Investment in Corn-Ethanol Plants in the Midwestern United States*

Karen E. Thome C.-Y. Cynthia Lin Lawell

October 9, 2015

Abstract

Ethanol has attracted considerable policy attention both for its use as a gasoline substitute, and as a way to enhance profits in rural areas. In this paper we examine how economic factors, government policies, and strategic interactions affect decisions about whether and when to invest in building a new ethanol plant. We model the decision to invest in ethanol plants at the county level using both reduced-form discrete response models and a structural model of a dynamic game. We focus on investment in corn-ethanol plants in the Midwestern United States, where the majority of corn in the US is grown, over the period 1996-2008. According to the results of the structural model, our preferred model, the intensity of corn production; government policies, particularly the MTBE ban and the 2007 Renewable Fuel Standard (RFS2); and private information shocks all have significant effects on ethanol investment payoffs and decisions. We use the estimated structural parameters to simulate counterfactual policy scenarios to disentangle the impacts of state and national policies on the timing and location of investment in the industry. We find that, of the policies analyzed, the MTBE ban and the RFS2 led to most of the investment during this time period.

Keywords: ethanol, biofuels, investment timing game
JEL codes: Q16, L13

*Thome: USDA Economic Research Service; karen.thome@ers.usda.gov. Lin Lawell: University of California at Davis; cclin@primal.ucdavis.edu. We thank Bruce Babcock, Timothy Fitzgerald, and Rich Sexton for helpful comments. We also benefited from comments from conference participants at the Association of Environmental and Resource Economists Conference, the American Economic Association annual meeting, the International Industrial Organization Conference, the Berkeley Bioeconomy Conference, and the Stanford Berkeley Industrial Organization Conference; and seminar participants at Duke University and Rice University. This research was supported by a UC-Davis Chevron Research Grant. Lin Lawell is a member of the Giannini Foundation of Agricultural Economics. The views expressed are those of the authors and should not be attributed to the Economic Research Service or USDA. All errors are our own.

1 Introduction

Ethanol has attracted considerable policy attention both as an environmentally-friendly alternative to imported oil, and as a way to boost farm profits and improve rural livelihoods. Several policies, including the Renewable Fuel Standard and state-level tax credits, actively promote ethanol production and have coincided with a boom in the construction of corn-ethanol plants beginning in the mid-1990s. This paper focuses on the investment decisions of ethanol-producing firms in the Midwestern United States during the period from 1996 to 2008. We use reduced-form and structural models to analyze how economic factors, government policy, and strategic interactions affect decisions about whether and when to invest in building new ethanol plants. We also use the estimated structural parameters to simulate the effects of various counterfactual policy scenarios on investment in ethanol plants.

Fuel ethanol can play different roles in the energy market, as an energy substitute for gasoline, or as an additive (oxygenate and/or octane booster) to gasoline. Fuel ethanol has been in use in the United States since the time of the Model T Ford, the original flex-fuel vehicle. The first US ethanol boom stemmed from the desire for more energy self-sufficiency resulting from the oil embargoes in 1973 and 1979. The resulting legislation, in the form of federal income tax credits and blenders' credits that continued through 2011 and the phase out of leaded gasoline, led to the construction of 153 new plants by 1985 (DOE, 2008). These plants were tiny by today's standards, with an average capacity of 8 million gallons per year, and by 1991 only 35 of these plants were still operational due to poor business judgment and bad engineering (DOE, 2008; Urbanchuck, 2006).

The second US ethanol boom began in the mid-1990s and hit full-stride by the early 2000s. The Clean Air Act of 1990 mandated use of oxygenates, which include ethanol, in gasoline. The subsequent phase out and ban of the oxygenate MTBE as an additive beginning in the late 1990s further increased demand for ethanol. Additionally, the Renewable Fuel Standard (RFS) was created under the Energy Policy Act of 2005 with the goal of accelerating the use of fuels derived from renewable sources (EPA, 2013). This initial RFS (RFS1) mandated that a minimum of 4 billion gallons be used in 2006, rising to 7.5 billion gallons by 2012. Two years later, the Energy Independence and Security Act of 2007 greatly expanded the biofuel mandate volumes, creating the RFS2, and extended the date through 2022. This 2007 Renewable Fuel Standard (RFS2) is a central component of US energy policy. It requires steadily increasing volumes of biofuel to be blended into the nation's fuel supply, reaching 37 billion gallons (bgal) a year by 2022. These federal policies coincided with increases in petroleum prices that made ethanol more competitive as an energy substitute for gasoline (Gallagher, 2009). Over this time period, the number of operational ethanol plants rose from 35 plants in 1991, to 50 plants in 1999, and to 192 plants in September of 2010, for a total capacity of 13 billion gallons per year.

We focus our analysis on investments made during the second US ethanol boom. The investment decision to build an ethanol plant is dynamic and may be affected by economic factors, government policies, and strategic

interactions. Because the payoff from investing in building a new ethanol plant depends on market conditions that vary stochastically over time, a potential entrant who hopes to make a dynamically optimal decision would need to account for the option value to waiting before making this irreversible investment (Dixit and Pindyck, 1994).

There are two sources of strategic interactions that add a strategic (or non-cooperative) dimension to potential entrants' investment timing decisions. The first is a competition effect: if there is more than one ethanol plant located in the same region, these plants may compete in the local feedstock input market or they may compete in the local fuel ethanol output market. The competition effect deters ethanol plants from entering in regions where there are other ethanol plants already present.

High transportation costs in both the feedstock and ethanol markets may be one reason for localized competition among neighboring plants. Feedstock is approximately 70% of the cost of producing corn-ethanol, and transportation costs account for bulky grains constitute a significant share (Whittington, 2006). As a consequence, the distance from a plant to the feedstock production area is extremely important. For example, Sarmiento and Wilson (2007) find that competition in feedstock procurement can lead to a negative competition effect in localized corn markets, and that a shift in demand from a new plant could increase corn feedstock prices. Thus, owing to high transportation costs, neighboring plants may compete in the local feedstock input market.

Fuel ethanol transportation is more difficult, and thus is more expensive, than gasoline transportation because ethanol can easily absorb water during the transportation process, and consequently must be transported using specialized tank trucks, unlike gasoline, which can be transported via pipelines. Rail is the primary form of transport used to ship ethanol from the Midwestern US to each coast. Rail transport has increasingly become congested given the growth in domestic crude oil production (EIA, 2015) and the small number of firms who operate most national rail routes. Neighboring ethanol plants may therefore compete over access to transportation for their ethanol output, leading to higher marketing costs for fuel ethanol. Thus, owing to high transportation costs, neighboring plants may also compete in the local fuel ethanol output market.

The second source of strategic interaction is an agglomeration effect; if there are several ethanol plants located in the same region, the existing plants may have developed transportation and marketing infrastructure and/or an educated work force from which entering plants can benefit (Goetz, 1997; Ellison and Glaeser, 1999; Lambert et al., 2008).

Because of the potential competition and agglomeration effects, the dynamic decision-making problem faced by the potential investors is not merely a single-agent problem but rather can be viewed as a multi-agent investment timing game in which plants behave strategically and base decisions on other investors' strategies. Because the investment decisions of others affect future values of state variables and the future payoff from investing in a new plant, potential investors must anticipate the investment strategies of others in order to make a dynamically optimal decision. Uncertainty over whether a plant might be constructed nearby is another reason there is an option value to waiting to

invest, and which makes the decision dynamic rather than static (Dixit and Pindyck, 1994).

We estimate two econometric models of potential investors' dynamic investment strategies. First, we estimate a reduced-form discrete response model that allows for comparison with other studies analyzing ethanol and manufacturing investment decisions and location choice. Second, we estimate a structural econometric model of the ethanol investment timing game that incorporates both the strategic and dynamic aspects of the ethanol investment decision. Additionally, we use the estimated parameters from the structural model to run counterfactual simulations to explore the policy factors driving industry growth and location.

The structural model has several advantages over the reduced-form model. First, the structural model explicitly models the dynamic investment decision, including the continuation value to waiting. As seen in the theoretical model, a potential entrant invests if the payoff from investment exceeds the continuation value from waiting.

A second advantage of the structural model is that we are able to estimate the effect of each state variable on the expected payoff from investing in an ethanol plant. In the reduced-form model, we estimate the effect of state variables on the per-period probability of investment; the parameters therefore represent the relative difference between the payoff from investment and the continuation value from waiting. As a consequence, the parameters in reduced-form models are confounded by continuation values. In contrast, in the structural model, we model the structural relationship between the continuation value from waiting and the payoff from investment, and use it to estimate parameters in the payoff from investing in the ethanol plant.

A third advantage of the structural model is that we are able to better estimate the strategic interaction between potential entrants. In the reduced-form model, we are unable to structurally model the effect of other potential entrants on a potential entrant's payoffs since we are unable to structurally capture a potential entrant's beliefs about other potential entrants. In the structural econometric model, however, potential entrants base their decisions in part on expectations of the future, including their expectations of how many plants will be built by the next year, which depend on what they expect other potential entrants to do in a given period.

A fourth advantage of the structural model is that the parameter estimates from the structural model can be used to simulate counterfactual scenarios. We use the estimated parameters from the structural model to run counterfactual simulations to explore the effects of alternative policies on ethanol investment.

