The Consequences of Natural Gas Subsidy Reform: The Case of Air Pollution in Thailand

Thanicha Ruangmas*

June 20, 2018

Abstract

Previous work has shown that policies that encourage substitution from gasoline and diesel to compressed natural gas (CNG) reduce air pollution. Less is known about consumer responses to such policies. This paper investigates policies that encourage CNG adoption and its impact on air pollution in Thailand from 2008 to 2015. First, I analyze the mechanisms underlying CNG adoption by estimating the effect of various fuel prices and CNG availability on CNG sales from retail fueling stations. Then, I estimate the effects of various fuel prices and CNG availability on air pollution in Thailand. By integrating simulated data from an air pollution model, I quantify the sources of air pollution measured at each air pollution monitor. I find that an increase in CNG availability leads to an increase in CNG sales. As a result, the doubling of CNG stations decreases SO2 concentrations by one percent.

*University of Wisconsin - Madison. ruangmas@wisc.edu. Funding from Economy and Environment Program for Southeast Asia (EEPSEA) and University of Wisconsin - Madison's Global Health Institute (GHI) is gratefully acknowledged. I would like to thank Corbett Grainger, Dominic Parker, Dan Phaneuf, and Jamie Schauer for their guidance. I would also like to thank Andrew Schreiber for his technical assistance. Any errors are my own.

1 Introduction

According to the International Monetary Fund, the elimination of energy subsidies can cut global CO2 emissions by 20 percent, cut pre-mature air pollution deaths by more than half, and raise global economic welfare by 1.8 trillion dollars a year (Coady et al. 2015). With global pressure to remove fossil fuel subsidies, some policymakers argue that compressed natural gas (CNG) subsidies could be beneficial. CNG fueled vehicles emit fewer conventional air pollutants than gasoline and diesel vehicles (Krupnick 2011). The adoption of CNG could reduce air pollutants which could reduce the health damages and mortality attributed to pollution.

This paper analyzes the benefits of CNG adoption in terms of its effects on air pollution. The Thai government introduced CNG as an alternative fuel choice for cars in 2003. CNG prices have been regulated in two separate parts. First, the government mandated that PTT Plc., the sole distributor of CNG in Thailand, sell CNG to fueling stations at prices lower than the cost. The Thai government began repaying PTT for some of their losses. Second, the Thai government started fixing the retail price of CNG at a rate lower than the market price. The Thai government stopped repaying PTT in 2012 and ultimately deregulated its retail prices in 2016. Because PTT is selling CNG at lower prices than its costs, the loss in profit can exceed the gain in consumer surplus. This welfare loss can be offset if there are positive externalities from the policy. A cheap alternative fuel option can decrease income inequality. CNG adoption can also be justified if its air pollution benefits partially offset its costs.

Unlike other countries where the impact of CNG adoption on air pollution has been studied, the Thai government did not mandate that cars be fueled by CNG. Instead, the Thai government decreased its price and increased its availability to encourage voluntary adoption. A similar case to Thailand is Dhaka, which widened the gap between gasoline and CNG prices. It was found that CNG car owners drove more after the policy (Wadud 2014). The rebound effect could potentially offset the benefits of its low price and increase congestion. Therefore, it is essential that we understand how consumers respond to such policies and how their response impact air quality. Most studies have either studied the effects of fuel prices and fuel availability for alternative fuel adoption (Greene 1989; Bunch et al. 1993; Anderson and Sallee 2011; Anderson 2012; Du and Li 2016; Li et al. 2017) or studied the impact of alternative fuel adoption on air pollution (Goyal 2003; Kathuria 2004; Ravindra et al. 2006; Narain and Krupnick 2007; Suthawaree et al. 2011; Foster and Kumar 2011). To our knowledge, the only paper that has looked at the mechanisms of fuel adoption and its impact on air pollution is Auffhammer and Kellogg (2011). As policies that encourage CNG adoption are becoming increasingly common around the world, it is essential that a comprehensive study has to be done to study its effects.

Many papers have assessed the direct impact of CNG conversion on air pollution as many countries have also encouraged CNG adoption in vehicles. However, existing research has relied on engineering simulation models (Schifter 2000; Dholakia et al. 2013; Wadud 2013) which ignore the behavioral changes that come with fuel adoption, heterogeneity in emissions due to engine technology, driver behavior, traffic conditions, and meteorology. A significant amount of emissions from on-road vehicles do not come from tail-pipe emissions alone. Particulate matter can come from abrasion, and volatile organic compounds can come from leakage at fueling stations. Most studies that have assessed changes in air quality after a CNG conversion policy rely on a single air pollution monitor (Goyal 2003; Kathuria 2004; Ravindra et al. 2006; Narain and Krupnick 2007; Suthawaree et al. 2011).[1] This paper looks at not only the effects of CNG price and availability on air pollution but also the effect of price and availability on CNG sales. Moreover, air pollution data from more than 44 air pollution monitors are used.

I first estimate the effect of gasoline and diesel prices relative to CNG and CNG stations on province-level CNG sales. I then estimate the effects of gasoline and diesel prices relative to CNG and CNG availability on air pollution concentration at each air pollution monitor. To identify pollution sources of each air pollution monitor, a Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT; see Draxler and Hess 1997; Draxler and Hess 1998; Draxler 1999) is used. HYSPLIT is an air pollution simulation model which could identify sources of air pollution from each location to each air pollution monitor every day of the year. I use results from the HYSPLIT simulations to weigh the impact of each fueling station and province level fuel prices within each 120-square kilometer grid cell on each air pollution monitor.

Because CNG stations may not be documented properly and are built in areas with high demand, I have to control for measurement error and endogeneity. My primary source of data comes from the Department of Energy Business (DOEB) database which contains individual fueling station data in Thailand. For each fueling station, I have their address, date of opening, and which fuel type it sells. Because the data set does not include the closing of fueling stations, the number of CNG stations can be overestimated. Furthermore,

[1] An exception of this is Foster and Kumar (2007) which uses remote sensing data. These studies have found improvements in air pollution within the regulated area.

not all CNG stations have registered with the DOEB. This can lead to a measurement bias in either direction and increase the standard errors of the coefficient on CNG station.

CNG stations are built in areas with high demand and time-varying, market-specific demand patterns are unobserved. Because the Thai government explicitly states that CNG has been introduced to improve air quality, CNG fueling stations may be opened in highly polluted areas first. To correct for endogeneity and measurement error bias, I employ a two-stage-least-squares (2SLS) estimation strategy. From 2008 onward, CNG stations can only be built by licensed engineering companies. The licenses are given to companies that have gone through government training in CNG infrastructure and safety precautions. I assume that CNG stations are built by the nearest licensed engineering company. In my first regression with provincial CNG sales as a dependent variable, I use the interaction between the distance between the closest licensed engineering company to the province centroid and the lagged CNG stations in all other areas as an instrument for the number of CNG stations in each province. In my second regression with air pollution concentrations as dependent variables, I use the interaction between the distance of the closest licensed engineering company to each 120-square kilometer grid cell and the lagged CNG stations in all other areas as an instrument for the number of CNG stations surrounding an air pollution monitor.

CNG prices are fixed nationwide while gasoline and diesel prices vary at the province-brand level. Because there is very little cross-sectional variation between gasoline and diesel prices and CNG, identifying the effect of CNG prices on CNG sales is not possible. Instead, I find that when the number of CNG stations in a province doubles, CNG sales can increase 27 percent. Because CNG vehicles emit negligible quantities of SO2, when the number of CNG stations in an area is doubled, SO2 concentrations decrease by one percent.

This research finds that the availability of CNG plays a primary role in encouraging CNG adoption, which can improve SO2 concentrations. I find that the effect of CNG adoption on other air pollutants are insignificant. However, the improvement in air quality is minimal, such that the air pollution benefits gained from the adoption of CNG do not justify the subsidies.

The remainder of this paper is organized is follows: Section 2 describes fuel choices for vehicles and its regulation in Thailand. Section 3 reviews existing literature. Section 4 describes our data. Section 5 explains our identification strategy. Section 6 presents our estimation results. Section 7 presents results from similar estimation strategies for robustness checks. Section 8 discusses our conclusions.

2 Policy Context

2.1 Pollution Regulation in Thailand

Air pollution in Thailand is monitored by the Pollution Control Department which sets the same air quality standards as the United States National Ambient Air Quality Standards (US NAAQS) that was amended under the 1990 Clean Air Act.[2] Unlike the United States, where counties with pollution concentrations exceeding the standard would have to apply for the "Reasonably Available Control Technology" to abate air pollution, the Thai standard is a mere guideline.

Because no serious action has been taken to reduce outdoor air pollution in Thailand, the air quality is severe in some areas. In general, average air pollution concentration in

[2] The quantitative details of the standard are discussed in section 4 of this paper

Table 1: A comparison of monthly maximum pollution concentrations in Thailand from March 2008 to December 2015, similar statistics in LA, and the US standards

	Obs	Mean	Std. Dev.	Min	Max	US NAAQS	Similar statistics to the mean in LA
1 hr. maximum CO (ppm)	3,850	2.425	1.309	0.200	9.200	35.000	4.300
24 hr. maximum PM10 (ug/m3)	4,054	70.562	42.693	8.000	352.000	150.000	95.000
1 hr. maximum SO2 (ppb)	3,601	17.381	18.727	0.000	373.000	75.000	12.000
1 hr. maximum NO2 (ppb)	3,751	54.094	27.222	1.000	180.000	53.000	75.000

Note: Maximum monthly air pollution data in Thailand comes from the Pollution Control Department. The table only include observations used in regressions in Table 11 and 9. US NAAQS for CO are not to be exceeded more than once per year. The reported CO level in LA is the 2nd maximum in 2015. US NAAQS for PM10 are not to be exceeded more than once per year on average over 3 years. The reported PM10 level in LA is the 2nd maximum in 2015. The 99th percentile of 1-hour daily maximum SO2 concentration cannot exceed the US NAAQS reported in the table. The reported SO2 level in LA is the 99th percentile in 2015. The 98th percentile of 1-hour daily maximum NO2 concentration, averaged over 3 years cannot exceed the US NAAQS reported in the table. The reported NO2 level in LA is the 98th percentile in 2015.

Thailand is comparable to Los Angeles, but with a much larger variance. Table 1 compares the average maximum pollution levels in Thailand with the average maximum pollution levels in Los Angeles and the US NAAQS. Areas that exceed the national standards are clustered in Bangkok and the northern region. Figure 1 shows the average number of months in a year that the maximum pollution levels of CO, SO2, PM10, and NO2 from each monitor exceed air quality standards during the period of 2003 to 2015.

2.2 Fuel Options for Road Vehicles in Thailand

The main sources of natural gas today come from the Gulf of Thailand and Myanmar. The initial intention of using natural gas was to make liquefied petroleum gas (LPG) in order to replace wood burning in cooking. Natural gas is also used in electricity production, mining, and other manufacturing industries. High world oil prices in the early 2000s led households to start adopting LPG in their cars instead of gasoline or diesel. As LPG is highly reactive,

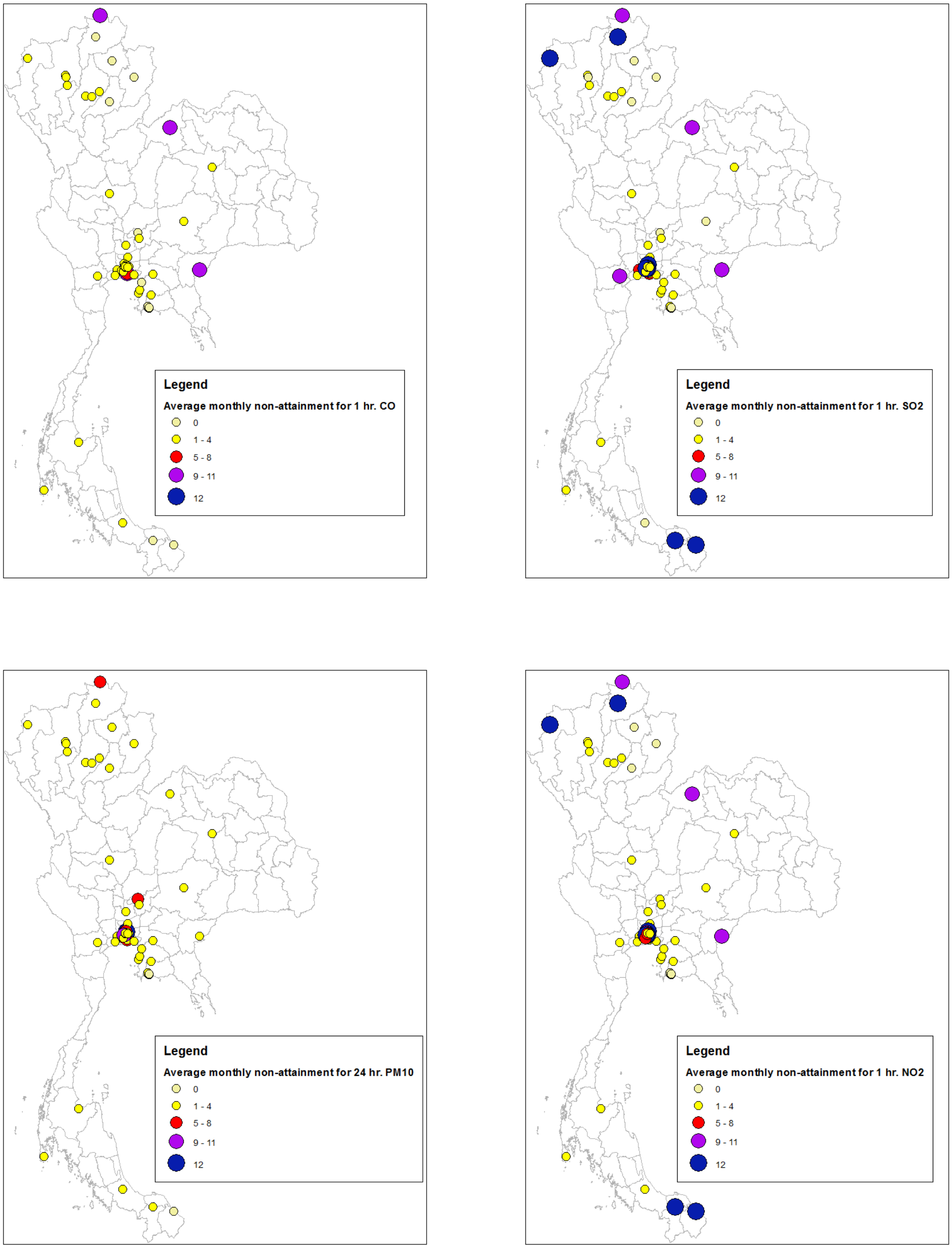

Figure 1: Average number of months in a year that pollution levels in each monitor exceed air quality standards during the period of 2003 to 2015

denser than air, it does not rise up when there is a leak. As a result, LPG in vehicles is very dangerous. Therefore, the Thai government then decided to introduce CNG in road vehicles. CNG decreases dependence on foreign fuel sources. CNG is much cheaper than gasoline and diesel. CNG cars also emit less pollutants than traditional fuel sources. Figure 2 summarizes the average emission factors (in grams per mile driven) of different vehicles and different fuel types in Thailand. The emission factors data are based on the United States Argonne National Laboratory's Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation Model (GREET; Cai, Burnham and Wang 2013).[3] Based on their estimation factors, CNG vehicles emit negligible amounts of SO2. This is because SO2 is emitted from the conversion of sulfur in the fuel to SO2. CNG have very small sulfur content so the SO2 emissions should be very small. On the other hand, CNG vehicles emit more PM10 than gasoline vehicles. This is because the combustion conditions are very different for CNG compared to gasoline in an internal combustion condition. PM10 emissions from gasoline engines are typically from the emissions of engine oil.

CNG was introduced in 2004, and the first CNG fueling station was built in 2006. PTT Plc., Thailand's sole distributor of CNG and a state-owned enterprise, distributes CNG to retail fueling stations under the market price.

Traditional fuel options for road vehicles in Thailand include gasoline and diesel fuels. Common alternative fuel options include LPG and CNG, which are generally adopted by retrofitting a gasoline-fueled car. Figure 3 shows the proportion of cars that use only CNG,

[3]Data on car age and registered cars by fuel sources is combined with emission factors (in grams of pollution emitted per mile) from the GREET model. I assume that cars that are released in Thailand have the same emission factors as cars that are released in the US during the same year.

