# The Higher Price of Cleaner Fuels: Market Power in the Rail Transport of Fuel Ethanol

Jonathan E. Hughes *

October 29, 2008

## Abstract

Demand for fuel ethanol is growing rapidly in response to U.S. environmental and energy policies. This paper provides convincing evidence of market power in the transportation of ethanol used in clean motor vehicle fuels. I estimate a reduced form model for railroad route-level prices. My identification strategy instruments for railroad entry, controls for selection and explicitly models capacity constraints using detailed geographic data (GIS). Ethanol shipment prices fall quickly with increased competition. For the first entrant, prices drop by as much as $104 or 3% of variable cost. For the second entrant, prices fall by $446 to $475 or 14% to 15% of variable cost. I also find that railroads are able to price discriminate based on environmental regulation at route destinations. Monopolist prices for shipments to carbon monoxide non-attainment areas are on average $148 or 4% higher than other destinations served by a single firm. The price premium falls sharply with increased competition.

---

[*] I thank Meghan Busse, Christopher Knittel, Douglas Miller, David Rapson and Victor Stango for helpful discussions and Michael Nicholas for assistance with GIS programming. I am grateful for financial support from the Sustainable Transportation Energy Pathways (STEPS) program and the UC Davis Institute of Transportation Studies.

Institute of Transportation Studies, University of California, Davis, One Shields Avenue, Davis, CA 95616; *email:* jehughes@ucdavis.edu

# 1 Introduction

Each year motor vehicles in the U.S. consume over 6.5 billion gallons of fuel ethanol.[1] Used primarily in ethanol blended gasoline, proponents argue that ethanol reduces air pollution and greenhouse gas emissions, supports local agriculture and in times of high oil prices, is a cost-effective alternative to petroleum. Because of these and other factors, the federal renewable fuel standard (RFS) seeks to double or triple ethanol consumption over the next decade.[2] In the near future, climate change policy may add even more momentum to the ethanol boom.

The vast majority of this fuel is produced in the U.S. Midwest and shipped via rail to demand centers in the eastern and western United States. Given the long distances, the transportation market is dominated by a small number of railroads who operate large multi-state networks. Most routes are highly concentrated with one or two firms competing at the route endpoints. Furthermore, ethanol policies themselves may exacerbate market power by providing opportunities for railroads to price discriminate among customers.

A number of previous authors provide evidence of market power in railroad markets. MacDonald (1987) and MacDonald (1989) find that grain shipment prices are positively correlated with railroad concentration after railroad deregulation in the 1980s. Schmidt (2001), using a cross-section of city-level data, finds that rail prices increase and shipments decrease as the number of firms serving route endpoints decreases. Numerous studies exist for coal markets. Atkinson and Kerkvliet (1986), estimate rents to railroads transporting low-sulfur coal during the early 1980s. Friedlaender (1992), and Kunce, Hamilton and Gerking (2005) estimate price cost margins. Busse and Keohane (2007) find evidence of price discrimination based on regulation and geography.

This paper investigates whether railroads exercise market power in the transportation of ethanol. I exploit a cross-section of shipment-level prices collected from the pubic ethanol tariffs of five of seven North American Class I railroads.[3] These prices are combined with detailed geographic information system (GIS) data that describe each firm's rail network, control variables as well as the locations and characteristics of ethanol plants.

---

[1] The U.S. Renewable Fuels Association (2008) estimates 2007 ethanol demand at 6.8 billion gallons. Domestic production in 2007 is estimated as approximately 6.4 billion gallons.

[2] The Energy Independence and Security Act of 2007 passed by the U.S. Congress establishes the RFS and requires production of 24 billion gallons of biofuel by 2017. The vast the majority of this new biofuel is expected to be ethanol.

[3] Railroad classes are a designation based on operating revenue assigned by the U.S. Surface Transportation Board. Additional information on the characteristics of the Class I railroads is provided in Section 2.

Two approaches are presented for analyzing whether railroads exercise market power in ethanol shipments. The first examines whether prices vary based on the level of competition at route origins and destinations. I use the number of firms participating at the route endpoints as an indication of the level of competition. This approach is similar to Schmidt (2001). However, Schmidt treats the number of firms as exogenous and ignores the possibility of capacity constraints as an explanation for the correlation between prices and firm participation. Because railroad participation may be endogenous, I instrument for the number of firms at the route endpoints using data from the 1900 U.S. census. To account for the possibility of railroad capacity constraints, I use detailed geographic data on rail traffic density as a measure of route congestion. In addition, I control for cost and demand characteristics using route distance, railroad effects, origin and destination state effects.

The second approach investigates whether firms are able to price discriminate based on environmental regulations at the route destinations. The paper adds to the empirical literature on price discrimination in the transportation sector beginning with Borenstein (1991) and Shepard (1991) in retail gasoline markets and Borenstein and Rose (1994) in airline markets.[4] The approach adopted here is similar in spirit to Stavins (2001) and Busse and Keohane (2007).

Stavins (2001) documents price discrimination in airline tickets that increases with the level of competition. The author uses a reduced form model to study discounts due to fare restrictions. Discounts are found to be smaller on routes with higher market concentration. However, route concentration is assumed to be exogenous. Busse and Keohane (2007) study price discrimination in rail shipments of low-sulfur coal to power plants under the 1990 Clean Air Act Amendments. The authors find that railroads do in fact practice price discrimination based on environmental regulation and geography. Prices increase for deliveries to regulated electricity plants relative to non-regulated plants. The price increases are less for electricity plants further from the low-sulfur coal mines where compliance options are more flexible. In contrast to Busse and Keohane who use panel data and exploit geographic variation, temporal and cross-sectional regulatory variation to identify price discrimination, I use cross-sectional variation in environmental regulation and competition at the route endpoints.

---

[4] Borenstein (1991) finds evidence of price discrimination based on search costs in the market for leaded gasoline during the early 1980s. Shepard (1991) investigates price discrimination based on service quality (self and full-service) in retail gasoline stations. Borenstein and Rose (1994) differentiate between price dispersion due to variation in costs and dispersion due to discriminatory pricing by regressing the Gini coefficient of fares on cost and market structure variables.

Increasing competition at the route origin by one firm decreases the price of an ethanol shipment by \$151 or 3.3% on average. When estimated using a flexible model that allows the effect of competition on price to be non-linear, increasing the number of firms participating at the route origin from monopoly to duopoly decreases the shipment price by \$104 or 3.3% of average variable costs. For two entrants, the shipment price decreases by \$446 or 14.2% of variable cost. At the route destination, prices do not decrease significantly with the first entrant. The second entrant reduces prices by \$475 on average or 15.1% of variable costs. These results are robust to a variety of alternate specifications.

I also find evidence of price discrimination based on environmental regulation. Since the 1990's, the Federal government has designated counties with high carbon monoxide levels "CO non-attainment areas." Many of these counties mandate the use of ethanol blended gasoline as part of a strategy to improve air quality.[5] I find that monopolist prices are, on average, \$148 or 3.8% higher at destinations located in federally designated CO non-attainment areas. The price premium in CO non-attainment areas decreases sharply with increased competition. For each additional entrant, the price premium falls by \$101 or 2.6%. It is unlikely that these effects are the result of railroad capacity constraints in non-attainment areas.

Finally, I use recent policy changes in the state of California to construct a "back of the envelope" estimate for the gains to railroads due to increased ethanol consumption. I find that increasing the ethanol content of California gasoline from 5.7% to 10% increases total railroad producer surplus by approximately \$26.5 million per year. Though only 0.2 cents per gallon of gasoline, this sum still represents a substantial transfer from ethanol consumers to the railroads. Furthermore, this estimate provides some sense of the magnitude of transfers likely to occur through more expansive changes in ethanol policy such as the federal RFS.

This paper also contributes to the existing literature on ethanol policy. Current studies focus primarily on the supply and demand for fuel ethanol, environmental impacts and the economic effects on related markets. Rask (1998) estimates reduced form models for ethanol supply and demand using state-level consumption data. Anderson (2008) estimates ethanol demand using station-level prices and quantities in Minnesota. He then uses his estimates to conduct a policy simulation of a renewable fuel standard. Farrel *et.al* investigate energy and greenhouse gas emissions

---

[5] Examples include: Los Angeles, CA; El Paso, TX; Denver, CO; Missoula, MT; and Phoenix, NV.

(2006) and Searchinger *et.al* (2008) estimate emissions of ethanol production processes. The costs of ethanol production have been studied by McAloon *et.al* (2000) and Kwiatkowski *et.al* 2006. McNew and Griffith (2005) study the impact of ethanol plant entry on local grain prices. While each of the areas is important to the evaluation of ethanol policy, none of these papers have focused on the interaction between ethanol consumption and rail markets. Evidence of market power could change the regulatory choices of policymakers. Estimation of the gains to firms highlights some of the distributional effects of ethanol policies.

The remainder of the paper is organized as follows. Section 2 provides an overview of ethanol use in the U.S., ethanol production, and the structure of the U.S. ethanol rail transportation market. Sections 3 and 4 describe the data and econometric framework used in this analysis. The main empirical results are presented in Section 5. Section 6 summarizes a variety of robustness checks. Section 7 explores the gains to railroads under an example ethanol policy and Section 8 concludes.

# 2 Ethanol in the U.S.

## 2.1 Drivers of demand

Historically, the use of ethanol as a transportation fuel has been driven by state and federal policies. In the U.S., fuel ethanol is seldom used on its own but rather in blends with gasoline. There are three main applications that can be classified according to the relative quantities of ethanol and gasoline used in each blend.

First, ethanol is used in reformulated gasoline as an “oxygenate” to reduce motor vehicle emissions, namely carbon monoxide (CO).[6] Typically, ethanol used for this purpose is blended with gasoline at several percent ethanol by volume. Federal reformulated gasoline containing 2% ethanol or other oxygenates was mandated until 2006 in approximately 130 counties.[7] An additional 65 counties choose to opt-in to this program.

Second, fuel ethanol is used as a gasoline replacement. In modern passenger vehicles, ethanol can be readily substituted for gasoline in quantities as high as 10% to 20%. The term “gasohol” is often used to describe blends with between 5.7% and 10% ethanol by volume. In many instances,

---

[6] Pollutants such as carbon monoxide (CO) are the result of incomplete fuel combustion. Oxygenates are compounds that when blended into fuels increase oxygen content and improve combustion.

[7] Oxygenation requirements for federal RFG were dropped in 2006 though many counties continue to use ethanol.

gasohol mandates or incentive programs have been created by the federal and state governments as a means of supporting local agriculture or reducing petroleum consumption.[8] Ethanol blends of this type may also be the result of so-called "economic blending" on the part of fuel suppliers when low ethanol prices relative to gasoline favor substitution.

Finally, ethanol is used as an "alternative fuel" in vehicles designed to burn ethanol instead of, or more commonly, in combination with a small amount of gasoline. The distinction between this and gasohol blends is the relatively high proportion of ethanol in the blended fuel. The standard formulation in the U.S. is E85 which contains 85% ethanol and 15% gasoline by volume. Flex fuel vehicles (FFV's) are designed to use E85, pure gasoline or a combination of the two fuels. However despite the availability of FFV's, E85 fueling stations are still relatively rare. According to the U.S. Department of Energy (2008) there were approximately 1,400 E85 fueling stations in the U.S. compared with around 150,000 gasoline fueling stations.

Climate change policy is a new and increasingly important driver of ethanol demand. Compared with conventional fuels, ethanol can result in lower emissions of carbon dioxide and other greenhouse gases. Policies such as the federal renewable fuel standard (RFS) are at least in part motivated by the desire to reduce environmental externalities due to climate change. In California and several other states, low carbon fuel standards (LCFSs) focus on reducing carbon emissions from the transportation sector by implicitly subsidizing cleaner fuels, (Holland, Hughes and Knittel (forthcoming)). These policies are likely to result in large increases in the quantity of ethanol consumed in both gasohol and E85 type blends over the coming years.

Most of these applications create demand for ethanol in cities with large numbers of motor vehicles and fueling stations. Air quality concerns in particular have tended to create centers of ethanol demand in large urban areas, typically on the east and west coasts. However, U.S. ethanol production is mainly centered in the Midwest. Over 95% of U.S. ethanol plants surveyed in Ethanol Producer Magazine (2008) use corn as a feedstock.[9] A major bi-product of the production process is animal feed in the form of distiller's grains. As a result, the majority of ethanol plants are located in the midwestern U.S. in areas of high corn and cattle production. Figure 1 shows the locations of

---

[8] For example, the U.S. government awards an ethanol blender's credit of 51 cents per gallon of ethanol used in transportation fuels. A partial list of state programs includes Minnesota which grants a wholesaler credit of 15 cents per gallon of ethanol used to make gasohol, Montana, Nebraska and North Dakota which grant credits of 30 cents, 18 cents and 40 cents per gallon, respectively.

[9] Kwiatkowski *et.al* (2006) estimate corn represents approximately 75% of the production cost of ethanol.

ethanol production facilities and destinations for ethanol shipments. The geographic distribution of supply and demand creates a major market for ethanol transportation.

## 2.2 Rail transportation

A recent U.S. Department of Agriculture (2007) estimates that approximately 60% of ethanol in the U.S. is transported by rail, 30% by truck and 10% by barge. Unlike petroleum, ethanol is difficult to transport via pipeline due to problems with corrosion and water absorption. Truck transport typically occurs over short distances or when relatively small quantities are required.[10] Barge transport on the Mississippi River, Hudson River and Great Lakes is an option for plants or rail terminals located near these bodies of water.[11] However for most routes, rail is the preferred mode of transportation and the role of outside options is limited.

The most likely destinations for ethanol shipments are petroleum distribution terminals. Gasoline is typically transported from the refinery to local distribution terminals by pipeline or barge. Because ethanol blended gasoline is difficult to transport via pipeline, blending of ethanol and gasoline usually occurs at the terminal.[12] The blended fuel is then delivered to retail stations by tanker truck.

Ethanol is transported by rail in 30,000 gallon tank cars. Shipments are arranged by the ethanol producer, customer or by a third party marketer. Both single car and unit train shipments are common. An ethanol unit train consists entirely of ethanol tankers and typically contains 80 to 90 cars traveling between a single origin (or gathering location) and destination.[13] Because this type of movement avoids the inefficiency of assembling or disassembling a train of cars from multiple locations, prices for unit train shipments are lower than single car shipments.[14] However, unit

---

[10] The typical capacities of each mode imply that 80 rail cars contain approximately the same quantity of ethanol as one barge or 300 tanker trucks.

[11] The USDA (2007) identifies major terminals located in: Albany, NY; Chicago, IL; Houston, TX; New Orleans, LA; and Sauget, IL. Unfortunately, I lack detailed data on barge shipments for the ethanol plants in my sample and cannot directly measure the effect of water transport on rail prices. Furthermore, not all ports are suitable for transporting ethanol. Including a dummy variable in the base model for whether an ethanol plant is within several miles of any port suggests rail prices are *higher* near navigable waterways but that the effect of railroad entry on rail prices is larger.

[12] The simplest method (splash blending) involves adding the appropriate quantity of ethanol to a tanker full of gasoline and allowing the two fuels to mix on route to the final destination.

[13] Sweeps or gathered trains are ethanol trains assembled from two origins delivered at a single destination.