According to the results of the structural model, our preferred model, the intensity of corn production; government policies, particularly the MTBE ban and the 2007 Renewable Fuel Standard (RFS2); and private information shocks all have significant effects on ethanol investment payoffs and decisions. We use the estimated structural parameters to simulate counterfactual policy scenarios to disentangle the impacts of state and national policies on the timing and location of investment in the industry. We find that, of the policies analyzed, the MTBE ban and the RFS2 led to most of the investment during this time period.

The balance of our paper proceeds as follows. In Section 2, we review the relevant literature. We present our

theory model in Section 3. We describe our data in Section 4. We present our reduced-form model and its results in Section 5. We present our structural model and its results in Section 6. We run counterfactual simulations in Section 7. Section 8 concludes.

2 Literature Review

The first branch of literature on which we build is that on reduced-form econometric models of business investment and location decisions (for reviews, see Goetz, 1997; Bartik, 1985). In this literature, investment in businesses, particularly manufacturing, is modeled as a function of output market prices and access, input costs and access, and the policy environment. In some papers, such as Goetz (1997), location decisions involve a two-step process in which investors first choose regions for broader consideration based on one set of criteria, and then narrow the choice within each region based on another set of criteria.

Sarmiento and Wilson (2007) use a cross-sectional discrete choice model to analyze the agricultural characteristics and spatial dimensions that determine plant location. Similarly, Lambert et al. (2008) use a cross-sectional discrete choice model with spatial clustering to look at factors that affect the presence of ethanol plants and proposed plants in a given county, and also isolate clusters that may attract investment. Haddad, Taylor and Owusu (2010) model state-by-state spatial determinants of plant location. Cotti and Skidmore (2010) estimate a model of investment in ethanol over time using aggregate state-level data on investments.

The location determinants identified in these studies provide a starting point for this analysis as far as identifying potentially important exogenous factors. The results of these studies are not always qualitatively similar, however, because of the different empirical specifications and regional focus. For example, Sarmiento and Wilson (2007) and Lambert et al. (2008) find that access to corn is an important location determinant. However, Haddad, Taylor and Owusu (2010) do not find access to corn to be significant, though they note that following location theory (e.g. Goetz, 1997), firms might first choose a region with a lot of corn production before subsequently making their location decision based on other factors, and that their study only models this second stage location decision conditional on firms already choosing a region with a lot of corn production.

None of these studies adequately addresses the effect of potential competition among plants in the location decision. Lambert et al. (2008) include plants established before 2000 as an explanatory variable and find a negative impact on the location of plants that entered between 2000 and 2007, though there is no analysis of timing of entry within that period. Sarmiento and Wilson (2007) employ a cross-sectional model of plant location with a spatially-lagged dependent variable in order to estimate the competitive effect among plants. They find a large negative effect of a nearby plant on the probability of another plant locating nearby, and furthermore, that this effect decreases with distance. However, these competitive effects only describe the relationship among existing plants; neither of these

models has a time element and without panel data it is not possible to analyze the effect of competition on entry.

We improve upon these previous models by estimating a dynamic model with panel data, by analyzing investment timing decisions, by directly estimating the effect of covariates on the payoff to investment, by analyzing strategic interactions in investment decisions, and by using the estimated structural parameters to simulate entry decisions and welfare under various counterfactual policy scenarios.

The real options literature takes a different approach to modeling how investment in ethanol plants responds to policy and economic factors. Like our structural model, the real options approach is based on an underlying dynamic investment decision under uncertainty. Real options models calculate trigger values for entry based on a specific plant technology and size in order to explore the sensitivity of these trigger values to changes in policy (e.g. Schmidt, Luo and Tauer, 2009) or market conditions (e.g. Maxwell and Davison, 2013). Our structural model and associated simulations have the advantages of being estimated across all entrants, instead of for a plant with a specific capacity and technology; and of enabling the investment strategies to be estimated econometrically from data.

The second branch of literature upon which we build consists of papers that develop and estimate structural econometric models of dynamic games. Some examples are Pakes, Ostrovsky and Berry (2007), who develop a structural econometric model of a dynamic entry/exit game in which the structural parameters are estimated semi-parametrically; Aguirregabiria and Mira (2007), who develop structural econometric methods for sequential estimation of dynamic games; and Bajari, Benkard and Levin (2007), who develop a structural econometric model of a dynamic game with continuous control variables. Ryan (2012) applies the model of Bajari, Benkard and Levin (2007) to the cement industry.

The structural econometric model we estimate in this paper follows Lin's (2013) structural model of the multi-stage investment timing game in offshore petroleum production, which builds on the work of Pakes, Ostrovsky and Berry (2007) on discrete games of entry and exit by modeling sequential investments with a finite horizon and by applying the model to actual data.

3 Theoretical Model

We model the dynamic and strategic decision faced by a potential investor (or entrant)[1] i of whether to invest in building an ethanol plant in county k in year t. Investment in an ethanol plant is irreversible and, in each year t, all investment decisions are made simultaneously. I_{ikt} is an indicator of whether potential investor i invests in building a new ethanol plant in county k in year t.

The state variables for county k at time t are given by $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$, where N_{kt} is the number of existing plants in the county; G_{kt} describes the policy environment; and X_{kt} are economic factors.

[1] Because we are modeling the decision to invest in building an ethanol plant, we use the terms 'investor' and 'entrant' interchangeably.

The payoff $\pi(\cdot)$ from investing in an ethanol plant in year t, representing the present discounted value of the entire stream of net benefits from investing in an ethanol plant, depends on the state variables Ω_{kt} at time t.

The value function for a potential investor i in county k in period t represents the expected present discounted value of the entire stream of net benefits to the potential investor from following the dynamically optimal investment policy, and can be written as:

$$V(N_{kt}, G_{kt}, X_{kt}) = max\{\pi(N_{kt}, G_{kt}, X_{kt}), \beta V^c(N_{kt}, G_{kt}, X_{kt})\},$$

where β is the discount factor. The continuation value $V^c(\cdot)$ is the expected value of the next period's value function, conditional on not building an ethanol plant in the current period, and is given by:

$$V^c(N_{kt}, G_{kt}, X_{kt}) = E[V(N_{kt+1}, G_{kt+1}, X_{kt+1})|N_{kt}, G_{kt}, X_{kt}, I_{ikt} = 0],$$

where expectations are taken over the values of the next period's state variables conditional on the value of the current period's state variables and conditional on not investing in the current year.

In a static model of investment, the statically optimal rule is to invest if the payoff $\pi(\cdot)$ from investing is greater than 0. However, when investments are irreversible and there is uncertainty over the future payoff from investment, then the static investment rule is not dynamically optimal. In particular, if the state variables $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ evolve stochastically over time then it is possible that the state variables may take on values in the future that yield a payoff $\pi(\cdot)$ that is high enough that the potential investor would do better in expected present discounted value to wait rather than make the investment now, even if the payoff $\pi(\cdot)$ now is positive. A potential investor holds an option to invest, which it loses when the irreversible investment is made. This lost option value is an opportunity cost that must be included as part of the cost of investment. A potential investor who hopes to make a dynamically optimal decision would therefore need to account for the option value to waiting before making this irreversible investment (Dixit and Pindyck, 1994).

The dynamically optimal investment policy is for the potential entrant to build an ethanol plant in year t if and only if the payoff $\pi(N_{kt}, G_{kt}, X_{kt})$ from investing exceeds β times the continuation value to waiting, $V^c(\cdot)$:

$$I_{ikt} = \begin{cases} 1 & \text{if } \pi(N_{kt}, G_{kt}, X_{kt}) > \beta V^c(N_{kt}, G_{kt}, X_{kt}) \\ 0 & \text{if } \pi(N_{kt}, G_{kt}, X_{kt}) \leq \beta V^c(N_{kt}, G_{kt}, X_{kt}). \end{cases}$$

Because the continuation value from waiting is positive, the dynamically optimal investment rule has a higher threshold for the payoff from investment to exceed before an investment is made compared to the static investment rule, whose threshold is 0. Because of the uncertainty in state variables measuring economic factors and government

policy, there is an option value to waiting, which means that potential entrants are more likely to delay their investments. This uncertainty combined with the irreversible nature of ethanol plant investment make a dynamic model more appropriate than a static model to model ethanol plant investment. Thus, our structural model, which is dynamic, is more appropriate than our reduced-form model, which which does not explicitly model the continuation values.

The dynamic decision-making problem faced by a potential investor is even more complicated when the investment payoff is affected not only by market conditions and government policies, but also by the existence of nearby plants. Due to competition effects and agglomeration effects, the presence of existing ethanol plants may affect the payoff from investing in an ethanol plant. As a consequence, a potential investor's investment decision depends on its conjecture about competitors' behavior. Uncertainty over whether a plant might be constructed and start production nearby is another reason there is an option value to waiting before investing that makes the decision dynamic rather than static (Dixit and Pindyck, 1994).