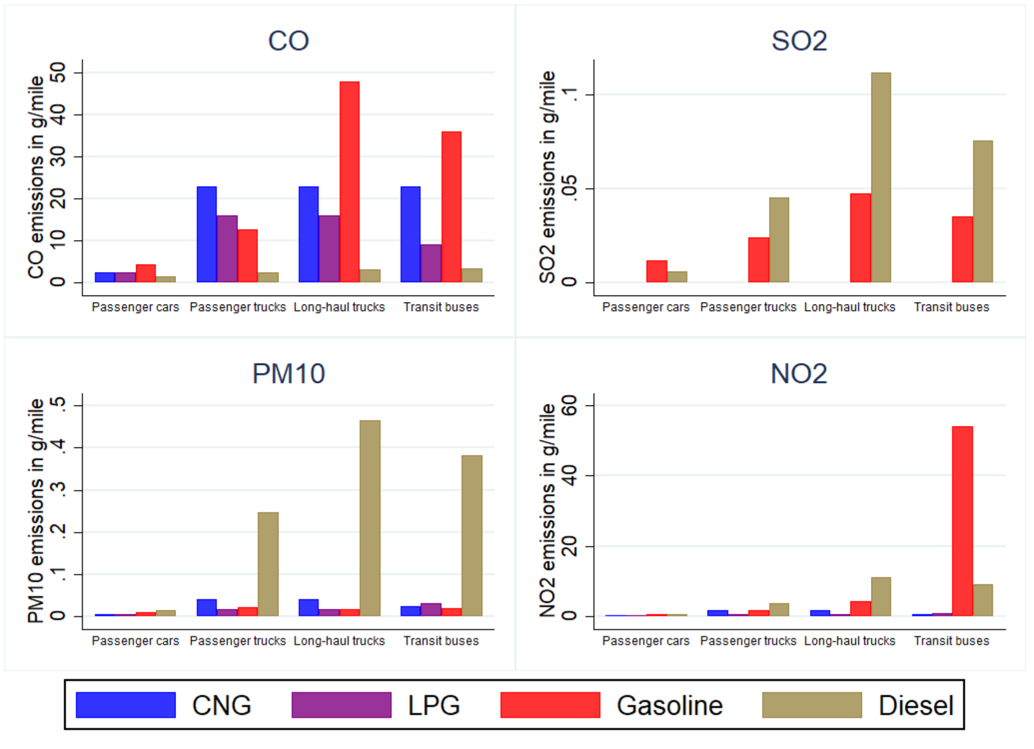

Figure 2: Emissions factor based on the GREET model

co-use CNG with gasoline, or co-use CNG with diesel.

Using the same data, I calculate potential emissions from 2010 to 2015 had each registered car driven one mile. Results are shown in Figure 4. Because I do not have data on vehicle miles traveled or miles per gallon, I cannot estimate the changes in emissions based on registered car and emissions factor data.

2.3 CNG Price Policies

Since 2007 CNG prices were fixed at 8.5 Baht per kilogram with plans to slowly lift the price ceiling and increase its price in 2011. In November 2007, the Thai government restructured the CNG market.[4] In short, CNG retail sellers have been mandated to sell

[4]See Appendix A for details of the CNG price restructuring policy.

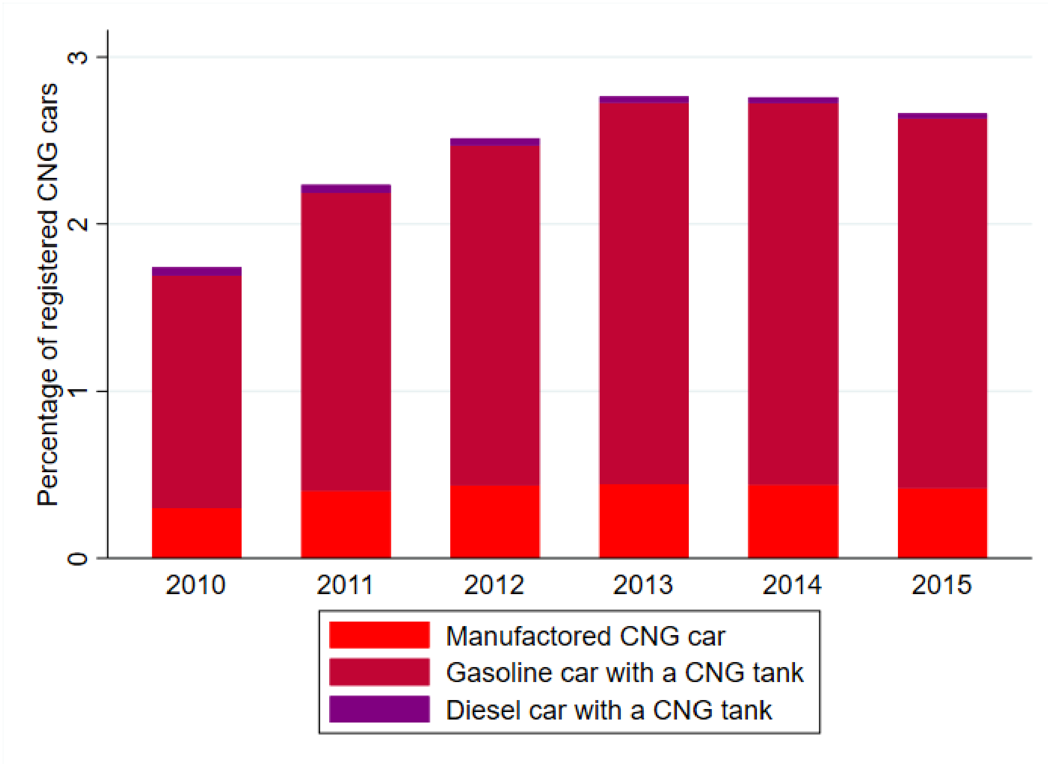

Figure 3: Percent of CNG cars relative to all registered cars in Thailand

CNG at a fixed price. PTT Plc., a state-owned-enterprise that extracts and distributes CNG, bears all the losses. To alleviate PTT's loss, the government paid PTT 2 Baht per kilogram of CNG sold since its restructuring. The government reduced their payment to PTT to 1 Baht per kilogram in 2011. In February 2012, the government their payments to PTT.

Figure 5 compares estimated CNG costs[5] and the regulated retail price of CNG. Low global natural gas prices caused the world price to fall below the fixed price in Thailand. This led to complete deregulation of CNG prices in March 2016. Since our period of study is

[5]This is the Henry Hub spot price for natural gas converted to Baht per kilogram using the Thai-US dollar exchange rate plus additional distribution and operating costs described in Appendix A. The additional costs are assumed to be the same as the costs in July 2013 (PTT 2013).

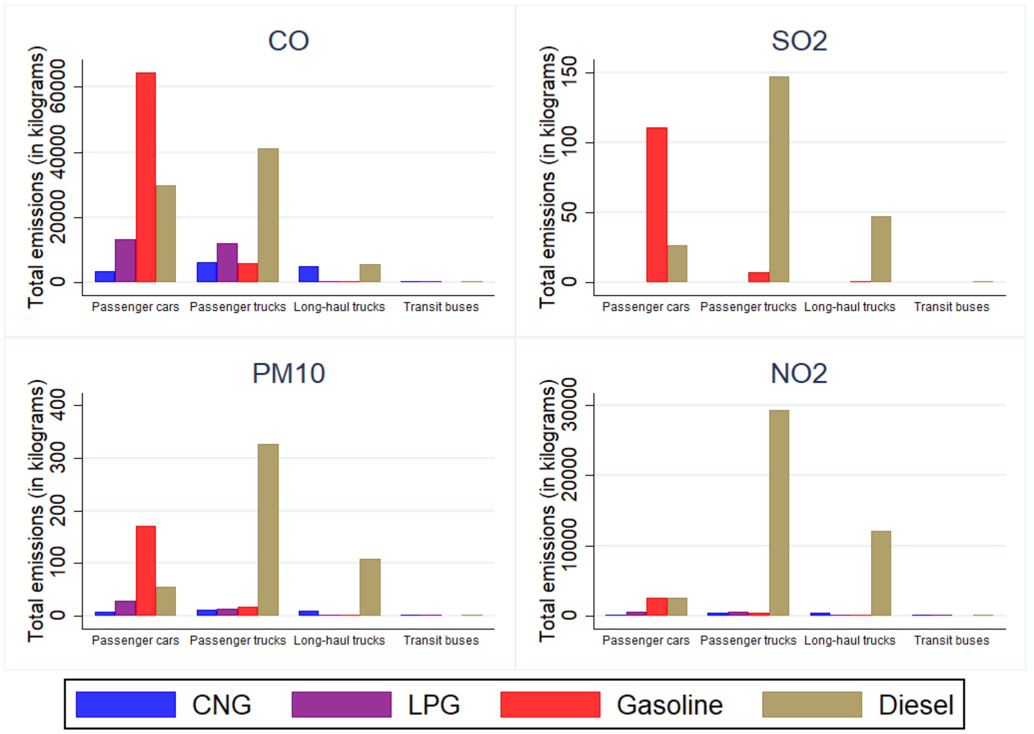

Figure 4: Total emissions from 2010 to 2015 had each registered car driven one mile

from 2008 to 2015 when there is almost no variation in CNG prices, I treat the retail CNG price in Thailand as exogenous.

Figure 6 compares the estimated loss from CNG subsidies (in Baht per kilogram) and the quantity of CNG sold each month (in tons per month.)[6] This is the difference in the estimated marginal costs and the fixed unit price at retail stations. This is the loss burdened by PTT and the government during the time that it subsidized PTT. Based on these estimates, the annual loss from the subsidy is during the time period focused on this paper (March 2008 to December 2015) is 6.45 billion Baht or about 200 million USD a year. This is about 0.05 percent of Thailand's GDP.[7]

[6]Details of quantity sold is available in section 4 of this paper.

[7]Thailand's GDP is about 400 billion USD a year. (Source: https://tradingeconomics.com/thailand/gdp)

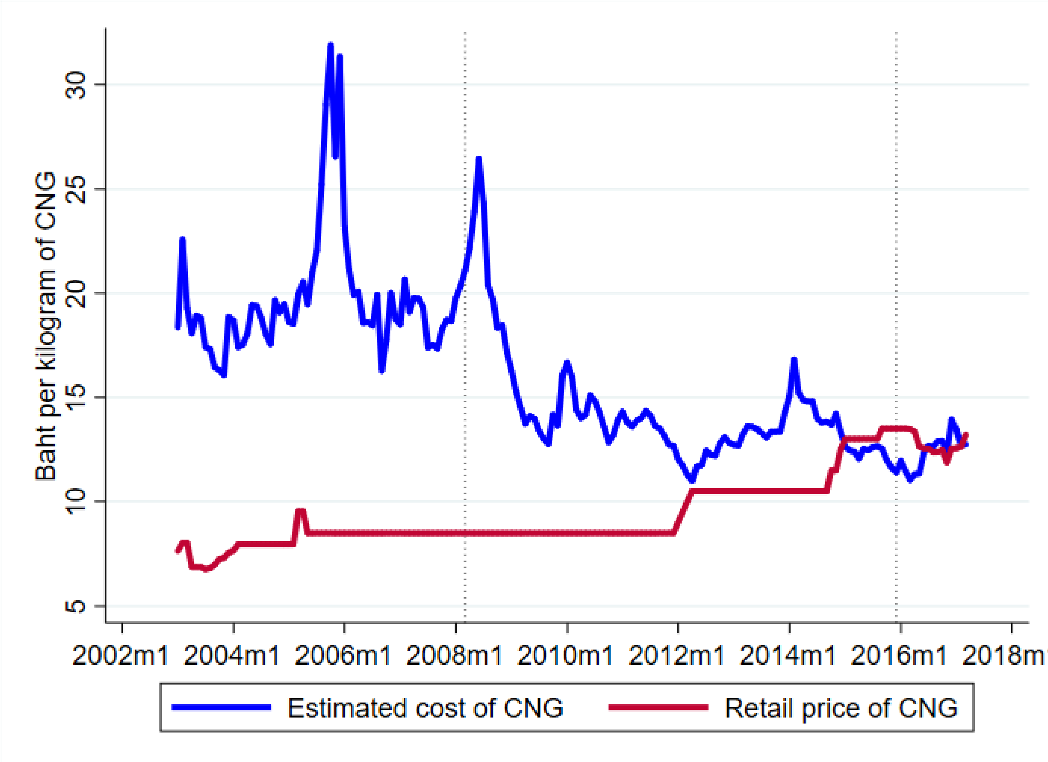

Figure 5: The figure compares domestic and global CNG price. The area between two dotted lines represents time period focused in this paper.

2.4 Price Policies of Other Fuels

Four main types of fuels used in Thailand's on road vehicles include gasoline,[8] diesel, LPG and CNG. Because gasoline and diesel are imported, the domestic prices closely follow the world prices. There is very little cross sectional variation in gasoline and its prices as it is subject to fixed transport costs and fixed mark ups at each fueling station company. Most fueling station companies sell fuel at the same markup across Thailand. There is no temporal variation in these markup costs.

LPG is also subsidized because of its initial intention to provide cheap and clean cooking

[8]Gasoline can be subdivided into different classes based on its octane value and ethanol content, including ULG91, ULG95, gasohol 91, gasohol 95, gasohol E20 and gasohol E85.

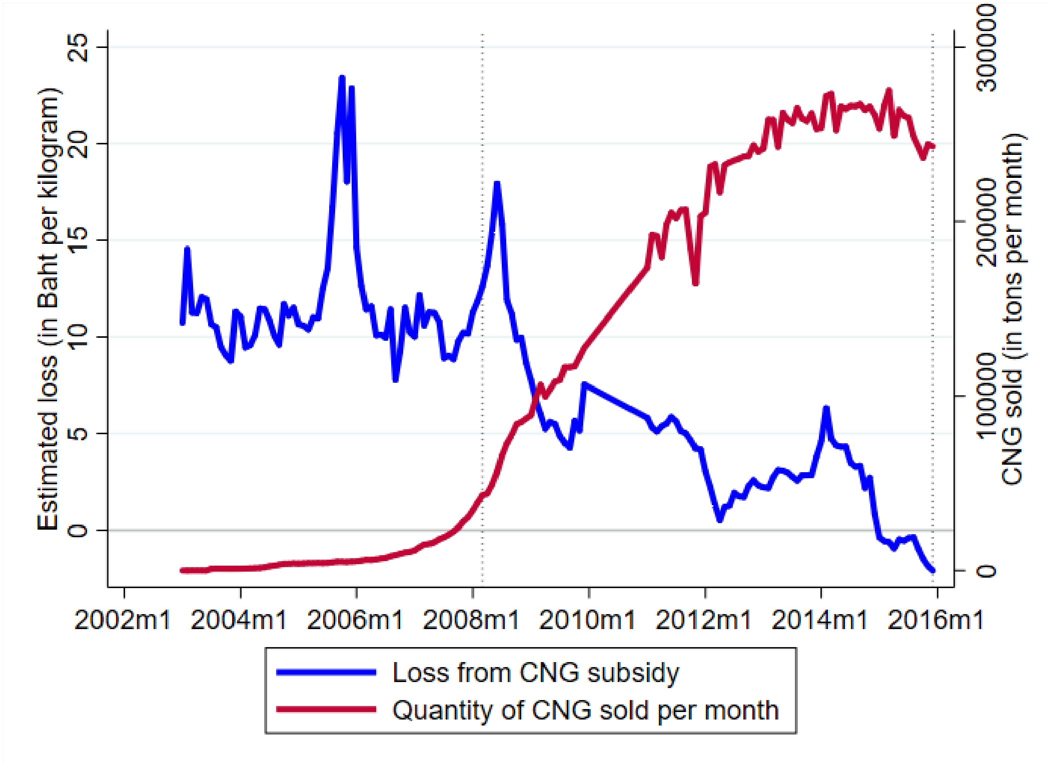

Figure 6: The figure shows estimated loss from CNG price ceiling compared to total quantities of CNG sold. The area between two dotted lines represents time period focused in this paper.

fuels for households. There is no spatial variation between CNG and LPG prices but for different reasons. CNG's market price is higher than the ceiling, so fueling stations must set prices equal to the ceiling. On the other hand, LPG's transport cost is subsidized which makes its price the equal across the country.

