Understanding Gasoline Price Dispersion*

Demet Yilmazkuday† and Hakan Yilmazkuday‡

May 26, 2016

Abstract

This paper models and estimates the gasoline price dispersion across time and space by using a unique data set at the gas-station level within the U.S.. Nationwide effects (measured by time fixed effects or crude oil prices) explain up to about 51% of the gasoline price dispersion across stations. Refinery-specific costs, which have been ignored in the literature due to using local data sets within the U.S., contribute up to another 33% to the price dispersion. While state taxes explain about 12% of the price dispersion, spatial factors such as local agglomeration externalities, land prices, distribution costs of gasoline explain up to about 4%. The contribution of brand-specific factors is relatively minor.

JEL Classification: L11, L81, R32, R41,

Key Words: Gasoline Prices; Gas-Station Level Analysis; Nighttime Lights; Land Prices, the United States

---

*The authors also would like to Shu Deng and James Bailey for excellent research assistance, Zillow Research for the housing price data, and seminar participants at various places for their feedback and suggestions. The usual disclaimer applies.

†Department of Economics, Florida International University, Miami, FL 33199, USA;
e-mail: dyilmazk@fiu.edu

‡Department of Economics, Florida International University, Miami, FL 33199, USA;
e-mail: hyilmazk@fiu.edu

1. Introduction

Retail prices of gasoline are considerably different across gas stations within the U.S. where consumer expenditure share of gasoline is about 5%. Since such price differences may be reflecting the frictions in the economy, understanding the reasons behind them is the key to an optimal policy that would improve the level and distribution of individual welfare. For example, consider a typical day in 2010-2011 when the retail-level gasoline price difference between any two gas stations within the U.S. was as high as $2.25 (followed by $2.19) per gallon of regular gas.[1] If you think that this price dispersion was due to differences in state-taxes per gallon, which ranged between 46.6 cents (for California) and 8 cents (for Alaska) in 2010 and between 49.6 cents (for Connecticut) and 8 cents (in Alaska) in 2011, you are only partially right, because, for a typical day of the very same sample period, the price difference between any two gas stations within any given state/district of the U.S. was as high as $1.57 (for Washington D.C. on October 15th, 2010) followed by $0.99 (for Iowa for October 10th, 2010).[2] Therefore, gasoline price differences not only exist across states but also within states; accordingly, a detailed spatial analysis (rather than a state-level analysis) is required to understand the details behind gasoline price dispersion.

We pursue such an investigation by testing the implications of a spatial model (of gasoline

[1] The gasoline price differences of both $2.25 and $2.19 were between Washington D.C. and Michigan on October 23rd, 2010 and December 16th, 2010, respectively

[2] The gasoline price dispersion was not due to outliers, either; because, for a typical day, the median price difference between any two gas stations within any given state/district of the U.S. was as high as $0.60 (for Hawaii) followed by $0.44 (for Washington D.C.) and by $0.41 (for Wyoming). Table 1 depicts the complete descriptive statistics.

consumption, production, and distribution) with a comprehensive gas-station level gasoline retail price data (with about half a million price observations) obtained from all around the U.S.[3] In the model, consumer-producers get utility out of their consumption of goods as well as gasoline which is required for transportation purposes. The necessity for transportation also depends on the distribution of economic activity around the individual who is assigned a specific spatial location. Spatially distributed gas stations maximize their profits based on the gasoline demand of consumer-producers who are involved in economic activity around the gas station. Gas stations also take into account the cost of gasoline (charged by nearest refiners) and other spatial factors (such as the size and the distribution of economic activity around the gas station, the distance of the gas station to the refinery, number of other gas stations in the vicinity). Optimization implies that the equilibrium price of gasoline at a particular gas station depends on spatial economic factors, the cost of gasoline at the refinery, legislative factors (such as state-level taxes), local competition, and brand-specific costs/benefits. We test the implications of this model on the gasoline price difference across gas stations using a unique data set that is introduced next.

The daily gasoline price data set covering the period between September 10th, 2010 and January 31st, 2011 involves brand information of gas stations as well as their location information at the exact address level. Combining this data set with the exact (address-level)

[3]Studies such as Chouinard and Perloff (2007) have used the state-level price observations and shown that (i) the gasoline price dispersion within the U.S. is due to changes in the price of crude oil; and (ii) across-state variation in prices can be attributed to demand variation, as well as differences in taxes, environmental regulations and market power. In addition to such studies, this paper attempts to explain the spatial gasoline price differences within states.

location information of oil refineries, state-level taxes, crude oil prices, land prices (at the zip-code level), and local agglomeration externalities of spatial gasoline demand (namely, the distribution of nighttime lights data across space), we decompose the effects of crude oil prices, refinery costs, distribution costs, brand-specific costs, state taxes, land prices, and local agglomeration externalities (i.e., spatial demand and number of competitors) through a spatial analysis that models and estimates the transportation needs of individuals and distribution of gasoline from refineries to gas stations. While the crude oil prices (or any other time-varying effect that is common across gas stations) are considered to capture time-varying effects that are common across all gas stations within the U.S., other cross-sectional variables are considered to capture the effects due to spatial factors.

The empirical results show that one percent of an increase in local agglomeration externalities reduces gasoline prices up to 1.92%, while one percent of an increase in the number of competitors can reduce gasoline prices as much as 0.19%. Distribution costs of gasoline from the nearest refiner to the gas station have positive and significant effects, where one percent of an increase in distribution costs leads to an increase up to about 0.41% in gasoline prices. It is also shown that one percent of an increase in land prices leads to an increase up to about 0.79% in gasoline prices. These results are robust to the consideration of brand fixed effects, time fixed effects, refinery fixed effects, and alternative spatial measures (for the matching between consumer-producers and gas stations). It is also shown that the proposed model explains the data well; the R-squared values for alternative spatial measures are all about 0.90.

We further investigate the contribution of each model component in explaining the gasoline price differences across gas stations. The corresponding variance decomposition of gasoline prices across time and space suggests that the highest contribution is by time fixed effects (or crude oil prices) capturing up to 51% (or 39%) of the price dispersion (i.e., nationwide effects are almost half of the overall effects), followed by refinery-specific costs capturing up to 33% of the price dispersion, state taxes capturing up to 12% of price dispersion, and spatial factors (such as local agglomeration externalities, land prices, distribution costs of gasoline) capturing up to 4% of price dispersion. The contribution of brand-specific costs is relatively minor. Within this picture, as discussed in details in the next section, the main contribution to the existing literature has been to show that refinery-specific costs explain a big part of the price dispersion across gas stations, which has been mostly ignored in the literature due to using local data sets covering either a couple of cities or states within the U.S..

2. Literature Review

This paper is connected to the empirical literature analyzing the determinants of gasoline price dispersion at the gas station level. The literature has focused on price control variables such as (i) local demographics or station location (LDSL), (ii) brand or contractual arrangements (BCA), (iii) station density or local concentration (SDLC), and (iv) physical station characteristics (PSC).[4]

In the literature, Shepard (1993) has used cross-sectional gasoline price data collected over a twelve week period in the first quarter of 1987 from over 1,100 gas stations from Eastern Massachusetts. Shepard has shown that prices across stations vary with the contractual form between refiners and stations, with the number, proximity and characteristics of competing retailers, and with station characteristics such as selling non-gasoline products or services; but

[4]See Eckert (2013) for a survey of empirical studies of gasoline retailing.

prices do not vary with market conditions such as traffic volume. Ning and Haining (2003) have used gasoline price data collected over eight biweekly period in mostly the first quarter of 1995 from 113 gas stations, together with price data collected over six weekly period in mostly the first quarter of 1997 from 96 gas stations, from Sheffield, England. They have shown that spatial competition and being attached to a supermarket are statistically significant, but the number of cars around the gas station and/or percentage of higher income households in the neighborhood of a station are not statistically significant. Van Meerbeeck (2003) has used gasoline price data covering 477 gas stations in Belgium on a daily basis between 1998 and 2001. This data set includes only two brands of gasoline (Shell and TotalFina). Van Meerbeeck has observed that prices vary with stations located along a highway always charge the maximum price determined by an agreement between the oil industry and the Belgian government, and prices are below the maximum price in markets with sufficient competition and that the number of local competitors does not have a large impact on retail gasoline prices. Barron et al. (2004) have used cross-sectional gasoline price data covering 3,197 gas stations in four metropolitan areas (Phoenix, AZ, Tuscon, AZ San Diego, CA, and San Francisco, CA) for only one day in 1997 (where that one day differs across metropolitan areas). They have found that an increase in station density consistently decreases both price levels and price dispersion across four geographical areas. They have also found that price vary with brand identifiers, with a variable indicating whether the primary image of the station is a convenience store or a repair station, with a variable indicating whether full-service gasoline is sold at the station, and with a variable reflecting the number of fueling positions. Eckert and West (2004) have used daily gasoline price data covering several gas stations over a period of eight months in 2000 from Vancouver and Ottawa metropolitan areas in Canada. They have shown that brand identities, concentration of gas stations, and the type of ownership are significant in explaining price differences across gas stations. Clemenz and Gugler (2006) have used daily gasoline price data collected over a five month period in 2000 and 2001 covering 1,603 gas stations from in Austria. They have found that prices vary with station density, but ownership concentration is not statistically significant. Pennerstorfer (2009) has used cross-sectional price data covering 400 gas stations in Lower Austria for September 2003. Pennerstorfer has shown that prices vary with branded versus unbranded gas stations. Chandra and Tapatta (2011) have used daily gas-station level data covering the states of California, Florida, Texas, and New Jersey covering the 18 months period between January 2006 to May 2007. They have shown that price dispersion increases with the number of firms in the market and consumer search costs.

Compared to the literature mentioned so far, this paper uses gasoline price data of an unbalanced daily panel over 144 days from 38,245 gas stations from all around the U.S. In addition to the literature, we have found that refiner fixed effects are econometrically significant in explaining gasoline prices, which is the key in understanding the spatial gasoline price dispersion within the U.S. and has been ignored by the existing literature focusing on local data sets. In particular, having a local-level analysis (as in the existing literature) would suppress the main determinants of gasoline prices such as refinery effects, because many gas stations within the same location would already face similar local input costs due to purchasing gasoline from the same (or closeby) refiners. In contrast, employing a nationwide spatial analysis in this paper has allowed for comparisons across locations, which have resulted in depicting the high contribution of refinery fixed effects (among others, as discussed, above) to the spatial gasoline-price dispersion.[5] Furthermore, considering gasoline price dispersion at the gas-station level across time and space within the U.S., we also show that time fixed effects (capturing crude oil price changes as well as other time-varying costs common across gas stations) contribute most, followed in descending order of magnitude by LDSL (i.e., the sum of the effects of refinery costs due to the station location, spatial demand, and land prices), legislative effects (i.e., state-level taxes), SDLC (i.e., number of competitors) and BCA (i.e., brand identities).