The effect of existing plants N_{kt} on the payoff from investment measures the net effects of the competition and agglomeration effects. Potential entrants may condition their investment decisions on both the current number of other plants in the county N_{kt} and their expectations on what future values of N_{kt} may be. Future values of N_{kt} may be different from current values if other potential entrants enter in a given year. In the data, the maximum number of plants in existence in any county is three. We therefore model each county as having a potential of up to three entrants. If there were no plants in existence in a county in the first year of the data set, then there are three potential entrants, and each potential entrant faces up to two other potential entrants. Once all potential entrants in a county invest and the number of existing plants in that country becomes three, there are no more potential entrants in that county. N_{kt} measures how many of the other potential investors have already invested.

The covariates in G_{kt} describe the policy environment faced by the corn-ethanol industry. State and federal policies can affect the expected payoff from investing in building a new ethanol plant through the cost of inputs, expected revenues, and building costs. At the federal level we include indicators for the two versions of the Renewable Fuel Standard (RFS1 and RFS2), which are implemented as blending mandates. At the state-level, we include the year the MTBE ban was implemented; MTBE was a popular oxygenate used to meet environmental regulations and also to boost octane level, and ethanol is a substitute for MTBE. We also include state-level production tax credits.

The covariates in X_{kt} include economic factors that affect the payoffs from investing in building an ethanol plant. On the revenue side, we include ethanol price; gasoline price; and proximity to cattle, which is a proxy for sales price of distillers' grains (DDGS, or distillers' dried grains with solubles, is a co-product of corn-ethanol production which is used for animal feed).[2] Gasoline price could have a positive or negative impact on investment depending on whether ethanol is viewed as as an energy substitute for gasoline or an oxygenate (additive), respectively.

[2]The co-product market is becoming more significant due to lower prices for ethanol (Dhuyvetter, Kastens and Boland, 2005). There is significant variability in participation in co-product markets (Perrin, Fretes and Sesmero, 2009). Participation is driven by mill type and plant age; wet mills (corn syrup) and dry mills (DDGS) produce different co-products (DOE 2008).

The vector X_{kt} also includes covariates describing the cost of ethanol production. One important factor is availability and cost of corn, the primary feedstock in the region of focus; local availability is important because transportation is costly (USDA, 2007). Corn is the largest variable cost in ethanol production (Kwiatkowski et al., 2006; Perrin, Fretes and Sesmero, 2009). In addition to corn availability and price, X_{kt} also includes soy availability and price; these variables help describe the intensity of a county's corn production. We include the natural gas price because it is a major energy source for milling corn. We also include electricity price; electricity is an important energy source in some plants.

We also control for the existence of a biodiesel plant because biodiesel and ethanol plants may compete indirectly in the feedstock market: while biodiesel plants use soy as a feedstock, much of the Midwest can be planted to soy or corn. Also, an ethanol plant may be built to satisfy a community need for crop value-added, and a biodiesel plant may compete for support.

We do not explicitly model transportation costs because data on transportation costs infrastructure is generally time-invariant, which means the impact cannot be identified in the reduced-form model as these variables are absorbed by the county fixed effects.[3] However, we do include a metro area indicator, which could capture proximity to market, as well as the potential costs of regulations.

4 Data

4.1 Time Frame and Focus Region

We focus on investments in corn-ethanol plants in the Midwestern United States over the period 1996 to 2008. While ethanol is produced throughout the United States using various feedstocks, 95% of the ethanol produced in this time frame is produced from corn. Focusing on corn-ethanol plants eliminates the need to consider feedstock choice in the model.[4] The majority of corn (and ethanol from corn) is produced in the Midwestern United States, so we focus on ethanol plant entry in this region, specifically in the following ten states: Iowa, Illinois, Indiana, Kansas, Minnesota, Missouri, Nebraska, Ohio, South Dakota, and Wisconsin.

We focus on the time period 1996 to 2008, which corresponds to the latest ethanol boom in the US. This time period is narrow enough to allow us to use one set of policy variables, as well as ensure similarity in plant technology. Starting the analysis earlier would also be difficult because plant startup and closure information is not readily available before this date.[5] Figure 1 shows the number of ethanol plants at the beginning and end of our study period.

[3] In cases where the transportation infrastructure is not time-invariant, then it is likely to be endogenous at the county level. The modeling of transportation infrastructure investment decisions, which has been studied elsewhere (Fatal et al., 2012), is beyond the scope of this paper.

[4] For structural econometric models of feedstock choice, see Yi and Lin Lawell (2015b), who model ethanol investment and feedstock choice in Europe; and Yi and Lin Lawell (2015a), who model ethanol investment and feedstock choice in Canada.

[5] Including the entrants during 2009 and 2010 would require accounting for plant closure due to the market crash and implosion of Verasun, a large producer. Many plants stopped production in late 2008 or early 2009 following Verasun's bankruptcy declaration on October 31, 2008.

Though the start-up month for new plants is available, we use annual observations for three reasons. First, the feedstock of focus, corn, has one growing season in the US. Second, construction of an ethanol plant takes significantly longer than a month, but usually less than a year, from the start of physical construction activities.[6] Finally, much of the data on other variables are publicly available at an annual level.

We eliminate completely non-agricultural counties within the ten states (e.g. northern Minnesota), as well as those with missing data on agricultural production, resulting in a sample with 855 unique counties. This results in potentially 11,115 county-year observations over the thirteen-year time period. We add another dimension to account for the number of potential entrants in each county-year.

4.2 Plant Variables

Our ethanol plant data set includes information about start-up date of new entrants, and nameplate capacity and ownership type for new and existing plants. The original list of operational plants was obtained online from the Renewable Fuels Association and Ethanol Producer magazine, including historical lists from the Renewable Fuels Association; these lists do not match perfectly. We were able to rectify inconsistencies between the two lists as well as collect additional information on plant owners by searching through plant websites, newspaper articles, and SEC filings.

The sample begins with 22 operational plants at the start of 1996, and ends with 149 operational plants with a total capacity of almost 10 billion gallons per year in 2008. Figure 1 maps the number of operational ethanol plants by county in the first and last years of our data set, respectively.

The investment variable I_{ikt} is an indicator of whether potential investor i invests in building a new ethanol plant in county k in year t. It has up to 3 observations per county-year and is equal to 1 if the plant enters in a given calendar year.[7] Once a potential investor i invests, it is no longer a potential investor and therefore exits the sample.

The number of existing plants N_{kt} in the county measures the number of operational plants in that county on January 1 of year t, and is therefore observable to any potential investor making a decision in year t. In an alternate specification of the reduced-form model, we define N_{kt} as a continuous variable of capacity of existing plants.[8] We also define a spatial lag of the existing plant variable for use in the reduced-form estimation; here, it is a count of the number of plants in the counties bordering a given county.

The dataset on biodiesel plants was constructed in the same manner as the ethanol plant variables. The original biodiesel plant lists were from the National Biodiesel Board and Biodiesel Magazine. Analogous to the

Operations were normal the rest of the year, and many of the shuttered plants have since restarted under new ownership. Prior to 2008, there was only one permanent closure (exit) in the sample; others closures were the result of accidents or buyouts, and the plants returned to normal operations. The exit phenomenon in a subject of ongoing work and is outside the scope of this model.

[6] There was a production bottleneck in 2007, when plants took 18-24 months to build (Koplow, 2007). We do not consider announcements of new plants, as other studies did, because many announced projects were never completed as investors fell through before construction began.

[7] Entry is the date of the first grind of corn, which is the first step in corn-ethanol production.

[8] Capacity is a good proxy for production because plants operate continuously at or near nameplate, except during regular maintenance (Kwiatkowski et al., 2006).

number of existing plants, we construct an indicator variable that signals the existence of a biodiesel plant in county k at the start of the calendar year t.

4.3 Policy Variables

We include state-level policy variables. The first state-level policy variable we use is an indicator of whether the state banned MTBE at any point in a given year. All the Midwestern states in our sample implemented MTBE bans by 2005, before the nationwide ban took effect in 2006.

The second state-level policy variable represents the state producer tax credits.[9] Defining this variable is complicated because each state places different contingencies on receiving these funds. For example, some states support only large-capacity plants, others only small or community-owned plants. Thus, even in states with tax credits, not all entering or incumbent plants qualify. In addition, some of the credits are available for a specified number or years, while others expire on a date unrelated to time of plant entry. Because of these differences, we represent these policies with a binary variable indicating if producer tax credit benefits were offered to plants that entered in that year, and test the robustness to that specification in the reduced-form model.

We specify two variables to capture the effects of the Renewable Fuel Standards (RFS).[10] The RFS was created under the Energy Policy Act of 2005 with the goal of accelerating the use of fuels derived from renewable sources (EPA, 2013). This initial RFS (RFS1) mandated that a minimum of 4 billion gallons of ethanol be blended into gasoline in 2006, rising to 7.5 billion gallons by 2012. Two years later, the Energy Independence and Security Act of 2007 greatly expanded the biofuel mandate volumes, creating the RFS2. The RFS2 requires steadily increasing volumes of biofuel to be blended into the nation's fuel supply, reaching 37 billion gallons a year by 2022. We model RFS1 with an indicator for the years 2005 and 2006 and RFS2 as an indicator for the years 2007 and 2008.