Because of the aforementioned reasons and Thailand is a small country globally, I assume that all fuel prices are exogenous. Compared to other fuel sources available for road-vehicles in Thailand, CNG is considered the cheapest. Figure 7 compares fuel prices available for on-road vehicles in Thailand in Baht per gallon with their gasoline equivalents.[9]

[9]CNG is priced in Baht per kilogram while LPG, diesel, and gasoline are priced in Baht per liter. Because the same quantity of fuel has different energy contents, I have converted it to the same energy content equivalent to a gallon of gasoline. Our conversion factors come from http://www.midwestenergysolutions.net/cng-

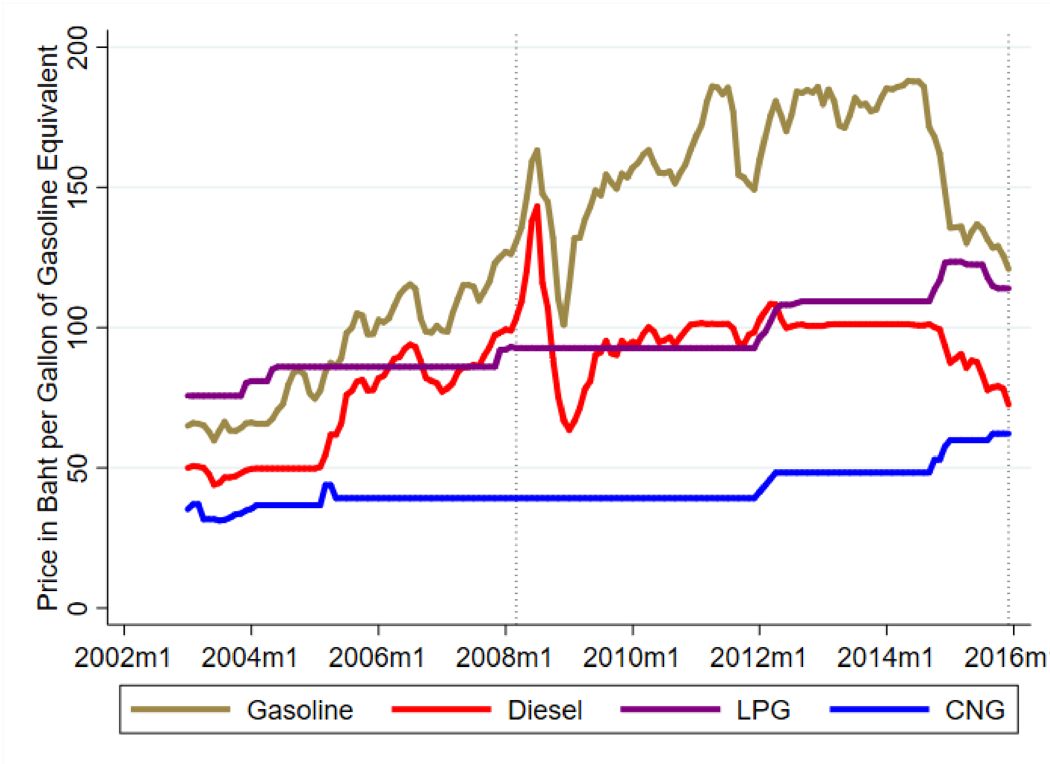

Figure 7: The figure shows retail fuel prices in Baht per gallon of gasoline equivalent. The area between two dotted lines represents time period focused in this paper.

3 Literature Review

Natural gas adoption in road vehicles is not a new policy. Outside Thailand, cities like New Delhi, Dhaka, and Mexico City have encouraged all taxis and buses to run on natural gas. Papers that have evaluated this policy can be divided into two streams based on their approaches, categorized as either bottom-up or top-down estimation. Atmospheric science papers rely on a bottom-up estimation. Bottom up approaches usually start with estimating the emission factors (quantities of emissions per unit of fuel used) for different types of vehicles with different types of fuel. Ideally, emission factors should be measured from the tailpipe of a vehicle. The Argonne National Laboratory's GREET model is one of the most

resources/energy-volume-weight.

comprehensive databases of tailpipe emissions. If we assume that the annual emissions of vehicles in Thailand release the same quantity of pollutants as vehicles in the United States, then the average emission factors, weighted by the age of vehicles in Thailand, can be shown in Figure 2. Bottom-up models then make assumptions about where vehicles travel to. Atmospheric models are used to simulate changes in ambient concentrations from vehicular sources.

There are several criticisms at the bottom-up approach. "The use of average emission factors is a significant simplification of reality. Emissions factors are highly sensitive not only to the engine technology but also to influences as diverse as driver behavior, vehicle loading, fuel quality, local topography, climatic conditions, and traffic conditions" (Reynolds and Kandlikar 2008). A significant amount of emissions from on-road vehicles does not come from tail-pipes alone. Particulate matter can come from abrasion, and volatile organic compounds can come from leakage at fueling stations.

The top-down approach uses existing data to analyze what has happened. Some papers have compared ambient concentrations before and after a natural gas adoption policy (Ravindra et al. 2006; Suthawaree et al. 2011). Specific contents of natural gas, such as polycyclic aromatic hydrocarbons (PAHs) (Ravindra et al. 2006), or specific characteristics of on-road emissions, such as CO to NOx ratios, or SO2 to NOx ratios, were used as indicators to ensure that natural gas adoption has attributed to changes in ambient concentrations (Goyal 2003; Suthawaree et al. 2011). Other papers have used econometric models to evaluate the impacts of the policy on air pollution (Kathuria 2004; Narain and Krupnick 2007; Foster and Kumar 2011). In general, the existing literature have found that CNG adoption reduces particulate matter, SO2, and NOx concentrations in cities.

However, previous papers do not consider behavioral changes. To fully capture the effects of a natural gas adoption policy, a number of factors have to be considered. Because the price of natural gas is less than gasoline, this policy could trigger a rebound in which natural gas fuel users would drive more miles, buy more cars (Wadud 2014), or buy less fuel-efficient cars resulting in a total increase in air pollution. If there is a decrease in air pollution during the same period that natural gas is introduced, it may be because of a decrease in fuel consumption due to an increase in oil price, improvement in vehicle technology, or an introduction of other types of cleaner fuel at the same time. The spatial distribution of the policy also has to be taken into account. Using remote sensing measures of air pollution, Foster and Kumar (2007) have found that natural gas regulations in New Delhi have increased air pollution outside the regulated areas. Because there is only one air pollution monitor in New Delhi and Dhaka, other studies have ignored this effect.

To understand why households adopt CNG and how such policies will change air pollution concentrations, I can rely on existing methodologies from research about consumer choices of alternative fuel. Initial studies of fuel choices in vehicles used discrete choice models and data from household surveys. It was found that the primary influencers of alternative fuel adoption include car prices, fuel prices, and fuel availability (Greene 1989; Greene 1997). Low fuel availability can be compensated for by lowering vehicle prices (Bunch et al. 1993).

The development of ethanol fuel and electric vehicles has resurrected and advanced this field of research concerning adoption of alternative fuel. Actual sales from individual fueling stations are used instead of surveys (Houde 2008; Manuszak and Moul 2009; Anderson 2012). Instead of using fuel prices as independent variables, the gap between the alternative fuel price and gasoline prices is used instead (Anderson 2012; Wadud 2014). The fact that alternative fuel is more available in areas that have a higher demand leads to a bias in the coefficient used for fueling stations. Li et al. (2017) estimated the network effect on the adoption of electric vehicles. Because the number of charging stations are endogenous, an interaction between the total number of charging stations in all other areas and the number of supermarkets in the same area are used as an instrument. More information about each fueling location, such as whether or not the fueling station is on each consumer's commuting path (Houde 2008) or the costs for stopping to refuel ethanol compared to gasoline (Salvo and Huse 2013), is taken into account.

4 Data

This paper is based on two main estimation strategies. First, I estimate how CNG sales are effected by the relative prices of various fuels compared to the price of CNG and CNG sales. The unit of observation of this estimation has been defined at the province-month level. Second, I estimate how the concentration of air pollution is effected by the relative prices of various fuels compared to the price of CNG and its availability. This estimation has been defined at the air pollution monitor-month level. From section 4.1 to 4.5, I describe province-level data. In section 4.6, I describe how I adapt data to the air pollution monitor level. In section 4.8 and 4.7, I describe additional data that are used to estimate air pollution.

4.1 Fueling Station Data

Three fuel specific data sets are available; one for CNG stations, one for LPG stations, and one for gasoline and diesel. Each data set does not cover the entire period of interest from 2008 to 2015. A fourth data set containing information about all retail stations (but not identifying which fuel types are being sold at each fueling station) is used to supplement each data set. I will call this data set the "general retail station data set". This section discusses how each data set is obtained, cleaned, and combined. Once each fueling station is spatially located, I calculate the number of fueling stations for each province-month for my first estimation.

The first data set that contains data on on individual CNG stations came from the Department of the Energy Business (DOEB). Before January 2013, CNG was considered a hazardous material. All CNG fueling stations were required to obtain a permit from the DOEB. This data set shows 476 CNG permits that were given out to fueling stations through August 2012. The data set includes the permit issue date,[10] the permit expiration date, the fueling station address, and whether it is connected to a pipeline. The data set shows that 5 CNG stations have been closed but not when each station was closed. Random checks of stations from the list show that some stations have closed in 2015. To omit inconsistencies resulting from CNG station closings, I only take into account when each gas station opened. Note that this assumption overestimates the number of CNG stations in my data set.

CNG stations that have been built after 2013 have been recorded in the "general retail station" data set. This data set is posted on the DOEB's website and includes updated information regarding the names of existing gas stations, their addresses, and the dates that the fueling station started their businesses. This data set does not specify which fuels are sold at each gas station, but CNG gas stations are identified as the businesses with "natural gas" in their name. I use the August 2015 version of this data set because it was the only version

[10]this represents the data in which each CNG fueling station started its business

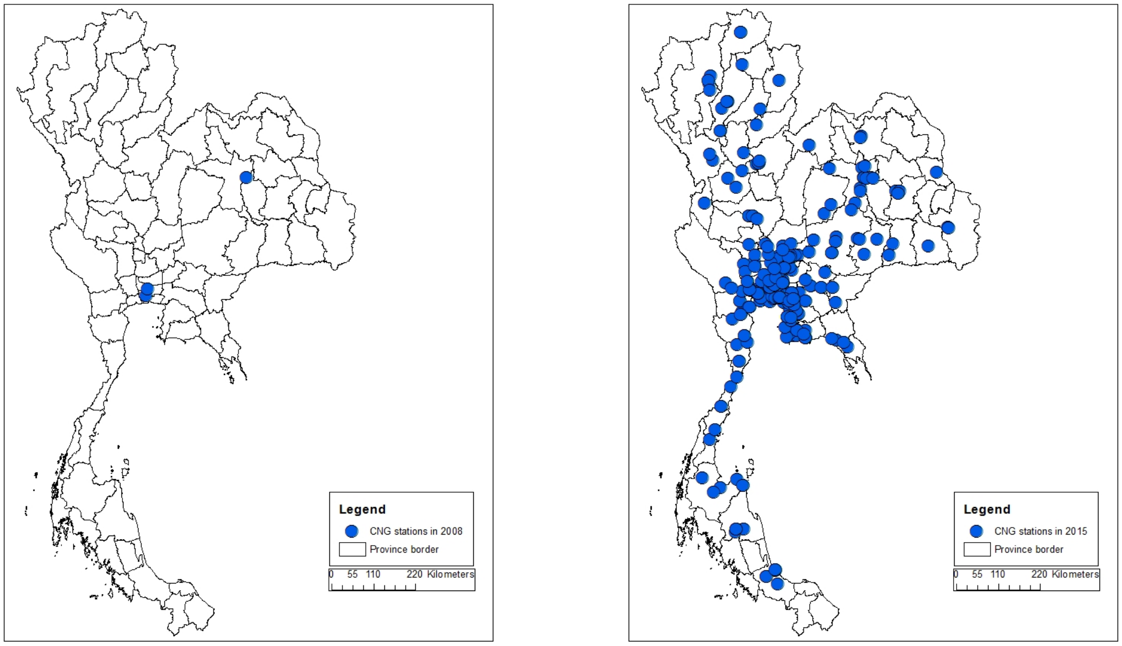

Figure 8: This figure compares CNG stations in 2008 (left) and 2015 (right). Each dot represents a CNG station. CNG station addresses are available from the Department of Energy Business. I independently geocoded each station using Google Maps.

available in the year 2015. I identified 40 new CNG stations that have opened between August 2012 and August 2015.

The LPG fueling station data from the DOEB contains a list of the existing LPG fueling stations and their addresses as of March 2015 but not the date that the businesses were started. No other LPG data sets are available from 2015. I merged business names with names from the "general retail station" data set to find the date that each LPG fueling station started its business. Of the 1,852 LPG fueling stations that existed, I was able to find the dates that the businesses started for 1,320 of them (72 percent of all the LPG stations) . Therefore, I only included the 1,320 LPG stations in this study. Note that this is an underestimation of the actual number of LPG stations because some stations were closed before March 2015, and I was not able to locate some stations geographically.

By using the August 2015 list of gas stations from the DOEB, I assume that all other stations that do not sell CNG or LPG sell gasoline and diesel fuel. Because I ignored the fact that some gas stations have closed, the number of regular gas stations in my estimation has been overestimated.

The DOEB also provides the name, address, and date that each factory converted from other types of fuel to natural gas. However, I do not know when each factory was set up and what types of fuel they used prior to converting to natural gas.

4.2 Fuel Sales

The DOEB also provides the province-month data of fuel sales for all fuel types except CNG. PTT provides the total CNG sales from each province and month. An anomaly is that there are CNG stations in provinces that the DOEB reports to have no stations.

4.3 Fuel Price

The average monthly fuel price data is available from the Energy Policy Planning Office (EPPO). The natural gas price is fixed across the country as described in the policy context section. Prices for other types of fuels are based on the daily fuel price, gas station brand, and district level fixed markup. Therefore, there is very low spatial variability in fuel prices.

Different mixes of gasoline with different blends of ethanol are available over time, at varying prices. I use the price of unleaded gasoline with an octane level of 95 to represent the price of gasoline as it was the only type of gasoline available from 2008 to 2015. Although this is the most expensive and least used type of gasoline, I find that its price strong correlates with other types of gasoline and ethanol.

4.4 Provincial Covariates

Province-specific variables include the annual population, annual average income, and total gross provincial product (GPP) in the manufacturing, utility, and automotive sectors and is available from Thailand's Office of the National Economic and Social Development Board. All prices are adjusted to 2015 price levels based on the consumer price index from the Bank of Thailand.

4.5 CNG License for Engineering Companies

Since 2008, it has been mandated that CNG stations only be built by licensed engineering companies. Licenses are given to companies that have gone through government training about CNG infrastructure and safety precautions. The DOEB provides a list of the engineering companies that have these licenses. The data set includes each company's address and the date each company received its license. From 2008 to 2015, there were a total of 92 companies with this license. Table 2 summarizes all the province-level variables used in my estimation.

Table 2: Province-level summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
CNG sales (in tons)	7,050	85.928	235.328	0.000	2200.650
Distance from licensed engineer to province centroid (in thousand kilometers)	7,050	0.274	0.198	0.000	0.690
Lagged CNG stations in all other provinces	7,050	196.809	222.691	1.000	498.000
CNG stations	7,050	2.730	10.807	0.000	134.000
LPG stations	7,050	12.734	22.308	0.000	234.000
CNG factories	7,050	6.617	26.003	0.000	231.000
Regular stations	7,050	217.551	161.230	13.000	1091.000
CNG price	7,050	10.021	2.082	7.924	14.349
LPG price	7,050	20.277	3.026	16.901	25.839
Diesel price	7,050	29.223	3.882	17.367	40.963
Gasoline price	7,050	42.654	6.813	24.721	53.857
Annual GPP in mining (in thousand Baht)	7,050	5658.849	34462.996	0.000	3.70e+05
Annual GPP in manufacturing (in thousand Baht)	7,050	45587.187	96496.368	198.775	6.21e+05
Annual GPP in construction (in thousand Baht)	7,050	4298.651	9696.004	509.421	1.06e+05
Annual GPP (in thousand Baht)	7,050	1.58e+05	4.37e+05	7358.546	4.72e+06
Annual GPP per capita (in Baht per person)	7,050	1.38e+05	1.39e+05	26675.844	1.10e+06
Population density (in thousand people per sq. km.)	7,050	0.281	0.686	0.016	5.510

Note: All prices are adjusted to 2013 price levels. GPP stands for gross provincial product.

4.6 HYSPLIT

To take into account sources of air pollution that would affect air pollution at each monitor, a Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT; see (Draxler and Hess 1997; Draxler and Hess 1998; Draxler 1999)) is used. I use back trajectory analysis in HYSPLIT to identify origins of air masses that influence the readings of each air pollution monitor. Each HYSPLIT simulation can trace sources of air that were emitted up to 12 hours before being detected by an air pollution monitor. HYSPLIT can trace origins of air masses at different gradient levels. In this paper, I divide my area of interest into 63 120-square-kilometer areas. I use HYSPLIT to quantify the impact of each 120-square-kilometer area on each air pollution monitor. The impact air masses within each 120-square kilometer area is assumed to have a uniform impact on an air pollution monitor within each HYSPLIT simulation.[11] Of all other area sizes I have tested, I find that the 120-square kilometer grid level to best fit my estimation strategy. In this paper, HYSPLIT is primarily used to quantify the impact of each fueling station on each air pollution monitor under the assumption that the fueling station is the source of emission. As cars can drive long distances after it is being fueled, I find that the 120-square-kilometer grid would best fit this analysis.

I run HYSPLIT every day, two times a day during a representative year. I find the average impact that each 120-square-kilometer area has on each air pollution monitor in each month. I then scale its impact such that all non-zero values for each month-monitor sum up to one. I call this the "HYSPLIT weights". The higher the HYSPLIT weights,

[11] I also ran HYSPLIT at other gradient levels; where air masses within a 30-square-kilometer area and 10-square-kilometer area are assumed to have a uniform impact on an air pollution monitor within each simulation. Estimation results from the 30-square-kilometer is presented in the Robustness Check section.