[14] The USDA (2007) estimates that a unit train can make a return trip 30 times a year as compared to 12 times a year for single car shipments. This reduces inventory carrying costs on the part of shippers and capital and operating costs for the railroads.

trains are only suitable for the largest volume producers and customers. In my conversations with railroads, representatives also described a rigorous certification process for unit train facilities to insure, among other things, sufficient capacity for handling and unloading tank cars.[15] Because the number of destinations served by unit train is still a small share of total destinations, this paper focuses on prices for single car shipments.[16]

Recent estimates for the volume of ethanol transported by rail are difficult to obtain. The Association of American Railroads (2008b) reports that 96,000 carloads of ethanol were shipped in 2006. The USDA estimate of 60% share suggests that approximately 3.9 billion gallons of ethanol were shipped by rail in 2007. At 30,000 gallons per car, this implies a total of 130,000 shipments in 2007.[17] At today's prices, this implies a market of approximately $600 million.

Of the more than 500 freight railroads operating in the U.S. in 2006, the vast majority of ethanol was shipped by a few firms. The Association of American Railroads (2008b) estimates that 80% to 85% of ethanol shipments originate on "Class I railroads." The seven Class I railroads are the largest carriers in terms of revenue as defined by the Surface Transportation Board. These firms accounted for approximately 70% of industry mileage and approximately 90% of industry revenue in 2006 according to the Association of American Railroads (2008a). This group includes Burlington Northern Sante Fe (BNSF), Canadian National (CN), Canadian Pacific Rail Service (CPRS), CSX Transportation (CSX), Kansas City Southern (KCS), Norfolk Southern (NS) and the Union Pacific Railroad (UP). Class I railroads typically own tracks in many states and operate long-haul routes between major cities. Given the large distances that ethanol is transported from plants in the Midwest to destinations on the Gulf, East and West Coasts, the Class I railroads are the most likely carriers for these shipments.

This paper focuses on ethanol shipments that occur on Class I railroads. Given that these firms control the majority of the market and that price data are more difficult to obtain for smaller railroads, this seems a reasonable simplification. Furthermore, the services that smaller firms

---

[15] For example, the Watson terminal in Southern California has the capacity to unload a 95-car (2.85 million gallon) unit train in a 24-hour period, http://www.us-dev.com/l_lomita_rail_terminal.htm. On average, this facility receives a new unit train every 3 to 4 days http://www.bnsf.com/media/news/articles/2004/09/2004_09_28a.html.

[16] A count of current or planned unit train facilities includes: Watson, Carson and Stockton, CA; Ft. Worth and Arlington, TX; Providence, RI; Albany, NY; Sewaren and Linden, NJ; and Baltimore, MD.

[17] Though a growing business, ethanol represents a small share of total rail shipments. Association of American Railroads (2008a) estimates put the total shipments of class I railroads at approximately 31.5 million per year implying that ethanol shipments are currently less than 1% of total traffic.

provide are in many respects different from those provided by the Class I railroads. Regional "short-line" railroads move shipments between ethanol plants and interchanges with Class I carriers. Smaller firms also participate in state or regional markets on shorter routes suited to their more limited rail networks.

## 3 Data

The data used in this analysis consist of a cross section of ethanol shipment prices for five North American Class I railroads. The unit of observation is a railroad-route pair. These data provide a snapshot of ethanol rail prices during the first quarter of 2008. Price data were combined with detailed geographic information on the U.S. railroad network including track length, location, traffic density and ownership. Additional explanatory variables were collected from a variety of U.S. government agencies.

### 3.1 Ethanol shipment price data

Single-car ethanol price data were collected from the public tariffs of five Class I North American railroads. For the two remaining railroads the locations of origins and destinations of ethanol shipments were known, though prices were not available.[18] The public tariffs represent the "advertised" price for transporting a single tanker car (30,000 gallons) between an origin (*e.g.*, Iowa City, IA) or origin group (*e.g.*, southeastern Iowa) and a given destination (*e.g.*, Los Angeles, CA) or destination group (*e.g.*, southern California).

When origin or destination cities are grouped, cities within the group are assigned the same prices.[19] In some cases, a railroad's tariffs are listed by destination state only and no prices for individual destinations are given.[20] In these cases, I assign prices to individual cities in the destination state in the following way. For railroad $i$ and state $j$, I define a set of possible destination

---

[18] The railroads that report public ethanol tariffs are: Burlington Northern Santa Fe (BNSF), Canadian National(CN); CSX Transportation (CSX); Norfolk Southern (NS); and the Union Pacific Railroad (UP). The Canadian Pacific Rail Service (CPRS) and Kansas City Southern (KCS) do not publish public ethanol tariffs.

[19] The implications of price grouping on the estimated standard errors are discussed in Section 6.3.

[20] In addition, BNSF uses a state group for origins in Illinois. Prices for this group are excluded from the analysis. Results are qualitatively similar to those presented below when ethanol plants and interchanges in Illinois are included as potential origins.

cities as all cities within state $j$ having a population greater than 100,000 persons. Then, I exclude destination cities that do not fall on railroad $i$'s rail network.[21]

Since the advent of the Staggers Rail Act of 1980, railroads have been permitted to negotiate private contracts for the shipment of goods. In my discussions with the railroads, firms differ in their use of private contracts with some firms relying exclusively on public prices, some negotiating only private contracts and some employing a mix of both types of prices. One manager stated that the western railroads, BNSF and UP, use only public prices while the eastern railroads, CN, CSX, KCS, NS and CPRS are more apt to utilize private contracts.

The lack of data on private contract posses two challenges for this analysis. First, private rates negotiated by individual shippers are presumably lower than those contained in the public tariffs. Therefore, the public tariffs represent an upper bound on the actual price paid for an ethanol shipment. Second, since only advertised prices are observed and not actual transaction prices, the existence of a public tariff does not guarantee that any ethanol is actually shipped at that price.[22]

To increase the likelihood that actual shipments occur on the routes with published public tariffs, I further restrict my sample by requiring the following: valid origin cities contain an ethanol production facility or serve as an interchange depot between rail carriers; and valid destination cities contain at least one petroleum fuel terminal or serve as an interchange depot.[23] Data on the locations of ethanol plants are taken from Ethanol Producer Magazine (2008). The locations of petroleum terminals are taken from the U.S. Environmental Protection Agency (2008). Interchange depots are those identified through conversations with representatives of the Class I railroads.

While some authors such as MacDonald (1989), have expressed reservations about the use of public railroad tariffs, these are the best and most extensive public data on ethanol prices today. Previous studies such as MacDonald (1987), have utilized confidential railroad price and quantity data from the Surface Transportation Board Waybill Sample. Unfortunately, I lack access to these

---

[21] Excluding destinations that are part of a destination state-group results in parameter estimates that are very similar to those presented below.

[22] This situation is similar in many ways to Internet price data. In their study of Internet book sales, Chevalier and Goolsbee (2003) show that sales weighted-average prices are significantly different from price averages that weight all books equally. These differences can affect the conclusions one draws from price dispersion online or from relative price differences between online and traditional retailers. Without ethanol quantity data, it is impossible to weight the observed public tariff data by the actual railroad-route shipments. Furthermore, the lack of reliable data on county or state-level ethanol demand prevents the use of ethanol consumption as a proxy for actual shipments.

[23] Ethanol plants may be operating, idle or under construction. Conversations with railroads indicate that prices may be published some time before a plant begins operation.

data. Another alterative is the public use file of the Waybill Sample, a subset of the full sample with the confidential data removed. However, these data are unsuitable because they include only a small number of ethanol shipments for a only few routes.[24] Therefore, despite the limitations, the public tariffs seem the most suitable available data.

### 3.2 Railroad network data

Geographic Information Systems (GIS) data on each railroad's rail network were obtained from the Bureau of Transportation Statistics (2007). The data include the physical location of tracks in latitude and longitude coordinates, track length, density of rail traffic on individual track segments, track ownership as well any leases or "trackage rights" provided to other firms. I define a given railroad's network as all segments of track owned by the railroad plus all tracks to which the railroad has trackage rights.[25]

The locations of origin and destination cities were added using latitude and longitude coordinates from the Bureau of Transportation Statistics (BTS). For cities not listed in the BTS rail data, latitude and longitude coordinates were obtained from http://www.batchgeocode.com/. The origin and destination cities for the routes listed in each railroad's public tariffs were checked against that railroad's network. Routes that contain cities that do not appear on the network were dropped from the sample.[26]

The rail distance for each route was calculated as the shortest-distance path between cities on a given carrier's network. The calculation was implemented using the Network Analyst routine in ArcView GIS version 3.3. Figure 2 depicts an example displaying the shortest-distance route between Iowa Falls, IA and West Sacramento, CA on Union Pacific's rail network. Of course, firms may optimize over other factors in addition to distance when selecting a route between origins and destinations. For example, transit time, if congestion or speed limits are a consideration on certain tracks.

---

[24] The most recent version (2006) contains approximately 160 ethanol shipments.

[25] For a given railroad and route, sections of owned track may be separated by sections owned by another railroad but operated through trackage rights. Therefore, in practical terms, the inclusion of leased tracks is often necessary for a continuous network. This definition results in network maps that are consistent with route maps published by the various carriers.

[26] One possible explanation is that portions of these routes are served by contract with a regional short-line carrier and therefore are not included in the carrier's network. A second explanation lies in the fact that some sections of track in the BTS rail data are not assigned ownership though they may if fact belong to a Class I carrier.

To get a sense for the validity of the shortest-distance assumption, the actual routing distances were obtained from the Burlington Northern Santa Fe for approximately 2,200 of the railroad's ethanol routes.[27] Figure 3 plots the kernel density estimate of the ratio of the shortest distance path on BNSF's network to BNSF's reported distances with mean and standard deviation of 0.96 and 0.029, respectively. This suggests that the shortest-distance route is a fair approximation of the actual distance, though on average the calculated distance is slightly smaller.[28]

### 3.3 Summary statistics

Summary statistics for the combined price and rail mileage data are shown in Table 1. The first column presents statistics for the full sample of price data. The second and third columns summarize the effects of the two types of restrictions imposed on the data. Limiting the sample to observations for which both the origin and destination cities appear on the listing carrier's network reduces the total number of price observations by 27% from approximately 25,000 to approximately 18,000 as shown in column 2. The mean price is approximately 2% higher for the remaining observations after the network restriction is imposed.

Column 3 adds the additional restrictions on origins and destinations requiring that route endpoints contain an ethanol plant, petroleum terminal or serve as a rail interchange. The total sample is reduced by approximately 79% to approximately 5,200 observations. Compared to the full sample, the mean price and rail mileage of the remaining observations are approximately 7% and 6% higher, at $4,632 and 1,207 miles respectively.[29] For individual firms in the sample, the combined effect of these restrictions varies between a 20% (NS) and 96% (CSX) reduction in the number of observations.[30] Summary statistics for all variables in the final sample are shown in Table 2.

---

[27] BNSF reports distances for the purpose of fuel surcharge calculations. These data are available at http://www.bnsf.com/bnsf.was6/RailMiles/RMCentralController.

[28] Calculated distances may also be smaller due to limitations in the spatial resolution of the GIS rail network data. These data break the actual tracks into discrete segments of between 0.01 miles and 10 miles which appear as straight lines in the data. On curved portions of track, a segment in the GIS data will appear as a straight-line between two measured points on the curve while the actual track covers a longer distance along the arc.

[29] A Wald test rejects the equivalence of the base model specification in samples with and without restrictions with an F-statistic of 8.39.

[30] For CSX, the vast majority of observations are dropped due to the origin and destination restrictions. This is because the price data for CSX contain many routes *between* likely ethanol demand centers, *e.g.* Washington, DC to Charlotte, NC or *from* likely demand centers *e.g.* Buffalo, NY to interchange terminals *e.g.* Chicago, IL. These types of shipments do not satisfy the origin and destination restricts and seem unlikely to occur frequently in practice.

Table 3 summarizes railroad participation at origins and destination in the final sample. Most routes are highly concentrated with the majority of origins and destinations having only one or two firms. Approximately 94% of observations have fewer than three railroads participating at either the route origin or destination.

## 4 Empirical approach

The basic empirical model "base model" regresses the price per ethanol car shipment on the number of railroads participating at the endpoints of each route. This specification is shown as Equation 1 below where $p_{ijk}$ is railroad $i$'s price for a single car shipment between origin $j$ and destination $k$, $Miles_{ijk}$ is the shortest-distance rail mileage, $Num\_RR\_org_j$ and $Num\_RR\_dest_k$ are the number of railroads that participate at origin $j$ and destination $k$, respectively. The term $\epsilon_i$ represents unobserved price factors that vary at the railroad level and $\epsilon_l$ and $\epsilon_m$ represent unobserved price factors that vary at the origin state and destination state levels, respectively.

$$\begin{aligned} p_{ijk} = \beta_0 + \beta_1 Miles_{ijk} + \beta_2 Miles_{ijk} \times \epsilon_i + \beta_3 Miles\_sq_{ijk} + \beta_4 Miles\_sq_{ijk} \times \epsilon_i + \\ \beta_5 Num\_RR\_org_j + \beta_6 Num\_RR\_dest_k + \beta_7 RTD_{ijk} + \beta_8 O\_interchange_j + \\ \beta_9 D\_interchange_k + \epsilon_i + \epsilon_l + \epsilon_m + \epsilon_{ijk} \end{aligned} \tag{1}$$

I model the unobserved railroad, origin and destination state effects as mean effects. This is denoted in Equation [2] by replacing the $\epsilon_i$, $\epsilon_l$ and $\epsilon_m$ with fixed-effects. I incorporate the interaction term $\beta_2 Miles_{ijk} \times \eta_i$ to capture differences in the relationship between price and miles that vary at the railroad level. Because the railroads in the sample each charge a different mileage-based fuel surcharge in addition to the ethanol shipment price, I expect significant differences in this relationship across firms.[31]

$$\begin{aligned} p_{ijk} = \beta_0 + \beta_1 Miles_{ijk} + \beta_2 Miles_{ijk} \times \eta_i + \beta_3 Miles\_sq_{ijk} + \beta_4 Miles\_sq_{ijk} \times \eta_i + \\ \beta_5 Num\_RR\_org_j + \beta_6 Num\_RR\_dest_k + \beta_7 RTD_{ijk} + \beta_8 O\_interchange_j + \\ \beta_9 D\_interchange_k + \eta_i + \eta_l + \eta_m + \epsilon_{ijk} \end{aligned} \tag{2}$$

In addition, I include the square of the shortest-distance rail mileage $Miles\_sq_{ijk}$ (in units of hundreds of miles squared) and railroad interactions to capture any scale economies that occur due

---

[31] Norfolk Southern (NS) does not assess a fuel surcharge to its ethanol shipments. The remaining railroads assessed the following fuel surcharges in April 2008: BNSF (\$0.54/mile); CN (\$0.207/mile); CSX (\$0.35/mile); and UP (\$0.26/mile).

to route distance. To capture price differences for routes ending or beginning at an interchange, $O\_interchng_j$ and $D\_interchng_k$ are dummy variables that equal one when the origin or destination is an interchange city.

To account for route congestion in the observed rail prices, I include the weighted-average rail traffic density $RTD$ for each route in the sample. One would expect highly congested routes to have higher costs than less congested routes. Furthermore, if congestion is correlated with railroad participation, failure to account for congestion would bias the estimated effect of competition on prices.[32] I obtain GIS data on rail traffic density from the Bureau of Transportation Statistics (2007). The Federal Railroad Administration assigns each section of railroad track to one of seven categories (class 1 - 7) based on the quantity of freight transported over the track each year. The variable $RTD$ is the mile-weighted average density over the entire route.