4.4 Other Data

Corn and soy prices are available annually from the National Agricultural Statistics Service of the USDA (NASS) at the state level. Corn and soy production and acreage are available annually by county from NASS. Because counties are different areas, we construct a county corn intensity variable, defined as the corn acreage divided by the total area of the county, to capture area-independent acreage using county acreage from the US Census.[11] We also construct a spatial lag of the corn intensity variable as well as a county-level soy intensity variable. Because corn price data are not publicly available at a county level, the local competition in the corn feedstock market is captured both by the

[9] The American Coalition for Ethanol (2007) provides detailed description and review of the policies. Cotti and Skidmore (2010) study state-level impacts of these policies.

[10] We do not include other federal-level policy variables such as tax credit or the small producer subsidy in the analysis because they do not vary enough in the time period to identify the effects.

[11] As a robustness test, we also run specifications defining corn intensity as production over area.

county-level corn intensity variable and by the state variable N_{kt} measuring the number of existing plants in the county.

To represent the potential market for distillers' grains (DDGS), a co-product of corn-ethanol production that is used for animal feed, we construct a county-level cow density variable using the number of cows per district-acre, where the number of cows is the count of 'all cattle', available from NASS, and districts are defined by the USDA.[12] The potential DDGS market also includes hogs, but data is not available at the district level for all states. However, because cattle use DDGS more efficiently than hogs, they represent the larger market for co-products (NASS, 2007).

The ethanol price is the free on board price in Omaha, and is published by the Nebraska Energy Office. We use state-level total gasoline rack prices from the Energy Information Administration. We do not include an E85 price in this analysis because the price series began much more recently than our time frame, and it lacks spatial variation. Natural gas (city gate) price and electricity price to industry are available annually from the EIA, also at state level.[13] We use the average urban CPI to deflate all the prices. The final variable, an indicator for metropolitan areas, is the US Census definition of counties in metropolitan statistical areas.

Because we do not have local variation in ethanol, gasoline, natural gas, or electricity prices, local competition in the ethanol and gasoline output markets and in the gasoline, natural gas and electricity input markets are captured by the state variable N_{kt} measuring the number of existing plants in the county.

5 Reduced-Form Model

We estimate a reduced-form discrete response panel model to compare to the literature on the location choice of ethanol plants and to help us choose the state variables to use in our structural model. We regress the probability of investment in an ethanol plant on the covariates using the following logit fixed effects model:

$$Pr(I_{ikt} = 1) = 1 - F(-(N'_{kt}\delta_N + G'_{kt}\delta_G + X'_{kt}\delta_X + Year_t{}'\gamma + \nu_k)), \tag{1}$$

where $F(\cdot)$ is the logistic cumulative distribution function. We also estimate the following linear probability fixed effects model:

$$Pr(I_{ikt} = 1) = N'_{kt}\delta_N + G'_{kt}\delta_G + X'_{kt}\delta_X + Year_t{}'\gamma + \nu_k. \tag{2}$$

There are several reasons to estimate both the fixed effects logit and fixed effects linear probability models. The linear probabiliy model is easier to implement and the estimates are consistent if we control for the heteroskedastic errors. The logit model in (1) is preferred, however. In this particular dataset, there are relatively few instances of investment. Because the probability of investment is relatively low, we are on the left side of the distribution, and it is

[12] A district is made of up to 120 counties and there are usually 6-8 districts per state.

[13] We use city gate natural gas price instead of price to industry because the complete series is available; these two price series trend together within a given state.

therefore advantageous to use the logit model.

In both reduced-form models, $Year_t$ is either a year effect or a time trend depending on the specification. ν_k is the county fixed effect that controls for unobservable county traits, such as size or promotion of business development, which remain fixed over time. The errors for the fixed effect logit are calculated with the observed information matrix. The errors for the linear probability model are clustered at the county level.

N_{kt} represents the number of existing ethanol plants open at the start of the period in which the investment decision is made and is not endogenous. Because of the time necessary to construct a plant, the potential investor necessarily observes previously existing plants before investing. Because the parameters in the reduced-form model are confounded by continuation values, we are unable to model the effect of other potential investors on a potential investor's payoffs. Instead, the coefficients on *existing plants* and *spatial lag of existing plants* tell us the impact of existing competitor plants on the probability of entry.[14]

The vector G_{kt} contains indicators of the different policies; the state policies *MTBE Ban* and *Tax Credit* are used in all specifications, and *RFS1*, *RFS2* are included in specifications with continuous $Year_t$. X_{kt} contains the following exogenous covariates: *corn price*, *soy price*, *corn intensity* and its spatial lag, *soy intensity*, *cow density*, *electric price*, *natural gas price*, *gasoline price*, *ethanol price*, and the indicator *biodiesel plant*. Like the RFS variables, we can only identify *ethanol price* in the specifications without individual year effects because there is no spatial variation in the ethanol price in our data.

The summary statistics for the explanatory variables used in the reduced-form analysis are presented in Table 1.

The results from the estimation of the base fixed effects logit model in equation (1) along with a comparison with the results from the linear probability fixed effects model in equation (2) are presented in Table 2.

The results of the logit estimation in (1) are in specifications A and B of Table 2. The coefficient on *existing plants* is large, negative, and significant; indicating that existing plants, on net, negatively impact the probability of investment in a new plant in a given county. The coefficient on the *spatial lag of existing plants* is positive, though insignificant, indicating a potential positive agglomeration affect regionally. The significant negative sign on *existing plants* that is only present within a county confirms the existence of localized competition also seen in Sarmiento and Wilson (2007) and Lambert et al. (2008).

One reason we see so few significant variables in this regression is because the fixed effects logit relies upon within-county variation for identification, meaning we estimate (1) for only the counties that had an entrant in the time period. Variables such as *corn intensity* vary more spatially than they do across time, hence this regression does not

[14]Because the reduced-form model does not model the strategic interaction among plants as a game, we do not need to model each player's expectation of the strategies of all other players in the game; as a consequence, we can include *spatial lag of existing plants* without having to model the expectations of all players in both the county and all its neighboring counties of the strategies of all other players in both the county and all its neighboring counties.

detect an impact on probability of entry. Goetz (1997) suggests a two-stage location-selection process where firms chose a region based on some factors, and then enter in a specific location based on others as in Haddad, Taylor and Owusu's (2010) regional model of ethanol plant location. Corn-ethanol plants are location in regions with high corn availability, but variation within the region (and in our case, over time) is not large. Instead, local and market factors drive location and entry decisions.

The magnitude, and sometimes sign, of some of the other coefficients depends on the specification of time $Year_t$ in the regression model. A continuous $Year_t$ controls for changes in technologies and preferences over time, while the individual $Year_t$ effects also capture events, policies, market conditions at the national level. The coefficients on *natural gas price*, *corn price*, and *soy price* change sign and magnitude across the regression, though none are significantly different from zero. These variables are all correlated and trend upwards over time, which may make their effects difficult to distinguish from the time trend.

The coefficient on *Tax Credit* is positive but insignificant in the regression with year effects (specification A of Table 2), and is larger and becomes significant in the regression with continuous time (specification B). We find no significant impact of *RFS1*, *RFS2*, or the *MTBE Ban*. Cotti and Skidmore (2010) found positive impacts of state ethanol tax credits on state ethanol capacity, suggesting that these policies drive regional location choices, but that perhaps national-level policies drive overall growth in the industry.

The only other significant coefficient is that on *gasoline price*, which also has a large, positive effect on the probability of investment in a new plant, indicating that investors in ethanol plants potentially view ethanol as a gasoline substitute. We explore this result further below.

The linear probability fixed effects model estimation of equation (2) is interesting for two reasons. First, the results in specifications C and D of Table 2 serve as a comparison to the logit fixed effects model in (1). The signs and significance levels of the linear probability model estimates in specifications C and D are qualitatively similar to the logit estimates. Like the logit model, the linear probability models in specifications C and D are estimated only for the counties k that have an entrant at some point in the period.

A second reason the linear probability model is informative is that we can include the full data sample and account for within and cross-sectional variation. As seen in specifications E and F of Table 2, we find more significant variables for the linear probability model when we use the full dataset and not just the counties who have entrants. While *existing plants* still have a negative and significant effect on entry, we see that the effect of *spatial lag of existing plants* is positive and significant. This indicates that the net negative competitive effect among plants not only dissipates with distance, but also becomes net positive, indicating possible agglomeration benefits in the ethanol industry.

We run several specification tests of the reduced-form models in equations (1) and (2). Their results are presented in Table 3. First, we use a Hausman test to choose between random effects and fixed effects. The Hausman

χ^2 statistics from the test on the restricted and full random effects models are all very large, with corresponding *p-values* of 0.000, indicating county unobservables are likely to be correlated with the regressors, and therefore that fixed effects is the appropriate specification.[15]

We test for potential endogeneity of *corn intensity* in the base specification using a Durbin-Wu-Hausman test. In the first-stage regression, the instruments for *corn intensity* are the time lags of *corn intensity* and *corn price*. The estimated coefficient on the first-stage residuals in the second stage regression is insignificant, indicating that we cannot reject the exogeneity of corn intensity in any specification.[16]

We do not anticipate endogeneity problems with the other variables such as *corn price* because they are observed on a more aggregate level, and thus would not be expected to respond to the addition of one ethanol plant at the county level. For example, McNew and Griffith (2005) find that while ethanol plants increase the basis for corn price, this effect is limited to around 50 miles from the plant, while the price variables in this analysis are measured at the state level. An additional argument for using contemporaneous prices rather than futures prices in our model is that while futures prices exist, they are at a national level, and therefore will be absorbed by the year effects in our reduced-form model. We model the evolution of price expectations in the dynamic structural model in the following section.