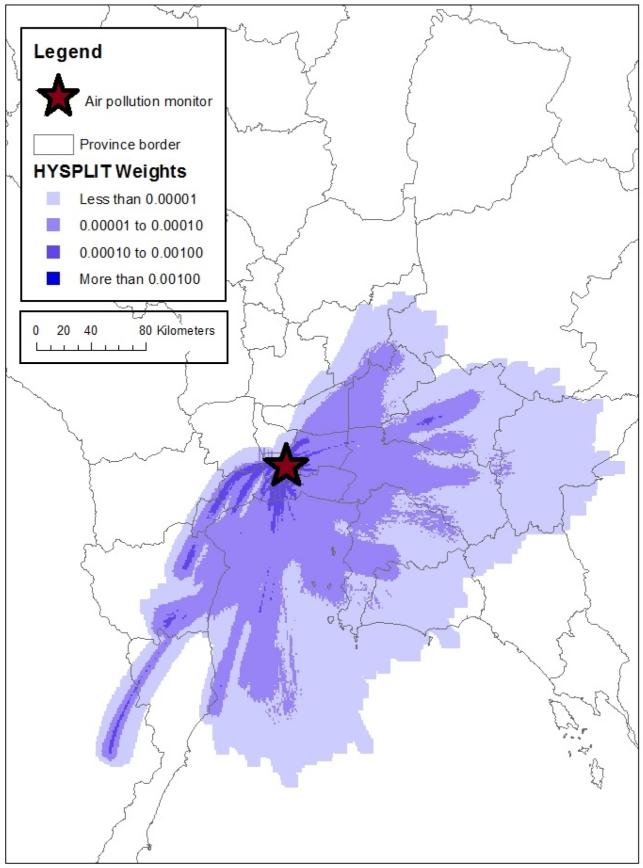

Figure 9: This figure shows shows the average HYSPLIT weights of simulations from a sample pollution monitor in Bangkok.

the higher the impact on each air pollution monitor. Figure 9 shows the average HYSPLIT weights of simulations from a sample pollution monitor in Bangkok.

To estimate the number of fueling stations surrounding each air pollution monitor, I first find the number of stations within each 120-square kilometer grid. Then the weighted average number of stations within each grid is calculated. The "HYSPLIT weights" are used to weigh the impact of each grid on each air pollution monitor. Table 4 summarizes HYSPLIT weighted fueling stations.

Table 3: Summary statistics of HYSPLIT-weighted fueling stations

Variable	Obs	Non-Zero Obs	Mean	Std. Dev.	Min	Max
CNG stations	5,992	4,116	0.384	0.703	0.000	3.227
LPG stations	5,992	5,845	1.101	1.033	0.000	4.172
CNG factories	5,992	3,760	0.816	1.011	0.000	4.159
Regular station	5,992	5,922	2.627	0.839	0.100	4.992

Note: The table summarizes data from all 120 sq. km. grids from March 2008 to December 2015. There are 63 grids in a month.

The distance to the licensed engineer is the distance from the closest licensed engineering firm to each 120-square-kilometer grid. The average distance to a licensed engineer around an air pollution monitor is then calculated by using the "HYSPLIT weights" for each grid.

Lagged stations in all other areas are defined as the number of stations in Thailand that do not affect the readings of air pollution monitors based on the HYSPLIT simulations.

Before finding province data at the monitor-level, I first have to generate data at the grid level. Grid-level data that are derived from province-level data such as fuel prices are weighted averages from province level prices. The fraction of each grid that is made up of each province is used as the weights. Mathematically, grid-level data derived from province-level data is expressed as

$$Price_{gt} = \sum_{p} \frac{area_{p \cap g}}{area_g} Price_{pt}$$

where $Price_p$ is fuel price is province p. $area_{p \cap g}$ is area of province p that is in grid g. $area_g$ is the size of grid g. The average fuel price in grid g is then the weighted average of prices from all provinces that intersect with grid g. An example of this can be found in Appendix B.

I then find the average weighted value of each data at the monitor level by scaling them with "HYSPLIT weights." For example, fuel price of monitor m and time t can be expressed as

$$Price_{mt} = \sum_{g} HYSPLIT_{mgt} Price_{gt}$$

where $HYSPLIT_{mgt}$ is the monthly aggregate impact of grid g on monitor m in time t. It is scaled such that $\sum_g HYSPLIT_{mgt} = 1$. $Price_{gt}$ is price in grid g at time t.

4.7 Weather Covariates

Weather data from 52 weather stations from the year 2008 to 2014 in Thailand have been obtained from the National Climatic Data Center (NCDO). Weather covariates includes monthly mean temperature, monthly mean minimum temperature, monthly mean maximum temperature, monthly extreme minimum temperature, monthly extreme maximum temperature, monthly total precipitation, and monthly extreme maximum daily precipitation. In addition, I have the monthly mean temperature interact with the monthly total precipitation.

Weather variables are also first calculated at the grid level. For each grid, I find the closest weather station to the grid centroid and use weather data from that weather station. If there is missing data in a specific month or year, data from the closest weather station is used. Table 4 summarizes HYSPLIT weighted weather and provincial covariates used in my second estimation.

4.8 Air Pollution Monitor

The Pollution Control Department releases the monthly average concentrations from the air pollution monitors. In 2008, there were 47 CO monitors, 46 PM10 monitors, 44 SO2

Table 4: Summary statistics for all other covariates at the 120 sq. km. grid level

Variable	Obs	Mean	Std. Dev.	Min	Max
Distance from licensed engineer to grid centroid (in thousand km.)	5,922	1.915	1.816	0.024	5.838
Lagged CNG stations in all other provinces	5,922	4.178	2.107	0.881	6.904
CNG price	5,922	8.784	2.514	5.311	15.440
LPG price	5,922	17.710	4.079	11.329	28.294
Diesel price	5,922	25.266	4.532	12.337	37.479
Gasoline price	5,922	37.049	8.126	16.651	58.313
Annual GPP in mining (in thousand Baht)	5,922	17155.033	49883.328	0.000	4.12e+05
Annual GPP in manufacturing (in thousand Baht)	5,922	1.14e+05	1.07e+05	208.487	5.34e+05
Annual GPP in construction (in thousand Baht)	5,922	9558.029	9070.732	612.036	90547.820
Annual GPP in utilites (in thousand Baht)	5,922	11959.252	11997.183	76.766	70253.789
Annual GPP in automotive (in thousand Baht)	5,922	70239.626	93425.365	508.159	8.72e+05
Annual GPP in transport (in thousand Baht)	5,922	38778.346	52579.781	307.205	4.44e+05
Annual GPP (in thousand Baht)	5,922	4.17e+05	4.41e+05	6733.833	4.01e+06
Annual GPP per capita (in Baht per person)	5,922	2.12e+05	1.72e+05	30805.355	1.23e+06
Population density (in thousand people per sq. km.)	5,922	0.708	0.840	0.016	5.109
Dummy variable indicating whether a monitor has been moved	5,922	0.048	0.213	0.000	1.000
Monthly mean temperature (Celcius)	5,922	28.460	1.972	19.170	33.362
Monthly mean minimum temperature (Celcius)	5,922	23.495	2.691	12.118	27.600
Monthly mean maximum temperature (Celcius)	5,922	33.274	1.761	24.980	40.586
Extreme minimum temperature (Celcius)	5,922	21.085	3.425	5.888	26.138
Extreme maximum temperature (Celcius)	5,922	35.818	1.820	28.372	42.524
Total precipitation (millimeters)	5,922	112.677	100.613	0.000	1072.920
Extreme maximum daily precipitation (millimeters)	5,922	34.321	26.622	0.000	240.789

Notes: The table summarizes data from all 120 sq. km. grids from March 2008 to December 2015. There are 63 grids in a month. All prices are adjusted to the 2013 price level. GPP stands for gross provincial product.

monitors, and 45 NO2 monitors across Thailand. By 2015, the total number of air pollution monitors increased to 63, with each monitor measuring multiple air pollution concentrations. Figure 10 shows the spatial distribution of all air pollution monitors in 2015.

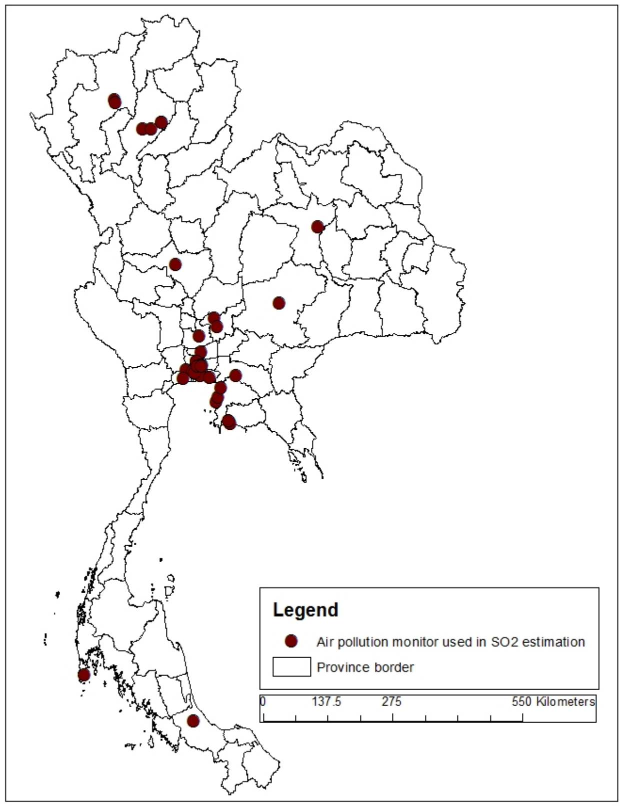

Figure 10: This figure shows spatial distribution of air pollution monitors that measures SO2 from 2008 to 2015. Addresses of air pollution monitors are available from the Pollution Control Department. I independently geocoded each air pollution monitor using Google Maps.

Older monitors have higher pollution concentrations than the newer monitors as they are placed in areas with the highest concern for air pollution. By including data from newer monitors, there may be a downward bias in my estimates as concentrations of new air

Table 5: Summary statistics for air quality

Variable	Obs	Mean	Std. Dev.	Min	Max
1 hr. average CO (ppm)	3,855	0.654	0.372	0.000	3.810
24 hr. average PM10(ug/m3)	4,059	42.326	23.346	5.000	184.000
1 hr. average SO2 (ppb)	3,601	2.771	2.065	0.000	20.600
1 hr. average NO2 (ppb)	3,757	15.986	9.843	0.000	68.00

Note: Average monthly air pollution concentration from the Pollution Control Department from March 2008 to December 2015. The table only include observations used in regressions in Table 11 and 9.

pollution monitors may be positively correlated with the set up of new CNG gas stations. In my preferred specifications, data from air pollution monitors that were built after January 2008 have been dropped.

Besides observations about the old and new monitors, there are additional missing observations. In some months, data is not released as some monitors may be broken. Many observations are missing in early 2012 as the central region of the country was struck by a floods. Table 5 summarizes air pollution concentrations measured by all monitors and compares its level to the national air quality standard.

5 Empirical Methodology

An ideal experiment would be one that looks at air quality in four areas: one with a subsidized price and a fueling station , one with a subsidized price and no fueling stations, one with no subsidy and a fueling station, and one without both. However, such social experiments do not exist.

Because the retail price of CNG is uniform, there is no spatial variation that I can exploit to explain the differences in air quality at different monitors. Instead, I exploit the fact that there are costs associated with the search for fuel (Houde 2008, Manuszak and Moul 2009). I have the fuel price and various proxies that represent the costs associated with the search for fuel. I interact fuel price with fuel availability and use this as the variable of focus in my reduced form estimation.

My estimation is divided into two parts. First, I estimate the effect of CNG prices and CNG availability on CNG, gasoline, and diesel sales. Second, I estimate the effect of CNG prices and CNG availability on air pollution including CO , SO2, NO2, and PM10.

5.1 Fuel Availability and Price on CNG Sales

The equation below shows my preferred specification for when CNG sales act as the dependent variable. Each observation is at the province-month level. Q_{pt} stands for the daily retail sales of CNG, averaged at the month level. X_{mt} are annual provincial covariates. Because there is very little temporal variation in CNG and LPG prices, consumers must make fuel choice decisions based on the relative price of diesel and gasoline fuel. To add spatial variation to fuel prices, I have the fueling stations interact with the relative fuel prices. Because CNG is the main substitute for gasoline, I have the CNG stations interact with the relative price of gasoline compared to that of CNG. I include a linear time trend and region-year fixed effects to control for national and regional changes in preferences for CNG. I also include province fixed effects to control for time-invariant factors that vary across provinces but do not vary over time across provinces. Standard errors are clustered at the province level. There are 75 clusters.

$$
\begin{aligned}
Q_{pt} = {} & \alpha_0 + \alpha_1 CNGStations_{pt} + \alpha_2 CNGStations_{pt} \times \frac{GasolinePrice_t}{CNGPrice_t} \\
& + \alpha_3 LPGStations_{pt} + \alpha_4 RegularStations_{pt} + \alpha_5 \frac{GasolinePrice_t}{CNGPrice_t} \\
& + \alpha_6 \frac{DieselPrice_t}{CNGPrice_t} + \alpha_7 \frac{GasolinePrice_t}{LPGPrice_t} \\
& + \alpha_8 RegularStations_{pt} \times \frac{GasolinePrice_t}{CNGPrice_t} \\
& + \alpha_9 RegularStations_{pt} \times \frac{DieselPrice_t}{CNGPrice_t} \\
& + \alpha_{10} RegularStations_{pt} \times \frac{GasolinePrice_t}{LPGPrice_t} \\
& + \alpha_{11} LPGStations_{pt} \times \frac{GasolinePrice_t}{LPGPrice_t} \\
& + X'_{pt}\gamma + t + \eta_{region,year} + \xi_p + \epsilon_{pt}
\end{aligned}
$$

There are potentially unobserved factors that could affect CNG demand. The fact that PTT Plc. announced that they will distribute CNG to fueling stations that are near natural gas pipelines and are near major highways could imply that there are some unobservable factors. First, natural gas pipelines are initially built to supply natural gas as an energy source for factories. Industrial areas are more likely to use long-distance transport vehicles to supply their inputs and distribute their outputs, leading to a larger demand for CNG. To control for this, I add the annual provincial output from the mining, manufacturing, and construction sector as control variables.[12]

Provinces with long distances of major roads and highways are also likely to demand more CNG. Because there was no large construction of major roads and highways during

[12] I have tried including annual provincial output from other sectors also, but it does not affect our results. Additional results can be provided upon request.

our period of interest, I include province fixed effects in the regressions.[13]

As CNG fuel is more likely to be used by lower-income households, I also include annual provincial income per capita as an additional control. One may claim that households that live in a province may drive to another province for work, which will cause them to refuel CNG in a different province than where they live in. Because I have already included annual gross provincial product from each province, this unobservable is accounted for.

I also include region-year fixed effects to control for national and regional changes in preferences for CNG. Besides relative fuel prices and availability, all other controls do not vary by month, I include a linear time trend into our regressions.

Most importantly, I know that the number of CNG stations is not accurately reported. There are CNG stations that have not registered with the government. CNG stations that have reported to the government, but went out of business, are also not included in the data. From an interview, I know that the number of unreported CNG stations had reduced due to the enforcement of stricter regulations. From 2008 to 2015, the measurement error has been reduced.

To predict the direction of the bias, the following model assumes that the measurement error and unobserved demand are lumped into one variable, called the $OmittedVariable$.

$$Q_{pt} = \alpha_0 + \alpha_1 CNGStations_{pt} + ... + \theta(OmittedVariable_{pt}) + \epsilon_{pt} \tag{1}$$

If the $OmittedVariable_{pt}$ is uncorrelated with other variables once controlled for $CNGStations_{pt}$,

[13] I have tried including distances of major roads and highways within each province from a single year, 2008, into the regressions. The results are similar to adding province fixed effects. Because there may be other province-specific factors that affect province CNG demand, I opt for province fixed effects instead of the distances of major roads and highways in a province. Additional results can be provided upon request.

then I can write the linear projection of the $OmittedVariable_{pt}$ onto $CNGStations_{pt}$,

$$OmittedVariable_{pt} = \delta_0 + \delta_1 CNGStations_{pt} + \rho_{pt} \tag{2}$$

where ρ_{pt} is uncorrelated with other variables. When I substitute the second equation 2 into 1, the following equation results:

$$Q_{pt} = \alpha_0 + (\alpha_1 + \theta\delta_1)CNGStations_{pt} + ... + \theta\delta_0 + \theta\rho_{pt} + \epsilon_{pt}$$

Let the error $\theta\delta_0 + \theta\rho_{pt} + \epsilon_{pt}$ have a zero mean and be uncorrelated with each regressor (Woolridge 2002). Because more CNG stations will induce more sales, $\alpha_1 > 0$ is the equation used. Because the $OmittedVariable_{pt}$ includes stations that are not reported and have unobserved demand, $\theta > 0$ holds true. If the measurement error decreases as the number of CNG stations increases over time, $\delta_1 < 0$ holds true. Let $\hat{\alpha_1}$ be a coefficient for $CNGStation$ from an OLS estimation.

$$\hat{\alpha_1} = \alpha_1 + \theta\delta_1 < \alpha_1$$

Therefore, the true effect of CNG stations on CNG sales should be greater than the effect obtained from an OLS regression. To obtain the true effect, I employ a two-stage-least-squares (2SLS) estimation strategy. A CNG station has to be built by a licensed construction company that has passed a government training about CNG infrastructure.[14] I use distance from each province centroid to the nearest licensed engineering company as part of the instrumental variable for the number of CNG stations in each province.

