station specific regressions							
R2	2.2%	4.5%	36.3%	59.2%	60.0%	73.5%	61.6%
Number of observations	531	531	531	526	526	505	509
Number of regressors	3	4	40	45	46	91	49
Mean value of dependent variable	0.72	0.72	0.72	0.72	0.72	0.73	0.72
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 11 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.01**	**0.08**	**0.75**	**0.28**	**0.25**	**0.15**	**0.28**
	(0.23)	**(0.37)**	**(0.44)**	**(0.24)**	**(0.25)**	**(0.26)**	**(0.26)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.08**	**0.08**	**0.11**	**-0.01**	**-0.02**	**-0.01**	**-0.01**
	(0.08)	**(0.08)**	**(0.11)**	**(0.07)**	**(0.07)**	**(0.07)**	**(0.07)**
Mean across 11 station specific regressions							
R2	0.7%	1.1%	29.8%	57.5%	57.7%	74.1%	58.8%
Number of observations	534	534	534	530	530	512	515
Number of regressors	3	4	40	45	46	92	50
Mean value of dependent variable	0.95	0.95	0.95	0.94	0.94	0.95	0.95

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXIII. Robustness test: Dependent variable is log concentration. Mean fuel mix effects and precision on ln([O_3]+0.001) averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, at different times of the day.

Table SXIV. Robustness tests: (a) to (e) (see text). Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, 13:00 to 16:00 readings.

Table SXV. Robustness tests: (f) to (j) (see text). Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, 13:00 to 16:00 readings.

Table SXVI. Robustness tests: (k) to (o) (see text). Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, 13:00 to 16:00 readings.

Table SXVII. Robustness tests: (p) on (see text). Mean fuel mix effects and precision on O_3 averaged over 12 regressions, one for each of the 12 O_3-monitoring stations, for non-holiday weekdays, 13:00 to 16:00 readings.

Table SXIII. Robustness: Predicting Ozone, log of (0.001 + measure in μg/m^3), non-holiday weekday

Specification:	I	II	III	IV	V	VI	VII
Time of day: 01:00 to 06:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-0.3**	**-1.8**	**-3.5**	**-2.2**	**-2.2**	**-2.1**	**-2.3**
	(0.5)	**(0.8)**	**(1.0)**	**(0.5)**	**(0.5)**	**(0.5)**	**(0.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.3**	**-0.2**	**-0.2**	**-0.2**	**-0.2**	**-0.1**	**-0.3**
	(0.2)	**(0.2)**	**(0.3)**	**(0.2)**	**(0.2)**	**(0.2)**	**(0.2)**
Mean across 12 station specific regressions							
R2	1.4%	3.8%	19.3%	44.1%	44.2%	53.5%	45.1%
Number of observations	1957	1957	1957	1943	1943	1890	1896
Number of regressors	3	4	44	49	50	97	54
Mean value of dependent variable	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Time of day: 07:00 to 10:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.1**	**-1.1**	**-1.8**	**-1.6**	**-1.7**	**-1.4**	**-1.7**
	(0.3)	**(0.4)**	**(0.5)**	**(0.3)**	**(0.3)**	**(0.3)**	**(0.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.1**	**0.1**	**0.0**	**-0.1**	**-0.1**	**-0.2**	**-0.1**
	(0.1)	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**
Mean across 12 station specific regressions							
R2	2.3%	5.3%	40.2%	51.7%	51.8%	57.2%	52.7%
Number of observations	1562	1562	1562	1547	1547	1521	1493
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	1.7	1.7	1.7	1.7	1.7	1.7	1.7
Time of day: 10:00 to 13:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.1**	**0.1**	**-0.7**	**-0.9**	**-1.0**	**-1.0**	**-1.0**
	(0.2)	**(0.3)**	**(0.3)**	**(0.2)**	**(0.2)**	**(0.2)**	**(0.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**0.1**	**0.1**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	(0.1)	**(0.1)**	**(0.1)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**
Mean across 12 station specific regressions							
R2	1.6%	2.7%	36.8%	65.8%	66.0%	72.5%	66.7%
Number of observations	1556	1556	1556	1538	1538	1514	1493
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	3.6	3.6	3.6	3.6	3.6	3.5	3.5
Time of day: 13:00 to 16:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.1**	**0.3**	**-0.3**	**-0.6**	**-0.6**	**-0.6**	**-0.6**
	(0.2)	**(0.2)**	**(0.3)**	**(0.2)**	**(0.2)**	**(0.1)**	**(0.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.1**	**-0.1**	**-0.1**	**0.0**	**0.0**	**0.0**	**0.0**
	(0.1)	**(0.1)**	**(0.1)**	**(0.0)**	**(0.0)**	**(0.0)**	**(0.0)**
Mean across 12 station specific regressions							
R2	1.4%	2.4%	21.0%	70.5%	70.6%	77.9%	71.1%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Time of day: 17:00 to 20:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.5**	**1.0**	**-0.3**	**-0.9**	**-0.9**	**-1.1**	**-0.8**
	(0.3)	**(0.4)**	**(0.4)**	**(0.3)**	**(0.3)**	**(0.3)**	**(0.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.3**	**-0.3**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	(0.1)	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**
Mean across 12 station specific regressions							
R2	2.2%	3.2%	37.0%	52.4%	52.5%	62.8%	53.9%
Number of observations	1587	1587	1587	1578	1578	1544	1535
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Time of day: 21:00 to 00:00 readings only							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**0.5**	**-0.4**	**-1.7**	**-1.5**	**-1.5**	**-1.9**	**-1.5**
	(0.3)	**(0.5)**	**(0.5)**	**(0.4)**	**(0.4)**	**(0.4)**	**(0.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**-0.4**	**-0.4**	**-0.2**	**-0.2**	**-0.2**	**0.0**	**-0.1**
	(0.1)	**(0.1)**	**(0.2)**	**(0.1)**	**(0.1)**	**(0.1)**	**(0.1)**
Mean across 12 station specific regressions							
R2	2.5%	4.5%	19.0%	38.8%	38.9%	47.5%	40.2%
Number of observations	1585	1585	1585	1575	1575	1537	1531
Number of regressors	3	4	43	48	49	95	54
Mean value of dependent variable	1.8	1.8	1.8	1.8	1.8	1.7	1.8