In Table 4, we estimate the logit model (1) with alternate specifications of the corn and soy variables (*corn price*, *soy price*, *corn intensity*, *soy intensity*). We construct ratios of *corn to soy price* and *corn to soy intensity* and include them in the regressions in place of, and as well as, the previously specified variables. The hypothesis is that perhaps the relative prices and production intensities may capture more variation in entry probability than the levels. However, the results are not qualitatively different from the base specification.

In Table 5, we explore the large positive effect of *gasoline price* further by estimating the logit model (1) with alternate specifications of the *ethanol price* and *gasoline price* variables. We construct a ratio of the *ethanol to gasoline price* and include in the regression in place of, and as well as, the individual price variables. One advantage of this alternate specification is that we can control for the ethanol price regardless of the specification of $Year_t$. While *ethanol price* is measured at the national level, the ethanol-gasoline price ratio is at the state level.

The alternate specification of *gasoline price* and *ethanol price* does not have any qualitative effects on other coefficient estimates, except for the coefficients on *RFS1* and *RFS2*. The estimates of the RFS impacts are larger, and significant, when *ethanol to gasoline price ratio* is included in the regression. Additionally, we detect a positive impact of *ethanol price* in the specifications with continuous $Year_t$. In these specifications, the coefficient on *ethanol to gasoline price* is small and insignificant. In the specifications with year effects instead of a continuous $Year_t$, the

[15] The restricted random effects model includes the same regressors as the fixed effects model, while the full model includes the time-invariant regressors, allowing accounting for potential efficiency gain from their inclusion (Wooldridge, 2010).

[16] As a robustness check, we estimate the models with a time lagged corn intensity variable instead of contemporaneous corn intensity. There is no significant difference in the other estimates (results not reported).

effect of *ethanol to gasoline price* on the probability of entry is large, negative, and significant, which supports the view of ethanol and gasoline as substitutes. Babcock (2012) discusses the relative cost of gasoline and ethanol in a policy context, and finds market scenarios in which ethanol can be viewed as an energy substitute for gasoline, and others in which ethanol is viewed as an additive.

In specifications T and U of Table 6, we estimate the logit model (1) with an alternate specification for *Tax Credit*: we model the effect of the expected lifetime value of the tax credit instead of an indicator for the existence of the policy. In specifications V and W of Table 6, we estimate the logit model (1) with an alternate specification for *existing plants*: we model capacity instead of count of other plants. In all cases, there are no qualitative differences in the results.

The primary advantages of our reduced-form model are that we can use continuous variables without having to discretize them and, because state-space constraints are less of a concern, we can include many covariates. However, the reduced-form models only estimate the per-period probability of building a plant, and therefore do not have a clear structural interpretation. Because the payoffs from investing in building an ethanol plant depend on economic factors, government policies, and strategic interactions that vary stochastically over time, a potential entrant who hopes to make a dynamically optimal decision would need to account for the option value to waiting before making an irreversible decision to invest in building an ethanol plant (Dixit and Pindyck, 1994). The parameters in the reduced-form models are therefore confounded by continuation values. We now develop a structural econometric model of the ethanol investment timing game which better and more explicitly captures the dynamic and strategic nature of ethanol plant investment decisions.

6 Structural Model

A structural model has several advantages over a reduced-form model: it explicitly models the dynamic investment decision, including the continuation value to waiting; it enables us to estimate the effect of each state variable on the expected payoff from investing in an ethanol plant; it enables us to estimate the strategic interaction between potential entrants; and it generates parameter estimates that can be used to simulate counterfactual scenarios. The structural model is therefore our preferred model.

In the structural model, we estimate the parameters in the theoretical model presented in Section 3. As in the theoretical model, the decision of a potential investor i of whether to invest in building an ethanol plant in each county k in year t depends on the publicly observable state of the county $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$, a vector of discrete and finite-valued state variables that are observed by all the potential investors in county k as well as by the econometrician. The state variables N_{kt}, G_{kt}, and X_{kt} evolve according to a first-order controlled Markov process and summarize the direct effect of the past on the current environment.

The state variable N_{kt} captures the strategic component of the investment decision, which, unlike in the reduced-form model, includes the existence of competing plants as well as the expectation about how many other plants will be built in the next year. In this model, N_{kt} is represented by the variable *existing plant*, which is an indicator of whether another investor has built an ethanol plant in a given county k in year t.

The state variables in G_{kt} are the same policy variables used in the reduced-form estimation: *MTBE Ban*, *Tax Credit*, *RFS1*, and *RFS2*.

The state variables in X_{kt} include *corn price*, *soy price*, *corn intensity*, *cow density*, the indicator *biodiesel plant*, as well as a constant. We also use alternate specifications for the energy price variables: either *electric price*, *natural gas price*, or the combined *input price indicator*; and either *gasoline price*, *ethanol price*, or the combined *output price indicator*.

We discretize each of the continuous variables in our data into discrete and finite-valued state variables, as detailed in Table 7. For our base specification, we discretize the continuous variables into two bins each. In some cases we aim for equally-sized bins (*natural gas price*, *electricity price*, *gasoline price*, *ethanol price*, *corn intensity*). For other variables, owing in part to their skewed distribution, we create bins that put higher weight on the lower (*corn price*) or higher (*cow density*) part of the continuous variable. We also construct alternate bins to test the robustness of our model to different break points, including discretizing the continuous variables into three instead of two bins. Summary statistics for the discretized state variables used in our structural model are in Table 8.[17]

Each state of the county $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ is represented by a combination of discretized state variables. The number of potential states of the county is the product of the number of bins of each state variable. Dimensionality is an important consideration for the simulations we perform using the structural estimates. For example, when we simulate removing a policy, we must observe the rest of the variables describing the state of the world Ω_{kt} with and without the policy. Thus, our preferred specification has fewer bins and covariates, thus fewer potential states of the world Ω_{kt} that we must identify and observe to conduct simulations.

Because the main objectives of this paper are to learn about strategic interactions among potential investors and the effects of policy on investment, we are most concerned with the other covariates to the extent that they can fully describe the state of input and output markets. The indicator variables we construct for output and energy input prices allow us to control for prices in the state of the world, while freeing up dimensions to focus on and identify different policies in our simulations. The variable *energy input price* is an indicator that is one when both the electricity and natural gas prices are high. The variable *output price* indicator is one when both the gasoline and ethanol price is high. Our reduced-form analysis showed that both *gasoline price* and *ethanol price* and positive, large, and sometimes significant impacts on the probability of investment. The preferred specification includes only *natural gas price* and

[17] The sample size here is slightly larger than in the reduced-form model because some counties were missing one or two years of data for corn production or cow density. For the structural model, we imputed these values as the lowest bin in those cases because the reported values in these counties for other years were all very small.

not *electricity price* or the *input price indicator* because the reduced form model indicated that *electricity price* did not have a significant impacts on the probability of entry.

In addition to the observable state variables $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$, the decision of a potential investor i of whether to invest in building an ethanol plant in each county k in year t also depends on a shock ε_{ikt}, which is private information to the potential investor and unobserved by either other potential investors or by the econometrician. Such private information may include, for example, a shock to the cost of building an ethanol plant. We assume the error term is independently and identically distributed exponentially with mean σ, which is among the parameters to be estimated.

The payoff $\pi(N_{kt}, G_{kt}, X_{kt}, \varepsilon_{ikt}|\theta)$ from investing in an ethanol plant in county k in year t can be separated into a deterministic component and a stochastic component as follows:

$$\pi(N_{kt}, G_{kt}, X_{kt}, \varepsilon_{ikt}|\theta) = \pi_0(N_{kt}, G_{kt}, X_{kt}|\theta) + \varepsilon_{ikt},$$

where the deterministic component $\pi_0(\cdot)$is linear in the state variables:

$$\pi_0(N_{kt}, G_{kt}, X_{kt}|\theta) = N'_{kt}\gamma_N + G'_{kt}\gamma_G + X'_{kt}\gamma_X,$$

and where $\theta = (\gamma_N, \gamma_G, \gamma_X, \sigma)$ denotes the parameters to be estimated. The coefficients γ_N, γ_G, and γ_X measure the effects of the state variables N_{kt}, G_{kt}, and X_{kt}, respectively, on the payoff to investing in building a new ethanol plant.

The coefficient γ_N on the number of existing plants N_{kt} in the county measures the effect on a potential investor's payoff from investing of having a competitor that has already invested, and identifies the net effect of other investor's decisions on a potential investor's payoff. This measure includes the agglomeration and competition effects, and is a net measure indicating whether ethanol plants interact strategically on net. A positive coefficient γ_N indicates that the agglomeration effect dominates. A negative value for γ_N indicates that the competition effect dominates.