3. The Economic Environment

We would like to model the details of (i) the retail chain in the U.S. gasoline market (to the degree that our data allow us), (ii) price-setting behavior of gas stations, and (iii) the spatial distribution of economic activity among individuals. Regarding the retail chain in the U.S. gasoline market, at least 90% of the gasoline sold at retail-level gas stations is produced by refining crude oil (because up to 10% of the retail-level gasoline may be ethanol in the U.S.). The gasoline produced at the refinery is then either directly transported to gas stations by truck or to local distribution/bulk terminals by pipeline, barge or rail. In the latter case, local distribution terminals add some additives such as ethanol or brand-specific ingredients (e.g., Chevron including *Techron* into its gasoline) and sell the final version of gasoline to

[5] To further understand the effects of refiner fixed effects on the spatial distribution of gasoline prices, one may also consider the refinery-related price jumps in California in October 2012, where gasoline prices were as high as \$5 at the gas-station level, while the nationwide average was only about \$3.82. Please visit http://moneyland.time.com/2012/10/08/so-much-for-cheaper-gas-pump-prices-go-haywire-in-california/ for more details.

retail-level gas stations where transportation is mostly achieved by trucks. Studies such as Borenstein et al. (1997) and Kleit (2005) have depicted more details about the gasoline production and distribution within the U.S. Within this picture, since (i) our primary data set is for retail gasoline prices at the gas station level, (ii) our secondary data set is for the location of refineries, and (iii) we do not have any data on distribution terminals, we only model the production of gasoline at the refineries and the distribution of gasoline from refineries to the gas stations; i.e., we skip modeling local distribution terminals and include the possible injection of additives in the production function of gas stations. Accordingly, as in Chouinard and Perloff (2007) or Doyle and Samphantharak (2008), we assume that each gas station purchases gasoline from the nearest refiner.

We model the gasoline demand of individuals/producers in a spatial context. A typical individual s produces a particular amount of good s to sell; this is the only source of income for individual s. To match individuals with space, we also denote the location/space that is used for the economic activity of the individual by s. Individual s consumes gasoline and other goods purchased from other individuals around her. The transportation needs of individual s also depends on the distribution of economic activity around her; e.g., a dispersed city like Miami, FL may require more transportation than a compact city like San Francisco, CA.

Spatially distributed gas stations maximize their profits based on the gasoline demand of individuals, the cost of gasoline (charged by nearest refiners), and other spatial factors. Optimization implies that the price of gasoline at a particular gas station depends on spatial economic factors (such as the size and the distribution of economic activity around the gas station, the distance of the gas station to the refinery, number of other gas stations in the vicinity), the cost of gasoline at the refinery (reflecting the changes in oil prices), legislative factors (such as state-level taxes), local competition, and brand-specific costs/benefits.

Although the model is mostly static in technical terms, the time dimension comes into picture through state-level taxes, land prices, and time-varying crude oil prices. To keep the notation simple, we will skip the time dimension until the end of this section.

3.1. Individuals

The unique individual s at location $s \in \left[S_s^{lat}, S_s^{lon}\right]$, with S_s^{lat} and S_s^{lon} representing the latitude and longitude intervals, has the utility given by:

$$U_s \equiv \log C_s - \omega Y_s \tag{3.1}$$

where C_s is an aggregate index of consumption and ωY_s is the disutility due to producing Y_s amount of good s. Consumption of goods is combined with usage of gasoline that is assumed to depend on the spatial distribution of economic activity around individual s. Accordingly, C_s is defined by the following constant elasticity of substitution (CES) aggregator:

$$C_s = \left((Q_s)^{\frac{\eta-1}{\eta}} + (D_s)^{-\frac{1}{\eta}} (G_s)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

where Q_s is an aggregate index of goods, G_s is the consumption of gasoline, and D_s represents the spatial distribution of economic activity around individual/location s. Therefore, individual s has a higher preference toward gasoline (i.e., a higher value of D_s) if she is living in a neighborhood with a more dispersed economic activity.[6]

[6] This modeling approach considers distance-related costs in an indirect way through the spatial distribution of economic activity. An alternative approach may be to consider the distance-related disutility of individuals from going to any gas station in a direct way. However, there is no empirical evidence in the

Individual s sells her product Y_s to other individuals at a price of χ_s^s, which is further used to maximize utility subject to the following budget constraint:

$$C_s P_s \leq Y_s \chi_s^s \tag{3.2}$$

where P_s (to be further defined, below) is the price index for C_s. Utility maximization results in the following expression for the amount of good s produced:[7]

$$Y_s = \frac{1}{\omega} \tag{3.3}$$

The optimal allocation of expenditure yields the following demand functions for Q_s and G_s:

$$Q_s = \left(\frac{\chi_s}{P_s}\right)^{-\eta} C_s \tag{3.4}$$

and

$$G_s = D_s \left(\frac{R_s}{P_s}\right)^{-\eta} C_s \tag{3.5}$$

where χ_s and R_s are the prices of Q_s and G_s, respectively. It is implied that P_s is connected to χ_s and R_s as follows:

$$P_s \equiv \left((\chi_s)^{1-\eta} + D_s \left(R_s\right)^{1-\eta}\right)^{\frac{1}{1-\eta}} \tag{3.6}$$

Individual s can satisfy her aggregate gasoline demand of G_s from different gas stations around her through another CES aggregator:

$$G_s = \left(\sum_{g \in S_s} (N_s)^{-\frac{1}{\eta}} \left(G_s^g\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}} \tag{3.7}$$

literature for such a consideration; moreover, we do not have any data (that match individuals with gas stations) that can be used to test the implications in such a case. In sum, we believe that we have the best available modeling approach given the data and empirical evidence in the literature.

[7] This is also the expression one can get with a modeling strategy which endows each individual with $\frac{1}{\omega}$ unit of product.

where $g \in S_s$ represents the set of gas stations around individual s, G_s^g is the gasoline demand of individual s at gas station g (which is around individual/location s), and N_s is the number of gas stations around individual/location s. The optimal allocation of gasoline expenditure yields the following demand function of individual s at gas station g:

$$G_s^g = \frac{1}{N_s}\left(\frac{R^g}{R_s}\right)^{-\eta} G_s \tag{3.8}$$

$$= \frac{D_s X_s \left(R^g\right)^{-\eta}}{N_s}$$

where $X_s = (P_s)^\eta C_s$ is a measure of the *size/scale* of economic activity of individual s, and R^g is the retail price of gasoline (per gallon) at gas station g that satisfies:

$$R_s \equiv \left(\sum_{g \in S_s} \frac{\left(R^g\right)^{1-\eta}}{N_s}\right)^{\frac{1}{1-\eta}} \tag{3.9}$$

3.2. Gas Stations

Since we do not model local distribution terminals, we include the possible injection of additives (e.g., Chevron injecting *Techron* into its gasoline) in the production function of gas stations. In particular, gas station g of brand b located at location g produces Y^g gallons of gas according to:

$$Y^g = \frac{\left(K^g\right)^\alpha \left(B^g\right)^\phi \left(H^g\right)^{1-\alpha-\phi}}{A^g}$$

where K^g is the amount of gasoline purchased from the nearest refinery, B^g is the brand-specific additives included into the gasoline, H^g is the land/area, and A^g is the productivity which is assumed to be subject to local agglomeration/Marshallian externalities:

$$A^g = \frac{\left(\widetilde{G^g}\right)^\gamma \left(\widetilde{N^g}\right)^{1-\gamma}}{\alpha^\alpha \phi^\phi \left(1-\alpha-\phi\right)^{1-\alpha-\phi}}$$

where $\widetilde{G^g}$ represents the size and spatial distribution of economic activity around gas station g, $\widetilde{N^g}$ is the number of gas stations around gas station g, and γ and $(1-\gamma)$ are the parameters governing the strength of production externalities.[8] The gas station chooses K^g taking the price set by the nearest refinery as given. The cost minimization results in the following marginal cost of production:

$$M^g = \frac{\left(W_r\left(1+\tau_r^g\right)\right)^{\alpha}\left(b^g\right)^{\phi}\left(h^g\right)^{1-\alpha-\phi}}{\left(\widetilde{G^g}\right)^{\gamma}\left(\widetilde{N^g}\right)^{1-\gamma}} \tag{3.10}$$

where W_r is the price of gasoline charged by the nearest refinery r, τ_r^g represents transport cost of gasoline from the nearest refinery r to gas station g, b^g is the price of brand-specific ingredients/additives, and h^g is the price of land. On top of the marginal cost of production, gas station g also pays (net) gasoline taxes of t^g. Considering such additional costs, gas station g sets its price by maximizing its profit of:

$$\pi^g = \left(R^g - M^g\left(1+t^g\right)\right)Y^g$$

subject to the demand for gasoline at gas station g.

We need to match individuals with gas stations to calculate the demand for gasoline at gas station g. Accordingly, we will model individuals purchasing gasoline from gas stations around them. Although "around them" is an uncertain concept, we will define it in detail and provide alternative measures for it in the Appendix covering the details of the data. For now, we will simply consider that any individual s around gas station g will have a demand for gasoline at gas station g given by Equation 3.8. The number of individuals, however, will change with respect to the spatial measure of location considered. As a result, the total

[8]For simplicity, the denominator has been selected to get rid of the constant terms in the calculation of marginal cost, below.

demand is given by:

$$\sum_{s \in S_g} G_s^g = (R^g)^{-\eta} \sum_{s \in S_g} \frac{D_s X_s}{N_s}$$

where $s \in S_g$ represents the set of individuals around gas station g. Substituting this expression into the market clearing condition, we have $Y^g = \sum_{s \in S_g} G_s^g$, which implies that the optimal retail price of gasoline for gas station g can be obtained as:

$$R^g = \mu M^g (1 + t^g) \tag{3.11}$$

where M^g is already given in Equation 3.10, and $\mu \left(= \frac{\eta}{\eta - 1}\right)$ represents the gross markup.

3.3. Refiners

Refiner r produces Y_r units of gasoline according to a linear production function:

$$Y_r = O_r$$

where O_r is the amount of crude oil used as an input. The refinery chooses O_r taking its price as given. The cost minimization results in the following marginal cost of production:

$$M_r = W^o (1 + \tau_r^o) (1 - T) \tag{3.12}$$

where W^o is the price of crude oil at the oil field, τ_r^o represents net transport costs of crude oil from the oil field to the refinery, and T represents subsidies/transfers received. Since we do not have any price data for refiners, rather than complicating the model with unnecessary details, we simply assume that the refiner sets its price equal to its marginal cost:[9]

$$W_r = W^o (1 + \tau_r^o) (1 - T) \tag{3.13}$$

[9] In the empirical analysis below, we will use refiner fixed effects to capture any other detail (that is common through time) of the gasoline prices charged by refiners that have not been modeled here.

As in the case of gas stations, although we accept the location of refiners given in our empirical investigation, we assume that the location-selection problem of refiners has already been achieved in an upper-level optimization (that has resulted in a zero-profit condition under the simple assumption of free entry).