### 4.1 Econometric issues

A major concern in identifying the effect of competition on the price of ethanol rail shipments is the potential endogeneity of the number of firms participating at the route endpoints. Specifically, one might expect the number of railroads participating at an origin or destination to increase with positive price shocks. In this case, $Num\_RR\_org$ and $Num\_RR\_dest$ would be correlated with the error term and the OLS coefficients on these terms would be biased toward zero.[33]

To address this issue I instrument for $Num\_RR\_org$ and $Num\_RR\_dest$ using county-level data on agricultural production, manufacturing and population from the 1900 U.S. Census of Population and Housing. Since the total size of the U.S. railroad network peaked in the early 1900's and existing tracks operate on historical right-of-ways, I expect historical rail service to be correlated with present railroad participation. Because I do not observe actual historical railroad participation, I use county-level population, the value of manufactured goods, the value of agricultural crops and the value of livestock production in 1900 as predictors of railroad participation at each origin or destination. Given the long time lag, it seems reasonable to assume that these variables are also uncorrelated with rail prices in 2008. However, one might imagine the existence of factors such as

---

[32] If for example entry results in the construction of new tracks that reduce congestion.

[33] It is difficult to know how readily railroads may enter new geographic markets. On one hand, the large fixed costs of building new tracks poses a significant barrier to entry. However, if trackage rights (leases) are easily negotiated, barriers to entry may be much lower.

geography or long-lived policies that cause rail prices today to be correlated with the 1900 census instruments. To control for factors affecting population, manufacturing and agriculture, and in turn rail service that persist over many years, I also include current data on population (U.S. Bureau of Economic Analysis 2005), agricultural production (U.S. Census of Agriculture 2002) and manufacturing (U.S. Census of Manufacturing 2002) as reported in U.S. Census Bureau (2008).

A second complication arises if ethanol producers also take rail prices into account seeking locations (origins) with negative price shocks. Anecdotal evidence from the citing of recent ethanol plants supports this hypothesis.[34] Therefore, the method used to select origins, namely requiring that origins be operating ethanol plants, may bias estimates of [2]. In many ways this situation is similar to Olley and Pakes (1996) who address the problems of endogeneity *and* selection due to unobserved productivity in the telecommunications industry. However, here the number of firms participating at the route endpoints may be endogenous *and* the selection of observations depends on the error term in rail prices.

I account for the possibility of selection bias by adopting a control function approach. To motivate this choice, consider the following stylized model of the ethanol plant operation decision. Assume that the single period profits of ethanol plant $p$ in period $t$ are given by Equation 3, where $q_{p,t}$ is the quantity of ethanol produced, $c(q_{p,t})$ are variable costs excluding transportation, and $\omega_{p,t}$ represents components of profit not observed by the econometrician. Further assume that the ethanol plant will continue to operate in period $t$ if the present value of future profit streams is positive. This is represented by the Bellman equation [4].

$$\pi_{p,t} = p(q_{p,t})q_{p,t} - Rail\_transport\_cost(q_{p,t}) - c(q_{p,t}) + \omega_{p,t} \tag{3}$$

$$V(q_{p,t}, \omega_{p,t}) = \max_{q_{p,t}} \left[\pi(q_{p,t}, \omega_{p,t}) + \beta EV(q_{p,t}, \omega_{p,t})\right] > 0 \tag{4}$$

Assuming that $\omega_{p,t}$ evolves according to a Markov process, I write the ethanol plant's condition to continue operations as Equation 5.

$$p(q_{p,t-1})q_{p,t-1} - Rail\_transport\_cost(q_{p,t-1}) - c(q_{p,t-1}) + \omega_{p,t-1} > 0 \tag{5}$$

Rearranging and recognizing that rail transportation costs depend on [2], I rewrite the continuation condition in terms of a function of the base model error [6]. The unobserved components of ethanol

---

[34] For example, see Zerschling (2008).

plant profits $\omega_{p,t-1}$ depend in part on $\epsilon_{ijk}$ and in part on factors that I assume are orthogonal to transportation costs. Therefore, to avoid confusion I drop $\omega_{p,t-1}$ from [6].

$$\begin{aligned} p(q_{p,t-1})q_{p,t-1} - f(\epsilon_{ijk}) - c(q_{p,t-1}) &> 0 \\ f(\epsilon_{ijk}) &< p(q_{p,t-1})q_{p,t-1} - c(q_{p,t-1}) \qquad (6) \end{aligned}$$

The right-hand side of Equation 6 depends on revenue, observable factors of ethanol production and unobserved factors. To a first-order approximation, ethanol prices are arguably independent of the ethanol location decision. Factors of production in $c(q_{p,t-1})$ include corn, energy (electricity, natural gas, coal) and labor plus the co-product animal feed (distiller's dried grains).[35] I control for corn prices using county-level data on corn production from the 2002 U.S. Census of Agriculture. Energy prices and labor are unlikely to vary substantially by location within a state and therefore should be captured by the origin state fixed-effects. Finally, I control for animal feed prices using county-level livestock production from the 2002 U.S. Census of Agriculture.

A similar selection problem arises if petroleum terminal operators take rail prices into account when establishing facilities for ethanol blending. The petroleum terminal analog to [6] contains cost factors related to gasoline transportation, energy and labor. I control for gasoline transportation cost using the distance between each petroleum blending terminal (destination) and the nearest petroleum refinery as reported by the U.S. Environmental Protection Agency (2008). Energy prices and labor are captured by the destination state fixed-effects.

A final issue relates to the standard errors of the estimated coefficients. As mentioned previously, several of the railroads in the sample group origins and destinations into price groups. For observations in the same price-group, the disturbance terms are likely correlated. One might also expect origin or destination-level common shocks for origins and destinations that are not part of price groups because of unobservables that vary at the city level. Because both route origins and destinations may be affected, I cluster on origin price-group and destination price-group.[36] Cities not reported as part of a price group are assigned to their own group for the purpose of calculating cluster-robust standard errors. Section 6.3 discusses these issues in further detail and presents alternate models for within-group correlations in the disturbance term.

---

[35] For an engineering analysis, see Kwiatkowski *et. al* (2006).

[36] I currently use a "one-way" clustering estimator and report the largest standard errors that result from clustering on either origin price-group or destination price-group. Initial attempts to cluster on both origin and destination group using the "multi-way" algorithm described by Cameron, Gelbach and Miller (2007) produced imaginary standard errors.

# 5 Market power in the rail transport of ethanol

Results from estimation of the base model [2] are shown in Table 4. The origin and destination state fixed-effects have been suppressed to simplify the presentation of results. Discussion of the first-stage regression results for the 2SLS estimates is deferred to Section 6.1. Column 1 shows OLS estimates of the full sample. Because historical data used to construct the instrumental variables are not available for the entire sample, the total number of observations used in the 2SLS estimation is reduced. Column 2 reports OLS estimates using the reduced sample. The third column of Table 2 presents 2SLS estimates of the basic model coefficients instrumenting for $Num\_RR\_org$ and $Num\_RR\_dest$ using the instruments outlined in Section 4.1.

The estimated coefficients on the railroad fixed-effects, distance terms and their interactions are in general statistically significant and similar in magnitude across the three models. The smaller sample used in the 2SLS estimates appears to have little effect on the parameter estimates which are not substantially different between columns 1 and 2 of Table 4. Therefore, I will focus my discussion on the 2SLS estimates in column 3.

Evaluated at the sample mean, an additional mile increases the price of an ethanol shipment between \$1.18 and \$1.91 depending on the railroad. This translates to between 1.2 and 1.9 cents per ton-mile. Busse and Koehane (2007) find that the delivered price of coal increases with distance at a rate of about 0.9 cents per ton-mile. Published mileage scales for BNSF and UP list the price per mile of an ethanol shipment at \$1.47 per mile to \$2.43 per mile or 1.5 cents to 2.5 cents per ton-mile. The coefficients on the interchange dummies are positive and significant suggesting that prices for routes with a destination or origin interchange are higher (1.9% to 5.0%) compared to other routes.

An increase in average density $RTD$ of one class is correlated with a decrease in price of approximately \$39. It is well known that railroads exhibit economies of density. For example see Ivaldi and McCollough (2001). Therefore, the negative coefficient on average density is not surprising.[37] However, one might still expect congestion on routes with very high density resulting in slower travel speeds and higher costs. Therefore, the combined effect on prices may be positive, negative or u-shaped. In unreported results I allow the effect of $RTD$ to be nonlinear by grouping

---

[37] Another possible explanation of the negative coefficient is that railroads may invest more heavily in track maintenance, signaling, and other infrastructure on higher-density routes resulting in lower costs and prices.

the route average rail traffic density into seven dummy variables. These results do show the u-shaped effect of density with a minimum at an average rail traffic density of six. However, the estimated coefficients on $num\_RR\_org$ and $num\_RR\_dest$ are quite similar to those reported in Table 4.

Increased competition leads to lower prices for ethanol shipments. For each additional railroad participating at the route origin, price decreases by approximately \$151, on average, or approximately 3.3% of the sample mean price. The effect of an additional railroad at the route destination is smaller at approximately \$6 and is not statistically significant. These results are robust across the three models summarized in Table 4 though the effect at the route origin is twice as large in the 2SLS estimates.

## 5.1 Non-parametric competition effects

In the previous section I estimated the effect of competition on ethanol shipment prices assuming a linear relationship between the number of firms competing at the route endpoints and prices. Bresnahan and Reiss (1991) show that competition in concentrated markets can increase quite dramatically with the entry of a small number of firms. In railroad markets, Schmidt (2001) shows that the assumption of a constant marginal effect for each additional entrant may be too restrictive. Therefore, in this section I estimate the effects of competition on prices using a more flexible specification that replaces $Num\_RR\_org$ and $Num\_RR\_dest$ with a set of zero-one dummy variables for the number of railroads participating at the route origin and destination. Specifically, $Two\_RR\_org$ is equal to one if there are two railroads participating at the route origin and zero otherwise. Similarly, $Three\_RR\_dest$ is equal to one if there are three railroads participating at the route destination and zero otherwise.

$$\begin{aligned} p_{ijk} = \beta_0 + \beta_1 Miles_{ijk} + \beta_2 Miles_{ijk} \times \eta_i + \beta_3 Miles\_sq_{ijk} + \beta_4 Miles\_sq_{ijk} \times \eta_i + \\ \beta_5 Two\_RR\_org_j + \beta_6 Three\_RR\_org_j + \beta_7 Two\_RR\_dest_k + \\ \beta_8 Three\_RR\_dest_k + \beta_9 Six\_RR\_dest_k + \beta_{10} RTD_{ijk} + \\ \beta_{11} O\_interchange_j + \beta_{12} D\_interchange_k + \eta_i + \eta_l + \eta_m + \epsilon_{ijk} \end{aligned} \tag{7}$$

Since the maximum number of railroads participating at an origin is three, I include $Two\_RR\_org$ and $Three\_RR\_dest$ and omit the dummy for origins with one railroad. Similarly, since the maximum number of railroad at any destination is six and there are no destinations in the sample served by four or five railroads, I include the variables $Two\_RR\_dest$, $Three\_RR\_dest$ and $Six\_RR\_dest$ and omit the dummy for destinations served by a single railroad. It is important to note that Chicago, IL is the only destination served by six railroads and therefore, $Six\_RR\_dest$ should be interpreted as a Chicago-specific effect.

The first column of Table 5 shows results estimated using OLS on the full sample. The second and third columns show OLS and 2SLS estimates on the 2SLS sample, respectively.[38] The origin and destination state fixed-effects, railroad and route distance effects have been suppressed to simplify the presentation of results. Focusing on the 2SLS estimates in column three we see that the coefficients on the origin railroad participation dummies are negative and significant ($p < 0.01$). The average decrease in price for a duopoly of firms competing at the origin relative to a monopoly is approximately \$104. For three railroads participating at the origin, price drops by \$446 or \$342 more than for two firms. At the destination, the price per ethanol shipment falls slightly by approximately \$11 dollars for two railroads participating at the destination relative to monopoly, though this effect is not statistically significant. For three railroads the price decrease is approximately \$475 or \$464 larger than for duopoly.[39]

To put these effects in perspective, relative to the origin mean monopoly ethanol tariff of \$4,647, prices decrease by 2.2% with two firms and 9.6% for three firms. At the destination, prices decrease by 0.2% with two firms and 11.1% with three firms relative to the destination mean monopoly price of \$4,285. In their analysis of the California market for tire dealers, Bresnahan and Reiss find that tire prices decrease about 8% for a market with five competitors relative to a monopoly and by about 20% for an urban (competitive) market. While the retail tire market is far from the best model for competition in rail shipments, the similarity between the two sets of estimates is at least reassuring.

A better indication is to compare the estimated price effects to estimates of rail margins for ethanol shipments. Variable cost estimates were obtained for each route from the Surface Transportation Board Uniform Rail Cost System (2008). The URCS is compiled from the confidential

---

[38] In the 2SLS models, I instrument for the railroad origin and destination participation dummies using the instrumental variables discussed previously.

[39] Standard errors for $two\_RR\_dest$ and $three\_RR\_dest$ are 84.23 and 118.42 when the robust variance estimator is clustered on destination price-group instead of origin price-group.

railroad waybill sample and provides firm-level distance-based variable cost estimates.[40] The most recent year for which costs estimates are available is 2006. Therefore, 2006 estimates are adjusted to 2008 costs using the Association of American Railroads Rail Cost Adjustment Factors (AAR 2008).[41]

For one entrant at the route origin, the average price change is approximately 3.3% of the average variable cost. For two entrants, the price change is approximately 14.2% of variable cost at the route origin and 15.1% of variable cost at the route destination. The sample mean price and variable cost suggest an average margin of 32% and an average markup of approximately 48%. Therefore these figures strike me as substantial, and suggest that prices do indeed fall fast with increased competition.

It is interesting to note that prices fall more substantially for the second entrant relative to the first. This result runs contrary to models of firm behavior in non-cooperative games and suggests the possibility of collusion between firms in setting rail prices. One might suspect that repeated interaction between firms along common routes may facilitate tacit collusion. To investigate this hypothesis I test whether prices are significantly different in duopoly markets where firms are "close competitors."

I focus on BNSF and UP who operate in the western U.S. and CSX and NS who operate primarily in the east. Both pairs of railroads serve many of the same ethanol origins and destinations. Approximately 30% of BNSF origins are also served by UP. Nearly 70% of BNSF destinations are also served by UP.[42] In the East, CSX and NSX share between 40% and 70% of origin cities and 60% and 80% of ethanol destinations. I add to the base model zero-one dummy variables $O\_close\_comp$ and $D\_close\_comp$ that are equal to one if the two firms at the route origin (destination) are either both BNSF and UP or CSX and NS. Results are summarized in Table 6 below.

Table 6 suggests that prices are higher when the two firms in a duopoly are close competitors, though the estimated coefficients are not statistically significant. A Wald test fails to reject the null hypothesis that the price effect is zero for a duopoly of close competitors at the route origin and destination with chi-squared values of 0.74 and 1.22, respectively. When firms are not close

---

[40] Since ethanol estimates are not reported separately, I use variable costs for chemicals transported in tanker cars greater than 22,000 gallons.