The value function for a potential entrant i in county k in period t can be written as:

$$V(N_{kt}, G_{kt}, X_{kt}, \varepsilon_{ikt}|\theta) = max\{\pi(N_{kt}, G_{kt}, X_{kt}\varepsilon_{ikt}|\theta), \beta V^c(N_{kt}, G_{kt}, X_{kt}|\theta)\}.$$

The potential entrant will invest if and only if the payoff from investing exceeds β times the continuation value to waiting. The continuation value, $V^c(\cdot)$, is the expected value of the next period's value function, conditional on not building an ethanol plant in the current period, and is given by:

$$V^c(N_{kt}, G_{kt}, X_{kt}|\theta) = E[V(N_{kt+1}, G_{kt+1}, X_{kt+1}, \varepsilon_{ikt+1}|\theta)|N_{kt}, G_{kt}, X_{kt}, I_{ikt} = 0].$$

Let $g(N_{kt}, G_{kt}, X_{kt}|\theta)$ denote the probability of investing in an ethanol plant at time t, conditional on the publicly available information $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ at time t, but not on the private information ε_{ikt}. The investment probability function $g(N_{kt}, G_{kt}, X_{kt}|\theta)$ represents a potential investor's perceptions of the probability that a competitor who has not yet invested will decide to invest at time t.

Using the exponential distribution for ε_{ikt} the continuation value $V^c(\cdot)$ reduces to:

$$\begin{aligned} V^c(N_{kt}, G_{kt}, X_{kt}|\theta) = E[&\beta V(N_{kt+1}, G_{kt+1}, X_{kt+1}, \varepsilon_{ikt+1}|\theta) \\ &+ \sigma g(N_{kt+1}, G_{kt+1}, X_{kt+1}|\theta)|N_{kt}, G_{kt}, X_{kt}, I_{ikt} = 0], \end{aligned} \tag{3}$$

and the investment probability $g(\cdot)$ reduces to:

$$g(N_{kt}, G_{kt}, X_{kt}|\theta) = exp\left(-\frac{\beta V^c(N_{kt}, G_{kt}, X_{kt}|\theta) - \pi_0(N_{kt}, G_{kt}, X_{kt}|\theta)}{\sigma}\right), \tag{4}$$

as shown by Lin (2013).

We employ a two-step semi-parametric estimation procedure following Pakes, Ostrovsky and Berry (2007) and Lin (2013). In the first step, the continuation value is estimated non-parametrically and this estimate is used to compute the predicted probabilities of investment. In the second step, the parameters $\theta = (\gamma_N, \gamma_G, \gamma_X, \sigma)$ are estimated by matching the predicted probabilities with the actual probabilities in the data using generalized method of moments (GMM).

For the first step in the estimation we construct a transition matrix M, which describes the evolution of the state variables N_{kt}, G_{kt}, and X_{kt} over time, conditional on not investing. The transition matrix M gives, for each combination of state variables in year t, the probability of transitioning to each combination of state variables in year $t+1$ conditional on not investing in year t. The element in each row r, column c is represented by: $M_{rc} = Pr(\Omega_{k,t+1} = c|\Omega_{kt} = r, I_{ikt} = 0)$. We estimate M non-parametrically using empirical averages. We therefore assume rational expectations on the part of potential ethanol plant investors, namely that their expectations about the evolution of state variables over the time period of our data set were consistent with the actual evolution realized.

Let $\bar{g}$ be the vectorized investment policy function, which is a vector whose length is the number of combinations of state variables and whose value at each component is the investment policy function $g(\cdot)$ evaluated at a particular combination of state variables. $\bar{g}$ gives the probability of investment in a new ethanol plant for every observed state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$. We estimate $\bar{g}$ using empirical averages:

$$\bar{g}(N_{kt}, G_{kt}, X_{kt}) = Pr(I_{ikt} = 1|N_{kt}, G_{kt}, X_{kt}).$$

From equation (3), the vectorized continuation value $\bar{V}^c$, which is a vector whose length is the number of combination of state variables and whose value at each component is the continuation value V^c evaluated at a particular combination of state variables, can be specified in vector form as $\bar{V}^c = M(\beta\bar{V}^c + \sigma\bar{g})$, where M is the empirical transition matrix, β is the discount rate, and $\bar{g}$ is the vector of empirical investment probabilities. Because this is an infinite horizon problem, we estimate the continuation value by solving for the fixed point $\hat{V}^c$, which, from Blackwell's Theorem, is unique. We then use the estimate $\hat{V}^c$ to form the predicted probability of investment in an ethanol plant, which from equation (4) can be specified in vector form as:

$$\hat{g}(N_{kt}, G_{kt}, X_{kt}|\theta) = -\frac{\beta\hat{V}^c - N'_{kt}\gamma_N - G'_{kt}\gamma_G - X'_{kt}\gamma_X}{\sigma}. \tag{5}$$

In the second step of the estimation procedure, we estimate the parameters $\theta = (\gamma_N, \gamma_G, \gamma_X, \sigma)$ by finding the parameters that best match the investment probability predicted by our model with the respective empirical investment probabilities in the data using GMM. We use the following moment function:

$$\psi = (\hat{g}(N_{kt}, G_{kt}, X_{kt}|\theta) - \bar{g}(N_{kt}, G_{kt}, X_{kt}))n(N_{kt}, G_{kt}, X_{kt}|I_{ikt-1} = 0),$$

where $n(N_{kt}, G_{kt}, X_{kt}|I_{ikt-1} = 0)$ counts the number of times each state $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ occurs where there is a potential investor. Thus, ψ is a vector where each row represents difference in the predicted and empirical probabilities of investment in ethanol plants for each of the possible states of the world Ω_{kt}, and is weighted by the number of times that state occurs in the data. The population moment condition is that in expectation, ψ equals zero. Additional moments are constructed by interacting the above moments ψ with the state variables Ω_{kt}.

The GMM estimator $\hat{\theta}$ is the solution to the problem:

$$min_\theta \left(\frac{1}{obs}\sum\psi\right) W_n^{-1} \left(\frac{1}{obs}\sum\psi\right),$$

where *obs* is the number of investor-county-year observations. Because the system is exactly identified, we use an identity matrix as the weight matrix W_n.[18]

We form standard errors by a nonparametric bootstrap. We randomly draw counties from the data with replacement to generate 250 independent panels of size equal to the actual sample size. The structural econometric model is run on each of the new panels. The standard error is then formed by taking the standard deviation of the estimates from each of the random samples.

The problem of spatially correlated unobservables can be addressed by interpreting the investment payoff in

[18] One challenge is determining whether the model has converged at a global or local minimum. We experimented with several combinations of starting values to initialize the parameters to be estimated. We found the model is robust to the starting value.

the model as expected payoff conditional on observables, where the expectation is taken over the correlated unobservables. The model is still able to separately identify the (expected) strategic interaction from the correlated unobservables. The online Appendix of Lin's (2013) Monte Carlo experiments analyzes the effect of a state variable that is observed by firms when they make their decisions but unobservable to the econometrician (i.e., a common shock), and show that the bias introduced by spatially correlated unobservables is small. This is consistent with Pakes, Ostrovsky and Berry (2007), who find that the bias from serially correlated common shocks is small. In the reduced-form model, time-invariant spatially correlated unobservables are absorbed in the county fixed effects.

The results from the structural estimation of the parameters are reported in Tables 9 and 10. The preferred specification, which we use for the counterfactual policy simulations, is specification (i). The additional specifications (ii)-(vi) in Table 9 show the robustness of the model to different price specifications. Table 10 has an alternate specification with additional covariates whose effects we cannot separately identify for the policy simulations (specification (vii)) and also shows the results with alternate bins (specifications (viii)-(ix)).

All of the policy variables have positive impacts on the payoff from investment in an ethanol plant, and two, *MTBE Ban* and *RFS2*, are significant. Because both the MTBE ban and the Renewable Fuel Standard can function as implicit blending mandates (De Gorter and Just, 2010; Anderson and Elzinga, 2012), the similar magnitude of the coefficients suggests similar implicit state blending levels. Further, the coefficient on *RFS1* is much smaller and is not statistically significant, which would suggest that the first version of the RFS was not big enough to induce investment.

The coefficient on *existing plant* is small and insignificant, unlike in the reduced-form model. The interpretation of the coefficients in the two models is different, however. In the reduced-form model, we are measuring the impact of an existing plant on the probability of entry, and the parameter estimates are confounded by continuation values. In the structural model, we are measuring the strategic effect of existing and potential plants on the payoff from investment. This small and insignificant coefficient on *existing plant* in the structural model indicates that the competition effect and the agglomeration effect do not have any net strategic effect on the payoff from investing in an ethanol plant.

On the input (cost) side, *corn intensity* has a positive impact on the payoff from investment, while *corn price* is not significant. This result is similar to the reduced-form literature on plant location, which finds that physical access to feedstock is a significant location determinant, but more aggregate feedstock prices are not important (e.g. Cotti and Skidmore, 2010).

On the revenue side, the coefficient on *output price indicator* is negative; this means when both ethanol and gasoline prices are high, there is a negative impact on the payoff from investing. In the alternate price specifications (iii) and (iv) in Table 9, we show that high ethanol and gasoline prices have negative impacts on the payoff from investment when modeled individually, though the effects are insignificant. This is in contrast to the reduced-form results, where high gasoline prices positively impacted the probability of entry, and where the ratio of ethanol to gasoline price has a

negative impact, indicating ethanol and gasoline are viewed as substitutes.