3.4. Closing the Model

For simplicity, we close the model by assuming that (i) gas stations and land are owned by refiners with equal shares among them (although they are managed independently so that they have their own optimization problems), and (ii) tax revenues are used to subsidize refiners. Accordingly, gas station profits, tax revenues, and land income are transferred to all refiners through T in Equation 3.12, where T, which is common across refiners, is chosen in such a way that the total amount of subsidies/transfers is equal to the total amount of gas station profits, tax revenues, and land income for each time period.

3.5. Time Dimension

The time dimension of the model comes from the price of crude oil W^o (which is an input for the refiner), taxes t^g (which can be different between legislative years), land prices h^g (subject to changes due to local markets that we do not model), gas station markups μ (determined on a daily basis, since marginal costs may change every day), and transfers to refiners T (capturing time-varying gas station profits, tax revenues, and land income). Accordingly, using Equations 3.10 and 3.13, the gasoline price expression for gas station g (i.e., Equation

3.11) can be rewritten as follows, this time with the time dimension:

$$R_t^g \quad = \quad \frac{\left(W_t^o\left(1+\tau_r^o\right)\left(1-T_t\right)\left(1+\tau_r^g\right)\right)^{\alpha}\left(b^g\right)^{\phi}\left(h_t^g\right)^{1-\alpha-\phi}\left(1+t_t^g\right)\mu_t}{\left(\widetilde{G^g}\right)^{\gamma}\left(\widetilde{N^g}\right)^{1-\gamma}} \tag{3.14}$$

This is the expression of which log version we estimate below, where $\widetilde{G^g}$ and $\widetilde{N^g}$ are assumed to be constant over time.

4. Empirical Analysis

This section depicts the estimation methodology, details of the variance decomposition analysis, and empirical results. We estimate the log version of Equation 3.14 by matching the details of the model with the data as detailed in the Appendix through Equations 7.1, 7.4, 7.5, and 7.6.

4.1. Estimation Methodology

Using the data definitions in the Appendix, the log gasoline prices can be written as follows:

$$\begin{aligned}\log R_t^g \quad = \quad & \underbrace{\alpha \log\left(W_t^o\left(1-T_t\right)\mu_t\right)}_{\text{Time Fixed Effects}} + \underbrace{\alpha \log\left(1+\tau_r^o\right)}_{\text{Refinery Fixed Effects}} + \underbrace{\alpha\delta \log\left(d_r^g\right)}_{\text{Distance to Refinery}} \\ & - \underbrace{\gamma\kappa \log\left(\sigma^g\right)}_{\text{S.D. of Nighttime Lights}} \underbrace{-\left(1-\gamma\right)\log\left(\widetilde{N^g}\right)}_{\text{\# of Nearby Stations}} + \underbrace{\phi \log\left(b^g\right)}_{\text{Brand Fixed Effects}} \\ & + \underbrace{\left(1-\alpha-\phi\right)\log\left(h_{t\in m}^g\right)}_{\text{Land Prices}} + \underbrace{\log\left(1+t_{t\in y}^{g,state,multiplicative}\right)}_{\text{State Taxes}} \\ & + \underbrace{\log\left(1+t_{t\in y}^{g,federal,multiplicative}\right)}_{\text{Federal Taxes}} + \underbrace{\varepsilon_t^g}_{\text{Residuals}}\end{aligned} \tag{4.1}$$

which can be estimated using time fixed effects to measure the time-varying costs of $\log\left(W_t^o\left(1-T_t\right)\right)$, refinery fixed effects to measure log transport costs of crude oil from the oil field to the refinery $\log\left(1+\tau_r^o\right)$, log distance to the refinery $\log\left(d_r^g\right)$ to measure transportation costs of gasoline from the refinery to the gas station, log standard deviation of nighttime lights $\log\left(\sigma^g\right)$ to measure the size and spatial distribution of economic activity around gas station g, log number of gas stations $\log\left(\widetilde{N^g}\right)$ around gas station g, brand fixed effects to measure brand-specific costs $\log\left(b^g\right)$, time-varying log Zillow Home Value Index at the zip-code level to measure land prices $\log\left(h_{t\in m}^g\right)$, time-varying state taxes calculated by Equation 7.5 for different years (e.g., 2010 versus 2011) to measure $\log\left(1+t_t^{state,multiplicative}\right)$, and federal taxes calculated by Equation 7.6 to measure $\left(1+t_t^{federal,multiplicative}\right)$; the unexplained part (i.e., residuals of ε_t^g) are assumed to represent the measurement errors. Since state and federal taxes have a coefficient of one in front of them, we use restricted least squares for the estimation.[10]

4.2. Variance Decomposition Analysis

Although the regression analysis depicts how the gasoline prices change with right hand side (RHS) variables, it cannot explain the contribution of each RHS variable on the variance of gasoline prices across gas stations, which is a natural measure of gasoline price dispersion. Note that the variance decomposition analysis also depicts the percentage of total sum of squares explained by each RHS variable; i.e., it decomposes the contribution of each RHS variable on the R-squared measure. Accordingly, using the estimated parameters and the

[10] It is important to emphasize that we cannot identify the individual parameters of α, δ or ϕ, while we can identify γ and κ. Nevertheless, their identification is not necessary for the variance decomposition analysis, below, where all we need is the overall fitted value of each component on the right hand side of Equation 4.1.

fitted values coming from the estimation of Equation 4.1, we will conduct a variance decomposition analysis where we will further investigate the role of each explanatory variable on the variance of gasoline prices across gas stations as follows:

$$
\begin{aligned}
var\left(\log R_{s,t}^{g}\right) \quad = \quad & \underbrace{cov\left(\alpha \log\left(W_t^o\left(1+T_t\right)\mu_t\right), \log R_t^g\right)}_{\text{Time Fixed Effects}} + \underbrace{cov\left(\alpha \log\left(1+\tau_r^o\right), \log R_t^g\right)}_{\text{Refinery Fixed Effects}} \\
& + \underbrace{cov\left(\alpha\delta \log\left(d_r^g\right), \log R_t^g\right)}_{\text{Distance to Refinery}} - \underbrace{cov\left(\gamma\kappa \log\left(\sigma^g\right), \log R_t^g\right)}_{\text{S.D. of Nighttime Lights}} \\
& - \underbrace{cov\left(\left(1-\gamma\right)\log\left(\widetilde{N^g}\right), \log R_t^g\right)}_{\text{\# of Nearby Stations}} + \underbrace{cov\left(\phi \log\left(b^g\right), \log R_t^g\right)}_{\text{Brand Fixed Effects}} \\
& + \underbrace{cov\left(\left(1-\alpha-\phi\right)\log\left(h_{t\in m}^g\right), \log R_t^g\right)}_{\text{Zillow Home Value Index}} \\
& + \underbrace{cov\left(\log\left(1+t_{t\in y}^{g,state,multiplicative}\right), \log R_t^g\right)}_{\text{State Taxes}} \\
& + \underbrace{cov\left(\log\left(1+t_{t\in y}^{g,federal,multiplicative}\right), \log R_t^g\right)}_{\text{Federal Taxes}} + \underbrace{cov\left(\varepsilon_t^g, \log R_t^g\right)}_{\text{Residuals}}
\end{aligned}
$$

where var is the operator of variance and cov is the operator of covariance.

4.3. Empirical Results

The estimation results of Equation 4.1 are given in Table 2 where different spatial measures (as introduced in the Data Appendix) have been considered. In all regressions, brand, refinery, and time fixed effects have been included. As is evident, local agglomeration externalities of spatial gasoline demand (measured by the standard deviation of nighttime lights) and the number of local competitors have negative and significant effects on gasoline prices as expected by the model. In terms of their magnitudes, one percent of an increase in local externalities can lower gasoline prices as much as 1.92%, while one percent of an increase in the number of competitors can lower gasoline prices as much as 0.19%. Distribution costs of gasoline from the nearest refiner to the gas station have positive and significant effects, also as expected; one percent of an increase in distribution costs leads to an increase up to about 0.41% in gasoline prices. The effects of land prices are also positive and significant (as expected) where one percent of an increase in land prices leads to an increase up to about 0.79% in gasoline prices. Also considering the unit coefficients in front of state and federal taxes (through restricted least squares), the R-squared values are all about 0.90, which is promising.

Although an R-squared value of 0.90 is a high measure of fit, we would like to know the contribution of each RHS variable on this value. Accordingly, we depict the variance decomposition of log gasoline prices (across time and space) in Table 3 where the percentage contribution of each RHS variable is depicted for different spatial measures. As is evident, the highest contribution to the gasoline price dispersion of about 51% is by time fixed effects, followed by refinery fixed effects contributing up to 33%, state taxes contributing about 12%, spatial demand contributing up to 2%, land prices contributing up to about 1.4%, the number of local competitors contributing up to 0.3%, and distribution costs contributing up to 0.1%. The contribution of brand fixed effects on the gasoline price dispersion, however, is very minor, ranging between -0.5% and -0.1% that can be considered as a smoothing effect of discount gasoline stations. Since additive federal taxes were common across all gas stations, their multiplicative version has another smoothing effect on gasoline price dispersion; federal taxes reduce gasoline price dispersion by more than 7%.

Overall, since time fixed effects can be thought as nationwide effects (because they capture any effect common across all gas stations within the U.S.), we can claim that about half of the gasoline price dispersion (across time and space) is due to nationwide effects, and the remaining half is due to spatial factors.

5. Discussion and Robustness

This section questions the RHS variables that have been included in the regression analysis, provides alternative RHS measures for robustness, and discusses the empirical results through comparisons with the existing literature.

Although the model implies that we should be including all the RHS variables in the regression analysis, is their inclusion statistically significant? To answer this question, for each spatial measure, we used several F-tests to compare the full version of the model with its restricted versions where we excluded one set of RHS variables (e.g., spatial gasoline demand, brand fixed effects or time fixed effects) at a time. Under the null hypothesis that the restrictions are valid, the F-test results suggested that $p-$values are virtually equal to zero; hence, all restrictions were rejected, or the full version of the model was accepted as the correct version of the model.