[14] It may be better to use the number of licensed engineering companies in their province, but because licensed engineering company are clustered in some provinces only, the large number of zero elements made the instrumental variable a weak instrument.

In addition, there is not much temporal variation in the distances to the nearest licensed construction company from each province, I interact the term with lagged CNG stations in all other provinces. Lagged CNG stations in all other provinces not only add temporal variation to my instrument, but they also capture national trends to investments in CNG stations. In conclusion, I use the interaction between the distance from the closest construction company to each population's centroid and lagged CNG stations in all other provinces to serve as the instrument for CNG stations. This should address endogeneity bias and measurement bias in my estimation.

To judge the validity of the exclusion restriction, the relationship between distance to the nearest licensed engineering company and potentially endogenous variables are considered. First construction companies that have applied for the government license to build CNG stations may be in areas with high demand for construction. I have controlled for this potential correlation by included annual gross provincial product from the construction sector in my regressions.

Second, engineering companies must have applied for the government license because of high CNG demand in their areas. Therefore, there may be unobserved demand factors that are correlated with the number of CNG stations near licensed engineering companies. Because I have included annual gross provincial product from each sector, average provincial income, provincial population density, province fixed-effects, region-year fixed effects, and a linear time trend as control variables, the error term ϵ_{pt} should be uncorrelated with the instrument.

Third, there may be some reason that some firms in the same province applied for a CNG construction license while other firms choose not to. If this unobserved factor is correlated with the instrument, this would invalidate the exclusion restriction of the instrumental variable strategy. To control for this possibility, we use an identification strategy proposed by Nevo and Rosen (2012) where the exclusion restriction is relaxed. The Nevo and Rosen (2012) strategy will give a range and confidential interval of the endogenous coefficient. However, this is strategy relies on two assumptions. First, the correlation between the instrument and the error term must have the same sign as the correlation between the endogenous variable and the error term. If we assume that the number of unreported CNG stations makes up most of the error term, then the error term is negatively correlated with the number of CNG stations[15] is negatively correlated with distance to the nearest licensed engineering company.[16]. Second, the instrument is less correlated with the error term than the endogenous variable. This is obvious since we should expect higher correlated with between the number of reported and unreported stations than other factors.

5.2 Fuel Availability and Price on Pollution

In this section, I describe how I estimate the effect of relative fuel prices on CNG and CNG availability on air pollution. The equation below shows my preferred specification for when monitor-level air pollution concentrations act as the dependent variable. AQ_{mt} is the air pollution concentration of monitor m at month-year t. X_{mt} are provincial covariates, weather covariates, and a dummy that indicates whether a monitor has been moved. The results are clustered at the district level. There are 39 clusters.

[15]When there were more CNG stations later in our period of interest, there was more regulation and the number of under-reported CNG stations is reduced. and the error term

[16]The more the number of unreported stations, the closer the licensed engineering company

$$
\begin{aligned}
AQM_{mt} = \beta_0 &+ \alpha_1 invsinh(CNGStations_{mt}) + \beta_2 invsinh(CNGStations_{mt}) \times \frac{GasolinePrice_t}{CNGPrice_t} \\
&+ \beta_3 invsinh(LPGStations_{mt}) + \beta_4 invsinh(RegularStations_{mt}) + \beta_5 \frac{GasolinePrice_t}{CNGPrice_t} \\
&+ \beta_6 \frac{DieselPrice_t}{CNGPrice_t} + \beta_7 \frac{GasolinePrice_t}{LPGPrice_t} \\
&+ \beta_8 invsinh(RegularStations_{mt}) \times \frac{GasolinePrice_t}{CNGPrice_t} \\
&+ \beta_9 invsinh(RegularStations_{mt}) \times \frac{DieselPrice_t}{CNGPrice_t} \\
&+ \beta_{10} invsinh(RegularStations_{mt}) \times \frac{GasolinePrice_t}{LPGPrice_t} \\
&+ \beta_{11} invsinh(LPGStations_{mt}) \times \frac{GasolinePrice_t}{LPGPrice_t} \\
&+ X'_{mt}\gamma + t + \tau_{year} + \eta_{monitor,month} + \zeta_{region} + \varepsilon_{mt}
\end{aligned}
$$

Instead of using the number of stations, I use the inverse hyperbolic sine of the number of stations for ease of interpretation. This is similar to taking a logarithm of a variable, but an inverse hyperbolic sine allows for elements with zero value to be included. A ten percent increase in CNG stations is the same as a ten percent increase in the HYSPLIT-weighted number of stations surrounding an air pollution monitor. $\beta_1/100$ is interpreted as the unit change in air pollution concentrations in response to a one percent change in CNG stations.

There are potentially unobserved factors that could affect air pollution. The fact that PTT Plc. announced that they will distribute CNG to fueling stations that are near natural gas pipelines and are near major highways could imply that there are some unobservable factors. First, firms that have converted their energy source to natural gas would also be the cause of an improvement in air pollution. To control for this, I add the annual provincial output from the mining, manufacturing, and construction sector as control variables. Areas with long distances of major roads and highways are more likely to have high levels of air pollution. Because there was no large construction of major roads and highways during our period of interest, I include region fixed effects in the regressions.

Furthermore, I include monitor-month fixed effects to control for seasonal factors. To control for common contemporaneous shocks, I include year fixed effects. I include region fixed effects to control for base-level ambient pollution concentrations in each region. Additionally, I included a linear time trend to control for unobservable factors that might improve air quality over time such as improvements in energy efficiency.

Most importantly, I know that the number of CNG stations is not accurately reported. Additional details about the direction of bias can be found in the Appendix C. To obtain the true effect of CNG stations, I employ a 2SLS estimation strategy. I use an instrument from the same data used in the previous estimation, the weighted average distance from each construction company to a 120-square-kilometer grid and its interaction with lagged CNG stations in all areas that do not affect that pollution monitor.

To judge the validity of the exclusion restriction, the relationship between distance to the nearest licensed engineering company to each grid centroid and the error term ϵ_{pt} is considered. All unobservable factors at the province level are addressed in the previous section. Again, there may be some reason that some firms in the same province applied for a CNG construction license while other firms choose not to. If this unobserved factor is correlated with the instrument, this would invalidate the exclusion restriction of the instrumental variable strategy. Again, To control for this possibility, we use an identification strategy proposed by Nevo and Rosen (2012) in which a range of possible range of coefficient is reported when the exclusion restriction is relaxed.

6 Results

6.1 The Effects of Price and Availability on CNG Sales

Because our instrument maybe imperfect, results from 2SLS regressions may not be superior than OLS regressions with a suite of fixed effects. However, the 2SLS regressions may have adjusted for the measurement error. Table 6 compares OLS to 2SLS results. First stage results are shown in Appendix D.

When the exclusion restriction is relaxed, the identification strategy proposed by Nevo and Rosen (2012) is tested. Because the strategy can be tested when only one variable is endogenous, the 2SLS specification in column (2) in Table 6 is tested. Using the strategy proposed by Nevo and Rosen (2012), I find that the bounds for the coefficient of CNG stations is between 5.090 and 8.524 and the confidence interval is between 2.419 and 12.408. The coefficient of CNG station from column (2) in Table 6 is 8.734 which is out of the range, but inside the confidence interval found when the exclusion restriction is relaxed.

Figure 11 shows the net effect of the fuel prices of CNG sales when other variables are fixed at the mean from the OLS and 2SLS regressions. The coefficients are the linear combinations of coefficients of different fuel prices in column (3) and (4) in Table 6. The number beside each coefficient represents a change in CNG sales (in tons) per change in fuel prices (in Baht per kilogram or liter). The numbers in the parentheses are percent changes in CNG sales due to a one percent change in price. I find that an increase in CNG prices decreases CNG sales. OLS results show that an increase in price by one percent can decrease CNG sales by almost one percent. However, the coefficient is not statistically different from zero in the 2SLS regression. This is because I do not have enough temporal variation in

Table 6: Second stage results for CNG demand

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
CNG stations	4.411***	8.734***	6.905***	0.788
	(1.388)	(2.380)	(1.840)	(5.375)
CNG stations × Gasoline price / CNG price			-0.547*	1.800*
			(0.286)	(1.045)
Regular stations × Diesel price / CNG price			-0.013	-0.006
			(0.018)	(0.013)
Regular stations × Gasoline price / CNG price			-0.007	-0.042
			(0.053)	(0.046)
Gasoline price / CNG price	20.379***	12.206**	23.632	9.079
	(7.049)	(5.766)	(15.344)	(12.706)
Diesel price / CNG price	-3.686	-0.052	1.441	4.186
	(2.741)	(2.506)	(4.283)	(4.167)
Gasoline price / LPG price	-37.393***	-30.223**	-59.053	-32.280
	(12.639)	(11.674)	(40.130)	(36.053)
Regular stations × Gasoline price / LPG price			-0.012	0.097
			(0.138)	(0.140)
LPG stations × Gasoline price / LPG price			1.707**	-0.138
			(0.681)	(1.323)
Observations	7050	7050	7050	7050
Adjusted R^2	0.983	0.980	0.983	0.980
Cragg-Donald F-stat		88.686		15.637

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, province fixed effects, region-year fixed effects, and a linear time trend. Standard errors clustered at the province level are in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

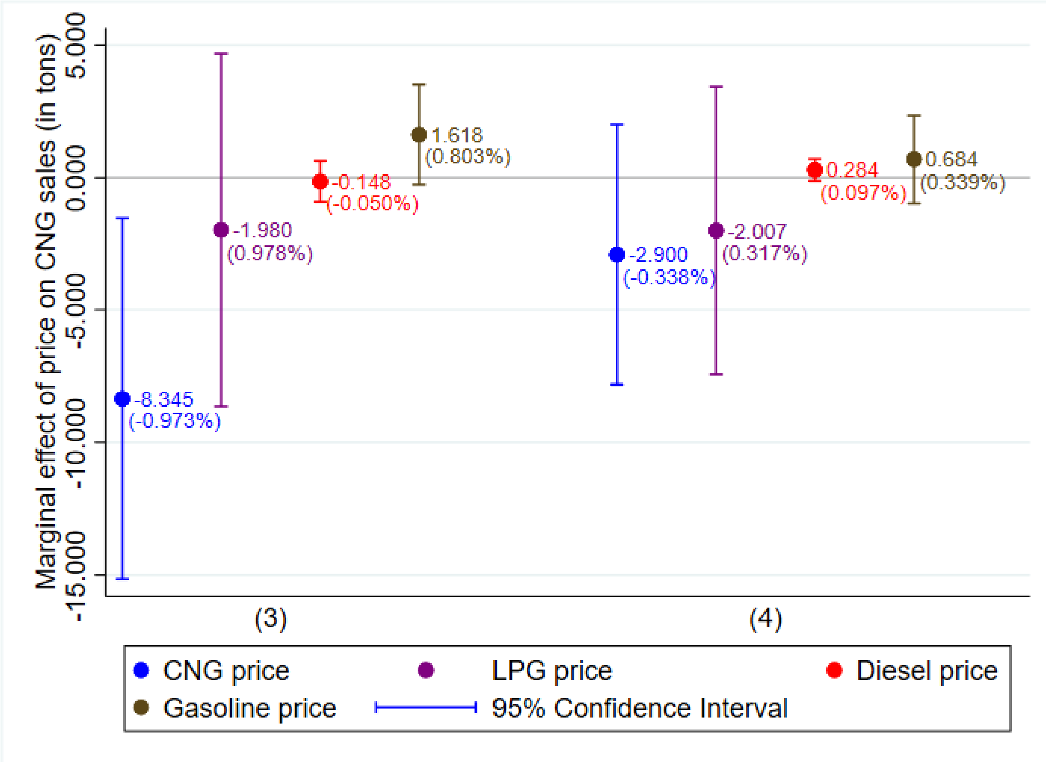

Figure 11: The figure shows the net effect of fuel prices on CNG sales. The numbers in the parentheses are percent changes in CNG sales due to a one percent change in price.

price to extract the direct effect of CNG prices.

Figure 12 shows the net effect of stations on CNG sales from OLS and 2SLS regressions. The coefficients are the linear combinations of coefficients of different fueling stations in column (3) and (4) in Table 6. The number beside each coefficient represents a change in CNG sales (in tons) per change in the number of fueling stations. The numbers in the parentheses are percent changes in CNG sales due to a one percent change in stations. An increase in CNG stations significantly increases CNG sales. OLS regression shows that an addition CNG station in a province can increase CNG sales by 4 tons a day. The coefficient from the 2SLS regression is larger than the coefficient from the OLS regression. This implies that the omitted variable is positively correlated with CNG sales, but the magnitude of

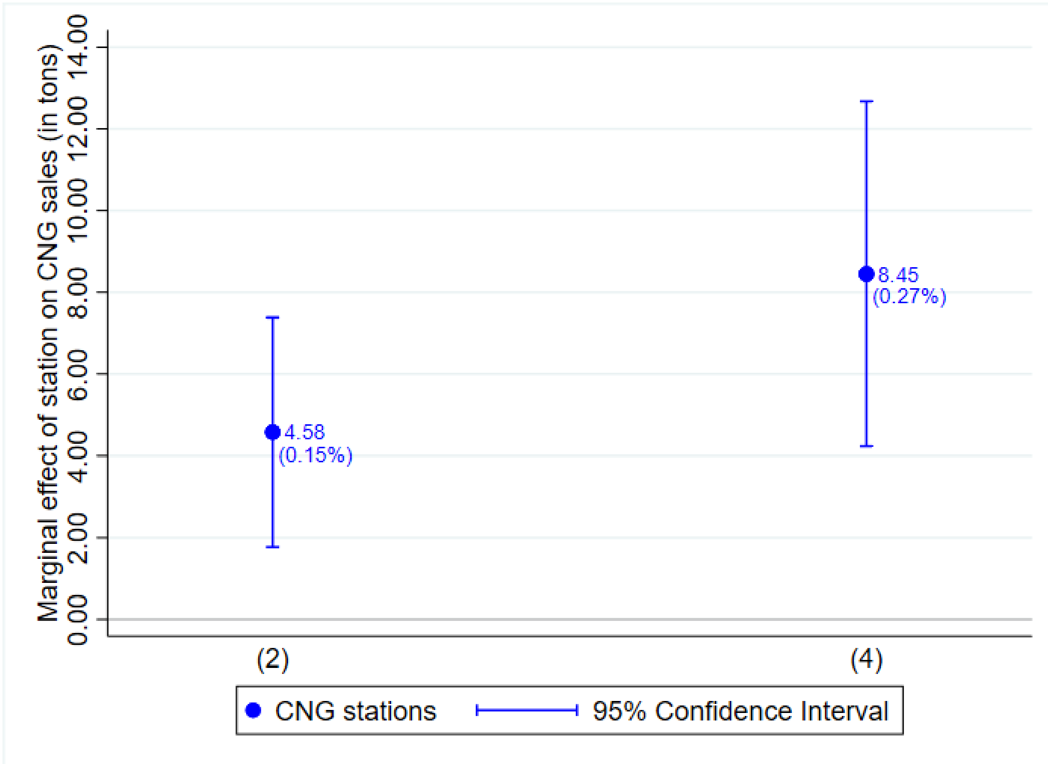

Figure 12: The figure shows the net effect of fueling stations on CNG sales. The numbers in the parentheses are percent changes in CNG sales due to a one percent change in stations.

the omitted variable decreases over time. The hypothesis concerning the direction of the measurement bias is confirmed.

6.2 The Effects of Price and Availability on Pollution

OLS and 2SLS results when the SO2 concentration is a dependent variable, are reported in Table 7. Table 8 show results from the preferred OLS regression specification (column (3) of Table 7) for all pollutants. Table 9 shows the regression results from my preferred 2SLS specification (column (4) from Table 7) for all pollutants.

When the exclusion restriction is relaxed, the identification strategy proposed by Nevo and Rosen (2012) is tested. Because the strategy can be tested when only one variable is endogenous, the 2SLS specification in column (2) in Table 7 is tested. Using the strategy proposed by Nevo and Rosen (2012), I find that the upper bound for the coefficient of CNG stations is is -1.081 which exactly matches our coefficient from the 2SLS regression from column (2) in Table 7. This shows that our results are valid even when the exclusion restriction assumption is relaxed.