Specifications (v)-(vi) in Table 9 show the robustness of the model to various specifications of the input price variables, none of which have significant impacts on the payoffs from investing in an ethanol plant.

The constant and the mean of the private shock σ are both significant determinants of the payoff from investing. The estimate of σ is similar in magnitude to the coefficients on *MTBE ban* and *RFS2*, indicating that this private information shock can be as important as the policies in determining investment payoff. The constant is large and negative, indicating there are significant fixed costs to investing in an ethanol plant.

Specification (vii) in Table 10 builds on the base specification by adding the additional covariates *metro area* and *biodiesel plant*. These variables have insignificant effects on the expected payoff from investing in an ethanol plant, and their inclusion does not lead to noticeable differences in the other estimates. Consequently, we do not include these covariates in our preferred specification.

Specifications (viii) and (ix) in Table 10 show the results of structural estimation with alternate bins and more covariates than our preferred specification (specification (i)). However, since dimensionality is an important consideration for the simulations we perform using the structural estimates, our preferred specification (i) has fewer bins and covariates, thus fewer potential states of the world Ω_{kt} that we must identify and observe to conduct simulations.

7 Policy Simulations

We use the parameter estimates from the structural model to simulate investment in ethanol plants under different counterfactual policy scenarios. We also conduct a replication exercise to validate the structural model and show how we can expect the simulations to perform.

The methodology for the simulations and replication exercise begins by using the observed data to construct the empirical investment probabilities $\bar{g}(\cdot)$ and a transition matrix M for each state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ using empirical averages; these are identical to the vectors used in the structural estimation. Substituting in the structural model estimate of $\hat{\sigma}$ into equation (3), we calculate the estimated continuation value for each state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ by solving for a fixed point $\tilde{V}^c(\cdot)$:

$$\tilde{V}^c(N_{kt}, G_{kt}, X_{kt}, |\hat{\theta}) = M(\beta\hat{V} + \hat{\sigma}\bar{g}(\cdot)). \tag{6}$$

This vector has one observation for each of the observed states of the world Ω_{kt}.

We next substitute the structural parameters estimates, $\hat{\theta} = (\hat{\gamma}_N, \hat{\gamma}_G, \hat{\gamma}_X, \hat{\sigma})$ from specification (i) of Table 9, and the estimated continuation value $\tilde{V}^c$ from equation (6) into the expression for the predicted probability of

investment from structural model equation (5) to form the estimated probability of investment $\tilde{g}(\cdot)$, where:

$$\tilde{g}(N_{kt}, G_{kt}, X_{kt}|\hat{\theta}) = -\frac{\beta\tilde{V} - N'_{kt}\hat{\gamma}_N - G'_{kt}\hat{\gamma}_G - X'_{kt}\hat{\gamma}_X}{\hat{\sigma}}.$$

Each observed state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ observed in that data is associated with an estimated probability of entry in the vector $\tilde{g}(\cdot)$.

We begin simulating investment at *t=1*, which corresponds to 1996, our first year of data. In the first step, for each county k, we evaluate the estimated probability of entry $\tilde{g}(N_{kt}, G_{kt}, X_{kt}|\hat{\theta})$ for the state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ for county k at time t. For the replication, we use the observed exogenous state variables to define Ω_{kt}.

In the second step of the simulation, we take a random draw, d, from a uniform distribution for each potential entrant i so that $d_{ikt} \sim U(0, 1)$. The entry rule is that a potential investor enters if the random draw is less than the estimated probability of investment in the current state of the world:

$$I^d_{ikt} = \begin{cases} 1, & \text{if } d_{ikt} < \tilde{g}(N_{kt}, G_{kt}, X_{kt}|\hat{\theta}) \\ 0, & \text{if } d_{ikt} \geq \tilde{g}(N_{kt}, G_{kt}, X_{kt}|\hat{\theta}). \end{cases}$$

Once a potential investor i makes an investment ($I_{ikt} = 1$), that investor exits the sample.

In the third simulation step, we update N_{kt} for year $t + 1$ to account for any investments made in each county k in year t. We use the observed data for the exogenous variables G_{kt} and X_{kt}. This implies that the evolution of these state variables is not dependent on the number of ethanol plants in a county, N_{kt}. Many of the variables, including all of the policies, are measured at aggregate levels that include many counties k and potential investors i. Thus while they move together, we would not expect a rare county-level event to affect their trajectory. Other than the number of competing plants, the only other variable measured at the county level is corn intensity.

We repeat steps 1-3 for each year through 2008 (the 13th year), updating N_{kt} for each period.

In the final step, after simulating entry for the period 1996-2008, we record the total number of entrants, E, and the number of entrants in each year t, E_t. We also calculate the total welfare of all entrants, W, and the welfare of each entrant, w_e, which is specified as:

$$w_e = E[\pi(N_{kt}, G_{kt}, X_{kt}, \varepsilon_{ikt}|\hat{\theta})(N_{kt}, G_{kt}, X_{kt}, \varepsilon_{ikt}|\hat{\theta})] = N'_{kt}\hat{\gamma}_N + G'_{kt}\hat{\gamma}_G + X'_{kt}\hat{\gamma}_X + \hat{\sigma}.$$

The expression for entrant welfare w_e incorporates both the deterministic part of the payoff from investing, $\pi_0(N_{kt}, G_{kt}, X_{kt}|\hat{\theta}) = N'_{kt}\hat{\gamma}_N + G'_{kt}\hat{\gamma}_G + X'_{kt}\hat{\gamma}_X$, as well as a measure of the mean private shock $E[\varepsilon_{ikt}|\hat{\theta})] = \hat{\sigma}$.

The mean welfare per entrant $\bar{w_e}$ is taken over all entrants e in all years for a given simulation, and is given

by:

$$\bar{w}_e = \frac{\sum_{e=1}^{E} w_e}{E}.$$

The standard deviation of the welfare per entrant, s_e, over all years of the simulation is:

$$s_w = \sqrt{\frac{1}{E-1}\sum_{e=1}^{E} |w_e - \bar{w}_e|^2}.$$

We normalize welfare so that the mean welfare per entrant of the No Policy scenario is equal to 1.

We conduct 50 rounds of the simulation, each with 13 years of draws d_{ikt}. We record the average of each of the previous statistics (E, E_t,W, $\bar{w}_e$, s_e) across the 50 rounds of simulation; this accounts for the randomness of the draws d_{ikt}.

We estimate the standard errors for the statistics E, E_t,W, $\bar{w}_e$, s_e using a nonparametric bootstrap. We randomly draw counties from the dataset with replacement to generate 250 independent panels of size equal to the actual sample size. These are the same datasets that we generated when bootstrapping the standard errors of the structural parameters. For each bootstrap sample, we run 50 simulations using the estimated parameters $\hat{\theta}$ and estimated probabilities of investment $\tilde{g}(\cdot)$ associated with the particular bootstrap draw, and then take the average of the statistics (E, E_t,W, $\bar{w}_e$, s_e) across the 50 rounds of simulation. The standard error is then formed by taking the standard deviation of the estimated statistics from each of the random samples.

In addition to a simulated replication using the observed exogenous state variables (the Base scenario), we simulate investment under several counterfactual policy scenarios. These counterfactual scenarios include No RFS1, No RFS2, No Tax Credit, No MTBE Ban, and No Policy, and are summarized in Table 11. The methodology for the simulations is the same as the Base replication, except we replace the indicators for the specified policy variables in G_{kt} with zero. For example, in the no RFS1 simulation, we set $RFS = 0$ for all observations.

For the No MTBE Ban scenario, we can only run the simulations for the pre-RFS period (1996-2004) because 2004 was the last year any state in our sample permitted the use of MTBE; we therefore never see cases in which there is no MTBE ban in combination with either RFS1 or RFS2, both of which were implemented after 2004. As a consequence, the effect of the MTBE ban is not identified after 2004.

One challenge in simulating alternate policy scenarios is that calculating the investment policy function $\tilde{g}(\cdot)$ relies on observing the corresponding state of the world in the data. Because entry in the simulation is random, we sometimes simulate states of the world that we do not observe in the data. We follow a rule to replace the missing value of $\tilde{g}(\cdot)$ for the simulated states of the world Ω^*_{kt} that we do not observe in the data. Many of the states of the world that are missing in the simulation occur because we draw an entrant in a county k that did not have an entrant in the data. Consequently, the first replacement rule is replace $\tilde{g}(N_{kt} = 1, G_{kt} = g, X_{kt} = x)$ with $\tilde{g}(N_{kt} = 0, G_{kt} = g, X_{kt} = x)$

when we do not observe $\Omega_{kt} = (N_{kt} = 1, G_{kt} = g, X_{kt} = x)$ in the data.

The second replacement rule to replace the missing value of $\tilde{g}(\cdot)$ for the simulated states of the world Ω^*_{kt} that we do not observe in the data is necessary because for some values of the state variables X_{kt}, we may not observe that value X_{kt} both when the policies G_{kt} are equal to zero and when the policies G_{kt} are not equal to zero. In the second rule, we find the state of the world $\Omega_{kt} = (N_{kt}, G_{kt}, X_{kt})$ where $\tilde{g}$ is defined, and where the variables in X_{kt} that have a statistically significant effect on the payoff from investing in building an ethanol plant and the policy variables G_{kt} match our simulated data.