[41] The RCAF is a real cost index of input prices adjusted for productivity. Variable costs range from 2.0 to 94 cents per ton-mile with mean 2.4 and standard deviation 1.6 cents per ton-mile.

[42] Similarly, approximately 30% of UP origins and 70% of UP destinations are served by BNSF.

competitors, prices drop more sharply with entry. For two firms, prices fall by \$357 and \$688 at the route origin and destination compared with the base model estimates of \$104 and \$11. While these results are consistent with collusion, the parameter estimates are quite noisy. In addition, confounding factors such as the geographic similarity of markets served by close competitors may contribute to or explain the observed price effects.

While the ability to detect collusion using the reduced form approach presented here is admittedly limited, the results do highlight an import area for future research. There are relatively few instances of tacit collusion documented in the literature. Furthermore, the potential policy implications of collusive behavior in rail markets extend far beyond ethanol policy. As a structural approach is preferable for detecting collusion, I am currently developing a dataset suitable for this analysis.

### 5.2 Price discrimination based on environmental regulation

In this section I explore whether railroads practice price discrimination based on environmental regulations at route destinations. Under the Clean Air Act of 1990, counties exceeding minimum standards on air quality for specific pollutants are designated as "non-attainment areas." These areas are required to take steps to reduce vehicle travel, adopt emissions controls or require cleaner burning gasoline. The types of actions taken depends in part on the specific pollutants contributing to non-attainment status. In regions designated as carbon monoxide (CO) non-attainment areas, a common approach has been to introduce oxygenated gasoline.[43] Since many states have phased out the use of MTBE, ethanol is the most common oxygenate.

In the past, winter oxygenation was required in 24 CO non-attainment areas according to the U.S. Environmental Protection Agency, 2005. Some of these regions have since relaxed this requirement listing oxygenation as a "contingency measure." Nevertheless, many of these regions may still be expected to use oxygenated gasoline. This is because gasoline used in these areas has been specially formulated to allow for blending with ethanol and meet state and federal gasoline content regulations.[44] Switching to another gasoline formulation often requires physical changes to refiner-

---

[43] Typically, oxygenation is most important during the winter months in order to help reduce CO levels while avoiding contributing to smog formation.

[44] Oxygenation with ethanol increases octane and aids in the reduction of toxic compounds. Without ethanol, cleaner gasoline with higher octane must be substituted in order to meet the necessary fuel specifications. On the

ies or additional transportation costs to import fuel from more distant refineries. These additional costs may prevent some regions from switching fuels despite changes to local regulations.[45]

In these areas where ethanol use is required, either explicitly by law or implicity due to high switching costs, ethanol demand is expected to be more inelastic compared to other destinations. As a result, railroads exercising market power may charge higher prices for shipments to CO non-attainment areas. Of course, it could also be the case that the cost of shipping ethanol to CO non-attainment areas is different than the cost to ship ethanol to other destinations.[46] Under these circumstances, an observed difference in prices would not necessarily be an indication of market power.

To differentiate between market power and costs effects of CO non-attainment status on ethanol rail prices, I make the following changes to the base model. I add the a zero-one dummy $CO\_NA$ that is equal to one if a given destination has ever been designation a CO non-attainment area. I interact this dummy with the destination competition variable $Num\_RR\_dest$ as shown in Equation [8] below. Under the cost hypothesis, one would expect the coefficient on $CO\_NA$ to be positive but the coefficient on the interaction term to be zero. This is because cost factors which increase rail prices should be independent of the number of firms participating at each destination. On the other hand, a positive coefficient on $CO\_NA$ and negative and statistically significant coefficient on the interaction term is consistent with market power. This approach is similar to Stavins (2001) though the present approach allows for the endogeneity of the competition variable $num\_RR\_dest$.

$$\begin{aligned} p_{ijk} = \beta_0 + \beta_1 Miles_{ijk} + \beta_2 Miles_{ijk} \times \eta_i + \beta_3 Miles\_sq_{ijk} + \beta_4 Miles\_sq_{ijk} \times \eta_i + \\ \beta_5 Num\_RR\_org_j + \beta_6 Num\_RR\_dest_k + \beta_7 RR\_dest\_CO\_NA_k + \\ \beta_8 CO\_NA_k + \beta_9 RTD_{ijk} + \beta_{10} O\_interchange_j + \\ \beta_{11} D\_interchange_k + \eta_i + \eta_l + \eta_m + \epsilon_{ijk} \end{aligned} \tag{8}$$

Using GIS data from the Bureau of Transportation Statistics (2008), I map ethanol route destinations to CO non-attainment areas as shown in Figure 4. Of the 151 destinations in the sample, 61

---

other hand, gasoline that is to be blended with ethanol must have a lower Reid Vapor Pressure (RVP) because adding ethanol increases the RVP of the blended fuel.

[45] In California for instance, oxygenation is not required during the summer months under California's state gasoline content regulations. Oxygenation was required under federal law in California until 2006. However since the oxygenation requirement was lifted, California has continued to use oxygenated fuel.

[46] For example, if the seasonality of oxygenate demand results in higher congestion or longer turnaround times for unloading tank cars.

or approximately 40% are within historical CO non-attainment areas.

Results of OLS and 2SLS estimates of the base model including the CO non-attainment dummy variable and interaction term are shown in Table 7 below. As before, the state and railroad fixed-effects and distance terms have been suppressed to simplify the presentation of results. Prices for destinations in non-attainment areas are higher than prices for other destinations. The magnitude of the price difference decreases as the number of railroads participating at the destination increases. The coefficient on the dummy for CO non-attainment status is positive and significant ($p < 0.01$) while the coefficient on the interaction term are negative and significant with ($p < 0.01$).[47] In both models, the coefficients on $Num\_RR\_org$ and control variables are similar to the base model results.

Focusing on the 2SLS estimates in the second column of Table 7, monopolist prices are approximately $148 or 3.8% higher on average in CO non-attainment areas compared to other destinations. For each additional entrant, the price premium falls by $101 or 2.6%. With two entrants, the average price for shipments to CO non-attainment areas is actually predicted to be $53 lower than for other destinations. These results are consistent with railroads practicing price discrimination based on environmental regulations at the route destination.

Given that CO non-attainment status is likely to increase ethanol shipments to these destinations, one might be particularly concerned about the role capacity constraints (congestion) may play in the observed prices. Higher prices would be expected if non-attainment increased shipments to a level near the capacity constraint. Likewise, prices would be expected to decrease with entry if entrants utilize new rail lines (or terminals) thus easing congestion. Furthermore, route-level rail traffic density may not adequately control for these effects at route destinations. To address these issues I also estimate Equation [9] below. Here, the route weighted-average rail traffic density $RTD$ is replaced with $Org\_density$ and $Dest\_density$, the rail traffic density at the origin and destination, respectively. This specification isolates the effects of capacity constraints at the route endpoints.

The third column of table 7 presents results using the origin and destination rail traffic density measures. The coefficients on rail traffic density at the route endpoints are both positive, though the origin effect is not statistically significant. At the destination, an increase in rail traffic density

---

[47] Results using only destinations that were once non-attainment but have since come into attainment are qualitatively similar. This suggests that some aspect of non-attainment designation which impacts ethanol demand persists after the designation has been lifted.

class is correlated with a price increase of approximately \$19. The coefficients on the railroad participation variables, CO non-attainment status and the interaction term are comparable to those presented in column two.[48] Taken together, these results suggest that the observed price behavior at destinations in CO non-attainment areas is not the result of railroad capacity constraints.

$$\begin{aligned} p_{ijk} = \beta_0 + \beta_1 Miles_{ijk} + \beta_2 Miles_{ijk} \times \eta_i + \beta_3 Miles\_sq_{ijk} + \beta_4 Miles\_sq_{ijk} \times \eta_i + \\ \beta_5 Num\_RR\_org_j + \beta_6 Num\_RR\_dest_k + \beta_7 RR\_dest\_CO\_NA_k + \beta_8 CO\_NA_k + \\ \beta_9 Org\_density_{ij} + \beta_{10} Dest\_density_{ik} + \beta_{11} O\_interchange_j + \\ \beta_{12} D\_interchange_k + \eta_i + \eta_l + \eta_m + \epsilon_{ijk} \end{aligned} \tag{9}$$

# 6 Robustness checks

This section investigates the robustness of the results along several dimensions. In Section 6.1, I vary the instrumental variables used in the 2SLS estimates. In Section 6.2, I investigate the possibility of bias due to omitted variables in the base model estimates. Specifically, I add controls for service differentiation ("downstream connectivity") and ethanol plant characteristics. I also explore the number of firms participating on a given route as an alternate indicator of railroad competition. Section 6.3 experiments with several alternate models for group-level correlations in the disturbance term.

## 6.1 Instrumental variables

The results of several alternate sets of instrumental variables specifications are presented in Table 8 below. Results from the corresponding first-stage regressions are presented in Table 9. The first two columns of Table 8 assume that the economies of the route origins are agricultural while the destinations are manufacturing focused.[49] Therefore, these estimates use only the agricultural instruments, crops and livestock production, plus population at the route origin and use only the value of manufactured products plus population at the route destination. The first column excludes current corn production and refinery distance as controls for the ethanol plant and petroleum blending terminal location decisions. The third and forth column use both the value of agricultural

---

[48] Including the route-level average density $RTD$ in addition to the endpoint densities produces very similar results.

[49] This approach seems reasonable given the general mix of economic activity of the origin and destination cities in 2008. However, in 1900 many cities had both significant agricultural and manufacturing production.

and manufactured products at both the route origin and destination. The third column excludes current corn production and refinery distance. The fourth column is the 2SLS base model of Section 4 and includes the complete set of instruments and controls.

In general, the estimated coefficients are very similar across the various specifications. However, the effect of an additional railroad at the route origin is smaller and not statistically significant when only agricultural instruments are used at the route origin and only manufacturing instruments are used at the route destination. Furthermore, the estimated effect is smaller than the OLS estimate. The effect of an additional railroad at the route destination is slightly smaller when both sets of instruments are used, though neither estimate is statistically significant. The railroad fixed-effects and interactions terms are similar across models though the marginal effects of route mileage on price for the various railroads are in general slightly larger in the models with both sets of instruments. Comparing the first with the second column and the third with the fourth column, including current corn production and refinery distance as controls for selection has only a minor effect on the estimated coefficients for $Num\_RR\_org$ and $Num\_RR\_dest$.

Looking at the first-stage results in Table 8 we see that in each of the models the instruments are correlated with the endogenous right-hand side variables with $F-statistics$ for the excluded instruments ranging from 74 to 78 for the $Num\_RR\_org$ equation and 804 to 1430 for the $Num\_RR\_dest$ equation. Partial $R-squarred$ values range from 0.06 to 0.10 for $Num\_RR\_org$ and range from 0.67 to 0.68 for $Num\_RR\_dest$.

In general, the estimated coefficients on the excluded instruments have the expected signs and magnitudes. In model 1, the coefficient on livestock production in the origin regression is positive and significant ($p < 0.01$). The magnitude of the estimate seems reasonable in that an increase of one standard deviation in the level of livestock production ($21 million) is correlated with an increase in railroad participation of 0.14.[50,51] The coefficient on origin population is positive and statistically significant. However, the magnitude of the estimated effect seems small such that the mean origin population of $78,000$ persons corresponds to approximately 0.05 railroads participating at the route origin. An increase in population of 50,000 persons corresponds to an increase in

---

[50] All data for the 1900, 2000 and 2002 U.S. Census have been adjusted for inflation (2008 dollars) to facilitate interpretation of the first-stage results.

[51] The value of a choice steer is between $1,000 and $1,500, so an increase of $21 million in livestock value implies perhaps 14 to 21 thousand additional steers. At 20 to 30 steers per car in a 1900-era stock car, this translates to perhaps 700 to 1000 shipments per year.

railroad participation of 0.03 railroads. The coefficient on crop production is negative, though the effect is small and not statistically significant.

In the model one (2SLS1) destination equation, the coefficient on manufacturing production at the route destination is positive and significant ($p < 0.01$). An increase in production of one standard deviation ($3.2 billion) corresponds to an increase in railroad participation of 1.0 at the route destination.

Since the parameter estimates for models one (2SLS1) and two (2SLS2) are quite similar and likewise for models three (2SLS3) and four (2SLS4), I will focus on comparisons between the second and fourth models. This approach highlights the differences between models with the full set of instruments used in Section 4 and those that use only agriculture and population instruments for the route origin and manufacturing and population for the route destination.

In model two, the coefficient on origin population (1900) for $Num\_RR\_org$ is positive (0.660) and significant with ($p < 0.01$). An increase of 50,000 persons in origin population on average is correlated with an increase in the number of railroads at the origin of 0.03. The coefficient on destination population (1900) for $Num\_RR\_dest$ is negative (-0.911) and significant. An increase of 50,000 persons in destination population is correlated with a decrease in the number of railroads at the destination of -0.05. In model four, the coefficient on origin population (1900) for $Num\_RR\_org$ is substantially larger. An increase of 50,000 persons in origin population on average is correlated with an increase in the number of railroads at the origin of 0.17. The coefficient on destination population (1900) is comparable to model two (-0.903).

The interpretation of the coefficients on the other excluded instruments is complicated by the addition of current corn production in models two and four and the addition of the complementary agricultural and manufacturing production quantities at the route endpoints in models three and four. For example, it seems plausible for the first-stage coefficient on agricultural production to be negative in the origin model if production at the destination is large relative to production at the origin.[52]

Finally, as an additional robustness check I have also estimated the various instrumental variable specifications using the generalized method of moments.[53] The results are very similar to the

---

[52] Of course, the argument for including production at both endpoints lies in the fact that cities may produce different goods that are transported via rail between one another and the other cities in the sample.

[53] Both the 2SLS and GMM estimates use the “ivreg2” module written by Baum, Schaffer and Stillman (2007).

estimates presented in Table 8 and Table 9.

## 6.2 Robustness checks - omitted variables and route specification

In this section I describe a number of robustness checks related to the possibility of omitted variable bias in the results presented in Section 5. These include adding controls for service differentiation or what I will call “downstream connectivity” and ethanol plant characteristics. In addition, I explore defining competition at the route level, rather than at the route endpoints. Results of the various specifications are presented in Table 10. The base model results are reproduced in the first column of Table 10 for comparison. In each case, the estimated parameters in the alternative specifications are very similar to the base model results.[54]

Rail service may be differentiated by the number of destinations that are reachable from a given origin. For example, a railroad that serves more destinations may charge premium prices for the ability to access more ethanol markets from a given origin. From the rail tariff data I construct a “downstream connectivity” variable $DSC$ that is sum of destinations by railroad that originate at a given location.[55] The second column of Table 10 summarizes estimation results when the downstream connectivity variable is included in the base model. The coefficient on $DSC$ is negative and significant with ($p < 0.05$).[56] However, the estimated coefficients on $Num\_RR\_org$ and $Num\_RR\_dest$ are very similar to the base model.

In the third and fourth columns, I explore the possibility of bias due to the omission of ethanol plant characteristics from the base model. If for example there are scale economies in ethanol rail shipments, prices for routes originating at larger plants may be lower than shipments from smaller plants. Larger producers or multi-plant firms may also be more able to invest in logistics or rail infrastructure that reduce shipping costs and are reflected in lower prices.[57] It also seems plausible

---

[54] In addition, I have estimated each of the alternate models using the non-parametric competitive effects discussed in Section 5.1 and the CO non-attainment area model in Section 5.2. In each case the results are quite similar to the results presented in Tables 5 and 6.