Table 7: Regression results for SO2

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
invsinh(CNG stations)	0.065	-1.081	-1.079*	-6.272**
	(0.234)	(0.645)	(0.626)	(2.618)
invsinh(CNG stations) × Gasoline price / CNG price			0.272*	1.167*
			(0.139)	(0.578)
invsinh(Regular stations) × Diesel price / CNG price			-0.066	0.178
			(0.216)	(0.279)
invsinh(Regular stations) × Gasoline price / CNG price			0.636	-0.732
			(0.625)	(0.740)
invsinh(Regular stations) × Gasoline price / LPG price			-1.326	1.953
			(1.640)	(2.005)
invsinh(LPG stations) × Gasoline price / LPG price			-0.282	-2.008*
			(0.373)	(1.015)
Gasoline price / CNG price	-0.783	-0.375	-3.033	0.361
	(0.628)	(0.658)	(1.932)	(2.153)
Diesel price / CNG price	0.005	-0.103	0.222	-0.341
	(0.119)	(0.143)	(0.500)	(0.583)
Gasoline price / LPG price	2.179	1.565	6.926	0.021
	(1.372)	(1.375)	(4.852)	(5.349)
Observations	3601	3601	3601	3601
Adjusted R^2	0.620	0.602	0.622	0.592
Cragg-Donald F-stat		384.598		155.960

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, temperature and precipitation, linear time trend, monitor-month fixed effects, region fixed effects, and year fixed effects.

Standard errors clustered at the district level are in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: OLS regression results for all pollutants

	(1) CO (ppm)	(2) PM10 (ug/m3)	(3) SO2 (ppb)	(4) NO2 (ppb)
invsinh(CNG stations)	0.103	4.045	-1.515**	2.812
	(0.101)	(5.369)	(0.605)	(1.785)
invsinh(CNG stations) × Gasoline price / CNG price	-0.025	-0.247	0.370***	-0.421
	(0.021)	(1.158)	(0.127)	(0.359)
invsinh(Regular stations) × Diesel price / CNG price	-0.010	-0.198	0.110	-0.535
	(0.019)	(1.266)	(0.194)	(0.384)
invsinh(Regular stations) × Gasoline price / CNG price	0.077	3.276	0.836	2.767*
	(0.091)	(3.942)	(0.601)	(1.423)
invsinh(Regular stations) × Gasoline price / LPG price	-0.147	-9.871	-2.074	-5.117
	(0.221)	(9.974)	(1.554)	(3.406)
invsinh(LPG stations) × Gasoline price / LPG price	-0.020	3.553	-0.473	0.987
	(0.050)	(2.602)	(0.375)	(1.188)
Gasoline price / CNG price	-0.288	-14.264	-3.600*	-9.879**
	(0.261)	(12.011)	(1.875)	(4.601)
Diesel price / CNG price	0.035	-1.131	-0.204	1.390
	(0.043)	(3.209)	(0.458)	(0.873)
Gasoline price / LPG price	0.689	47.288	9.063*	20.083*
	(0.590)	(28.127)	(4.648)	(10.357)
Observations	3855	4059	3601	3757
Adjusted R^2	0.633	0.747	0.616	0.826

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, temperature and precipitation, linear time trend, monitor-month fixed effects, region fixed effects, and year fixed effects. Standard errors clustered at the district level are in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 9: 2SLS regression results for all pollutants

	(1) CO (ppm)	(2) PM10 (ug/m3)	(3) SO2 (ppb)	(4) NO2 (ppb)
invsinh(CNG stations)	-0.478**	-13.199	-6.272**	-7.306
	(0.234)	(14.524)	(2.618)	(5.123)
invsinh(CNG stations) × Gasoline price / CNG price	0.101*	1.960	1.167*	1.782
	(0.058)	(3.192)	(0.578)	(1.304)
invsinh(Regular stations) × Diesel price / CNG price	0.032	0.497	0.178	0.040
	(0.025)	(1.388)	(0.279)	(0.436)
invsinh(Regular stations) × Gasoline price / CNG price	-0.071	-5.984	-0.732	0.063
	(0.126)	(5.284)	(0.740)	(1.960)
invsinh(Regular stations) × Gasoline price / LPG price	0.200	11.762	1.953	1.763
	(0.304)	(13.825)	(2.005)	(4.935)
invsinh(LPG stations) × Gasoline price / LPG price	-0.224**	-1.030	-2.008*	-2.596
	(0.083)	(5.341)	(1.015)	(2.190)
Gasoline price / CNG price	-0.009	12.793	0.361	-4.773
	(0.382)	(15.182)	(2.153)	(6.506)
Diesel price / CNG price	-0.045	-3.143	-0.341	0.400
	(0.050)	(3.026)	(0.583)	(0.715)
Gasoline price / LPG price	0.117	-10.400	0.021	8.019
	(0.843)	(36.014)	(5.349)	(14.273)
Observations	3855	4059	3601	3757
Adjusted R^2	0.622	0.698	0.541	0.786
Cragg-Donald F-stat	190.766	142.435	155.960	137.065

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, temperature and precipitation, linear time trend, monitor-month fixed effects, region fixed effects, and year fixed effects. Standard errors clustered at the district level are in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Figures 13 and 14 show the net effect of fuel prices on CO, PM10, SO2, and NO2 concentrations. This is represented by the linear combinations of coefficients of different fueling stations in column (3) and (4) in Table 7 but for all pollutants. The number beside each coefficient represents a change in pollution concentration (ppm for CO, ug/m3 for PM10, and ppb for SO2 and NO2) in response to a unit change in price. The numbers in parentheses are the percent changes in pollution concentration due to a one percent change in price. I see that fuel prices do not have a large effect on air pollution concentration. Perhaps it is because I do not have enough temporal variation between CNG and LPG prices. Because the instrument variables correct for the measurement error, the coefficient associated with the number of CNG stations from the 2SLS regression decreases.

Figure 15 shows the net effect of CNG stations on CO, PM10, SO2, and NO2 concentrations. It is the coefficient associated with linear combinations of the coefficients for the different fueling stations in column (3) and (4) in Table 7 but for all pollutants. The numbers beside each coefficient represent a change in pollution concentrations (ppm for CO, ug/m3 for PM10, and ppb for SO2 and NO2) in response to a percent change in the number of CNG stations. The numbers in the parentheses are the percent changes in pollution concentrations due to a one percent change in the number of CNG stations. From the 2SLS result, I find significant reductions in SO2 levels when the number of stations increases. When the number of CNG stations is doubled, SO2 levels can decrease by one percent. I do not find any significant effect with other pollutants.

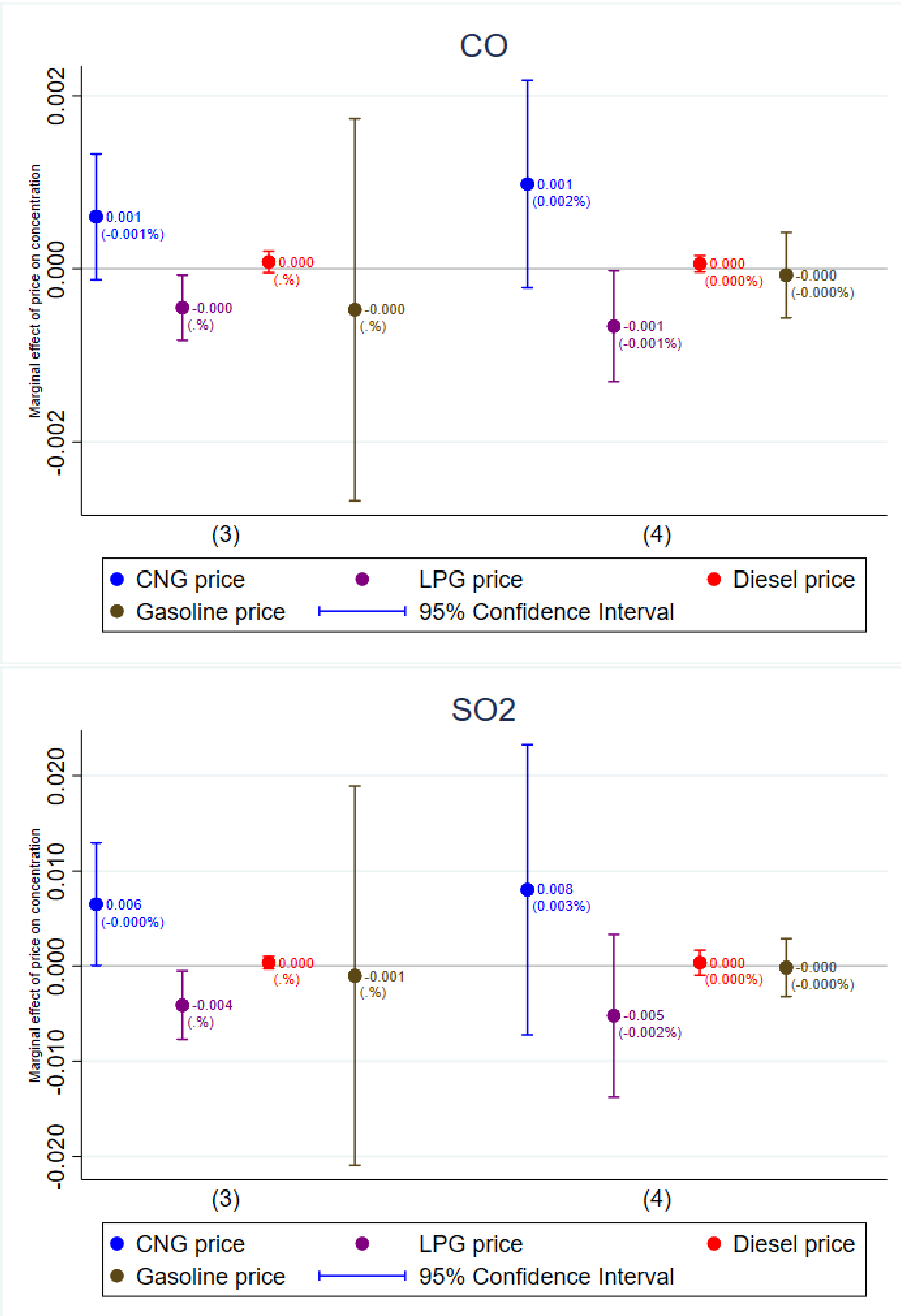

Figure 13: The figure shows the effect of fuel prices on pollution concentration. The numbers outside the parentheses are unit changes in concentration (ppm for CO and ppb for SO2) in response to a unit change in price. The numbers in the parentheses are percent changes in pollution concentrations due to a one percent change in price.

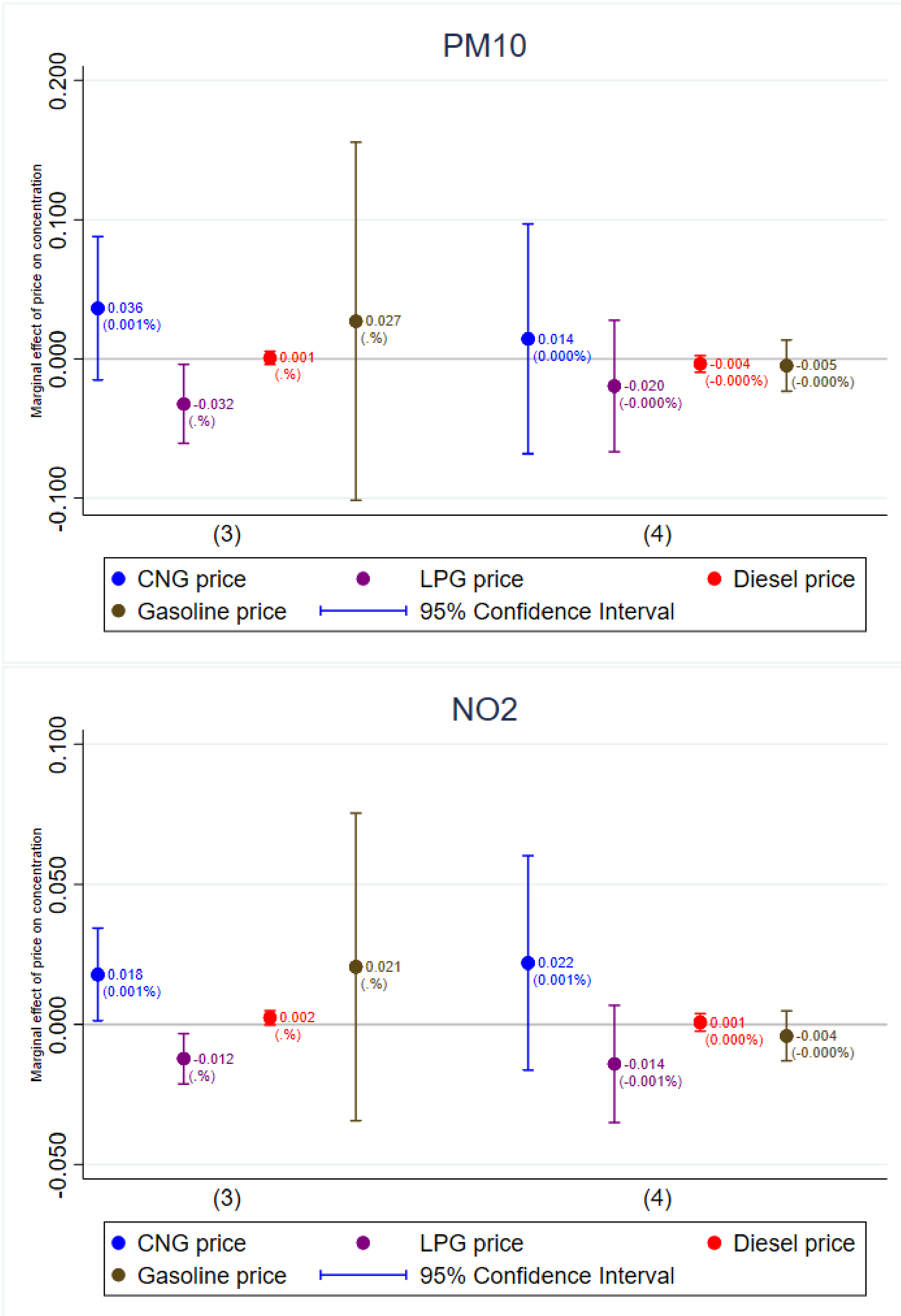

Figure 14: The figure shows the effect of fuel prices on pollution concentration. The numbers outside the parentheses are the unit changes in concentrations (ug/m3 for PM10 and ppb for NO2) in response to a unit change in price. The numbers in the parentheses are the percent changes in pollution concentrations due to a one percent change in price.

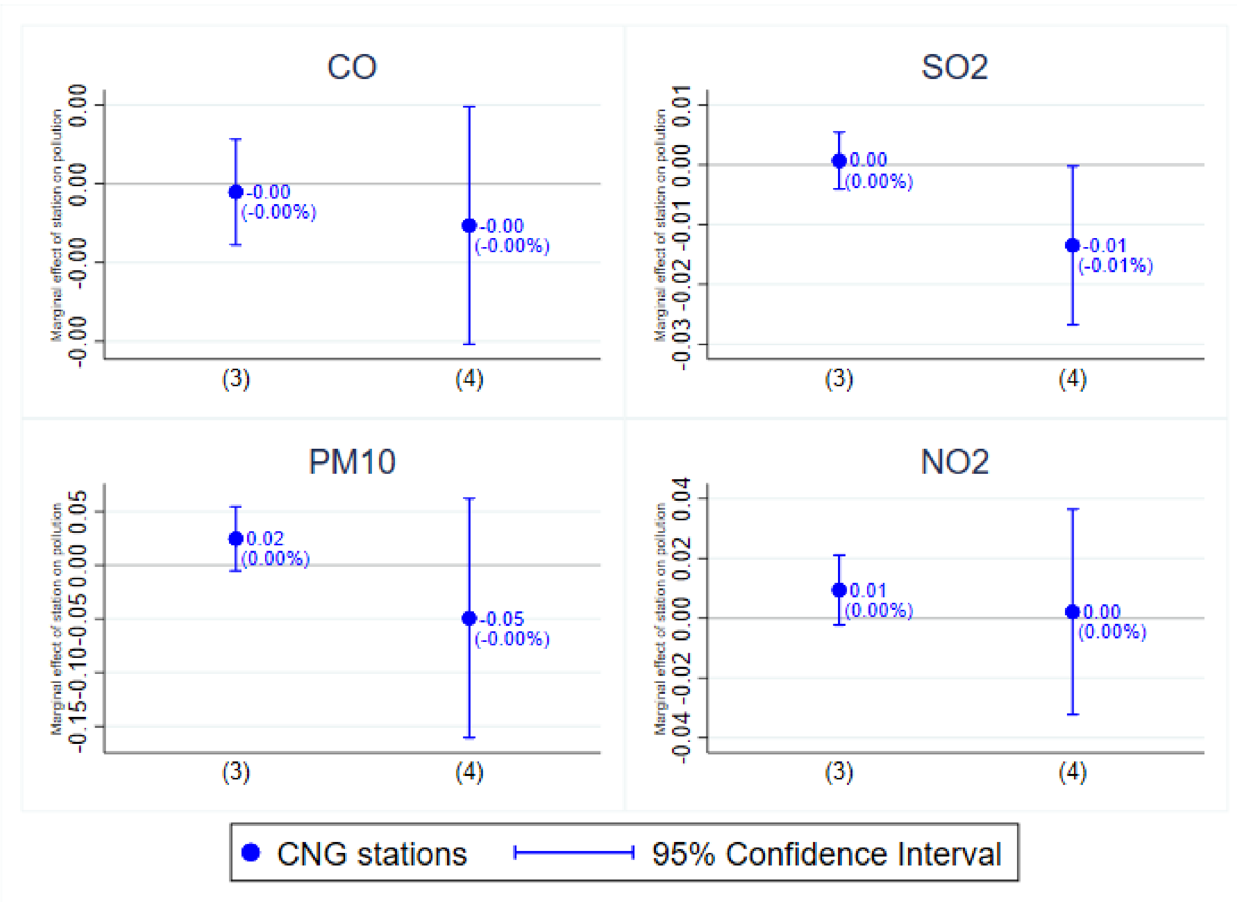

Figure 15: The figure shows the effect of CNG stations on pollution concentration. The numbers outside the parentheses are unit changes in concentrations (ppm for CO, ug/m3 for PM10, and ppb for SO2 and NO2) in response to a percent change in the number of stations. The numbers in the parentheses are percent changes in pollution concentrations due to a one percent change in the number of stations.