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXIV. Robustness: Predicting Ozone (μg/m^3), non-holiday weekday, 13:00 to 16:00 readings only

Baseline Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (a): Specify a quadratic trend in the date (rather than only a linear one)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**22.1**	**-7.0**	**-22.7**	**-25.5**	**-30.8**	**-21.4**
	(9.0)	**(13.0)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.8)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-10.6**	**-9.3**	**-4.2**	**-4.3**	**-2.8**	**-1.7**
	(4.0)	**(4.6)**	**(4.9)**	**(2.6)**	**(2.6)**	**(2.4)**	**(2.8)**
Mean across 12 station specific regressions							
R2	1.6%	3.6%	22.5%	69.5%	69.6%	76.0%	70.8%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	5	44	49	50	97	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (b): Specify meteorological covariates directly in their reported units (rather than the logarithmic transform)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-22.7**	**-25.3**	**-30.6**	**-21.6**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.6)**	**(8.5)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-2.6**	**-3.2**	**-1.4**	**-0.5**
	(4.0)	**(4.0)**	**(4.2)**	**(2.2)**	**(2.1)**	**(2.0)**	**(2.3)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.7%	69.9%	76.1%	71.1%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (c): Include atmospheric pressure in the set of meteorological covariates							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-23.8**	**-26.6**	**-30.7**	**-22.4**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(8.0)**	**(8.5)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.5**	**-4.1**	**-2.2**	**-1.7**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.1)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.3%	69.4%	76.2%	70.6%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	43	50	51	92	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (d): Specify traffic congestion covariates directly in their reported units (rather than the logarithmic transform)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-25.1**	**-29.1**	**-20.4**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.4)**	**(7.6)**	**(8.7)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.9**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.1%	75.7%	70.1%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (e): Additionally control for the median inverse traffic speed in the morning interacted with day-of-week fixed effects							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-25.4**	**-30.5**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(9.2)**	**(9.0)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-3.7**	**-2.5**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.4)**	**(2.1)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.8%	76.3%	70.4%
Number of observations	1574	1574	1574	1560	1408	1381	1515
Number of regressors	3	4	43	48	54	101	54
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.2	67.1	67.7