According to the empirical results, the gasoline price dispersion is mostly determined by time fixed effects. Although they have been included in the analysis to capture the effects of crude oil prices, they may be capturing any other time varying effect that is common across all gas stations as well (e.g., weekend, weekday, or before-holiday pricing strategies). When we compared estimated time fixed effects with daily crude oil prices, we obtained a correlation coefficient of about 0.87 (for all five spatial measures that we considered); hence, time fixed effects in fact mostly capture crude oil prices during the sample period. The difference between time fixed effects and crude oil prices can mostly be attributable to the transmission of crude oil prices to retail-level gasoline prices (or daily pricing strategies) as we discussed above. Nevertheless, for robustness, we had an alternative analysis where we replaced time fixed effects with crude oil prices. In such a case, the contribution of crude oil prices to the gasoline price dispersion (through our variance decomposition analysis) has been estimated as about 38% or 39% (depending on the spatial measure considered) where the difference between the contribution of time fixed effects (about 51%) and crude oil prices has shown up in the residuals (of which contribution increased from about 10% to about 22%).[11] Therefore, replacing time fixed effects with crude oil prices does not alter the result that the main contribution to the overall gasoline price dispersion during our sample period is by crude oil prices. This result is consistent with Chouinard and Perloff (2007) who show using a panel of monthly observations at the state level for the 1989–1997 period that almost the entire variation in national gasoline prices is due to a rise in the price of crude oil and that across-state variation can be attributed to demand variation, as well as differences in taxes. We also show that differences in spatial demand and taxes are important in explaining gasoline price dispersion, however, in addition to Chouinard and Perloff (2007), we have shown that refinery fixed effects also have significant effects (e.g., about 1/3 of the gasoline price dispersion is

[11] Compared to Table 3, the contributions of other RHS variables on the gasoline price dispersion were almost the same when time fixed effectes were replaced by daily crude oil prices.

due to refinery fixed effects).

Compared to the studies focusing on gas-station level gasoline prices (that we discussed during the literature review, above), showing the contribution of refinery fixed effects to the gasoline price dispersion is the main nuance of this paper. The main reason for other papers ignoring such effects were either due to the lack of a comprehensive data set and/or the lack of modeling such effects. Nevertheless, the high contribution of refinery fixed effects in this paper can be connected to studies for other countries at more aggregated levels such as the one by Sen (2003) who shows using a panel data set of monthly city-level prices for 11 Canadian cities over the 1991–1997 period that wholesale prices (captured by refinery fixed effects in this paper) play a larger role in retail price determination than city level market structure (captured by spatial demand or land prices in this paper).

Since the number of gas stations around gas station g measured by $\widetilde{N^g}$ in our gasoline price data set may not be capturing the actual number of gas stations, as an alternative, we used the average nighttime lights m_s to proxy for $\widetilde{N^g}$ as discussed, above. The corresponding analysis suggested that although the average nighttime lights a have negative and significant coefficient estimate, its contribution to the gasoline price dispersion is still at most about 1.5%. Moreover, although their coefficient estimates were found to be significant, the contributions of brand fixed effects and spatial demand to the gasoline price dispersion were found to be minor (in both benchmark and alternative analyses) as well. However, recall that the studies focusing on gas-station level gasoline prices (mentioned above) have mostly emphasized the effects of the brand of gasoline (measured by brand fixed effects in this paper), spatial gasoline demand, and local competition on the gasoline price dispersion. One reason for this difference may be the focus on local or small markets by such other studies, because, by focusing on the overall U.S. market, this paper can capture the big picture of the gasoline market. To be more specific, consider a special case of our model where we would focus on a very local market for a specific day where all gas stations purchase gasoline from the very same refinery. In such a case, gas stations would have very similar gasoline prices with each other, because they would be subject to very similar time fixed effects, refinery fixed effects, state taxes, land prices, and gasoline distribution costs; for sure, the only remaining factors to capture the price dispersion between these local gas stations would then be the brand fixed effects, spatial demand, and local competition. Therefore, focusing on the overall U.S. market is in fact the key here.

6. Conclusion

Using a unique comprehensive gasoline price data at the gas-station level, this paper has achieved a spatial analysis to explain the determinants of the gasoline prices and their dispersion across space and time within the U.S.. The results suggest that about half of the gasoline price dispersion (across time and space) is due to nationwide effects, and the remaining half is due to spatial factors. The change in gasoline prices through time is mostly explained by crude oil prices, and the spatial gasoline price dispersion is mostly explained by refinery fixed effects followed by state-level taxes.

Compared to the existing literature that emphasize the role of gasoline brands and local density of stations on gasoline prices, we have shown that the contribution of such effects are relatively minor after controlling for other/spatial determinants of gasoline prices (e.g., spatial gasoline demand, distribution costs of gasoline, and land prices). The main reason for the results of this paper being different compared to the existing literature is due to having a nationwide gasoline price data within the U.S., whereas existing studies (at the gas station level) for the U.S. have mostly focused on local markets with only a couple of cities/states involved in the spatial analysis. As discussed in more details in the text, having such a local-level analysis would suppress the main determinants of gasoline prices, such as refinery effects, because many gas stations within the same location would already face similar local input costs due to purchasing gasoline from the same (or closeby) refiners. In contrast, employing a nationwide spatial analysis in this paper has allowed for comparisons across locations, which have resulted in depicting the high contribution of refinery fixed effects (among others, as discussed, above) to the spatial gasoline-price dispersion.

Our analysis is not without caveats. Due to the lack of available data, unlike Ning and Haining (2003), Barron et al. (2004), Eckert and West (2004, 2005), or Hosken et al. (2008), the model of this paper has not considered any gas station characteristics (e.g., having a car wash or repair service) either. Also, we have mostly focused on the spatial (rather than a time series or a "full" panel) dimension of gasoline price determination due to the unbalanced panel nature of the gasoline price data, and we have not considered any dynamics, such as spatial variables interacting with time fixed effects in the regression analysis; considering such interactions through appropriate modeling for unbalanced panels would result in richer dynamics, although they were not the focus of this paper.

References

[1] Atkinson, B (2008) On retail gasoline pricing websites: potential sample selection biases and their implications for empirical research. Review of Industrial Organization 33: 161–175.

[2] Atkinson, B (2009) Retail gasoline price cycles: evidence from Guelph, Ontario using bi-hourly, station-specific retail price data. Energy Journal 30: 85–109.

[3] Barron, JM, Taylor, BA and Umbeck, JR (2004) Number of sellers, average prices and price dispersion. International Journal of Industrial Organization 22: 1041–1066.

[4] Bernard, A, Eaton, J, Jensen, JB, Kortum, S (2003). Plants and Productivity in International Trade. American Economic Review 93(4): 1268-1290.

[5] Borenstein, S, Cameron, AC, Gilbert, R (1997) Do gasoline prices respond asymmetrically to crude oil price changes? Quarterly Journal of Economics 112: 305–339.

[6] Byrne, DP, Leslie, GW, Ware, R. (2015) How do Consumers Respond to Gasoline Price Cycles? Energy Journal 36(1): 115-147.

[7] Chandra, A, Tappata, M (2011) Consumer search and dynamic price dispersion: an application to gasoline markets. RAND Journal of Economics, 42(4), 681-704.

[8] Chouinard, HH, Perloff, JM (2007) Gasoline price differences: taxes, pollution regulations, mergers, market power, and market conditions. The B.E. Journal of Economic Analysis and Policy 7(1), Article 8.

[9] Clemenz, G, Gugler, K (2006) Locational choice and price competition: some empirical results for the Austrian retail gasoline market. Empirical Economics 31: 291–312.

[10] Croft, TA (1978) Night-time Images of the Earth from Space. Scientific American, 239: 68-79.

[11] Doll, CNH, Muller, JP, Morley, JG (2006) Mapping Regional Economic Activity from Night-time Light Satellite Imagery. Ecological Economics 57(1): 75-92.

[12] Doyle, JJ, Samphantharak, K (2008) $2.00 Gas! Studying the effects of a gas tax moratorium. Journal of Public Economics 92: 869–884

[13] Ebener, S, Murray, C, Tandon, A, Elvidge, CD (2005) From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. International Journal of Health Geographics, 4(5): 1-17.

[14] Eckert, A (2013) Empirical Studies of Gasoline Retailing: A Guide to the Literature. Journal of Economic Surveys. Journal of Economic Surveys, 27(1): 140-166.

[15] Eckert, A and West, DS (2004) A tale of two cities: price uniformity and price volatility in gasoline retailing. Annals of Regional Science 38: 25–46.

[16] Eckert, A and West, DS (2005) Price uniformity and competition in a retail gasoline market. Journal of Economic Behavior and Organization 56: 219–237.

[17] Edelstein, P, and Kilian, L (2009). How Sensitive are Consumer Expenditures to Retail Energy Prices? Journal of Monetary Economics 56(6): 766-779.

[18] Elvidge, CD, Baugh, KE, Kihn, EA, Kroehl, HW, Davis, ER, Davis, CW (1997). Relation between satellite observed visible-near infrared emissions, population, and energy consumption. International Journal of Remote Sensing, 18: 1373-1379.

[19] Espey, M (1998) Gasoline demand revisited: an international meta-analysis of elasticities, Energy Economics, 20: 273–295.

[20] Foote, CL, Little, JS (2011). Oil and the Macroeconomy in a Changing World: A Conference Summary. Federal Reserve Bank of Boston, Public Policy Discussion Papers.

[21] Ghosh, T, Anderson, S, Powell, RL, Sutton, PC, Elvidge, CD (2009). Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagery. Remote Sensing, 1(3): 418-444.

[22] Hamilton, JD (1983) Oil and the Macroeconomy Since World World II. Journal of Political Economy. 91: 228-48.

[23] Hosken, DS, McMillan, RS, Taylor, CT (2008) Retail gasoline pricing: what do we know? International Journal of Industrial Organization 26: 1425–1436.

[24] Houde, J.-F. (2012) Spatial differentiation and vertical mergers in retail markets for gasoline. American Economic Review, 102(5): 2147-82.

[25] Kilian, L (2010). Explaining Fluctuations in U.S. Gasoline Prices: A Joint Model of the Global Crude Oil Market and the U.S. Retail Gasoline Market. Energy Journal 31(2): 87-104.

[26] Kleit, AN (2005) The economics of gasoline retailing: petroleum distribution and retailing issues in the U.S. Energy Studies Review 13: 1–28.

[27] Manuszak, MD (2010) Predicting the impact of upstream mergers on downstream markets with an application to the retail gasoline industry. International Journal of Industrial Organization 28: 99–111.

[28] Ning, X, Haining, R (2003) Spatial pricing in interdependent markets: a case study of petrol retailing in Sheffield. Environment and Planning A 35: 2131–2159.

[29] Noel, MD (2007) Edgeworth price cycles: evidence from the Toronto retail gasoline market. Journal of Industrial Economics 55: 69–92.

[30] Pennerstorfer, D (2009) Spatial price competition in retail gasoline markets: evidence from Austria. Annals of Regional Science 43: 133–158.

[31] Sen, A (2003) Higher prices at canadian gas pumps: international crude oil prices or local market concentration? An empirical investigation. Energy Economics 25: 269–288.

[32] Shepard, A (1993) Contractual form, retail price, and asset characteristics in gasoline retailing. RAND Journal of Economics 24: 58–77.

[33] Sutton, PC, Costanza, R (2002) Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. Ecological Economics, 41(3): 509-27.

[34] Sutton, PC, Elvidge, CD, Ghosh,T (2007) Estimation of Gross Domestic Product at Sub-National Scales Using Nighttime Satellite Imagery. International Journal of Ecological Economics and Statistics, 8(S07): 5-21.