[55] The number of destinations with published prices varies substantially by origin even within a single railroad. Therefore, the effect $DSC$ is identified relative to the firm fixed-effects. While in principle any ethanol destination on a railroad’s network is reachable from a given origin, in practice railroads do not publish prices for all possible destinations.

[56] Since DSC is a measure of the size of each firm’s network that is accessible from a given origin, this result is consistent with economies of network size in ethanol shipments.

[57] For example full time schedeuling personnel that more efficiently coordinate shipments with the railroads or large rail sidings and equipment located at the ethanol plant used to assemble individual tank cars into trains for shipment.

that the number of railroads serving a plant may be correlated with the capacity of the plant or market share of the firm.[58] If these effects do in fact exist, omitting ethanol plant characteristics in the base model would tend to overestimate the effect of competition at the route origin.

I use $EtOH\_cap_j$, the total rated production capacity of the ethanol plant at $origin_j$, as a measure of plant size. Relative firm size is approximated by $EtOH\_mkt\_shr_j$, the firm-level market share of the ethanol plant at $origin_j$.[59] In instances when there is more than one ethanol plant at an origin, the capacity-weighted average production capacity and market share are used. Because this approach further restricts the rail tariff sample by excluding origins that are interchanges, I present two sets of results. Column three shows the base 2SLS model estimates using only origins where an ethanol plant is present. Column four shows 2SLS estimates from the model including the additional explanatory variables $EtOH\_cap_j$ and $EtOH\_mkt\_shr_j$.

There is very little change in the estimated coefficients when ethanol plant characteristics are included. The coefficient on $Num\_RR\_org$ decreases somewhat when ethanol plant capacity and ethanol producer market share are included. The coefficients on ethanol plant characteristics are negative, though not statistically significant. Given this small difference in the estimated coefficients combined with loss of generality and the significant reduction in sample size resulting from focusing only on origins that contain ethanol plants, I reject this specification in favor of the base model.

Finally, in the fifth column of Table 10 I explore the number of firms participating on a given route as an alternate measure of competition. The variable $Num\_RR\_route$ is the number of railroads that post prices between a given origin and destination pair. To account for the possibility of endogeneity I instrument for $Num\_RR\_route$ with the same instruments used to identify $Num\_RR\_org$ and $Num\_RR\_dest$. The coefficient on $Num\_RR\_route$ is positive and significant with ($p < 0.01$).[60] The estimated effect implies that on average, price decreases by \$270 when the number of railroad participating on a route increases by one. The remaining coefficients are very similar to the base model results.

---

[58] If for example larger plants or firms are more successful citing plants in locations served by multiple railroads.

[59] Data on the characteristics of ethanol production facilities were obtained from Ethanol Producer Magazine (2008).

[60] Clustering on destination price-group, the standard error on the coefficient for $Num\_RR\_route$ increases to 113.5.

### 6.3 Variance estimators

Table 11 presents estimates of the standard errors of the main parameters of the CO non-attainment model using robust variance estimators clustered at different levels. The first column shows standard errors corrected for within-route correlations. The second and third columns present the base model standard errors which correct for correlations within origin price-group and within destination price-group, respectively. The remaining three columns present results clustering on origin state, destination state and railroad. Although the model contains state and railroad fixed-effects, clustering may still be present at the state or railroad level.[61]

Statistical inferences related to the main parameters of interest are robust to alternate models for group-level correlations in the data. Clustering at the route or railroad level results in standard error estimates that are quite similar to clustering on origin or destination price-group. The standard error estimates do increase when the data are assumed to be correlated at either the origin or destination state-level. However, with the exception of CO non-attainment coefficients, the main parameters remain statistically significant despite the extremely small number of groups.

## 7 Gains to railroads example: California reformulated gasoline

Railroad exercising market power are likely to benefit from state and federal ethanol policies. To get a sense of the gains to railroads, I construct a simple example using California's recent move to increase the ethanol content of gasoline from 5.7% to 10%. Because I lack detailed cost and quantity data, I am unable to calculate rail profits directly. Instead I construct "back of the envelope" estimates for railroad producer surplus before and after the policy change.[62] Though admittedly stylized, this example highlights the gains to railroads under the new policy.

In August of 2008 the California Air Resources Board approved amendments to the state's reformulated gasoline program. The new regulations effectively increase the allowable ethanol content of gasoline sold in California from 5.7% to 10%.[63] This change aims to reduce motor

---

[61] In many ways this is analogous to a state-year panel with state and year fixed effects instead of state and railroad fixed-effects.

[62] This analysis is similar in spirit to Busse and Keohane (2007) who develop similar estimates for coal deliveries under the 1990 Clean Air Act Amendments.

[63] The maximum allowable ethanol content in both the old and new regulations is 10% by volume. However, under the amended rules, new standards for permeative emissions favor increasing the ethanol content of oxygenated gasoline to the allowable limit.

vehicle emissions while increasing the market for ethanol in California, the later assisting refiners in achieving the federal government's renewable fuel standard.

Under the new regulation the quantity of ethanol shipped to California will increase. Prices may also change as quantity moves up on the ethanol transportation supply curve or due to price discrimination under the new regulations. I estimate the amount of ethanol shipped by rail to California before and after the policy change as 5.7% and 10% of California 2007 gasoline demand. This approach seems reasonable since the vast majority of gasoline sold within the state is oxygenated and since barge transport and in-state production are small. As a first-order approximation, I assume that shipments are equally distributed across all firms and routes terminating in California.[64] Unfortunately, the new regulation does not take effect until December 31, 2009 and any price changes are, as of yet, unobserved. As a simplification, I use the observed prices from the first quarter of 2008 for both before and after the policy shift. The resulting change in revenue therefore, is entirely due to the increase in the quantity of shipments.

Variable profits are calculated as the difference between route price and variable cost. Variable cost estimates are from the Uniform Rail Cost System described previously. Total producer surplus is the product of variable profits and shipments summing over all routes. The estimated annual revenue and producer surplus for ethanol rail shipments to California are shown in Table 12. Under the 5.7% ethanol gasoline blend, total ethanol rail revenues are approximately $168.1 million per year with producer surplus of approximately $35.1 million per year. Under the 10% ethanol blend, railroad revenues grow to $294.9 million. Producer surplus increases by $26.5 million to $61.6 million. To put these numbers in perspective, 2007 operating income for Class I RR's serving California were between $3.5 and $3.4 billion. On a per gallon gasoline basis, the additional gains to railroads as a result of the shift to E10 amount to less than 0.2 cents per gallon. Though small on a per gallon basis, this sum still represents a substantial transfer from ethanol consumers to the railroads. Furthermore, this estimate provides a sense of the magnitude of transfers likely to occur through more expansive changes in ethanol policy such as the federal RFS.

---

[64] In this case, the number of shipments on any route is just the total number of shipments divided by the total number of routes.

# 8 Conclusions

The goal of this paper is to investigate whether railroads exercise market power in the transportation of fuel ethanol. Historically, ethanol demand in the U.S. has been driven by state and federal policies. New policies such as the RFS seek to dramatically expand the use of ethanol. Understanding the effect of market power provides valuable information for the evaluation of these policies. Evidence of market power could change the regulatory choices of policymakers.

I find compelling evidence that railroads do exercise market power in the transportation of ethanol. The data show that prices depend on the level of competition at route origins and destinations. In addition, firms are able to price discriminate based on environmental regulations at the route destinations. Specifically, prices for ethanol delivered to cities within CO non-attainment areas are significantly higher than prices for other destinations. The price difference falls sharply with increased competition. Finally, railroads gain from policies that increase ethanol consumption. In the example developed for California, increasing the ethanol content of gasoline results in considerable transfers from ethanol consumers to railroads.

There are several implications of these findings that pertain to new ethanol policies such as renewable fuel standards or low carbon fuel standards. First, because rail carriers are able to exercise market power, policies that promote ethanol may incur additional social costs. To the extent that ethanol policies are aimed at increasing welfare, the effect of market power on prices and consumption must be considered. Second, the new policies themselves may exacerbate market power. For example, CO non-attainment designation allows firms in concentrated markets to raise prices at the expense of customers whose demand for ethanol is inelastic. New policies may enable similar price discrimination. Finally, prices are reduced substantially in markets with even a small number of competitors. Policies that minimize exposure to highly concentrated rail transportation markets can help mitigate the effects of market power.

# References

[1] Anderson, Soren T. 2008. "The Demand for Ethanol as a Gasoline Substitute." Mimeo, Michigan State University.

[2] Association of American Railroads. 2008a. "Class I Railroad Statistics." Policy and Economics Department. http://www.aar.org/ /media/AAR/Industry%20Info/Statistics.ashx, accessed on August 1, 2008.

[3] Association of American Railroads. 2008b. "Railroads and Ethanol." Policy and Economics Department. http://www.aar.org/IndustryInformation/ /media/AAR/BackgroundPapers/Railroads%20and%20Ethanol.ashx, accessed on August 1, 2008.

[4] Baum, C.F., Schaffer, M.E., Stillman, S. 2007. ivreg2: Stata module for extended instrumental variables/2SLS, *GMM* and *AC/HAC*, *LIML* and *k*-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

[5] Borenstein, Severin. 1991. "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins." *The RAND Journal of Economics*, 22(3): 354-369.

[6] Borenstein, Severin and Nancy Rose. 1994. "Competition and Price Dispersion in the U.S. Airline Industry." *The Journal of Political Economy*, 102(4): 653-683.

[7] Bresnahan, Tim and Peter Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy*, 99(5): 977-1009.

[8] Burlington Northern Santa Fe Railway. 2008. "April 2008 Ethanol Rate Adjustment." http://www.bnsf.com/markets/agricultural/ag_news/year2008/pricing08/p03-06-08a.html, accessed April 28, 2008.

[9] Busse, Meghan R. and Nathaniel O. Keohane. 2007. "Market Effects of Environmental Regulation: Coal, Railroads, and the 1990 Clean Air Act." *The RAND Journal of Economics*, Winter 2007, 38(4): 1159-1179.

[10] Cameron, Colin A., Johan B. Gelbach and Douglas L. Miller. 2007. "Robust inference with Multi-Way Clustering." NBER Technical Working Paper #T0327.

[11] Canadian National. 2008. “Price Documents.” http://www.cn.ca/customer_centre/shipping_tools/prices_servic
accessed April 18, 2008.

[12] Chevalier, Judith and Austan Goolsbee. 2003. “Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com.” *Quantitative Marketing and Economics*, 1: 203-222.

[13] CSX Corporation. 2008. “Full Text Price Lists.” http://www.csx.com/?fuseaction=customers.pricing_lists, accessed April 23, 2008.

[14] Ethanol Producer Magazine. 2008. “Plant List and Online Plant Map.” Ethanol Producer Magazine. http://www.ethanolproducer.com/plant-list.jsp?country=USA&view=, accessed on June 3, 2008.

[15] Farrell, Alex E., Richard J. Plevin, Brian T. Turner, Andrew D. Jones, Michael O’Hare, and Daniel M. Kammen. 2006. “Ethanol Can Contribute to Energy and Environmental Goals.” *Science*, 311(5760): 506-508.

[16] Friedlaender, Ann F. (1992). “Coal Rates and Revenue Adequacy in a Quasi-Regulated Industry.” *The RAND Journal of Economics*, 23(3): 376-394.

[17] Holland, Stephen P., Jonathan E. Hughes and Christopher R. Knittel “Greenhouse Gas Reductions under Low Carbon Fuel Standards?” Forthcoming in *American Economic Journal: Economic Policy.*

[18] Ivaldi, M. and G. J. McCullough. 2001. “Density and Integration Effects on Class I U.S. Freight Railroads.” *Journal of Regulatory Economics*, 19(2): 161-182.

[19] Kunce, M., S. Hamilton, and S. Gerking. 2005.“Marketable Permits, Low-Sulfur Coal and the Behavior of Railroads.”Mimeo, University of Wyoming.

[20] Kwiatkowski, Jason R., Andrew J. McAloon, Frank Taylor and David B. Johnston. 2006. “Modeling the Process and Costs of Fuel Ethanol Production by the Corn Dry-Grind Process.” *Industrial Crops and Products*, 23: 288-296.

[21] MacDonald, James M. 1987. “Competition and Rail Rates for the Shipment of Corn, Soybeans, and Wheat.” *The RAND Journal of Economics*, 18(1): 151-163.

[22] MacDonald, James M. 1989. “Railroad Deregulation, Innovation, and Competition: Effects of the Staggers Act on Grain Transportation.” *Journal of Law and Economics*, 32(1): 63-95.

[23] McNew, Kevin, and Duanne Griffith. 2005. “Measuring the Impact of Ethanol Plants on Local Grain Prices.” *Review of Agricultural Economics*, 27(2): 164-180.

[24] Norfolk Southern Corporation. 2008. “Public Price Publications.” http://www.nscorp.com/mktgpublic/publicprices, accessed April 17, 2008.

[25] Olley, Steven and Ariel Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica*, 64(6): 1263-1297.

[26] Rask, Kevin N. 1998. “Clean Air and Renewable Fuels: The Market for Fuel Ethanol in the U.S. from 1984 - 1993.” *Energy Economics*, 20: 325-345.

[27] Renewable Fuels Association. 2008. “Industry Statistics.” http://www.ethanolrfa.org/industry/statistics/, accessed August 1, 2008.

[28] Schmidt, Stephen. 2001. “Market Structure and Market Outcomes in Deregulated Freight Rail Markets.” *International Journal of Industrial Organization*, 19: 99-131.

[29] Searchinger, Timothy, *et.al.* 2008. “Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions From Land-Use Change.” *Science*, 319: 1238-1240.

[30] Shepard, Andrea. 1991. “Price Discrimination and Retail Configuration.” *The Journal of Political Economy*, 99(4): 30-53.

[31] Stavins, Joanna. 2001. “Price Discrimination in the Airline Market: The Effect of Market Concentration.” *The Review of Economics and Statistics*, 83(1): 200-202.

[32] Union Pacific Railroad. 2008. “Price and Transit Time Inquiry.” http://c02.my.uprr.com/cdm/pricedocument/price_document.jas, accessed April 17, 2008.

[33] U.S. Bureau of Transportation Statistics. 2007. “National Transportation Atlas Database 2007.” U.S. Department of Transportation. http://www.bts.gov/publications/national_transportation_atlas_database/2007/, accessed on May 15, 2008.

[34] U.S. Census Bureau. 2008. “1900 U.S. Census of Population and Housing.” U.S. Census Bureau. http://www.census.gov/prod/www/abs/decennial/1900.htm, accessed June 30, 2008.

[35] U.S. Census Bureau. 2008. “USA Counties Data Files.” U.S. Census Bureau. http://www.census.gov/support/DataDownload.htm, accessed June 30, 2008.

[36] U.S. Congress. 2007. “Energy Independence and Security Act of 2007: H.R.6.” U.S. Library of Congress. http://thomas.loc.gov/cgi-bin/bdquery/z?d110:h.r.00006:, accessed on August 1, 2008.

[37] U.S. Department of Energy. 2008. “Alternative Fuels and Advanced Vehicles Data Center, E85 Fueling Station Locations.” Energy Efficiency and Renewable Energy. http://www.eere.energy.gov/afdc/ethanol/ethanol_locations.html, accessed on August 21, 2008.