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXV. Robustness: Predicting Ozone (μg/m³), non-holiday weekday, 13:00 to 16:00 readings only

Baseline Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (f): Expand the sample to include weekdays during the school vacation fortnight (typically starts December 24							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**6.8**	**20.4**	**-4.4**	**-21.3**	**-23.0**	**-28.3**	**-19.2**
	(9.4)	**(14.0)**	**(16.1)**	**(7.7)**	**(7.8)**	**(7.3)**	**(7.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.7**	**-6.3**	**-5.3**	**-3.9**	**-4.2**	**-2.1**	**-1.5**
	(4.2)	**(4.2)**	**(4.0)**	**(2.4)**	**(2.4)**	**(2.1)**	**(2.6)**
Mean across 12 station specific regressions							
R2	1.5%	2.5%	20.9%	67.9%	68.0%	74.9%	69.2%
Number of observations	1660	1660	1660	1646	1646	1617	1599
Number of regressors	3	4	45	50	51	97	56
Mean value of dependent variable	68.0	68.0	68.0	67.9	67.9	67.9	67.8
Robustness (g): Expand the sample to include the colder months of June to September							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**24.6**	**29.8**	**-25.4**	**-19.6**	**-21.3**	**-26.3**	**-17.7**
	(8.8)	**(11.3)**	**(14.6)**	**(7.7)**	**(8.0)**	**(7.6)**	**(8.2)**
Three month period with gasoline E20, $e20_t^{gas}$	**-3.8**	**-4.4**	**-5.4**	**-4.8**	**-5.1**	**-3.5**	**-2.9**
	(3.9)	**(3.9)**	**(4.1)**	**(2.4)**	**(2.4)**	**(2.1)**	**(2.6)**
Mean across 12 station specific regressions							
R2	2.0%	2.7%	25.4%	70.3%	70.3%	75.6%	70.9%
Number of observations	2347	2347	2347	2333	2333	2304	2277
Number of regressors	3	4	60	65	66	113	71
Mean value of dependent variable	64.5	64.5	64.5	64.4	64.4	64.4	64.3
Robustness (h): Additionally control for the real price of diesel							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**14.4**	**18.0**	**-5.2**	**-21.7**	**-24.1**	**-27.8**	**-19.9**
	(11.7)	**(12.9)**	**(17.3)**	**(8.8)**	**(8.7)**	**(8.3)**	**(9.1)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-8.6**	**-4.7**	**-3.6**	**-3.8**	**-0.6**	**-0.9**
	(4.2)	**(4.4)**	**(5.9)**	**(3.0)**	**(2.9)**	**(2.6)**	**(3.1)**
Mean across 12 station specific regressions							
R2	2.3%	3.2%	22.2%	69.4%	69.5%	76.0%	70.7%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	96	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (i): Control for ridership on the public transport system in the São Paulo metropolis							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.1**	**19.9**	**-10.7**	**-19.9**	**-22.7**	**-30.9**	**-20.2**
	(8.9)	**(12.7)**	**(17.9)**	**(9.1)**	**(9.2)**	**(8.8)**	**(9.3)**
Three month period with gasoline E20, $e20_t^{gas}$	**-4.6**	**-5.1**	**-4.2**	**-4.5**	**-5.0**	**-2.7**	**-2.1**
	(4.0)	**(4.1)**	**(4.5)**	**(2.4)**	**(2.4)**	**(2.2)**	**(2.7)**
Mean across 12 station specific regressions							
R2	2.3%	3.2%	22.1%	69.1%	69.3%	75.8%	70.5%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	96	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (j): Control for physical industrial production in the state of São Paulo							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**16.3**	**19.6**	**7.9**	**-22.0**	**-24.8**	**-27.7**	**-19.9**
	(11.7)	**(12.9)**	**(17.3)**	**(8.8)**	**(8.7)**	**(8.3)**	**(9.1)**
Three month period with gasoline E20, $e20_t^{gas}$	**-4.9**	**-5.9**	**3.0**	**-4.2**	**-4.8**	**-1.5**	**-1.6**
	(4.2)	**(4.4)**	**(5.9)**	**(3.0)**	**(2.9)**	**(2.6)**	**(3.1)**
Mean across 12 station specific regressions							
R2	2.1%	2.8%	22.8%	69.3%	69.4%	75.9%	70.6%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	96	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step and clustered on date for the second step).