[35] Van Meerbeeck, W (2003) Competition and local market conditions on the Belgian retail gasoline market. De Economist 151: 369–388.

7. Appendix: Data

The details of the data required to estimate the log version of 3.14 are depicted in this section which literally connects Equation 4.1 to Equation 3.14 in the main text.

7.1. Data for Prices

We collected the regular gasoline prices per gallon R_t^g covering 50 states of the U.S., together with District of Columbia, from www.gasbuddy.com on a daily basis between September 10th, 2010 and January 31st, 2011 (i.e., for 144 days). GasBuddy.com is a network of web sites in the United States that operates under many local domain names such as MiamiGasPrices.com or NewYorkGasPrices.com. Individuals post the gasoline prices that they have seen around them to these web sites; in return, they earn points that can be used to enter raffles for prizes such as $250 gas cards. Atkinson (2008) has cross checked the validity of the price data collected from GasBuddy.com by using other independent gasoline price data and concluded that the data obtained from GasBuddy.com can in fact be used to examine spatial price competition, both within and across markets.

We downloaded the data at around 8pm each and every day during the sample period. This is the first paper introducing and using this data set. These are regular gasoline prices at the gas station level where the location information is as good as the actual address of the gas station in the U.S..[12] The brand of the gas station is also available (which we use to create brand fixed effects to measure b_s^g's in the regression analysis). A typical gasoline price observation that we have is a price of \$2.73 on October 15th, 2010 from the gas station brand of "Murphy USA" located at "8922 N Terre Haute Rd, Paris, IL". Since the data on the web page have been submitted by actual individuals, there were some problems, such as typos in the addresses, or digit errors made while typing the prices. For the former, we have filtered the data by matching the addresses with the Google Map addresses; when they matched (i.e., when the actual address could be found by Google Map with an accuracy of 7, 8 or 9, where 9 is the best level of accuracy at the exact location name and address level, 8 is the accuracy at the address level, and 7 is the accuracy at the level of intersection of roads according to the standards of the Google Map system), we kept the price observation and further obtained the latitude and longitude, together with zip codes (which we use to match the gasoline price data with land prices), of each and every gas station in our sample. For the latter, we filtered the price data by ignoring the price observations that are at least two times higher or lower than the average gas prices for a specific day. After such a data cleaning process, we ended up with 469,408 (i.e., close to half a million) price observations.

[12] In other contexts, studies such as by Byrne et al. (2015) have also used daily gasoline price data sets for a couple of cities obtained from similar web pages, but the data set of this paper is the most comprehensive one to our knowledge, because rather than just a couple of cities, it covers gas stations from all states of the U.S..

These gasoline price observations belong to 38,245 different gas stations; hence, on average, we have about 12 days of gasoline price observation per gas station. The spatial distribution of gas stations can also be found on a map of the U.S. in Figure 1.

The descriptive statistics of the gasoline price data are given in Table 1 where details at the state-level are provided for presentational purposes; these are the numbers providing the motivation of this paper. The left panel of Table 1 (after the first column) provides information on gasoline prices for the pooled sample of all locations and all time periods. Accordingly, during the sample period, the highest price has been recorded as \$4.26 from Washington D.C. and the lowest price has been recorded as \$2 from various states. The median gasoline price has been \$2.79. The highest (lowest) standard deviation has been recorded as 0.25 (0.05) from Washington D.C. (Utah), followed by 0.21 (0.08) from Alaska (Montana). Among the minimum (maximum) prices across states, the standard deviation is 0.28 (0.30), while, among the median (mean) prices across states, the standard deviation is 0.19 (0.18).

The middle panel of Table 1 provides further information on the price dispersion across states. In particular, the maximum gasoline price difference between any two gas stations is calculated for each state and for each day as follows:

$$\overset{state}{\overleftrightarrow{R_t^g}} = \max\left(R_t^g \mid state\right) - \min\left(R_t^g \mid state\right)$$

Then, across time, the other statistics are calculated according to $\overset{state}{\overleftrightarrow{R_t^g}}$; e.g., the column *min* for any state is calculated by finding the minimum value of $\overset{state}{\overleftrightarrow{R_t^g}}$ across t. Therefore, the maximum price difference for a typical state in a typical day was as high as \$1.57 (for Washington D.C.) and as low as \$0 (for Alaska). The maximum price difference on a typical day, which can be measured for any given state by the median of the maximum price difference across the whole sample is also given in Table 1, where Hawaii has the highest value of \$0.60, while states such as Florida, Minnesota, New Hampshire, New Jersey and Wisconsin share the lowest value of \$0.08. Such differences across states are also shown in Figure 2 on the U.S. map for the maximum price difference on a typical day. As is evident, Mountain and West North Central States (defined according to Census Regions and Divisions of the United States) have higher price differences across gas stations. According to Table 1, the standard deviation (across time) of the maximum price difference was as high as 0.60 (for Hawaii) and as low as 0.09 (for New Hampshire). In sum, there were high levels of gasoline price dispersion, even within states/districts, during the sample period.

7.2. Data for Economic Activity and Spatial Measures

To capture the part of the local agglomeration externalities through $\widetilde{G^g}$ in Equation 3.14 (that have negative effects on gas-station-level gasoline prices), since we have a spatial analysis of gasoline prices, we need corresponding spatial data for the *size* and *spatial distribution* of economic activity around gas station g. Accordingly, we use the standard deviation of the nighttime lights around gas station g as a proxy for $\widetilde{G^g}$:

$$\widetilde{G^g} = (\sigma^g)^\kappa \tag{7.1}$$

where σ^g represents the standard deviation of nighttime lights (across space) around gas station g. It is important to emphasize that standard deviation is a measure that can capture both the size/scale and the distribution of nighttime lights[13]; hence, it is a perfect match to

[13] Because, as is well known in statistics, standard deviation is a measure that captures both the scale and the variation in a distribution. If one would like to have a pure measure of spatial distribution of

capture the expected negative effects of $\widetilde{G^g}$. Our motivation comes out of many studies that have shown the almost-perfect correlation between economic activity and nighttime lights (e.g., see Croft 1978, Elvidge et al. 1997, Sutton and Costanza 2002, Ebener et al. 2005, Doll et al. 2006, Sutton et al. 2007, and Ghosh et al. 2009, among many others). However, this is the first paper we are aware of that uses the spatial distribution of nighttime lights data to measure the spatial distribution of economic activity. The average (rather than standard deviation of) nighttime lights would definitely be a measure of economic activity, but it cannot capture the spatial distribution of economic activity; we rather save the average measure of nighttime lights to measure the economic activity of individuals and for a robustness analysis, where we use it as an alternative proxy for the number of gas stations.

To capture the *size* of economic activity of individual s, X_s, we use the nighttime light of location s (i.e., the location of individual s) as a proxy:

$$X_s = (n^s)^\kappa \tag{7.2}$$

where n_s is the nighttime light of location s. Similarly, to capture the *spatial distribution* of economic activity around individual s, D_s, we use the coefficient of variation (defined as standard deviation over mean) of the nighttime lights around individual s as a proxy:

$$D_s = \left(\frac{\sigma^s}{m^s}\right)^\kappa \tag{7.3}$$

where σ^s represents the standard deviation and m^s represents the average nighttime lights (across space) around individual s.

economic activity (that is independent of the size/scale), one should better consider a measure like coefficient of variation defined as the standard deviation over mean, which is accepted to be a normalized measure of dispersion.

In technical terms, the nighttime lights are measured for each 30-arc-second grids, spanning -65 to 75 degrees latitude and -180 to 180 degrees longitude around the world; obviously, we only use the part of the nighttime lights for the U.S. To give the reader a better idea, although the exact measure is based on the exact spatial location (due to the spherical shape of the earth), on average, one 30-arc-second grid correspond to a space about $25 (= 5 \times 5)$ square blocks or $0.25\,(= 0.5 \times 0.5)$ square miles in the U.S.; therefore, these 30-arc-second grids are natural minimum measures in our analysis to create different spatial location concepts. Accordingly, to connect the data to the model, we assume that each 30-arc-second grid is occupied by a potential individual, say, individual s in the model. In Figure 3, an approximate typical example of one 30-arc-second grid in Chicago, IL is given by the area/space between "166 W Washington St, Chicago, IL 60602" and "520 S Michigan Ave, Chicago, IL 60605" which corresponds to the location $s \in \left[S_s^{lat}, S_s^{lon}\right]$ where $S_s^{lat} = [41.8753797, 41.8834904]$ represents the latitude interval and $S_s^{lon} = [-87.6330631, -87.6245221]$ represents the longitude interval of the area between the two addresses.

In the model, it is the individuals who travel for economic activity and purchase gasoline. Although we match each 30-arc-second grid with a potential individual, we also need to match the travel patterns of individuals with spatial measures of locations to identify the phrases that we have been using so far, "around individual s" and "around gas station g". Accordingly, the spatial location measures we consider in the empirical analysis is to capture the travel patterns of individuals. For example, if we would use only one 30-arc-second grid as our spatial measure, "around individual s" or "around gas station g" would correspond to 25 square blocks (or 0.25 square miles) in the U.S. on average, and if we would use hundred

30-arc-second grids as our spatial measure, they would correspond to $2,500 (= 50 \times 50)$ square blocks (or 25 square miles) in the U.S. on average. When "around individual s" and "around gas station g" intersect with each other on the map, it means that individual s has a demand for gasoline at gas station g given by Equation 3.8; hence, the gas stations located within the travel patterns of individuals are potential suppliers of gasoline.

In formal terms, we consider five different measures of spatial location.

Spatial Measure #1: Since we compute the spatial distribution of economic activity by the coefficient of variation of nighttime lights across 30-arc-second grids in Equation 7.1, we need nighttime light data from several grids to obtain a standard deviation measure. So, to find the intersection of "around individual s" and "around gas station g" on the map, for each gas station g, after finding the 30-arc-second grid at which gas station g is located (by matching the latitude and longitude of the gas station with the same information from the nighttime lights data), we consider 5 additional 30-arc-second grids (about 2.5 miles) to the north, south, west, and east of the gas station's grid. This would correspond to a space consisting of 121 30-arc-second grids around gas station g. It is also important to emphasize that we have the nighttime light data for every 30-arc-second grid; hence, we have enough number of observations to calculate the standard deviation of nighttime lights σ^g across these grids.

Spatial Measure #2: A larger version of Spatial Measure #1, this time with 10 additional 30-arc-second grids (about 5 miles) to the north, south, west, and east of the gas station's 30-arc-second grid. This would correspond to a space consisting of 441 30-arc-second grids or 110 square miles around the gas station considered.

Spatial Measure #3: A larger version of Spatial Measure #1, this time with 25 additional 30-arc-second grids (about 12.5 miles) to the north, south, west, and east of the gas station's 30-arc-second grid. This would correspond to a space consisting of 2,601 30-arc-second grids or 650 square miles around the gas station considered.