[38] U.S. Department of Energy. 2008. “Oxygenate Production.” Energy Information Administration. http://tonto.eia.doe.gov/dnav/pet/pet_pnp_oxy_dc_nus_mbbl_a.htm, accessed on August 1, 2008.

[39] U.S. Environmental Protection Agency. 2008. “Locational Data Improvement Project.” Envirofacts database. http://www.epa.gov/enviro/html/locational/, accessed May 16, 2008.

[40] U.S. Environmental Protection Agency. 2007. “Reformulated Gasoline Where You Live.” http://www.epa.gov/otaq/rfg/whereyoulive.htm, accessed August 1, 2008.

[41] U.S. Environmental Protection Agency. 2005. “State Winter Oxygenated Fuel Program Requirements for Attainment of Maintenace of CO NAAQS.” EPA420-B-05-113, November 2005.

[42] Zerschling, Lynn. 2008. “Railroad official knew ethanol plant would overload tracks.” Sioux City Journal. http://www.siouxcityjournal.com/articles/2006/03/18/news/local/caaa51a008aad434862571350 accessed on August 29, 2008.

# 9 Figures

Figure 1: Locations of ethanol plants and shipment destinations

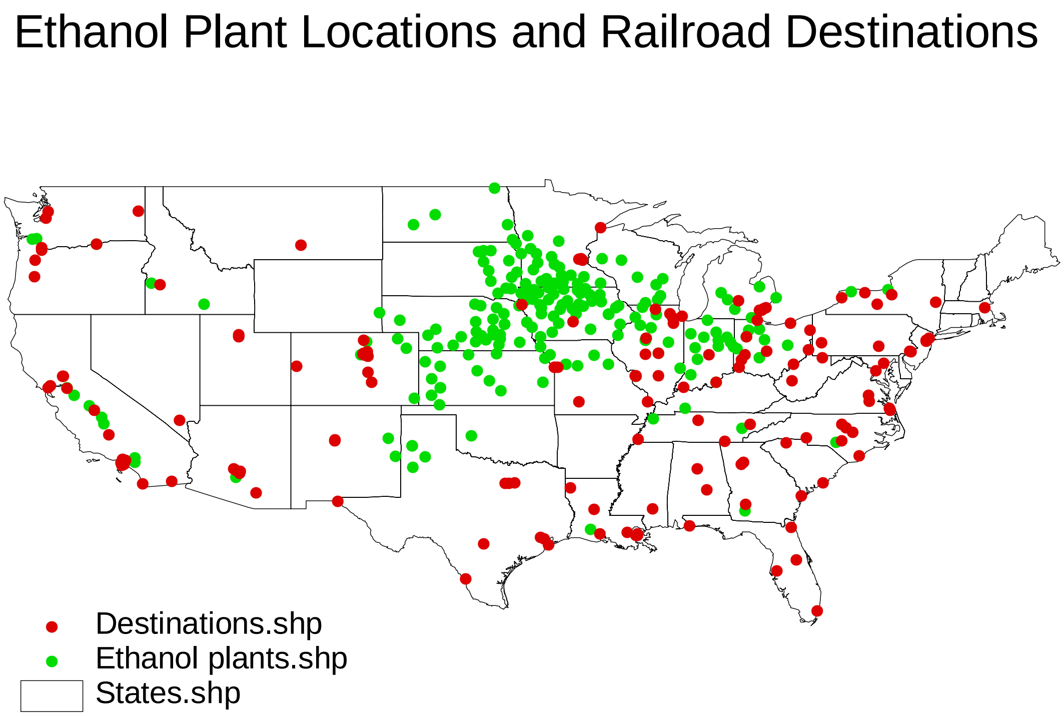


Figure 2: Shortest distance route from Iowa Falls, IA to West Sacramento, CA on Union Pacific's rail network

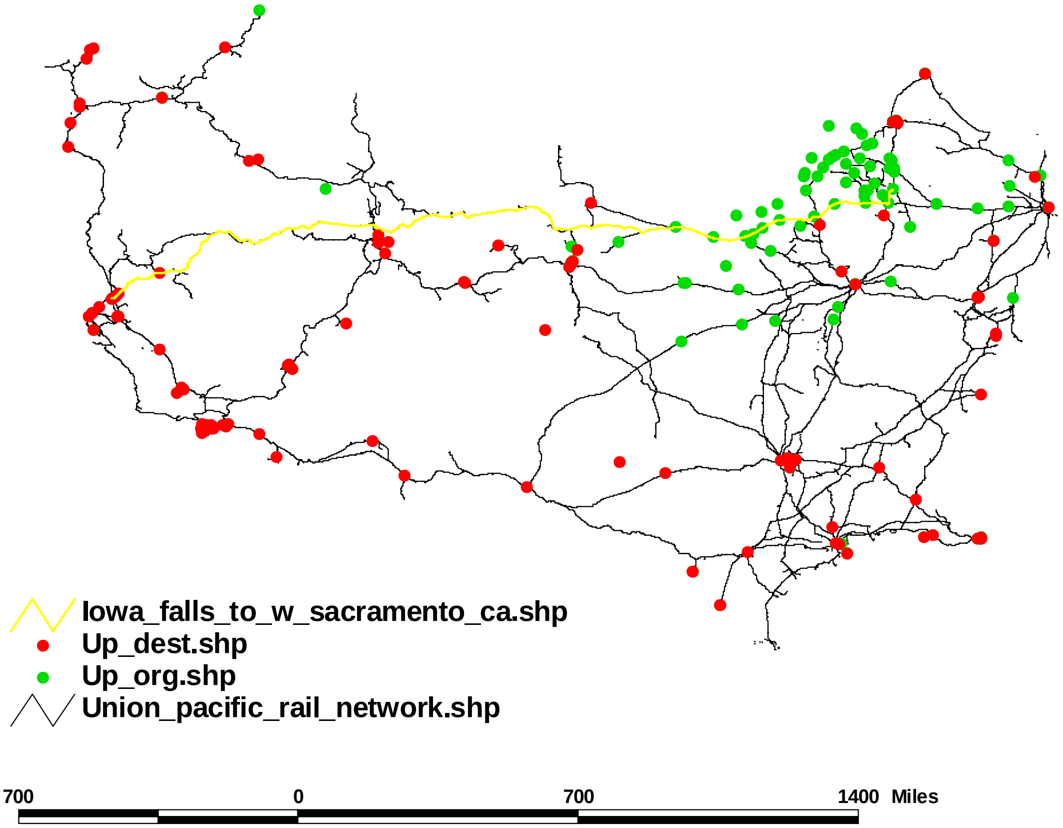


Figure 3: Ratio of calculated miles to reported rail miles

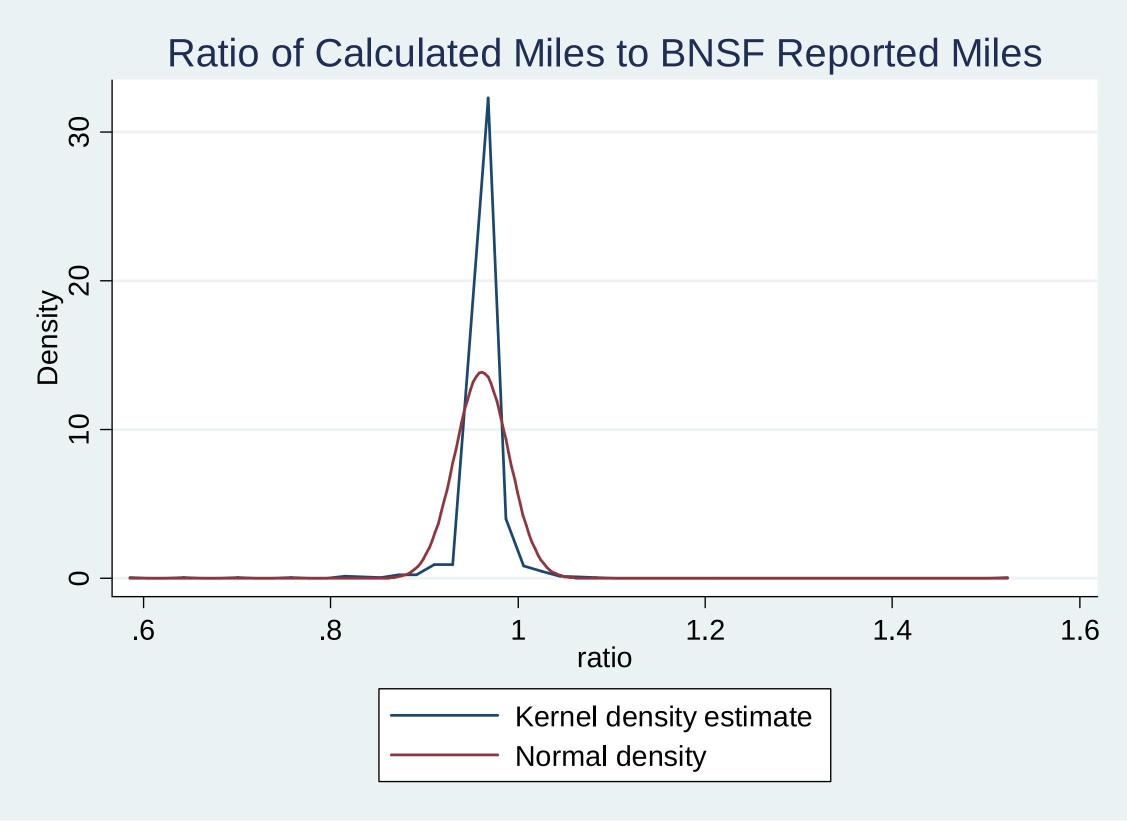
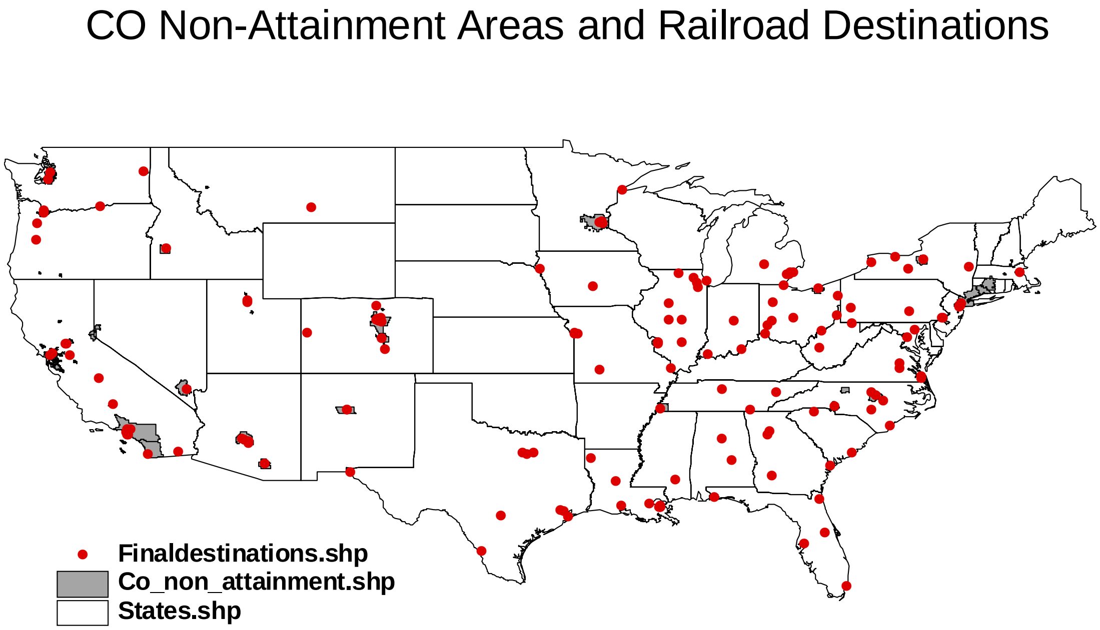


Figure 4: CO non-attainment areas and ethanol destinations



# 10 Tables

Table 1: Summary statistics for sample restrictions

| | Full Sample | On Rail Network | Final Sample |
|---|---|---|---|
| Num. Obs. | | | |
| BNSF | 11,280 | 8,059 | 2,597 |
| CN | 76 | 58 | 36 |
| CPR | 4 | 4 | 4 |
| CSX | 7,224 | 4,828 | 318 |
| KCS | 20 | 20 | 8 |
| NS | 86 | 84 | 69 |
| UP | 6,432 | 5,266 | 2,145 |
| Total Obs. | 25,122 | 18,319 | 5,177 |
| Price | | | |
| num. obs. | 25,098 | 18,295 | 5,165 |
| mean | $4,336.06 | $4,415.23 | $4,631.87 |
| std. dev. | $1,219.52 | $1,227.70 | $1,130.96 |
| min | $803.00 | $803.00 | $816.00 |
| max | $7,206.00 | $7,070.00 | $7,025.00 |
| Miles | | | |
| num. obs. | 18,401 | 18,319 | 5,177 |
| mean | 1139.4 | 1144.5 | 1207.2 |
| std. dev. | 592.7 | 589.1 | 597.4 |
| min | 0.0 | 1.5 | 3.6 |
| max | 2670.5 | 2670.5 | 2670.5 |

Notes: Price data not available for CPR and KCS

Table 2: Summary statistics

| Variable | Obs. | Mean | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Price | 5165 | 4631.87 | 1130.96 | 816.00 | 7025.00 |
| Miles | 5165 | 1208.40 | 597.49 | 3.59 | 2670.50 |
| Miles_sq | 5165 | 1817149.00 | 1490376.00 | 12.88 | 7131576.00 |
| Num_RR_org | 5165 | 1.32 | 0.57 | 1.00 | 3.00 |
| Num_RR_dest | 5165 | 1.68 | 0.86 | 1.00 | 6.00 |
| O_interchng | 5165 | 0.15 | 0.36 | 0.00 | 1.00 |
| D_interchng | 5165 | 0.10 | 0.31 | 0.00 | 1.00 |
| Num_RR_route | 5165 | 1.08 | 0.28 | 1.00 | 2.00 |
| RTD | 5165 | 4.77 | 0.82 | 1.00 | 6.93 |
| DSC | 5165 | 45.60 | 7.93 | 1.00 | 53.00 |
| EtOH_cap | 4447 | 68.50 | 52.98 | 4.00 | 345.07 |
| EtOH_mkt_shr | 4447 | 0.02 | 0.03 | 0.00 | 0.08 |
| CO_NA | 5165 | 0.64 | 0.48 | 0.00 | 1.00 |
| O_pop_1900 | 5068 | 0.08 | 0.27 | 0.00 | 1.84 |
| O_vcrops_1900 | 5068 | 49.28 | 30.74 | 0.44 | 195.35 |
| O_vlstk_1900 | 5068 | 41.05 | 21.01 | 7.11 | 117.05 |
| O_vman_1900 | 4968 | 626.23 | 3007.83 | 0.00 | 20044.48 |
| O_corn_2002 | 5076 | 14.81 | 10.96 | 0.00 | 45.74 |
| O_pop_2000 | 5165 | 0.25 | 0.88 | 0.00 | 5.30 |
| O_vcrops_2002 | 5165 | 72.29 | 46.03 | 0.00 | 238.92 |
| O_vlstk_2000 | 5165 | 104.72 | 175.54 | 0.00 | 1076.44 |
| O_vman_2002 | 5165 | 4636.31 | 14384.99 | 0.00 | 78208.32 |
| D_pop_1900 | 4634 | 0.13 | 0.30 | 0.00 | 1.84 |
| D_vcrops_1900 | 4634 | 65.19 | 49.73 | 2.36 | 184.96 |
| D_vlstk_1900 | 4634 | 35.54 | 19.13 | 0.03 | 87.82 |
| D_vman_1900 | 4634 | 1024.87 | 3233.55 | 2.63 | 20044.48 |
| D_refinery_dist | 5165 | 19.81 | 31.71 | 0.28 | 194.18 |
| D_pop_2000 | 5152 | 0.80 | 1.50 | 0.01 | 5.30 |
| D_vcrops_2002 | 5152 | 83.04 | 63.05 | 0.00 | 389.23 |
| D_vlstk_2002 | 5152 | 66.17 | 118.08 | 0.00 | 1076.44 |
| D_vman_2002 | 5152 | 13679.18 | 26625.81 | 0.00 | 78208.32 |