Table SXVI. Robustness: Predicting Ozone (μg/m^3), non-holiday weekday, 13:00 to 16:00 readings only

Baseline Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (k): Control for the São Paulo metropolitan area's "economically active" population							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**28.7**	**18.9**	**-0.1**	**-21.3**	**-24.9**	**-31.9**	**-20.8**
	(10.9)	**(12.9)**	**(16.8)**	**(8.2)**	**(8.2)**	**(8.0)**	**(8.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**-6.4**	**-6.0**	**-7.4**	**-4.0**	**-4.4**	**-2.5**	**-1.6**
	(4.0)	**(4.0)**	**(4.4)**	**(2.5)**	**(2.4)**	**(2.2)**	**(2.6)**
Mean across 12 station specific regressions							
R2	3.9%	4.8%	22.4%	69.3%	69.5%	76.0%	70.7%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	96	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (l): Control for the São Paulo metropolitan area's work force							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**8.0**	**23.9**	**-7.3**	**-21.7**	**-24.5**	**-30.7**	**-20.2**
	(11.6)	**(12.8)**	**(16.2)**	**(8.4)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.5**	**-5.6**	**-4.4**	**-3.9**	**-4.6**	**-3.3**	**-2.0**
	(4.0)	**(4.0)**	**(4.3)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	2.2%	3.4%	22.3%	69.4%	69.5%	76.0%	70.8%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	97	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (m): Control for the São Paulo metropolitan area's mean real earnings							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**10.7**	**19.7**	**-4.7**	**-29.0**	**-30.9**	**-33.4**	**-26.3**
	(9.1)	**(12.8)**	**(18.4)**	**(9.9)**	**(9.8)**	**(8.9)**	**(10.0)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.5**	**-5.9**	**-5.2**	**-4.4**	**-4.9**	**-3.0**	**-2.3**
	(4.0)	**(4.0)**	**(4.4)**	**(2.3)**	**(2.3)**	**(2.0)**	**(2.5)**
Mean across 12 station specific regressions							
R2	2.1%	3.0%	22.0%	69.2%	69.3%	75.8%	70.6%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	4	5	44	49	50	97	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (n): Base the gasoline share on predictions of a multinomial logit model (rather than a multinomial probit)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**6.5**	**12.7**	**-5.5**	**-20.3**	**-23.1**	**-28.3**	**-19.6**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.5**	**-6.0**	**-5.2**	**-3.6**	**-4.1**	**-2.4**	**-1.6**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.5%	2.4%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (o): Base the gasoline share on aggregate monthly reports of gasoline and ethanol shipments by wholesalers							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**-34.4**	**-52.4**	**-37.3**	**-33.9**	**-36.0**	**-41.8**	**-30.4**
	(18.9)	**(24.0)**	**(25.5)**	**(13.0)**	**(13.0)**	**(11.5)**	**(13.4)**
Three month period with gasoline E20, $e20_t^{gas}$	**-4.5**	**-4.4**	**-6.5**	**-4.9**	**-5.4**	**-3.6**	**-2.6**
	(3.9)	**(3.9)**	**(4.2)**	**(2.3)**	**(2.3)**	**(2.1)**	**(2.5)**
Mean across 12 station specific regressions							
R2	1.7%	3.1%	22.0%	69.0%	69.1%	75.7%	70.3%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step -- except for specification (o) where the share is based on data -- and clustered on date for the second step).