Spatial Measure #4: A larger version of Spatial Measure #1, this time with 50 additional 30-arc-second grids (about 25 miles) to the north, south, west, and east of the gas station's 30-arc-second grid. This would correspond to a space consisting of 10,201 30-arc-second grids or 2,550 square miles around the gas station considered.

Spatial Measure #5: A larger version of Spatial Measure #1, this time with 100 additional 30-arc-second grids (about 50 miles) to the north, south, west, and east of the gas station's 30-arc-second grid. This would correspond to a space consisting of 40,401 30-arc-second grids or 10,100 square miles around the gas station considered.

One may be curious about why we have considered a spatial measure up to about 10,100 square miles around the gas station. Our motivation comes from the National Household Travel Survey (for the year of 2009) that has collected data on both long-distance and local travel by the American public. According to this survey, on average, an individual has traveled (one-way) about 36 miles per day, and an individual has driven (one-way) about 29 miles per day. Since an individual has the option to purchase gasoline from any location on her (one-way) route, considering spatial measures up to 10,100 square miles (i.e., one-way routes up to 50 miles) around the gas station just covers the demand patterns of a typical gasoline consumer in the U.S.

The nighttime lights data have been obtained from the web page of the National Geophysical Data Center.[14] The data set is the part of "Version 4 DMSP-OLS Nighttime Lights

[14]The nighttime light files are cloud-free composites made using all the available archived DMSP-OLS

Time Series" for 2010. Within this data set, we use two different measures of nighttime lights: (i) Average Visible, Stable Lights, and Cloud Free Coverages; and (ii) The nighttime lights product known as "avg_lights_x_pct" which is derived from the average visible band digital number of cloud-free light detections multiplied by the percent frequency of light detection.[15] The empirical results that we depict have been obtained by using the latter; but, the results were almost exactly the same when we used the former.

7.3. Data for Number of Gas Stations

To capture the part of the local agglomeration externalities through $\widetilde{N^g}$ in Equation 3.14 (that have negative effects on gas-station-level gasoline prices), we have calculated the number of gas stations $\widetilde{N^g}$ around gas station g (in Equation 3.14) as the number of gas stations for which we have price data at least once during our sample period of 144 days. The number of gas stations for each state/district (for the pooled sample through time) is given in the first column of Table 1 where the numbers range between 1,609 (for Ohio) and 88 (for Washington D.C.) with a total of 38,245. The number of gas stations per day ranges between 3,480 and 3,840.

Many earlier studies have also investigated the contribution of local station density (measured by $\widetilde{N^g}$ in this paper) by considering different measures of spatial locations. For example, Hosken et al. (2008) and Barron et al. (2004) have considered the number of stations within

smooth resolution data for the calendar year of 2010.

[15] In the latter, the inclusion of the percent frequency of detection term normalizes the resulting digital values for variations in the persistence of lighting. For instance, the value for a light only detected half the time is discounted by 50%. Note that this product contains detections from fires and a variable amount of background noise. This is the product used to infer gas flaring volumes from the nighttime lights.

a 1.5 mile radius, while Eckert and West (2005) have considered a 2 kilometer radius, and Van Meerbeeck (2003) has considered the number of stations within the same municipality. To be consistent with our spatial measures introduced above, we calculated the number of gas stations $\widetilde{N^g}$ in our sample (by using their latitudes and longitudes from Google Map) depending on our different spatial measures, above. Therefore, compared to the existing literature, we focus on more than one spatial measure to calculate the number of station.

We are well aware that $\widetilde{N^g}$ (i.e., number of gas stations for which we have price data at least once during our sample period of 144 days) may not be capturing the actual number of gas stations around gas station g; hence, as an alternative, in the robustness/regression analysis, we also used the average nighttime lights m^g around gas station g as a proxy for $\widetilde{N^g}$ because of the trivial conjecture that the number of gas stations is correlated with the economic activity around gas station g.

7.4. Data for Refiners

Since we haven't modeled the petroleum bulk stations and distribution terminals (for simplicity, because of the lack of available location data for such stations/terminals, and because of the lack of data on possible contracts between terminals and gas stations), we assumed that gas stations purchase gasoline from the nearest refinery as in Chouinard and Perloff (2007) or Doyle and Samphantharak (2008). We considered the possibility of gas stations purchasing gasoline from Canadian refiners as well due to the open nature of trade between the two countries, as stipulated by the North American Free Trade Agreement (NAFTA). Accordingly, we obtained the list of gasoline refineries in the U.S. and Canada from www.globalenergyobservatory.org. Then, we found the exact address, together with latitude and longitude, of each and every refinery by using various sources on the internet including Google Maps, Google Earth, and individual refinery web pages. Using such location information of refineries, we found the nearest refinery for each and every gas station in our sample (i.e., we identified r in W_t^r) and calculated the corresponding great-circle distance of transportation of gasoline from the nearest refinery to each gas station (d_r^g) to be connected to the gasoline transportation costs $(1 + \tau_r^g)$ as follows:

$$(1 + \tau_r^g) = (d_r^g)^\delta \tag{7.4}$$

where δ is the elasticity of distance to capture the relation between distance and gasoline transportation costs. The spatial distribution of refiners can also be found on a map of the U.S. in Figure 1 where there are 143 refiners from the U.S. and 4 refiners from Canada; Canadian refiners are the nearest refiners to some gas stations (in our sample) in Northeastern United States. Note that we also use refinery fixed effects in the regression analysis, so that possible border effects (if any) caused by purchasing gasoline from Canadian refiners are captured by such fixed effects.

7.5. Data for Land Prices

The land prices (h_t^g's in Equation 3.14) were proxied by the Zillow Home Value Index (ZHVI) at the zip-code level. These monthly data have been obtained from www.zillow.com for the sample period of the gasoline price data. We believe that ZHVI is one of the best available proxies for land prices because of its spatial coverage (e.g., Zillow appraises about 75% of the houses in the U.S. several times a week and aggregates these house-level valuations into indexes at the ZIP code level).[16] To match the data with the model, we used the following

[16] It is important to notice that a house need not to be sold to be included in ZHVI.

expression:

$$h^g_{t\in m} = ZHVI_m$$

where $h^g_{t\in m}$ is the land price faced by gas station g at day t that belongs to month m (e.g., September, 2010), and $ZHVI_m$ is the Zillow Home Value Index for month m. In other words, the time dimension of land prices are due to changes in monthly home prices.

7.6. Data for Taxes

Since we have the address of each and every gas station, it was trivial to identify the state in which the gas station operates and to obtain gasoline taxes at the state level from www.taxfoundation.org for both 2010 and 2011. The state-level taxes per gallon of gasoline, which include taxes paid both to the state governments and other local governments, are given in the right panel of Table 1 where they range between 8 and 46.6 cents for 2010 and between 8 and 49.6 cents for 2011. The standard deviation of taxes across states is 8.5 for 2010 and 9.5 for 2011. The federal tax on gasoline (which is common across states) for both 2010 and 2011 was 18.4 cents per gallon.

However, the obtained gasoline taxes (at both state and federal levels) are in additive (rather than multiplicative) terms. Hence, to connect the data to our model (where taxes are in multiplicative terms), we have to use a transformation of taxes from additive to multiplicative form. Accordingly, for state-level taxes, we used the following transformation for any gas station:

$$t_t^{state,multiplicative} = \frac{t_t^{state,additive}}{R_t^g - t_t^{federal,additive} - t_t^{state,additive}} \tag{7.5}$$

and, for federal taxes, we used the following transformation:

$$t_t^{federal,multiplicative} = \frac{t_t^{federal,additive}}{R_t^g - t_t^{federal,additive} - t_t^{state,additive}} \tag{7.6}$$

where both $t_t^{state,multiplicative}$ and $t_t^{federal,multiplicative}$ may differ across gas stations (even within any state), since they represent taxes as percentage of retail gasoline prices (that may be different across gas stations) for a specific day. Finally, to connect the model to the multiplicative state-level and federal taxes, we used the following expression for gas station g:

$$\left(1 + t_{t\in y}^{g}\right) = \left(1 + t_{t\in y}^{g,state,multiplicative}\right)\left(1 + t_{t\in y}^{g,federal,multiplicative}\right)$$

where $t_{t\in y}^{g,state,multiplicative}$ and $t_{t\in y}^{g,federal,multiplicative}$ are the tax rates faced by gas station g at time t that belongs to year y.

An alternative would be to subtract the summation of state-level and federal additive taxes from the retail gasoline prices at the gas station level and conduct the empirical analysis without taxes; however, in such a case, it would not be possible to have a variance decomposition analysis to show the contribution of taxes to the gasoline price dispersion, since such an analysis requires linearity (either in logs or levels) to be tractable.

7.7. Data for Crude Oil Prices

We use time fixed effects to capture the effects of daily crude oil prices in the regression analysis. Although we have a rough justification through our model that $(1 - T_t)$ captures time-varying subsidies/transfers to refiners, the empirical reason for using time fixed effects rather than using the actual crude oil prices is that crude oil prices may be reflected by gas stations after a certain period of time. For example, Borenstein et al. (1997) show that one cent of an increase in crude oil prices leads to an increase of 0.55 cents in gasoline prices in the first two weeks and further 0.12 cent increase in the next two weeks, and for a 0.67 cents increase after four weeks; moreover, the response of retail gasoline prices may be different between an increase and a decrease in crude oil prices or the transmission of shocks from crude oil prices to gasoline prices can be affected differently by supply and demand factors as in Kilian (2010). Since the focus of this paper is the spatial analysis of gasoline prices, we simply ignore the dynamic transmission mechanism of crude oil price changes into the retail-level gasoline prices and use time fixed effects to capture daily crude oil prices. Nevertheless, for robustness, we alternatively replace time fixed effects with daily crude oil prices (of West Texas Intermediate obtained from www.eia.gov) to compare the estimated time fixed effects with daily crude oil prices.