7 Conclusion

As CNG subsidies are widely used in developing country to reduce fuel prices and improve air quality, it is important that the causes and consequences of CNG adoption are studied. By using a 2SLS estimation strategy and leveraging on backward HYSPLIT simulations, this research tries to estimate the causes and consequences of CNG adoption as accurately as possible given available data.

This research suggests that increasing the availability of CNG, via the setup of natural gas fueling stations, is the largest driver of CNG consumption. Because CNG vehicles do not emit SO2, a one percent increase in CNG availability significantly decreases the average SO2 concentrations by 0.01 percent. SO2 can harm the human health. Once SO2 is inhaled, it can dissolve and oxidize in the mucous membranes to sulfurous acid and damage respiratory functions (Jacobson 2012). Long-term epidemiological studies have shown that lowest observed effect of SO2 was judged to be 35 ppb (Agency for Toxic Substances and Disease Registry 1998). As for the case of Thailand, the highest monthly average SO2 level observed is 20.6 ppb, there are no health damages from SO2. A reduction in SO2 of one percent when the number of fueling station doubles will not have any direct health benefit. Because SO2 is a precursor to PM2.5, there may be health benefit of a reduction in PM2.5. However, PM2.5 is rarely measured by the Pollution Control Department in Thailand.

During my period of study, March 2008 to December 2015, the annual losses from CNG subsidies were about 6.45 billion Baht or about 200 million USD a year. This loss, which is equivalent to 0.05 of Thailand's GDP, is a loss when CNG adoption or air pollution levels. However, there are other benefits to CNG adoption, such as an affordable fuel source for the poor that this research does not take into account.

References

Agency for Toxic Substances and Disease Registry. 1998 "Toxicological profile for Sulfur Dioxide" U.S. Department of Health and Human Services, Public Health Service

Anderson, S. T. and Sallee, J. M. 2011. "Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards" *American Economic Review*, 1375-1409

Anderson, S. T. 2012. "The demand for ethanol as a gasoline substitute". *Journal of Environmental Economics and Management* 63(2), 151-168.

Auffhammer, M. and Kellogg R. 2011. "Clearing the Air? The Effects of Gasoline Content Regulation and Air Quality". *American Economic Review* 101:2687-2722.

Bunch, D.S., M. Bradley, T.F. Golob, R. Kitamura, and G.P. Occhiuzzo. 1993. "Demand for Clean-Fuel Personal Vehicles in California: A Discrete-Choice Stated Preference Study" *Transportation Research*, Vol. 27A, pp.297-253.

Cai, H., Burnham A., and Wang, M. 2013 "Updated Emission Factors of Air Pollutants from Vehicle Operations in GREET Using Moves." Systems Assessment Section, Energy Systems Division, Argonne National Laboratory.
https://greet.es.anl.gov/files/vehicles-13

Coady D., Parry, I., Sears, L., and Shang B. 2015. "How Large Are Global Energy Subsidies" IMF Working Paper WP/15/105.

Dholakia, H. H., Purohit, P., Rao, S., and Garg, A. 2013. "Impact of current policies on future air quality and health outcomes in Delhi, India". *Atmospheric Environment* 75, 241-248.

Draxler, R.R., and G.D. Hess. 1997. "Description of the HYSPLIT4 modeling system". NOAA Tech. Memo. ERL ARL-224, NOAA Air Resources Laboratory, Silver Spring, MD, 24 pp.

Draxler, R.R., and G.D. Hess. 1998. "An overview of the HYSPLIT4 modeling system of trajectories, dispersion, and deposition". Aust. Meteor. Mag., 47, 295-308.

Draxler, R.R 1999. "HYSPLIT4 user's guide". NOAA Tech. Memo. ERL ARL-230, NOAA Air Resources Laboratory, Silver Spring, MD.

Energy Policy and Planning Office 2011. "Natural Gas Price Manual"
http://www.eppo.go.th/ petro/pt-KPC2544-01.html.

Greene, D. L. 1989. "Fuel choice for Dual-Fuel Vehicles: An Analysis of the Canadian Natural Gas Vehicles Survey" *SAE Techical Paper Series No. 892067*

Houde, S. 2008. "Spatial differentiation in retail markets for gasoline". *American Economic Review*. 102(5), 1-47.

Jacobson, M. Z. 2012 "Air Pollution and Global Warming: History, Science, and Solutions" Cambridge University Press, Cambridge

Krupnick, A. J. 2011. "Will Natural Gas Vehicles Be in Our Future?" Retrieved April 5, 2015, from http://www.rff.org/Publications/Resources/Pages/181-Natural-Gas-Vehicles.aspx

Kumar, N., and Foster, A. D. 2007. "Have NGV Regulations in Delhi Done Their Job?" *Economic and Political Weekly*. 42(51), 48-58

Li, S., Xing, J., Tong, L., and Yiyi Z. 2017. "The Market for Electric Vehicles: Indirect Network Effects and Policy Design" *Journal of the Association of Environmental and Resource Economists*. doi=10.1086/689702

Manuszak, M. D. and Moul, C. C. November 2009. "How Far for a Buck? Tax Differences and the Location of Retail Gasoline Activity in Southeast Chicagoland" *The Review of Economics and Statistics*, 59 (2010) 91(4): 744-765

Narain, U., and Krupnick A. 2007. "The Impact of Delhi's CNG Program on Air Pollution". *Resources for the Future Discussion Paper*

Nevo, A., and Rosen A. 2012. "Identification with Imperfect Instruments" *Review of Economics and Statistics*. 94 (3) 659-71.

PTT Public Limited Company 2015. "Conditions and Details of Natural Gas Distribution in the Transportation Sector"
http://www.pttplc.com/th/Products-Services/Consumer/Services/Documents/

Ravindra, K., Wauters, E., Tyagi, S. K., Mor S., and Van Grieken, R. 2006. "Assessment of Air Quality after the Implementation of Compressed Natural Gas (CNG) as Fuel in Public Transport in Delhi, India". *Environmental Monitoring and Assessment*

Reynolds, C. O., and Kandlikar M. 2008. "Climate Impacts of Air Quality Policy: Switching to a Natural Gas-Fueled Public Transportation System in New Delhi". *Environmental Science and Technology*

Salvo, A., and Huse, C. 2013. "Build it, but will they come? Evidence from consumer choice between gasoline and sugarcane ethanol" *Journal of Environmental Economics and Management* 66(2), 251-279. doi:10.1016/j.jeem.2013.04.001

Schifter, I., Diaz, L., Lopez-Salinas, E., and Avalos, S. 2000. "Potential Impacts of Compressed Natural Gas in the Vehicular Fleet of Mexico City". *Environmental Science & Technology* 34(11), 2100-2104.

Suthawaree, J., Sikder, H. A., Jones, C. E., Kato, S., Kunimi H., Kabir, A. N. M. H., and Kajii, Y. "Influence of Extensive Compressed Natural Gas (CNG) Usage on Air Quality". *Atmospheric Environment*

Wadud, Z. 2014. "(Unintended) Transport impact of an energy-environment policy: The case of CNG conversion of vehicles in Dhaka".*Transportation Research Part A Policy and Practice* doi=10.1016/j.tra.2014.04.01

Wadud, Z., and Khan, T. 2013. "Air Quality and Climate Impacts Due to NGV Conversion of Motor Vehicles in Dhaka , Bangladesh".*Environmental Science & Technology*

Wooldridge, Jeffrey M. 2002. "Econometric analysis of cross section and panel data". *Cambridge, Mass: MIT Press*

Appendix A

For every unit of CNG sold at fuelling stations, revenue received by fuelling station retailers are calculated from the following equation (Energy Policy and Planning Office 2011)

$$P_{CNG} = WH_{Pool2} \times (1.0175) + Td_{Zone1+3} + Tc + OilFund + StationCosts + MarketingCosts$$

where WH_{Pool2} is well-head gas price (in Baht per million British Thermal Unit) obtained from the Gulf of Thailand, Myanmar and imported liquefied natural gas (LNG). Td is fixed costs for pipeline usage. It includes investment costs in past and future pipeline construction that are adjusted such that it is equal every year. Td is adjusted when new pipeline plans are approved. For CNG, Td only includes infrastructure investments for pipelines in Zone 1 and 3, which include investment costs for the national onshore pipeline gird and offshore pipelines in Rayong. Tc is unit transport cost in Baht per million BTU of gas transported.

$OilFund$ is a pool of government revenue collected from taxing diesel and gasoline fuel sales and is used to subsidize ethanol and CNG. CNG was first subsidized at 2 Baht per kilogram, but was reduced to 1 Baht per kilogram in 2011. CNG is no longer subsidized after February 2012.

Station cost differs between a "mother" station next to a pipeline, and a "daughter" station in which gas has to be transported via trucks. Station costs is equal to 1.12 and 1.00 Baht per kilogram of gas for mother and daughter CNG stations. Daughter station within 50 kilometers of a major station also bears an additional costs of 1.20 Baht per kilogram of transport costs, and additional cost of 0.012 Baht per kilogram outside the 50 kilometer radius. Marketing costs 1.73 to 2.33 Baht per kilogram depending on the size and type of CNG station. PTT's "conventional" station costs is equal to 3.33 Baht per kilogram. Other "conventional" station costs 2.26 Baht per kilogram.

In short, CNG station owners get P_{CNG} as discussed while a CNG consumer pays the regulated price of CNG $P_{Regulated}$. It is suspected that $P_{Regulated} \geq P_C CNG$ and that PTT, the wholesaler of CNG bears this cost.

In July 2015, the estimated P_{CNG} is 13.26 Baht per kilogram. Given that Henry Hub natural gas price plus 1.0175 is 4.82 Baht per kilogram (PTT 2015), $Td_{Zone1+3} + Tc + StationCosts + MarketingCosts$ is estimated to be 8.44 Baht per kilogram.

Appendix B

The equation below illustrates province-level data from three provinces, Karnchanaburi (KB), Nakorn Pathom (NP), and Rachaburi (RB), are combined to create grid-level data in grid 676. Figure 16 helps visualize the data.

$$Price_{676} = \frac{area_{KB \cap 676}}{area_{676}} Price_{KB} + \frac{area_{NP \cap 676}}{area_{676}} Price_{NP} + \frac{area_{RB \cap 676}}{area_{676}} Price_{RB}$$

$Price_{676}$ represents price level in grid 676 in Figure 16. $area_{KB\cap 676}$ is the size of the area than Karnchanaburi province intersects with grid 676. $area_{676}$ is the size of grid 676 which is 120 sq. km. Therefore, $\frac{area_{KB\cap 676}}{area_{676}}$ is the percentage of the impact that Karnchanaburi province has on grid 676. The sum of all weighted prices from each province that intersects with grid 676 is the generated prices for grid 676.

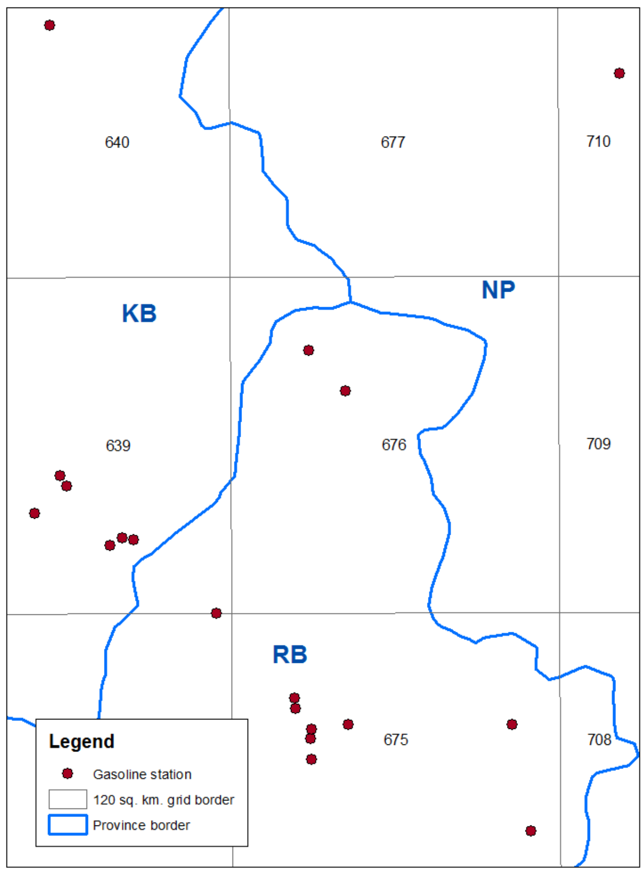

Figure 16: The figure shows intersection between province-level covariates and fueling station locations. Data from three provinces, Karnchanaburi (KB), Nakorn Pathom (NP), and Rachaburi (RB), are combined to create grid-level data in grid 676.

Appendix C

Below is a model with monitor-level pollution concentration as the dependent variable, and the measurement error and unobserved demand is lumped into one variable.

$$AQM_{mt} = \beta_0 + \beta_1 CNGStations_{mt} + ... + \kappa(OmittedVariable_{mt}) + \varepsilon_{mt} \tag{3}$$

If the $OmittedVariable_{mt}$ is uncorrelated with other variables once controlled for $CNGStations_{mt}$, then I can write the linear projection of the $OmittedVariable_{mt}$ onto $CNGStations_{mt}$,

$$OmittedVariable_{mt} = \delta_0 + \delta_1 CNGStations_{mt} + \mu_{mt} \tag{4}$$

where μ_{mt} is uncorrelated with other variables. When I substitute equation 4 into 3, I have the following equation:

$$AQM_{mt} = \beta_0 + (\beta_1 + \kappa\delta_1) CNGStations_{mt} + ... + \kappa\delta_0 + \kappa\mu_{mt} + \varepsilon_{mt}$$

Let the error $\kappa\delta_0 + \kappa\mu_{mt} + \varepsilon_{pt}$ have a zero mean and be uncorrelated with each regressor (Woolridge 2002). Because more CNG stations will induce more CNG adoption and decrease pollution concentrations, $\beta_1 < 0$ holds true. Because the $OmittedVariable_{mt}$ includes stations that are not reported as well as unobserved demand, I assume that $\kappa < 0$. If the measurement error decreases as the number of CNG stations increases overtime, I assume that $\delta_1 < 0$. Let $\hat{\beta_1}$ be a coefficient for $CNGStation_{mt}$ from an OLS estimation.

$$\hat{\beta_1} = \beta_1 + \kappa\delta_1 > \beta_1$$

Therefore, in order to see the true effect of CNG stations on pollution concentration, β_1 should be lower than the effect obtained from an OLS regression.

Appendix D

The first stage regression for column (2) and (3) in Table 6 is shown in Table 10. Column (1) in Table 10 is the first stage regression for column (2) in Table 6. Columns (2) and (3) in Table 10 are the first stage regression for column (4) in Table 6. The smaller the distance from licensed engineers to a CNG station, the greater the number of stations exist. Therefore, my instrument is negatively correlated with my endogenous variable.

The first stage regression when SO2 is used in the second stage is shown in Table 11. The first stage results are slightly different for different pollutants as the number of observations are different. Column (1) in Table 11 is the first stage regression for column (2) in Table 7. Columns (2) and (3) in Table 11 are the first stage regression for column (4) in Table 7. Again, my instrument is negatively correlated with the number of CNG stations. As the distance to licensed engineers becomes shorter, the greater number of stations there are in each 120-square-kilometer grid.