Table SXVII. Robustness: Predicting Ozone (μg/m^3), non-holiday weekday, 13:00 to 16:00 readings only

Baseline Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (p): Standard errors computed from 500 rather than 200 bootstrap samples (on both sets of data)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**9.6**	**19.3**	**-5.9**	**-21.9**	**-24.7**	**-30.2**	**-20.7**
	(9.2)	**(13.4)**	**(16.1)**	**(8.8)**	**(8.9)**	**(8.2)**	**(8.9)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.3**	**-3.9**	**-4.5**	**-2.8**	**-1.9**
	(4.0)	**(4.0)**	**(4.3)**	**(2.4)**	**(2.4)**	**(2.1)**	**(2.5)**
Mean across 12 station specific regressions							
R2	1.6%	2.5%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	43	48	49	96	54
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7
Robustness (q): Assume consumers store a larger volume of fuel in their vehicle tanks (7 rather than 4 days)							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**7.9**	**15.9**	**-5.7**	**-21.6**	**-24.5**	**-29.8**	**-20.7**
	(9.0)	**(12.9)**	**(16.2)**	**(8.3)**	**(8.3)**	**(7.9)**	**(8.6)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.6**	**-6.1**	**-5.2**	**-3.6**	**-4.2**	**-2.4**	**-1.6**
	(4.0)	**(4.0)**	**(4.2)**	**(2.3)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.4%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
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
Robustness (r): Instrument for the predicted gasoline share using the ethanol-to-gasoline price ratio							
Main variables of interest (mean estimates across 12 station-specific regressions)							
Proportion BFVs burning gasoline E25 over ethanol E100, s_t^{gas}	**13.9**	**29.7**	**-4.5**	**-23.0**	**-25.8**	**-33.0**	**-22.1**
	(8.8)	**(13.0)**	**(16.2)**	**(8.3)**	**(8.4)**	**(7.6)**	**(8.7)**
Three month period with gasoline E20, $e20_t^{gas}$	**-5.9**	**-6.6**	**-5.3**	**-3.9**	**-4.5**	**-2.9**	**-1.9**
	(4.0)	**(4.0)**	**(4.1)**	**(2.2)**	**(2.2)**	**(2.0)**	**(2.4)**
Mean across 12 station specific regressions							
R2	1.6%	2.4%	21.8%	69.0%	69.2%	75.7%	70.4%
Number of observations	1574	1574	1574	1560	1560	1531	1515
Number of regressors	3	4	44	49	50	96	55
Mean value of dependent variable	67.8	67.8	67.8	67.7	67.7	67.7	67.7

Notes: See notes to Table I in the main text. Estimated coefficients and standard errors (in parentheses, corrected for sampling variation in the first step -- except for instrumental variables specification (r) -- and clustered on date for the second step).

References.

1 *Please see http://www.isosugar.org/*.

2 Mukherjee, A. in *Wall Street Journal* (Dow Jones & Company, Inc., New York, 2009).

3 Economist. in *The Economist* (The Economist, London, 2011).

4 *Please see http://www.anp.gov.br/preco/prc/Resumo_Por_Municipio_Index.asp*.

5 *Please see http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&c=1737*.

6 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).

7 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).

8 *Please see http://www.shell.com/home/content/bra/products_services/on_the_road/fuels/shell_v power_ethanol/*.

9 Stephanes, R. in *Portaria MAPA No. 7* (ed Pecuária e Abastecimento Ministério da Agricultura) (Diário Oficial da União, Brasília, 2010).

10 *Please see http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&c=2938*.

11 Jank, M. S. Os Desafios do Setor Sucroenergético e o "Movimento + Etanol" [The Challenges facing the Sugar-Energy Sector and the "More Ethanol Movement"] Portuguese. (União da Indústria da Cana-de-Açúcar - UNICA, São Paulo, 2012).

12 *Please see http://www.imf.org/external/data.htm*.

13 *Please see http://www.seade.gov.br/produtos/pib/2009/cre2009_11.htm*.

14 *Please see http://www.detran.sp.gov.br/*.

15 Sindipeças. Levantamento da frota circulante brasileira [Study of the Brazilian circulating fleet] Portuguese. (Sindicato Nacional da Indústria de Componentes para Veículos Automotores
São Paulo, 2011).

16 de Paiva Cardoso, C. E. Qual o número de veículos que circula em São Paulo? [What is the number of vehicles circulating in São Paulo?] Portuguese. *Revista dos Transportes Públicos* **31**, 9 (2009).

17 Izidoro, A. in *Folha de S.Paulo* (São Paulo, 2009).

18 Duranton, G. & Turner, M. A. The Fundamental Law of Road Congestion: Evidence from US Cities. *Am Econ Rev* **101**, 2616-2652, doi:DOI 10.1257/aer.101.6.2616 (2011).