Table 3: Railroad participation at route endpoints

| | Num. Obs. | Percent of Total | Std. Dev. |
|---|---|---|---|
| Origins | | | |
| One RR | 5,165 | 73.8% | 0.44 |
| Two RR | 5,165 | 20.6% | 0.40 |
| Three RR | 5,165 | 5.6% | 0.23 |
| Four RR | 5,165 | 0.0% | 0.00 |
| Five RR | 5,165 | 0.0% | 0.00 |
| Six RR | 5,165 | 0.0% | 0.00 |
| Destination | | | |
| One RR | 5,165 | 44.8% | 0.50 |
| Two RR | 5,165 | 49.3% | 0.50 |
| Three RR | 5,165 | 3.6% | 0.19 |
| Four RR | 5,165 | 0.0% | 0.00 |
| Five RR | 5,165 | 0.0% | 0.00 |
| Six RR | 5,165 | 2.3% | 0.15 |

Table 4a: Base model results

| | OLS | OLSb | 2SLS |
|---|---|---|---|
| Num_RR_org | -70.05*** | -60.46*** | -150.8*** |
| | (7.318) | (7.515) | (26.43) |
| Num_RR_dest | -17.82** | -11.24 | -6.357 |
| | (7.005) | (7.190) | (8.548) |
| RR==CN | -2344.0*** | -2302.5*** | -2242.6*** |
| | (187.0) | (172.2) | (174.2) |
| RR==CSX | -1080.9*** | -1191.1*** | -946.5*** |
| | (148.4) | (153.4) | (154.0) |
| RR==NS | -1301.2*** | -1249.0*** | -1078.9*** |
| | (276.9) | (267.6) | (262.1) |
| RR==UP | -163.4*** | -159.4*** | -142.0** |
| | (53.65) | (60.72) | (61.10) |
| Miles | 1.108*** | 1.060*** | 1.063*** |
| | (0.0647) | (0.0659) | (0.0664) |
| (RR==CN)*miles | 3.285*** | 3.039*** | 2.908*** |
| | (0.825) | (0.767) | (0.752) |
| (RR==CSX)*miles | 0.883*** | 0.905*** | 0.915*** |
| | (0.246) | (0.259) | (0.257) |
| (RR==NS)*miles | 2.622** | 1.730* | 2.023** |
| | (1.076) | (1.035) | (0.998) |
| (RR==UP)*miles | 0.813*** | 0.826*** | 0.782*** |
| | (0.0938) | (0.100) | (0.0995) |
| Miles_sq | 0.645*** | 0.467** | 0.479** |
| | (0.224) | (0.217) | (0.216) |
| (RR==CN)*miles_sq | -15.26* | -12.96* | -11.13 |
| | (8.246) | (7.845) | (7.640) |
| (RR==CSX)*miles_sq | -1.729 | -1.281 | -1.260 |
| | (1.711) | (1.872) | (1.855) |
| (RR==NS)*miles_sq | -11.11 | -2.024 | -5.299 |
| | (9.464) | (8.997) | (8.704) |
| (RR==UP)*miles_sq | -2.307*** | -2.478*** | -2.230*** |
| | (0.357) | (0.378) | (0.376) |
| RTD | -54.09*** | -44.59*** | -39.34*** |
| | (8.089) | (8.893) | (9.287) |
| O_interchng | 127.8*** | 137.2*** | 232.7*** |
| | (18.77) | (18.46) | (25.65) |
| D_interchng | 96.74*** | 85.97*** | 90.38*** |
| | (22.93) | (21.43) | (21.67) |
| state effects | Yes | Yes | Yes |
| Observations | 5165 | 4413 | 4413 |
| $R^2$ | 0.958 | 0.963 | 0.963 |

Standard errors clustered on origin price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4b: Base model results - continued

| | OLS | OLSb | 2SLS |
|---|---|---|---|
| O_pop_2005 | | | -95.86** |
| | | | (41.84) |
| O_vcrops_2002 | | | -90.03 |
| | | | (233.1) |
| O_vlstk_2002 | | | 19.76 |
| | | | (56.00) |
| O_vman_2002 | | | 3.874 |
| | | | (2.569) |
| D_pop_2005 | | | -149.4*** |
| | | | (27.32) |
| D_vcrops_2002 | | | -321.4 |
| | | | (206.1) |
| D_vlstk_2002 | | | -82.81* |
| | | | (46.91) |
| D_vman_2002 | | | 9.146*** |
| | | | (1.704) |
| O_corn_2002 | | | -399.4 |
| | | | (776.1) |
| D_refinery_dist | | | -0.435** |
| | | | (0.182) |
| Constant | 3096.7*** | 3154.2*** | 2946.1*** |
| | (93.82) | (101.6) | (107.1) |
| state effects | Yes | Yes | Yes |
| Observations | 5165 | 4413 | 4413 |
| $R^2$ | 0.958 | 0.963 | 0.963 |

Standard errors clustered on origin price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5a: Nonparametric competition effects

| | OLS | OLSb | 2SLS |
|---|---|---|---|
| Two_RR_org | -42.41*** | -27.63*** | -104.0*** |
| | (9.410) | (9.465) | (40.06) |
| Three_RR_org | -201.6*** | -192.2*** | -445.5*** |
| | (21.96) | (23.35) | (150.2) |
| Two_RR_dest | -17.86* | -3.054 | -10.64 |
| | (9.311) | (10.22) | (87.53) |
| Three_RR_dest | -80.08** | -67.15* | -474.5*** |
| | (39.06) | (37.46) | (127.9) |
| Six_RR_dest | -79.14* | -60.06 | -8.929 |
| | (43.93) | (43.20) | (58.68) |
| RTD | -53.10*** | -42.91*** | -36.23*** |
| | (8.082) | (8.948) | (11.50) |
| O_interchng | 106.3*** | 110.8*** | 177.6*** |
| | (19.21) | (19.03) | (48.40) |
| D_interchng | 99.58*** | 92.62*** | 144.8*** |
| | (24.30) | (22.51) | (27.75) |
| Mile X RR effects | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes |
| Observations | 5165 | 4413 | 4413 |
| $R^2$ | 0.958 | 0.963 | 0.961 |

Standard errors clustered on origin price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5b: Nonparametric competition effects - continued

| | OLS | OLSb | 2SLS |
|---|---|---|---|
| O_pop_2005 | | | -20.54 |
| | | | (83.56) |
| O_vcrops_2002 | | | -258.5 |
| | | | (270.4) |
| O_vlstk_2002 | | | 16.54 |
| | | | (55.91) |
| O_vman_2002 | | | 0.443 |
| | | | (4.221) |
| D_pop_2005 | | | -149.4*** |
| | | | (37.51) |
| D_vcrops_2002 | | | 24.28 |
| | | | (363.0) |
| D_vlstk_2002 | | | -138.9** |
| | | | (68.28) |
| D_vman_2002 | | | 9.355*** |
| | | | (2.001) |
| O_corn_2002 | | | 4.805 |
| | | | (861.9) |
| D_refinery_dist | | | -0.547 |
| | | | (0.384) |
| Constant | 3028.1*** | 3094.9*** | 3238.7*** |
| | (98.08) | (106.8) | (169.5) |
| Mile X RR effects | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes |
| Observations | 5165 | 4413 | 4413 |
| $R^2$ | 0.958 | 0.963 | 0.961 |

Standard errors clustered on origin price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Price effects for close competitors in duopoly markets

| | Base Model | Close Comp. |
|---|---|---|
| Two_RR_org | -104.0*** | -687.9* |
| | (40.06) | (368.3) |
| O_close_comp | | 854.3 |
| | | (547.4) |
| Three_RR_org | -445.5*** | -656.1* |
| | (150.2) | (349.8) |
| Two_RR_dest | -10.64 | -795.5 |
| | (87.53) | (636.0) |
| D_close_comp | | 1167.5 |
| | | (933.9) |
| Three_RR_dest | -474.5*** | -1523.6* |
| | (127.9) | (867.4) |
| Six_RR_dest | -8.929 | -127.3 |
| | (58.68) | (133.7) |
| Mile X RR effects | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes |
| state effects | Yes | Yes |
| Observations | 4413 | 4413 |
| $R^2$ | 0.961 | 0.917 |

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7a: Price discrimination based on CO non-attainment status

| | OLS | 2SLS | 2SLS |
|---|---|---|---|
| Num_RR_org | -60.16*** | -146.6*** | -152.3*** |
| | (11.90) | (47.36) | (47.83) |
| Num_RR_dest | -10.34 | -5.113 | 4.226 |
| | (6.425) | (6.136) | (6.249) |
| CO_NA | 147.4*** | 248.8*** | 276.6*** |
| | (24.11) | (49.28) | (49.20) |
| RR_dest_CO_NA | -31.72*** | -100.5*** | -90.97*** |
| | (11.09) | (27.57) | (27.33) |
| RTD | -43.10*** | -39.20*** | |
| | (9.698) | (10.37) | |
| O_interchng | 137.2*** | 228.8*** | 238.7*** |
| | (28.24) | (42.98) | (42.24) |
| D_interchng | 96.27*** | 112.8*** | 110.5*** |
| | (13.67) | (12.52) | (12.65) |
| Org_density | | | 2.589 |
| | | | (2.547) |
| Dest_density | | | 19.17*** |
| | | | (1.750) |
| Mile X RR effects | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes |
| Observations | 4413 | 4413 | 4413 |
| $R^2$ | 0.963 | 0.963 | 0.963 |

Standard errors clustered on destination price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7b: Price discrimination based on CO non-attainment status - continued

| | OLS | 2SLS | 2SLS |
|---|---|---|---|
| O_pop_2005 | | -96.17 | -106.8 |
| | | (71.40) | (73.35) |
| O_vcrops_2002 | | -90.08 | -59.43 |
| | | (427.2) | (422.8) |
| O_vlstk_2002 | | 17.02 | 33.39 |
| | | (103.4) | (103.0) |
| O_vman_2002 | | 3.906 | 4.621 |
| | | (4.660) | (4.793) |
| D_pop_2005 | | -142.0*** | -115.8*** |
| | | (25.38) | (25.02) |
| D_vcrops_2002 | | -416.3** | -300.9 |
| | | (209.9) | (205.4) |
| D_vlstk_2002 | | -77.18* | -111.8*** |
| | | (40.37) | (39.46) |
| D_vman_2002 | | 8.704*** | 6.686*** |
| | | (1.628) | (1.607) |
| O_corn_2002 | | -410.3 | -501.1 |
| | | (1345.8) | (1347.7) |
| D_refinery_dist | | -0.292** | 0.0649 |
| | | (0.115) | (0.127) |
| Constant | 3149.7*** | 2944.6*** | 2666.0*** |
| | (116.1) | (121.8) | (109.7) |
| Mile X RR effects | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes |
| Observations | 4413 | 4413 | 4413 |
| $R^2$ | 0.963 | 0.963 | 0.963 |

Standard errors clustered on destination price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8a: Two-stage least squares models

| | 2SLS1 | 2SLS2 | 2SLS3 | 2SLS4 |
|---|---|---|---|---|
| Num_RR_org | -55.15 | -57.23 | -155.6*** | -150.8*** |
| | (37.87) | (37.00) | (27.14) | (26.43) |
| Num_RR_dest | -13.15 | -14.60 | -5.268 | -6.357 |
| | (8.826) | (8.920) | (8.487) | (8.548) |
| RR==CN | -2290.9*** | -2295.3*** | -2235.4*** | -2242.6*** |
| | (161.2) | (159.0) | (176.5) | (174.2) |
| RR==CSX | -956.1*** | -942.5*** | -957.6*** | -946.5*** |
| | (152.2) | (152.2) | (154.0) | (154.0) |
| RR==NS | -1096.1*** | -1087.9*** | -1082.2*** | -1078.9*** |
| | (263.2) | (265.0) | (260.8) | (262.1) |
| RR==UP | -152.2** | -154.5** | -136.4** | -142.0** |
| | (61.27) | (62.02) | (60.47) | (61.10) |
| Miles | 1.130*** | 1.136*** | 1.058*** | 1.063*** |
| | (0.0668) | (0.0667) | (0.0666) | (0.0664) |
| (RR==CN)*miles | 3.011*** | 3.040*** | 2.879*** | 2.908*** |
| | (0.734) | (0.727) | (0.759) | (0.752) |
| (RR==CSX)*miles | 0.883*** | 0.875*** | 0.920*** | 0.915*** |
| | (0.258) | (0.257) | (0.258) | (0.257) |
| (RR==NS)*miles | 1.957* | 1.963* | 2.004** | 2.023** |
| | (1.005) | (1.009) | (0.995) | (0.998) |
| (RR==UP)*miles | 0.843*** | 0.849*** | 0.772*** | 0.782*** |
| | (0.0998) | (0.101) | (0.0983) | (0.0995) |
| Miles_sq | 0.459** | 0.451** | 0.481** | 0.479** |
| | (0.226) | (0.225) | (0.217) | (0.216) |
| (RR==CN)*miles_sq | -12.93* | -13.14* | -10.90 | -11.13 |
| | (7.569) | (7.518) | (7.695) | (7.640) |
| (RR==CSX)*miles_sq | -1.544 | -1.532 | -1.272 | -1.260 |
| | (1.865) | (1.858) | (1.862) | (1.855) |
| (RR==NS)*miles_sq | -5.075 | -5.232 | -5.033 | -5.299 |
| | (8.735) | (8.762) | (8.690) | (8.704) |
| (RR==UP)*miles_sq | -2.546*** | -2.577*** | -2.188*** | -2.230*** |
| | (0.377) | (0.384) | (0.372) | (0.376) |
| RTD | -52.66*** | -52.63*** | -39.39*** | -39.34*** |
| | (9.381) | (9.453) | (9.253) | (9.287) |
| O_interchng | 143.7*** | 144.0*** | 237.5*** | 232.7*** |
| | (36.64) | (35.98) | (25.98) | (25.65) |
| D_interchng | 96.79*** | 99.45*** | 88.82*** | 90.38*** |
| | (23.88) | (23.80) | (21.83) | (21.67) |
| state effects | Yes | Yes | Yes | Yes |
| Observations | 4543 | 4509 | 4447 | 4413 |
| $R^2$ | 0.960 | 0.960 | 0.963 | 0.963 |