Table 1 - Descriptive Statistics

Variable		Gasoline Prices					Maximum Gasoline Price Difference					State-Level Taxes	
Sample		**Pooled Sample**					**Across Time (A Typical Day)**					**Cents per Gallon**	
STATE	**# of Stations**	**Min**	**Max**	**Median**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Median**	**Mean**	**Std. Dev.**	**2010**	**2011**
Alabama	822	2.27	2.95	2.63	2.67	0.15	0.04	0.22	0.09	0.04	0.10	20.9	30.9
Alaska	155	3.23	4.03	3.38	3.41	0.09	0.00	0.64	0.25	0.13	0.31	8.0	8.0
Arizona	581	2.29	2.89	2.61	2.62	0.12	0.08	0.30	0.13	0.04	0.14	19.0	19.0
Arkansas	628	2.00	2.95	2.64	2.67	0.15	0.06	0.78	0.13	0.07	0.14	21.8	21.8
California	1,267	2.00	3.19	2.97	2.98	0.14	0.05	0.94	0.11	0.12	0.14	46.6	49.1
Colorado	767	2.00	2.85	2.59	2.61	0.10	0.07	0.58	0.14	0.05	0.15	22.0	22.0
Connecticut	651	2.09	3.27	3.04	2.99	0.19	0.08	0.84	0.14	0.09	0.15	41.9	49.6
Delaware	216	2.00	3.39	2.89	2.87	0.19	0.06	0.75	0.17	0.08	0.18	23.0	23.5
Washington DC	88	2.57	4.26	3.08	3.09	0.25	0.01	1.57	0.44	0.31	0.49	23.5	23.0
Florida	1,509	2.09	3.04	2.72	2.76	0.16	0.03	0.78	0.08	0.10	0.11	34.5	34.5
Georgia	950	2.00	2.94	2.63	2.66	0.15	0.06	0.51	0.11	0.06	0.13	12.4	29.2
Hawaii	245	3.15	4.14	3.47	3.53	0.21	0.14	0.76	0.60	0.06	0.60	44.4	47.4
Idaho	463	2.42	3.29	2.87	2.86	0.10	0.14	0.64	0.26	0.09	0.28	25.0	25.0
Illinois	1,275	2.08	3.08	2.79	2.82	0.14	0.06	0.71	0.11	0.07	0.13	39.0	41.2
Indiana	1,356	2.16	3.09	2.75	2.76	0.16	0.05	0.31	0.12	0.04	0.12	34.1	39.7
Iowa	855	2.00	3.09	2.79	2.82	0.13	0.12	0.99	0.23	0.11	0.26	22.0	22.0
Kansas	605	2.43	3.05	2.69	2.73	0.14	0.05	0.26	0.12	0.04	0.12	25.0	25.0
Kentucky	762	2.42	3.04	2.73	2.75	0.14	0.08	0.39	0.17	0.06	0.18	22.5	27.8
Louisiana	644	2.10	2.91	2.63	2.66	0.14	0.08	0.55	0.14	0.05	0.15	20.0	20.0
Maine	447	2.51	3.17	2.96	2.92	0.17	0.05	0.18	0.10	0.03	0.10	31.0	31.5
Maryland	662	2.29	3.03	2.77	2.77	0.17	0.06	0.37	0.13	0.04	0.13	23.5	23.5
Massachusetts	822	2.25	3.05	2.84	2.79	0.18	0.06	0.53	0.12	0.05	0.13	23.5	23.5
Michigan	1,560	2.00	3.05	2.80	2.80	0.14	0.04	0.91	0.10	0.12	0.13	35.0	40.8
Minnesota	1,168	2.00	3.11	2.78	2.82	0.14	0.04	0.89	0.08	0.07	0.09	27.2	27.2
Mississippi	810	2.30	2.95	2.65	2.68	0.15	0.08	0.39	0.16	0.05	0.17	18.8	18.8
Missouri	1,288	2.00	2.89	2.58	2.61	0.14	0.04	0.61	0.10	0.08	0.12	17.3	17.3
Montana	353	2.70	3.29	2.89	2.90	0.08	0.17	0.48	0.31	0.06	0.32	27.8	27.8
Nebraska	597	2.51	3.32	2.89	2.91	0.13	0.11	0.49	0.26	0.07	0.27	27.7	27.2
Nevada	527	2.00	3.07	2.85	2.85	0.13	0.10	0.86	0.16	0.08	0.18	33.1	33.1
New Hampshire	361	2.31	3.05	2.85	2.81	0.17	0.05	0.30	0.08	0.03	0.09	19.6	19.6
New Jersey	1,160	2.15	2.95	2.73	2.68	0.19	0.04	0.34	0.08	0.05	0.10	14.5	14.5
New Mexico	540	2.14	2.97	2.69	2.71	0.10	0.13	0.75	0.24	0.09	0.26	18.8	18.9
New York	1,109	2.00	3.26	2.99	2.96	0.19	0.05	0.92	0.14	0.12	0.17	44.6	49.5
North Carolina	1,110	2.00	2.99	2.69	2.73	0.15	0.04	0.69	0.10	0.08	0.13	30.2	35.3
North Dakota	312	2.08	3.39	2.94	2.94	0.16	0.15	0.98	0.41	0.09	0.42	23.0	23.0
Ohio	1,609	2.37	3.02	2.72	2.74	0.14	0.05	0.34	0.12	0.05	0.13	28.0	28.0
Oklahoma	665	2.37	2.88	2.59	2.62	0.13	0.04	0.24	0.11	0.04	0.11	17.0	17.0
Oregon	501	2.59	3.15	2.89	2.90	0.11	0.12	0.26	0.18	0.03	0.18	25.0	31.0
Pennsylvania	951	2.17	3.09	2.85	2.83	0.19	0.04	0.60	0.10	0.08	0.12	32.3	32.5
Rhode Island	277	2.44	3.55	3.00	2.97	0.19	0.11	0.53	0.22	0.08	0.23	33.0	33.0
South Carolina	819	2.27	2.87	2.57	2.61	0.15	0.05	0.44	0.10	0.05	0.11	16.8	16.8
South Dakota	289	2.49	3.29	2.89	2.92	0.14	0.14	0.54	0.32	0.08	0.34	24.0	24.0
Tennessee	991	2.00	2.89	2.59	2.65	0.15	0.05	0.78	0.10	0.08	0.12	21.4	21.4
Texas	1,594	2.00	2.85	2.55	2.59	0.15	0.04	0.56	0.11	0.09	0.14	20.0	20.0
Utah	622	2.30	2.87	2.71	2.72	0.05	0.07	0.47	0.11	0.05	0.12	24.5	24.5
Vermont	295	2.54	3.34	3.04	2.99	0.19	0.06	0.30	0.17	0.04	0.17	24.5	26.6
Virginia	731	2.06	2.93	2.61	2.64	0.16	0.07	0.73	0.16	0.07	0.16	19.5	20.0
Washington	532	2.07	3.13	2.95	2.94	0.10	0.06	0.87	0.13	0.09	0.15	37.5	37.5
West Virginia	402	2.54	3.19	2.89	2.92	0.14	0.06	0.35	0.15	0.05	0.16	32.2	32.2
Wisconsin	1,048	2.55	3.05	2.79	2.83	0.13	0.04	0.49	0.08	0.05	0.09	32.9	32.9
Wyoming	254	2.42	3.09	2.69	2.73	0.13	0.26	0.58	0.41	0.05	0.41	14.0	14.0
Min	88	2.00	2.85	2.55	2.59	0.05	0.00	0.18	0.08	0.03	0.09	8.0	8.0
Max	1,609	3.23	4.26	3.47	3.53	0.25	0.26	1.57	0.60	0.31	0.60	46.6	49.6
Median	662	2.25	3.05	2.79	2.80	0.14	0.06	0.56	0.13	0.07	0.14	24.0	25.0
Mean	750	2.27	3.14	2.81	2.82	0.15	0.07	0.59	0.17	0.07	0.18	26.0	27.5
Std. Dev.	403	0.28	0.30	0.19	0.18	0.03	0.05	0.27	0.11	0.04	0.11	8.5	9.5

Notes: Min stands for minimum, Max for maximum, Std. Dev. for standard deviation. Pooled sample includes all gas stations from all 144 days in the sample period. Maximum price difference has been calculated for each state/district for each day; the numbers in the table show the summary statistics of the maximum price difference across time.

Table 2 - Estimation Results

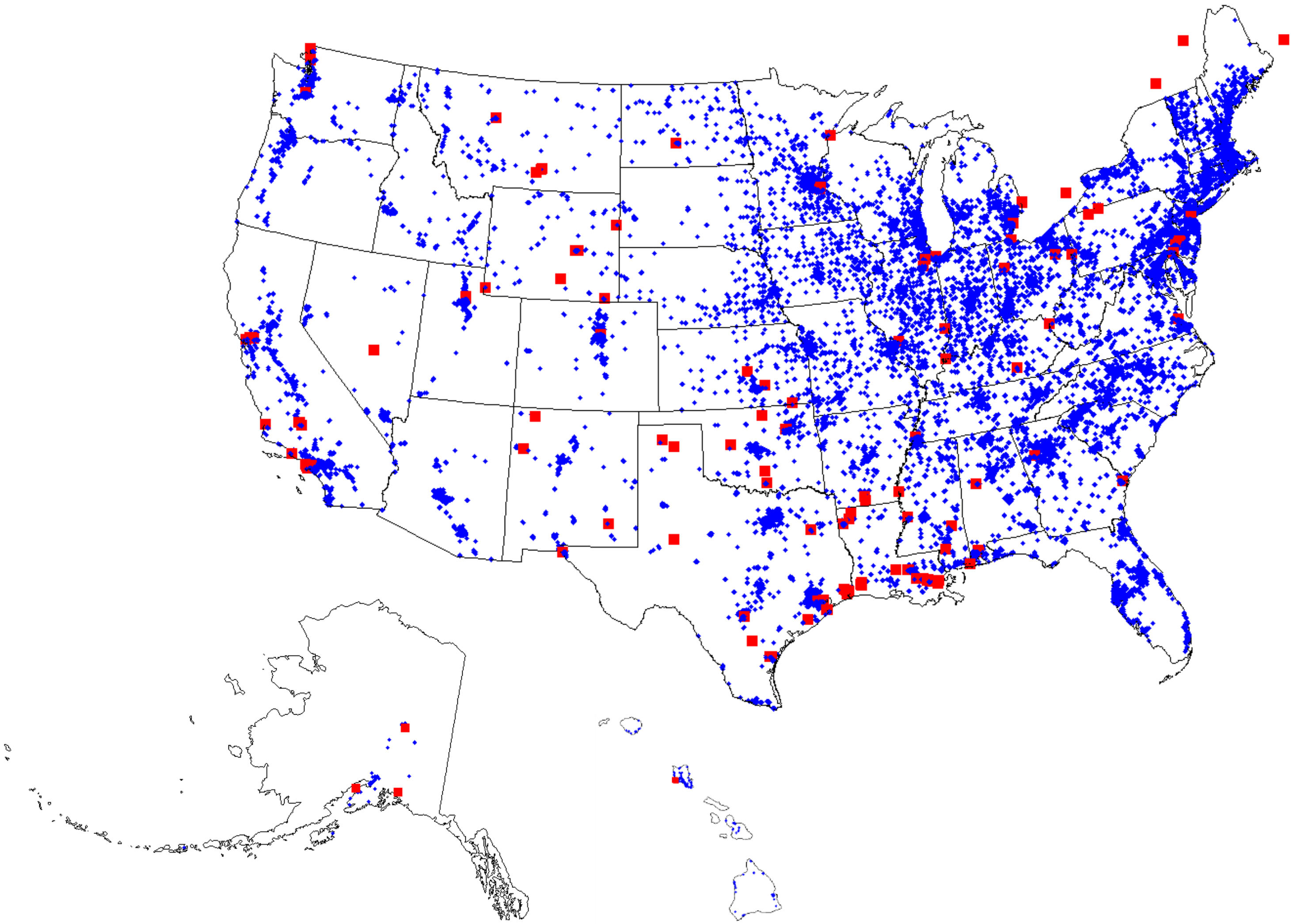
Independent Variable	Spatial Measure #1	Spatial Measure #2	Spatial Measure #3	Spatial Measure #4	Spatial Measure #5
S.D. of Nighttime Lights	-0.12 [-0.13,-0.11]	-0.23 [-0.24,-0.21]	-0.82 [-0. 85,-0. 78]	-1.64 [-1.73,-1.55]	-1.92 [-2.06,- 1.78]
# of Nearby Stations	-0.05 [-0.08 ,-0.04]	-0. 12 [-0.13 ,-0. 09]	-0.14 [-0.15,-0.12]	-0.02 [-0.06,0.03]	-0.19 [-0.24,-0.13]
Distribution Costs	0. 41 [0.39,0.43]	0. 41 [0.39,0.43]	0. 36 [0. 34,0.38]	0. 25 [0.23,0.27]	0.17 [0.16,0.19]
Land Prices	0.70 [0.66,0.74]	0.65 [0.61,0.69]	0.66 [0.62,0.70]	0.76 [0.72,0.80]	0.79 [0.75,0.83]
Brand Fixed Effects	YES	YES	YES	YES	YES
Refinery Fixed Effects	YES	YES	YES	YES	YES
Time Fixed Effects	YES	YES	YES	YES	YES
R-Squared	0.9016	0.9006	0.9008	0.9006	0.9004

Notes: Dependent variable is log gasoline prices. The 5% confidence intervals are in brackets. The estimates and the corresponding intervals have been multiplied by 100 for presentational purposes. State and federal taxes have restricted coefficients of one that are not shown.