19 MMA. Primeiro Inventário Nacional de Emissões Atmosféricas por Veículos Automotores Rodoviários [First National Inventory of Atmospheric Emissions by Road Auto Vehicles] Portuguese
(Secretaria de Mudanças Climáticas e Qualidade Ambiental, Brasília, 2011).

20 NHTSA. Vehicle Survivability and Travel Mileage Schedules. Report No. DOT HS 809 952, (Lu, S. (National Center for Statistics and Analysis), 2006).

21 Meghan, B., Knittel, C. R. & Zettelmeyer, F. Are Consumers Myopic? Evidence from New and Used Car Purchases. *Working Paper, Northwestern University and MIT* (2012).

22 INMETRO. in *INMETRO (Instituto Nacional de Metrologia, Normalização e Qualidade Industrial)* (2009).

23 INMETRO. in *INMETRO (Instituto Nacional de Metrologia, Normalização e Qualidade Industrial)* (2010).

24 *Please see http://cbn.globoradio.globo.com/home/HOME.htm*.
25 Goolsbee, A. & Petrin, A. The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* **72**, 351-381, doi:DOI 10.1111/j.1468-0262.2004.00494.x (2004).
26 Dabelstein, W., Reglitzky, A., Schütze, A. & Reders, K. in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag GmbH & Co. KGaA, 2000).
27 Dzyaloshinskii, I. E., Lifshitz, E. M. & Pitaevskii, L. P. *Adv. Phys.* **10**, 165 (1961).
28 Knight, C. A. *J. Geophys. Res.* **101**, 12921-12928 (1996).
29 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. *Atmospheric Environment* **36**, 403-410, doi:10.1016/s1352-2310(01)00508-8 (2002).
30 Al-Hasan, M. Effect of ethanol-unleaded gasoline blends on engine performance and exhaust emission. *Energy Conversion and Management* **44**, 1547-1561, doi:10.1016/s0196-8904(02)00166-8 (2003).
31 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. *Journal of Automobile Engineering* **223**, 941 (2009).
32 Mulawa, P. A. *et al.* Effect of Ambient Temperature and E-10 Fuel on Primary Exhaust Particulate Matter Emissions from Light-Duty Vehicles. *Environmental Science and Technology* **31**, 1302-1307, doi:10.1021/es960514r (1997).
33 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. *Atmospheric Environment* **36**, 3495-3503, doi:10.1016/s1352-2310(02)00288-1 (2002).
34 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. *Journal of Hazardous Materials* **123**, 29-34, doi:10.1016/j.jhazmat.2005.03.046 (2005).
35 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. *Renewable Energy* **31**, 2534-2542, doi:10.1016/j.renene.2006.01.004 (2006).
36 Brodowicz, P., Carey, P., Cook, R. & Somers, J. Motor vehicle related air toxics study. (USEPA, 1993).
37 Lide, D. *CRC Handbook of Chemistry and Physics*. 86 edn, (CRC Press, 2005).
38 Annamalai, K. & Puri, I. K. in *Computational Mechanics and Applied Analysis* (ed J. N. Reddy) (CRC Press, Boca Raton, 2007).
39 Ragland, K. W. & Bryden, K. M. *Combustion Engineering*. 2 edn, (CRC Press, 2011).
40 Graham, L. A., Belisle, S. L. & Baas, C.-L. Emissions from light duty gasoline vehicles operating on low blend ethanol gasoline and E85. *Atmospheric Environment* **42**, 4498-4516, doi:10.1016/j.atmosenv.2008.01.061 (2008).
41 Lynd, L. R. Overview and Evaluation of Fuel Ethanol from Cellulosic Biomass: Technology, Economics, the Environment, and Policy. *Annual Review of Energy and the Environment* **21**, 403-465, doi:doi:10.1146/annurev.energy.21.1.403 (1996).
42 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. *Atmospheric Environment* **37**, 949-957, doi:10.1016/s1352-2310(02)00973-1 (2003).