Standard errors clustered on origin price-group in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8b: Two-stage least squares models - continued

| | 2SLS1 | 2SLS2 | 2SLS3 | 2SLS4 |
|---|---|---|---|---|
| O_pop_2005 | -48.19*** | -48.13*** | -90.18** | -95.86** |
| | (10.44) | (10.46) | (42.21) | (41.84) |
| O_vcrops_2002 | -313.3** | -241.6 | -159.3 | -90.03 |
| | (127.8) | (182.2) | (139.1) | (233.1) |
| O_vlstk_2002 | -31.59 | -30.24 | 17.29 | 19.76 |
| | (39.83) | (37.70) | (51.61) | (56.00) |
| D_pop_2005 | -107.1*** | -113.3*** | -140.3*** | -149.4*** |
| | (24.98) | (24.95) | (27.04) | (27.32) |
| D_vman_2002 | 6.768*** | 7.096*** | 8.653*** | 9.146*** |
| | (1.580) | (1.580) | (1.688) | (1.704) |
| O_corn_2002 | | -355.4 | | -399.4 |
| | | (669.4) | | (776.1) |
| D_refinery_dist | | -0.449** | | -0.435** |
| | | (0.195) | | (0.182) |
| O_vman_2002 | | | 3.595 | 3.874 |
| | | | (2.571) | (2.569) |
| D_vcrops_2002 | | | -257.1 | -321.4 |
| | | | (202.4) | (206.1) |
| D_vlstk_2002 | | | -93.37** | -82.81* |
| | | | (46.69) | (46.91) |
| Constant | 2979.3*** | 2983.3*** | 2941.0*** | 2946.1*** |
| | (106.6) | (106.8) | (107.3) | (107.1) |
| state effects | Yes | Yes | Yes | Yes |
| Observations | 4543 | 4509 | 4447 | 4413 |
| $R^2$ | 0.960 | 0.960 | 0.963 | 0.963 |

Standard errors clustered on origin price-group
in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9a: First stage: two-stage least squares models

| | 2SLS1 org. | 2SLS1 dest. | 2SLS2 org. | 2SLS2 dest. | 2SLS3 org. | 2SLS3 dest. | 2SLS4 org. | 2SLS4 dest. |
|---|---|---|---|---|---|---|---|---|
| RTD | 0.0345*** | -0.0482*** | 0.0300*** | -0.0492*** | 0.0485*** | -0.0505*** | 0.0446*** | -0.0509*** |
| | (0.0115) | (0.0110) | (0.0114) | (0.0112) | (0.0116) | (0.0114) | (0.0116) | (0.0117) |
| O_interchng | 0.748*** | -0.00311 | 0.746*** | -0.00475 | 0.746*** | -0.0121 | 0.739*** | -0.0128 |
| | (0.0234) | (0.0278) | (0.0236) | (0.0270) | (0.0214) | (0.0288) | (0.0219) | (0.0279) |
| D_interchng | 0.00367 | 0.248*** | 0.00322 | 0.249*** | -0.000554 | 0.246*** | -0.000595 | 0.247*** |
| | (0.0274) | (0.0434) | (0.0275) | (0.0463) | (0.0282) | (0.0446) | (0.0285) | (0.0475) |
| O_pop_1900 | 0.640*** | 0.00395 | 0.660*** | 0.00751 | 3.418*** | 0.142 | 3.452*** | 0.148 |
| | (0.0620) | (0.0603) | (0.0619) | (0.0602) | (0.358) | (0.304) | (0.358) | (0.301) |
| O_vcrops_1900 | -0.291 | 0.199 | -0.584 | 0.178 | -4.965*** | 0.651 | -5.675*** | 0.632 |
| | (0.456) | (0.427) | (0.497) | (0.448) | (0.667) | (0.521) | (0.773) | (0.613) |
| O_vlstk_1900 | 6.532*** | -0.121 | 6.977*** | -0.0956 | 8.417*** | -0.419 | 9.243*** | -0.398 |
| | (0.580) | (0.594) | (0.617) | (0.615) | (0.662) | (0.614) | (0.748) | (0.682) |
| D_pop_1900 | -0.0711 | -0.552*** | -0.0502 | -0.911*** | -0.0562 | -0.684** | -0.0500 | -0.903*** |
| | (0.139) | (0.195) | (0.145) | (0.225) | (0.179) | (0.276) | (0.183) | (0.277) |
| D_vman_1900 | 0.00578 | 0.301*** | 0.00393 | 0.331*** | 0.00457 | 0.313*** | 0.00392 | 0.331*** |
| | (0.0128) | (0.0177) | (0.0133) | (0.0203) | (0.0159) | (0.0238) | (0.0163) | (0.0241) |
| O_corn_2002 | | | -0.485 | -0.0349 | | | -3.320*** | 0.208 |
| | | | (0.917) | (1.024) | | | (1.106) | (1.227) |
| D_refinery_dist | | | 0.000100 | -0.00270*** | | | 0.0000609 | -0.00277*** |
| | | | (0.000261) | (0.000385) | | | (0.000269) | (0.000386) |
| O_vman_1900 | | | | | -0.0604 | -0.0351 | -0.0598 | -0.0347 |
| | | | | | (0.0422) | (0.0290) | (0.0423) | (0.0287) |
| D_vcrops_1900 | | | | | 0.0747 | -0.132 | 0.0853 | -0.507 |
| | | | | | (0.294) | (0.491) | (0.298) | (0.400) |
| D_vlstk_1900 | | | | | -0.152 | 2.020*** | -0.141 | 2.162*** |
| | | | | | (0.533) | (0.627) | (0.540) | (0.587) |
| Mile X RR effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4543 | 4543 | 4509 | 4509 | 4447 | 4447 | 4413 | 4413 |
| $R^2$ | 0.523 | 0.837 | 0.524 | 0.842 | 0.543 | 0.837 | 0.544 | 0.843 |
| F-stat excl. instr. | 73.6 | 1430.1 | 75.22 | 1336.05 | 77.99 | 863.05 | 75.15 | 804.00 |
| $R^2$ excl. instr. | 0.056 | 0.671 | 0.058 | 0.676 | 0.102 | 0.670 | 0.103 | 0.675 |

Standard errors clustered on origin price-group in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9b: First stage: two-stage least squares models - continued

| | 2SLS1 org. | 2SLS1 dest. | 2SLS2 org. | 2SLS2 dest. | 2SLS3 org. | 2SLS3 dest. | 2SLS4 org. | 2SLS4 dest. |
|---|---|---|---|---|---|---|---|---|
| O_pop_2005 | -0.0285*** | -0.0109 | -0.0338*** | -0.0115 | -2.244*** | 0.251 | -2.269*** | 0.243 |
| | (0.00684) | (0.0131) | (0.00713) | (0.0134) | (0.229) | (0.153) | (0.243) | (0.161) |
| O_vcrops_2002 | 0.438** | -0.119 | 0.519* | -0.104 | 1.956*** | -0.263 | 2.716*** | -0.286 |
| | (0.176) | (0.210) | (0.265) | (0.257) | (0.229) | (0.280) | (0.401) | (0.416) |
| O_vlstk_2002 | -0.0726** | 0.0249 | -0.104*** | 0.0204 | -0.0902 | 0.0347 | -0.225*** | 0.0293 |
| | (0.0353) | (0.0416) | (0.0358) | (0.0394) | (0.0776) | (0.0801) | (0.0846) | (0.0908) |
| D_pop_2005 | -0.0159 | 0.0442 | -0.0138 | 0.00419 | -0.0239 | 0.0689 | -0.0219 | 0.0144 |
| | (0.0312) | (0.0385) | (0.0313) | (0.0375) | (0.0327) | (0.0420) | (0.0331) | (0.0418) |
| D_vman_2002 | 0.000905 | -0.00539** | 0.000805 | -0.00319 | 0.00132 | -0.00742*** | 0.00121 | -0.00433* |
| | (0.00188) | (0.00223) | (0.00189) | (0.00217) | (0.00196) | (0.00246) | (0.00197) | (0.00244) |
| O_vman_2002 | | | | | 0.108*** | -0.0127* | 0.109*** | -0.0123 |
| | | | | | (0.0112) | (0.00752) | (0.0119) | (0.00787) |
| D_vcrops_2002 | | | | | -0.155 | -0.358 | -0.145 | -0.498** |
| | | | | | (0.225) | (0.257) | (0.228) | (0.246) |
| D_vlstk_2002 | | | | | 0.0395 | 0.242*** | 0.0369 | 0.243*** |
| | | | | | (0.0731) | (0.0607) | (0.0738) | (0.0551) |
| Constant | -0.174 | 2.766*** | -0.155 | 2.814*** | 0.190 | 2.662*** | 0.209 | 2.713*** |
| | (0.110) | (0.150) | (0.110) | (0.147) | (0.136) | (0.157) | (0.137) | (0.152) |
| Mile X RR effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 4543 | 4543 | 4509 | 4509 | 4447 | 4447 | 4413 | 4413 |
| $R^2$ | 0.523 | 0.837 | 0.524 | 0.842 | 0.543 | 0.837 | 0.544 | 0.843 |
| F-stat excl. instr. | 73.6 | 1430.1 | 75.22 | 1336.05 | 77.99 | 863.05 | 75.15 | 804.00 |
| $R^2$ excl. instr. | 0.056 | 0.671 | 0.058 | 0.676 | 0.102 | 0.670 | 0.103 | 0.675 |

Standard errors clustered on origin price-group in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## Table 10a: Robustness checks

| | Base Model | RC DSC | EtOH Plants Only | RC EtOH Plants | RC route |
|---|---|---|---|---|---|
| Num_RR_org | -150.8*** | -159.7*** | -77.00*** | -64.09*** | |
| | (26.43) | (26.86) | (19.84) | (23.50) | |
| Num_RR_dest | -6.357 | -6.139 | -10.20 | -10.34 | |
| | (8.548) | (8.615) | (8.195) | (8.155) | |
| RTD | -39.34*** | -38.42*** | -41.89*** | -42.03*** | -46.13*** |
| | (9.287) | (9.261) | (10.11) | (10.16) | (9.218) |
| DSC | | -5.423** | | | |
| | | (2.188) | | | |
| EtOH_cap | | | | -0.117 | |
| | | | | (0.0722) | |
| EtOH_mkt_shr | | | | -212.4* | |
| | | | | (126.4) | |
| Num_RR_route | | | | | -269.5*** |
| | | | | | (79.68) |
| O_interchng | 232.7*** | 232.8*** | 89.06 | 125.6* | 223.3*** |
| | (25.65) | (25.53) | (68.97) | (73.36) | (35.85) |
| D_interchng | 90.38*** | 89.46*** | 93.90*** | 93.96*** | 98.53*** |
| | (21.67) | (21.78) | (21.93) | (21.85) | (18.45) |
| Mile X RR effects | Yes | Yes | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 4413 | 4413 | 3784 | 3784 | 4413 |
| $R^2$ | 0.963 | 0.963 | 0.968 | 0.968 | 0.961 |

Standard errors clustered on origin price-group in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10b: Robustness checks - continued

| | Base Model | RC DSC | EtOH Plants Only | RC EtOH Plants | RC route |
|---|---|---|---|---|---|
| O_pop_2005 | -95.86** | -85.86** | 1688.2*** | 1494.1*** | -130.3*** |
| | (41.84) | (42.27) | (185.6) | (201.5) | (44.17) |
| O_vcrops_2002 | -90.03 | -32.85 | -520.9** | -567.4** | -29.92 |
| | (233.1) | (232.2) | (233.7) | (236.8) | (251.2) |
| O_vlstk_2002 | 19.76 | 33.37 | -40.06 | -24.14 | 35.39 |
| | (56.00) | (57.68) | (52.62) | (53.60) | (56.63) |
| O_vman_2002 | 3.874 | 3.379 | -45.56*** | -38.76*** | 4.119 |
| | (2.569) | (2.556) | (5.433) | (6.127) | (2.701) |
| D_pop_2005 | -149.4*** | -154.3*** | -149.7*** | -148.1*** | -137.7*** |
| | (27.32) | (27.71) | (29.99) | (29.65) | (27.28) |
| D_vcrops_2002 | -321.4 | -343.8* | -618.5*** | -618.5*** | -347.7* |
| | (206.1) | (207.1) | (223.6) | (223.4) | (206.9) |
| D_vlstk_2002 | -82.81* | -80.38* | -22.59 | -22.28 | -66.10 |
| | (46.91) | (47.20) | (49.71) | (49.78) | (49.46) |
| D_vman_2002 | 9.146*** | 9.460*** | 8.943*** | 8.831*** | 8.430*** |
| | (1.704) | (1.729) | (1.842) | (1.819) | (1.737) |
| O_corn_2002 | -399.4 | -384.6 | 656.8 | 908.4 | -1225.9 |
| | (776.1) | (771.5) | (726.8) | (749.3) | (851.3) |
| D_refinery_dist | -0.435** | -0.441** | -0.318* | -0.319* | -0.526*** |
| | (0.182) | (0.184) | (0.192) | (0.191) | (0.177) |
| Constant | 2946.1*** | 3181.2*** | 3556.1*** | 2881.9*** | 3072.0*** |
| | (107.1) | (149.1) | (371.9) | (79.72) | (107.0) |
| Mile X RR effects | Yes | Yes | Yes | Yes | Yes |
| Mile_sq X RR effects | Yes | Yes | Yes | Yes | Yes |
| state effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 4413 | 4413 | 3784 | 3784 | 4413 |
| $R^2$ | 0.963 | 0.963 | 0.968 | 0.968 | 0.961 |

Standard errors clustered on origin price-group in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Alternate models for clustering

| | Cluster route | Cluster org. group | Cluster dest. group | Cluster orig. state | Cluster dest. state | Cluster railroad |
|---|---|---|---|---|---|---|
| num_RR_org | -146.6*** | -146.6*** | -146.6*** | -146.6*** | -146.6*** | -146.6*** |
| | (26.96) | (26.27) | (47.36) | (54.59) | (31.44) | (21.91) |
| num_RR_dest | -5.113 | -5.113 | -5.113 | -5.113 | -5.113 | -5.113 |
| | (6.772) | (8.872) | (6.136) | (6.105) | (16.68) | (10.05) |
| RR_dest_CO_NA | -100.5*** | -100.5*** | -100.5*** | -100.5*** | -100.5 | -100.5*** |
| | (26.43) | (32.02) | (27.57) | (30.91) | (69.85) | (18.96) |
| CO_NA | 248.8*** | 248.8*** | 248.8*** | 248.8*** | 248.8* | 248.8*** |
| | (47.95) | (57.79) | (49.28) | (41.56) | (130.0) | (58.59) |
| $N$ | 4413 | 4413 | 4413 | 4413 | 4413 | 4413 |
| N_clust | 4195 | 2846 | 2040 | 19 | 33 | 5 |

Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Gains to railroads: California example

| | Ethanol Quantity (million gallons) | Revenues ($ millon/year) | Total Variable Costs ($ millon/year) | Estimated Producer Surplus ($ millon/year) |
|---|---|---|---|---|
| California gasoline 5.7% ethanol | 893 | $ 168.1 | $ 132.9 | $ 35.1 |
| California gasoline 10% ethanol | 1,567 | $ 294.9 | $ 233.2 | $ 61.6 |
| Difference | 674 | $ 126.8 | $ 100.3 | $ 26.5 |
| Number of routes | 1462 | | | |