Table 3 - Variance Decomposition of Log Gasoline Prices

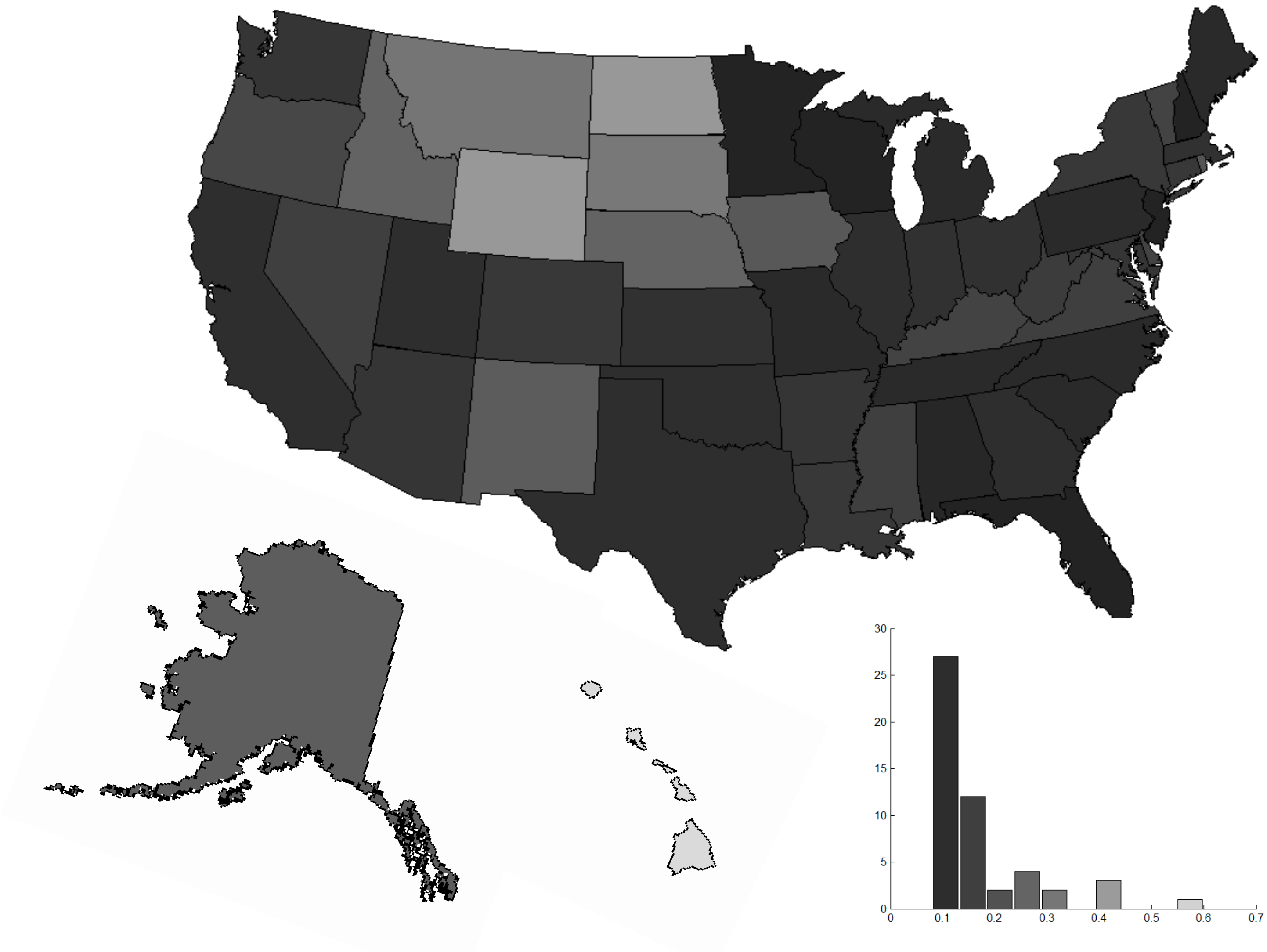
% Explained by	Spatial Measure #1	Spatial Measure #2	Spatial Measure #3	Spatial Measure #4	Spatial Measure #5
S.D. of Nighttime Lights	-0.35	-0.30	0.42	1.70	2.30
# of Nearby Stations	0.04	0.15	0.23	0.05	0.25
Distribution Costs	0.14	0.14	0.11	0.08	0.05
Land Prices	1.27	1.17	1.19	1.37	1.42
Brand Fixed Effects	-0.45	-0.33	-0.30	-0.16	-0.05
State Taxes	11.77	12.19	12.28	12.28	12.28
Federal Taxes	-6.87	-6.85	-6.84	-6.84	-6.84
Refinery Fixed Effects	33.44	33.10	32.25	30.86	29.91
Time Fixed Effects	51.17	50.78	50.73	50.73	50.72
Residuals	9.84	9.94	9.92	9.94	9.96

Figure 1 - Location of Gas Stations and Refiners



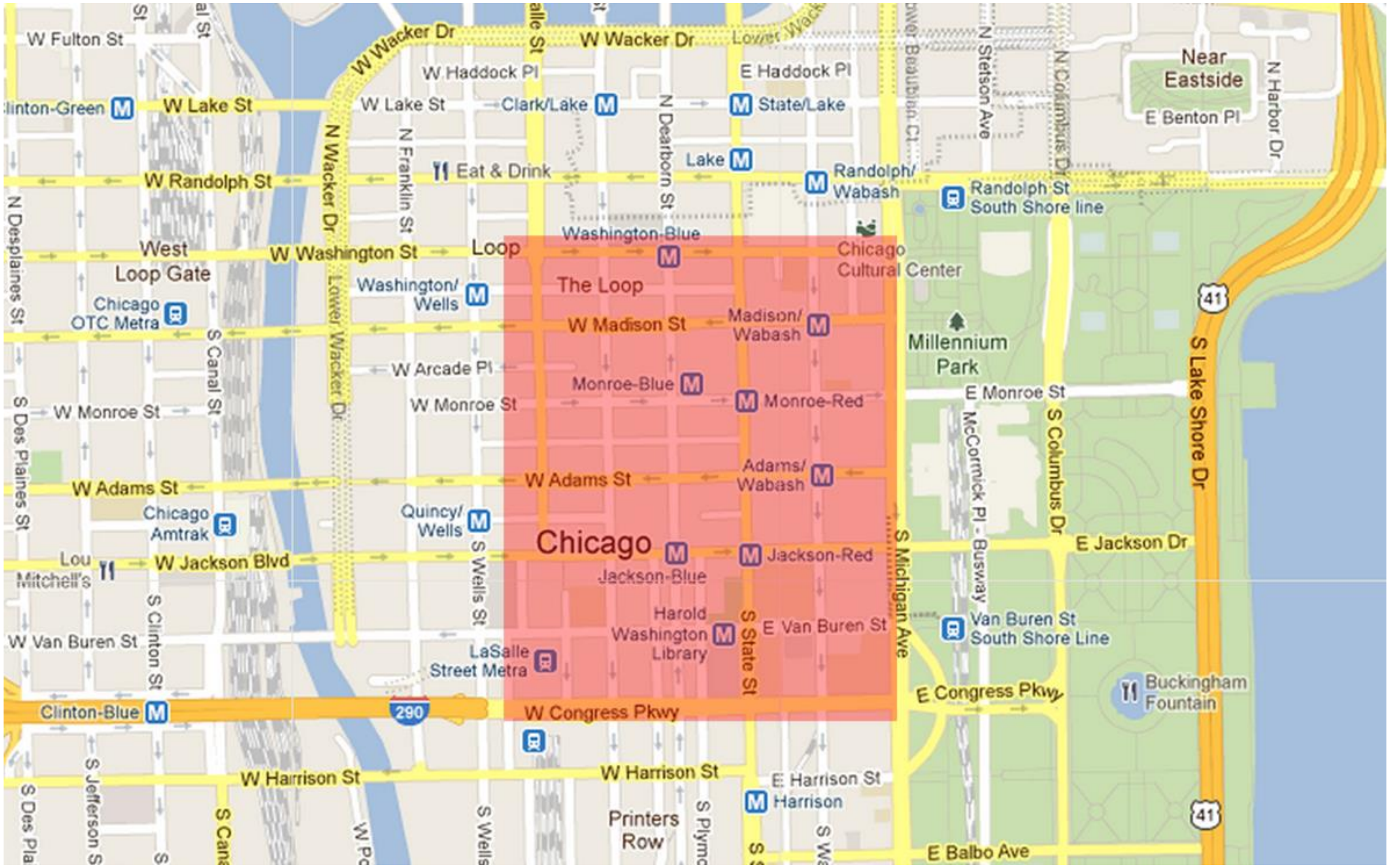
Notes: Gas stations are represented by blue small dots, and refiners are represented by red large squares. There are four Canadian refiners in the northeast of the map; these are the closest refiners to some gas stations (in our sample) in Northeastern United States.

Figure 2 - Maximum Price Difference on a Typical Day



Notes: For each day, the maximum price difference is calculated across gas stations within each state for each day. This map represents the median value (of the maximum price difference) in U.S dollars across all days in the sample.

Figure 3 - An Approximate Typical Example of One 30-Arc-Second Grid in Chicago, IL



Notes: The source is Google Maps. The transparent area in red represents one 30-arc-second grid.