43 Baukal, C. E. J. (CRC Press, Boca Raton, 2000).
44 Benedik, M. J., Gibbs, P. R., Riddle, R. R. & Willson, R. C. Microbial denitrogenation of fossil fuels. *Trends in Biotechnology* **16**, 390-395, doi:10.1016/s0167-7799(98)01237-2 (1998).
45 Poulopoulos, S. G., Samaras, D. P. & Philippopoulos, C. J. Regulated and unregulated emissions from an internal combustion engine operating on ethanol-containing fuels. *Atmospheric Environment* **35**, 4399-4406, doi:10.1016/s1352-2310(01)00248-5 (2001).
46 CETESB. Relatório Anual sobre a Qualidade do Ar no Estado de São Paulo [Annual Report on Air Quality in the State of São Paulo] Portuguese. (Companhia Ambiental do Estado de São Paulo, Sao Paulo, 2010).
47 Glassman, I. & Yetter, R. A. *Combustion*. 4 edn, (Elsevier, 2008).
48 Law, C. K. *Combustion Physics*. (Cambridge University Press, 2006).
49 Wu, C.-W., Chen, R.-H., Pu, J.-Y. & Lin, T.-H. The influence of air–fuel ratio on engine performance and pollutant emission of an SI engine using ethanol–gasoline-blended fuels. *Atmospheric Environment* **38**, 7093-7100, doi:10.1016/j.atmosenv.2004.01.058 (2004).
50 Yüksel, F. & Yüksel, B. The use of ethanol–gasoline blend as a fuel in an SI engine. *Renewable Energy* **29**, 1181-1191, doi:10.1016/j.renene.2003.11.012 (2004).
51 Abdel-Rahman, A. A. & Osman, M. M. Experimental investigation on varying the compression ratio of SI engine working on different ethanol-gasoline fuel blends. *International Journal of Energy Research* **21**, 31-40, doi:10.1002/(sici)1099-114x(199701)21:1<31::aid-er235>3.0.co;2-5 (1997).
52 Fenimore, C. P. Formation of nitric oxide in premixed hydrocarbon flames. *Symposium (International) on Combustion* **13**, 373-380, doi:10.1016/s0082-0784(71)80040-1 (1971).
53 da Silva, M. F., de Assunção, J. V., de Fátima Andrade, M. & Pesquero, C. R. Characterization of Metal and Trace Element Contents of Particulate Matter (PM10) Emitted by Vehicles Running on Brazilian Fuels-Hydrated Ethanol and Gasoline with 22% of Anhydrous Ethanol. *Journal of Toxicology & Environmental Health: Part A* **73**, 901-909 (2010).
54 Vertes, A. A., Qureshi, N., Blaschek, H. P. & Yukawa, H. (John Wiley & Sons, West Sussex, 2010).
55 Orlando, J. P., Alvim, D. S., Yamazaki, A., Corrêa, S. M. & Gatti, L. V. Ozone precursors for the São Paulo Metropolitan Area. *Science of the Total Environment* **408**, 1612-1620 (2010).
56 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 Pollution* **195**, 201-213, doi:10.1007/s11270-008-9740-x (2008).
57 Martins, L. & Andrade, M. Emission scenario assessment of gasohol reformulation proposals and ethanol use in the metropolitan area of São Paulo. *The Open Atmospheric Science Journal* **2**, 166-175 (2008).
58 Martins, L. D. A., Maria de Fátima; Ynoue, Rita Yuri; Albuquerque, Édler Lins de; Tomaz, Edson; Vasconcellos, Pérola de Castro. Ambiental volatile organic compounds in the megacity of São Paulo. *Química Nova* **31**, 2009-2013 (2008).