Appendix E

To test for robustness of the estimated impact of CNG stations, I tried estimating the impact of HYSPLIT-weighted CNG stations on various air pollution concentrations with different specifications. The net impact of CNG stations on air pollution concentration is shown in Figures 17 to 19. Each figure represents results from specifications in columns (2) and (4) in Table 7. Numbers in the plots represent a change in pollution concentration due to a percentage increase in the number of CNG stations. Each result is then subdivided into (a), (b), and (c) as labeled in each figure. Instead of a linear time trend, specification (a)

Table 10: First stage results for CNG demand

	(1) CNG stations	(2) CNG stations	(3) CNG stations × Gasoline price relative to CNG price
Distance to licensed engineer × Lagged CNG stations in all other provinces	-0.012***	-0.002	0.023
	(0.004)	(0.005)	(0.018)
Distance to licensed engineer × Lagged CNG stations in all other provinces × Gasoline price / CNG price		-0.003**	-0.018***
		(0.001)	(0.006)
Gasoline price / CNG price	2.297***	1.551	13.359**
	(0.578)	(1.147)	(5.552)
Diesel price / CNG price	-0.886***	-1.401***	-5.875***
	(0.218)	(0.448)	(1.916)
Gasoline price / LPG price	-2.812***	0.595	-13.635
	(0.871)	(2.778)	(11.210)
Regular stations × Diesel price / CNG price		0.001	0.000
		(0.002)	(0.007)
Regular stations × Gasoline price / CNG price		0.005	0.029
		(0.004)	(0.020)
Regular stations × Gasoline price / LPG price		-0.014	-0.086*
		(0.010)	(0.049)
LPG stations × Gasoline price / LPG price		-0.071	0.587***
		(0.061)	(0.212)
Observations	7050	7050	7050
Adjusted R^2	0.939	0.939	0.922

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, province fixed effects, region-year fixed effects, and a linear time trend. Standard errors clustered at the province level are in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 11: First stage results to estimate average SO2 concentration

	(1)	(2)	(3)
	invsinh(CNG stations)	invsinh(CNG stations)	invsinh(CNG stations) × Gasoline price relative to CNG price
Distance to licensed engineer × Lagged CNG stations in all other provinces	-0.348***	-0.253*	-0.456
	(0.092)	(0.129)	(0.501)
Distance to licensed engineer × Lagged CNG stations in all other provinces × Gasoline price / CNG price		-0.015	-0.213***
		(0.015)	(0.058)
invsinh(Regular stations) × Diesel price / CNG price		-0.056	-0.471**
		(0.037)	(0.175)
invsinh(Regular stations) × Gasoline price / CNG price		-0.288*	-0.572
		(0.167)	(0.689)
invsinh(Regular stations) × Gasoline price / LPG price		0.741	1.494
		(0.445)	(1.897)
invsinh(LPG stations) × Gasoline price / LPG price		-0.262***	0.221
		(0.055)	(0.205)
Gasoline price / CNG price	0.588***	1.503**	6.166**
	(0.162)	(0.580)	(2.341)
Diesel price / CNG price	-0.120***	-0.014	0.221
	(0.019)	(0.085)	(0.418)
Gasoline price / LPG price	-0.927***	-2.855*	-11.065*
	(0.304)	(1.442)	(6.003)
Observations	3601	3601	3601
Adjusted R^2	0.891	0.903	0.871

Note: Controls include number of LPG stations, regular fueling stations, CNG factories, annual gross provincial product from each sector, population density, temperature and precipitation, linear time trend, monitor-month fixed effects, region fixed effects, and year fixed effects.

Standard errors clustered at the district level are in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

includes year fixed effects to control for nation-wide shocks such as election years. Other controls included in the main specification including region fixed effects and monitor-month fixed effects are also included.

As monthly weather patterns may already be controlled by HYSPLIT, specification (b) use month-region fixed effects and monitor fixed effects separately to control for regional weather patterns. I also included to control for nation-wide shocks.

Because the placement of each monitor can directly impact air pollution concentration data, specification (c) includes monitor fixed effects by itself. I also included year-month fixed effects and region fixed effects to control for temporal and spatial trends.

Figure 17 shows the net impact of CNG stations on average air pollution concentrations with varying fixed effects and a linear time trend. Specification (a) in Figure 17 is the same result as shown in Figure 15. Figure 18 shows the net effect of CNG stations on the monthly average of pollution concentration with different fixed effects and a linear time trend with 30-square-kilometer HYSPLIT grids instead of 120-square-kilometer grids. Figure 19 shows the net effect of CNG stations on the monthly maximum of air pollution concentrations with varying fixed effects. Although I have pointed out that the maximum air pollution concentrations in Thailand is problematic in some areas, these results show that the number of CNG stations does not have an impact on improving the monthly maximum of air pollution concentrations.

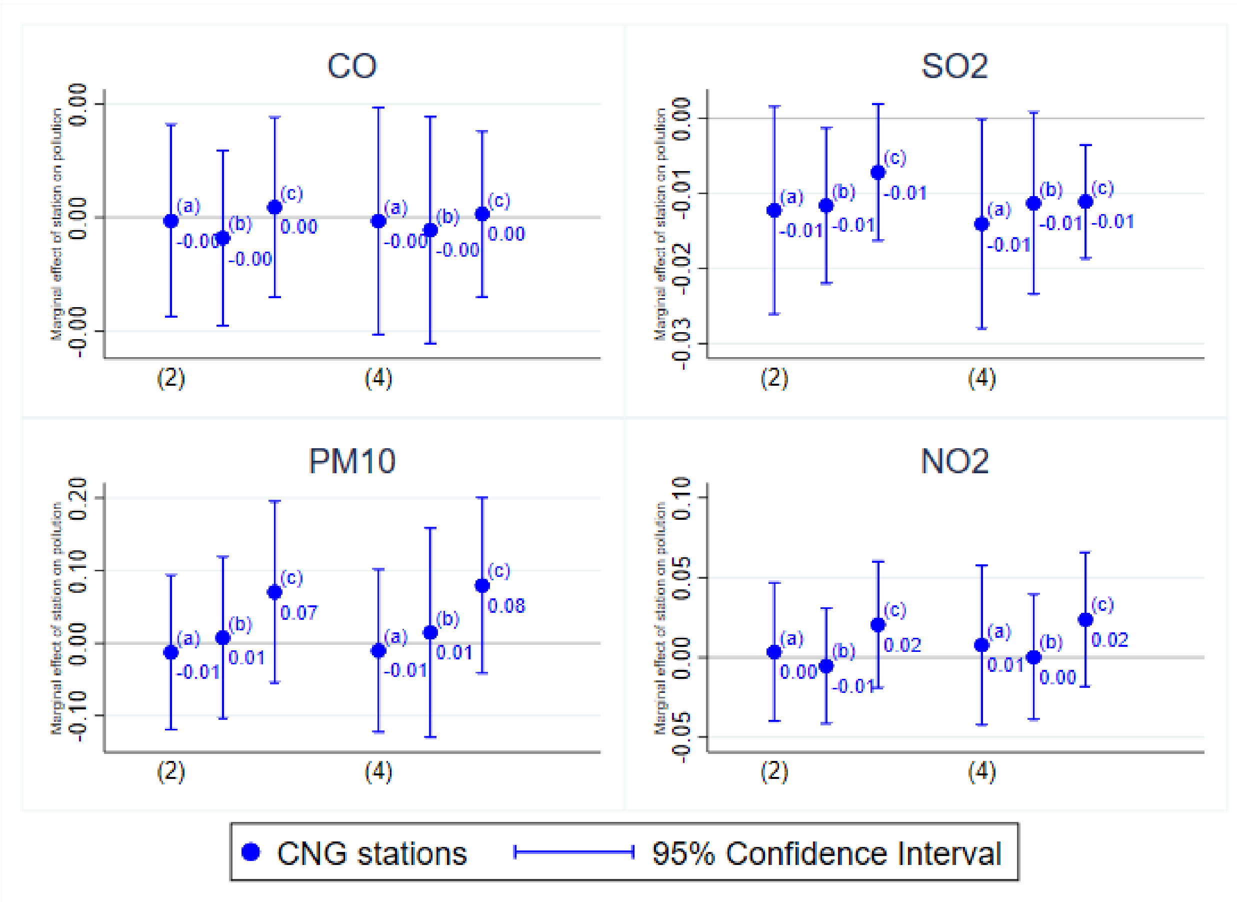

Figure 17: The plots show net effect of CNG station on monthly average pollution concentration with different fixed effects and a linear time trend. Specification (a) includes year fixed effects, region fixed effects, and monitor-month fixed effects. Specification (b) includes year fixed effects, month-region fixed effects, and monitor fixed effects. Specification (c) includes year-month fixed effects, region fixed effects, and monitor fixed effects. Numbers in the plots represent a change in pollution concentration due to a percent increase in the number of CNG stations.

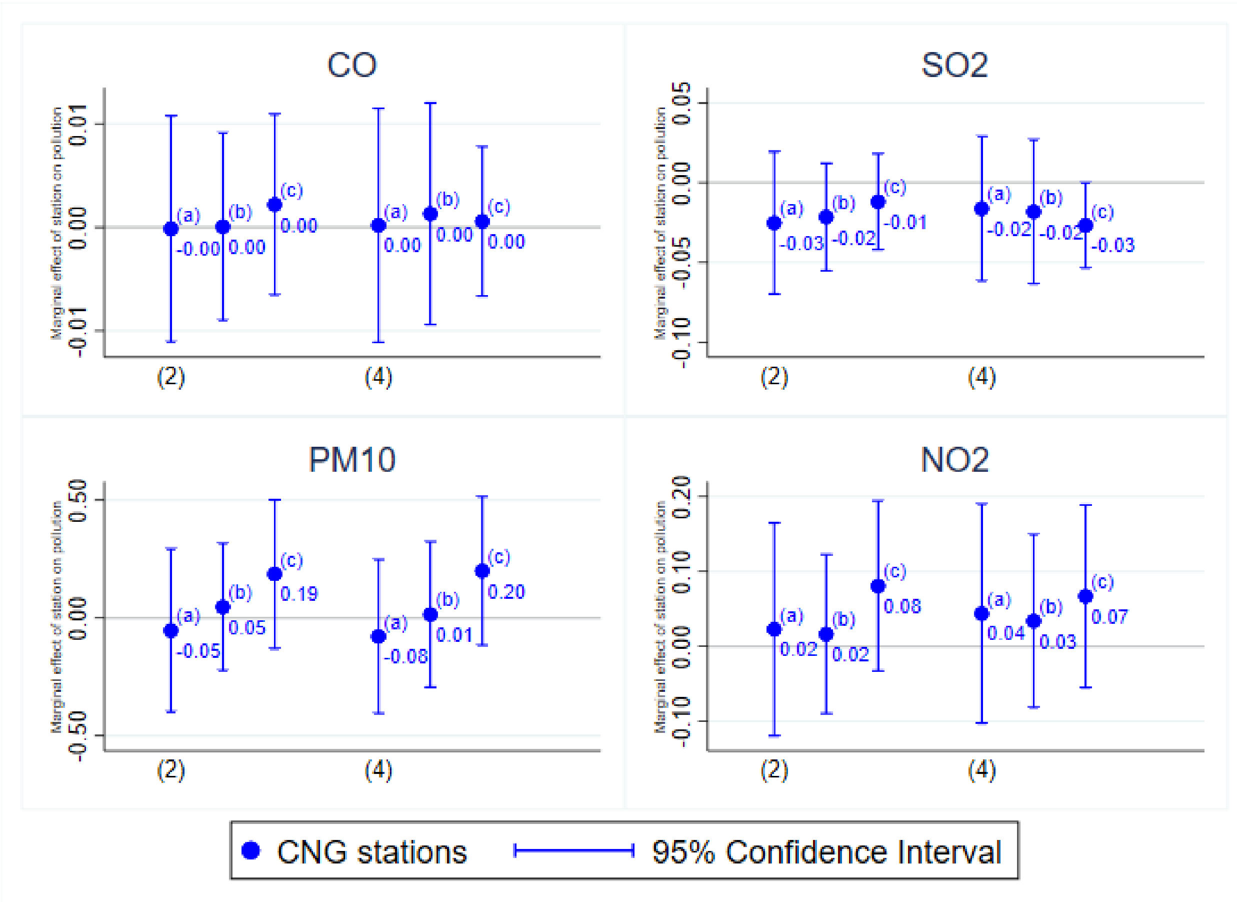

Figure 18: The plots show net effect of CNG station on monthly average pollution concentration with different fixed effects and a linear time trend with 30 sq. km. HYSPLIT grids. Specification (a) includes year fixed effects, region fixed effects, and monitor-month fixed effects. Specification (b) includes year fixed effects, month-region fixed effects, and monitor fixed effects. Specification (c) includes year-month fixed effects, region fixed effects, and monitor fixed effects. Numbers in the plots represent a change in pollution concentration due to a percent increase in the number of CNG stations.

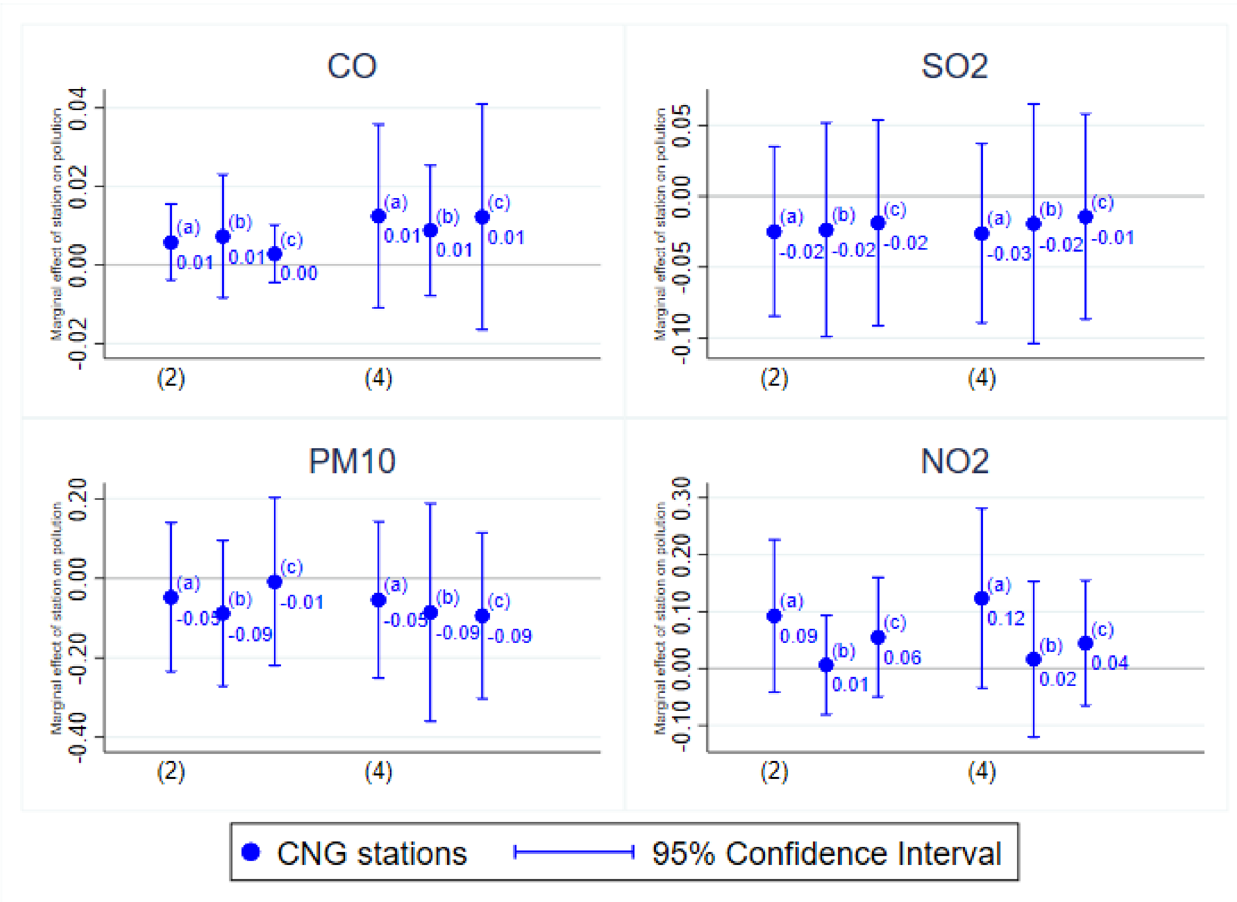

Figure 19: The plots show net effect of CNG station on monthly maximum pollution concentration with different fixed effects and a linear time trend. Specification (a) includes year fixed effects, region fixed effects, and monitor-month fixed effects. Specification (b) includes year fixed effects, month-region fixed effects, and monitor fixed effects. Specification (c) includes year-month fixed effects, region fixed effects, and monitor fixed effects. Numbers in the plots represent a change in pollution concentration due to a percent increase in the number of CNG stations.