59 Finlayson-Pitts, B. & Pitts, J. *Chemistry of the Upper and Lower Atmosphere: Theory, Experiments, and Applications*. (Academic Press, 2000).
60 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. *Atmospheric Environment* **44**, 1192-1199, doi:10.1016/j.atmosenv.2009.12.024 (2010).
61 Jacobson, M. Z. Effects of Ethanol (E85) versus Gasoline Vehicles on Cancer and Mortality in the United States. *Environmental Science and Technology* **41**, 4150-4157, doi:10.1021/es062085v (2007).
62 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. *Journal of the Brazilian Chemical Society* **15**, 646-651 (2004).
63 Hulsey, B. & Coleman, B. Clearing the Air with Ethanol: A review of the real world impact from fuels blended with ethanol. (Better Environmental Solutions and Renewable Energy Action Project, 2006).
64 Russell, A. G., Mccue, K. F. & Cass, G. R. Mathematical-Modeling of the Formation of Nitrogen-Containing Air-Pollutants .1. Evaluation of an Eulerian Photochemical Model. *Environmental Science & Technology* **22**, 263-270 (1988).
65 Romieu, I., Weitzenfeld, H. & Finkelman, J. Urban Air-Pollution in Latin-America and the Caribbean. *J Air Waste Manage* **41**, 1166-1171 (1991).
66 Anderson, L. G. Ethanol fuel use in Brazil: air quality impacts. *Energ Environ Sci* **2**, 1015-1037, doi:Doi 10.1039/B906057j (2009).
67 CETESB. Relatório Anual sobre a Qualidade do Ar no Estado de São Paulo [Annual Report on Air Quality in the State of São Paulo] Portuguese. (Companhia Ambiental do Estado de São Paulo, Sao Paulo, 2008).
68 *Please see http://www.cetesb.sp.gov.br/ar/qualidade-do-ar/32-qualar*.
69 Beirle, S., Platt, U., Wenig, M. & Wagner, T. Weekly cycle of NO2 by GOME measurements: A signature of anthropogenic sources. *Atmos. Chem. Phys.* **3**, 2225-2232 (2003).
70 Guardani, R., Aguiar, J. L., Nascimento, C. A. O., Lacava, C. I. V. & Yanagi, Y. Ground-level ozone mapping in large urban areas using multivariate statistical analysis: Application to the Sao Paulo Metropolitan Area. *J Air Waste Manage* **53**, 553-559 (2003).
71 Silva Júnior, R. S. d., Oliveira, M. G. L. d. & Andrade, M. d. F. Weekend/weekday differences in concentrations of ozone, nox, and non-methane hydrocarbon in the metropolitan area of São Paulo. *Revista Brasileira de Meteorologia* **24**, 100-110 (2009).
72 Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. (John Wiley & Sons, 1998).
73 *Please see http://cetsp1.cetsp.com.br/monitransmapa/agora/*.
74 CET. Relatório sobre a ampliação da base de vias monitoradas para medição da lentidão de trânsito na Cidade de São Paulo [Report on the expansion of the road grid that is monitored for traffic congestion in the City of São Paulo] Portuguese. 6 (Companhia de Engenharia de Tráfego, São Paulo, 2007).
75 *Please see http://cetsp1.cetsp.com.br/institucional/rodizio/default.asp*.
76 *Please see http://www.cetsp.com.br/consultas/rodizio-municipal/como-funciona.aspx*.

77 Hughes, J. E., Knittel, C. R. & Sperling, D. Evidence of a shift in the short-run price elasticity of gasoline demand. *Energ J* **29**, 113-134 (2008).
78 *Please see http://www.sptrans.com.br/a_sptrans/*.
79 *Please see http://www9.prefeitura.sp.gov.br/spMovimento/dadosoperacionais/caponibus.php*.
80 *Please see http://www.ibge.gov.br/home/estatistica/indicadores/industria/pimpfregional/defaulttab.shtm*.
81 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] Portuguese. (Centro de Estudos Integrados sobre Meio Ambiente e Mudanças Climáticas, Universidade Federal do Rio de Janeiro, São Paulo, 2005).
82 Almança, R. Anuário Estatístico de Energéticos por Município no Estado de São Paulo [Annual Energy Statistics by Municipality in the State of São Paulo] Portuguese. (Secretaria de Energia do Governo de São Paulo, São Paulo, 2010).
83 Negri, J. C. Balanço Energético do Estado de São Paulo [Energy Balance for the State of São Paulo] Portuguese. (Secretaria de Saneamento e Energia do Governo de São Paulo, São Paulo, 2010).
84 Schlenker, W. & Walker, W. R. Airports, Air Pollution, and Contemporaneous Health. *NBER Working Paper No. 17684* (2011).
85 *Please see http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&o=16&i=P&c=2040*.
86 *Please see http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&o=16&i=P&c=2050*.
87 *Please see http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?z=t&o=16&i=P&c=2187*.