The third and final replacement rule to replace the missing value of $\tilde{g}(\cdot)$ for the simulated states of the world Ω^*_{kt} that we do not observe in the data is to use the annual mean $\tilde{g}$ in place of the missing $\tilde{g}(\cdot)$.

Table 12 shows which replacement rule is used in each simulation. Each year there are a maximum of 2610 potential entrants (3 in each of the 870 counties), for a total of 33,930 potential observations.[19] Virtually all replacements were made in Rule 1 or Rule 2. The No Policy simulation was the most challenging in this respect because there were relatively few years and counties among which to find replacements.

Table 13 compares the observed statistics E, E_t,W, $\bar{w}_e$, s_e in the data with their simulated values under the Base replication. The Base replication does a good job of replicating the observed number of entrants and their welfare. The simulated number of entrants in the replication has a mean of 136, versus 132 in the data. The data and the Base replication also have similar values for the mean welfare per entrant $\bar{w}_e$ and for total welfare E.

In the first set of counterfactual policy scenarios, we remove each policy individually, and use two-sample t-tests to compare the results of these simulations to the those of the Base replication. As seen in the results in Table 14, removing the RFS2 significantly decrease the number of entrants compared to the Base, while removing RFS1 and the state tax credit have smaller but noticeable affects on the number of entrants as well. The removal of RFS2 also has the largest impact on the mean welfare per entrant $\bar{w}_e$ of the three policies, indicating RFS2 had the largest impact on entrant payoff of the four policies.

We also simulate the No Policy scenario, which removes all the policies, *MTBE ban*, *RFS 1*, *RFS 2*, and *Tax Credit*, that might promote investment in ethanol plants, and use two-sample t-tests to compare the results of each scenario to the those of the No Policy.

As seen in Table 14, there are two striking results that arise from comparing entrants and welfare in the Base and No Policy scenarios. First, the mean number of entrants with no policies is 37, versus 136 in the Base simulation. Together, the four policies led to most of the investment in plants over the 13 years of the simulation. The second important takeaway is that the mean welfare per entrant, $\bar{w}_e$, is significantly lower in the No Policy scenario than it is under the Base replication scenario. There is less entry because expected payoff from investment in an ethanol plant is much lower without the policies. The standard deviation of welfare per entrant is still large though; policy changes

[19] In practice, the total number of investor-year combinations is lower because once a potential investor invests, he exits the sample.

account for some, but not all, of the differences in profitability across space and time.

Table 15 shows the results of the No MTBE Ban scenario, as well as the Base, No Policy, and No Tax Credit scenarios for the pre-RFS period (1996-2004). We conducted these simulations through 2004 instead of through 2008 because it was not possible to identify states of the world with one of the RFS standards in place, but without the MTBE ban. In this period, there were 48 entrants in the Base replication (46 in the data: see Table 13), and 29 in the scenario with No MTBE Ban; this large difference is statistically significant. In this same time frame, there were 26 entrants in the No Policy scenario, and the difference between the No Policy scenario and the No MTBE Ban scenario is only marginally statistically significant.

In this same pre-RFS time period, the No Tax Credit scenario leads to fewer entrants that the Base replication, but this number is still more than the No MTBE Ban scenario. In aggregate, this results indicates that the MTBE Ban had a bigger effect on entry than the state tax credits in the pre-RFS era during which the effects of the two policies can be identified and compared.

We disaggregate these results by year in Table 16 to further explore the interactions among the policy effects. Viewing the simulated entrants by year is useful to begin to disentangle the effects of the MTBE Ban and the RFS. Figure 2 shows the cumulative number of entrants and the total cumulative welfare of entrants over time. Entry and total welfare of entrants increased faster in the later years of the analysis, particularly in the years during which the RFS2 was in effect (2007-2008).

As seen in Table 16, the No Tax Credit simulation yielded on average 9% fewer entrants per given year compared to the Base simulation. The impact was smaller in the earlier years of the simulation, when fewer states had policies in place. The No RFS1 simulation had a slightly larger impact on the number of entrants than the No Tax Credit simulation for the years when RFS1 was in effect (2005-2006), though the cumulative number of entrants was still greater under the no RFS1 scenario because it was in effect for fewer years.

The No RFS2 scenario led to a much more marked decrease in the number of entrants per year compared to the no RFS1 and No Tax Credit scenarios (Figure 2), though the number of entrants per year during the RFS2 period (2007-2008) was still greater than the beginning of our analysis period due to other favorable economic conditions (Table 16).

Though we can only identify the No MTBE Ban scenario in the pre-RFS era (before 2005), we find similar magnitude of impact on the number of entrants as the No RFS2 scenario, particularly as we get closer to 2005, when all the states in our analysis had banned MTBE.

In the No Policy scenario, entry was slow and relatively constant over time, ranging from 1.6 to 4.1 new plants each year. In the Base replication the number of entrants per year increased over time, with a maximum of 32.5 new plants in 2007 (the second to last year of the simulation).

Figure 3 shows how the mean welfare per entrant by year changed over time under each scenario. The lines

for the No RFS1 and the No Tax Credit scenario closely track the Base replication, indicating that these policies had relatively small impacts on profitability for entrants. However, both the No MTBE Ban and No RFS2 scenarios led to significantly lower welfare for entrants compared to the Base replication in respective the years when the MTBE ban and the RFS2 were in effect.

Welfare per entrant was lower in the pre-RFS era, which is why there were fewer entrants. The first states in our sample banned MTBE as early as 2000, when we see the welfare per entrant under the No MTBE Ban scenario drop significantly below that of the Base replication. During the period 2000-2004, which represents the period during which there were some MTBE bans but no RFS1 or RFS2, the MTBE ban accounted for 54% of the entrants in the period. Without the ban, there would have been 16 new plants instead of the 35 that entered in the Base scenario. The RFS2 had a larger impact in percentage and real terms. However, the level of entry in the Base replication was higher in later years due to the combination of policy and market factors.

We disaggregate the results by each of the 10 Midwestern states in Table 17. Different states provide better environment for entering ethanol plants, as well as implementing the MTBE ban and offering tax credits at different times. Figure 4 shows how entry compares across states and policy scenarios. Each bar in the graph shows the number of entrants in the pre-RFS period (1996-2004) in black, and the number of entrants in the post-RFS period (2005-2008) in grey, for each state and each policy scenario. Figure 5 presents the mean welfare per entrant for each scenario by state, for the full period (left panel) and for the pre-RFS period (right panel).

There are noticeable differences across states in the total number of entrants, in the timing of the entrants, and in the relative impact of the different policy scenarios on entry. First, some states attract much more entry of ethanol plants than others under all scenarios. In particular, Iowa and Nebraska have the most entrants. The total number of entrants does not exactly correspond with the mean welfare per entrant, however (Figure 3). The mean welfare per entrant is high in these two states, but overall, entrants had higher welfare from entry in Indiana and South Dakota in the Base replication; South Dakota had fewer entrants because only part of the state is at all suitable for ethanol production.

The second important difference across states is that some states had relatively more entrants in the pre-RFS era than others. Nebraska, for example, had over half of its plants enter before 2005. Minnesota also experienced mroe entry in the pre-RFS era. Both these states implemented MTBE bans early (in 2000), and also had state tax credits for plants that gave them more favorable conditions for entrants.

Different policies had different impacts on different states. The number of pre-RFS entrants in the Base and No MTBE Ban scenarios is directly proportional to the number of years the MTBE ban was in effect in each state, indicating that this policy made a large contribution to industry growth in the region. Likewise, the No RFS2 scenario led to fewer entrants in all states, indicating that the RFS2 was a driver of industry growth in the last two years of our analysis.

The No Tax Credit scenario had more mixed results. All the states except Ohio, Iowa, and Illinois had tax credits available to entrants at some point during the analysis, though the year these policies were in effect varied across states.

8 Conclusions

In this paper we examine how economic factors, government policies, and strategic interactions affect decisions about whether and when to invest in building a new ethanol plant. We model the decision to invest in ethanol plants at the county level using both reduced-form discrete response models and a structural model of a dynamic game. We focus on investment in corn-ethanol plants in the Midwestern United States, where the majority of corn in the US is grown, over the period 1996-2008. According to the results of the structural model, our preferred model, the intensity of corn production; government policies, particularly the MTBE ban and the 2007 Renewable Fuel Standard (RFS2); and private information shocks all have significant effects on ethanol investment payoffs and decisions. The competition effect and the agglomeration effect do not have any net strategic effect on the payoff from investing in an ethanol plant. We use the estimated structural parameters to simulate counterfactual policy scenarios to disentangle the impacts of state and national policies on the timing and location of investment in the industry. We find that, of the policies analyzed, the MTBE ban and the RFS2 led to most of the investment during this time period.

One possible reason the MTBE ban was effective in inducing investment in building ethanol plants is that it increased the demand for ethanol as an oxygenate in place of MTBE. Similarly, one possible reason the RFS2 was effective in inducing investment in building ethanol plants is that it increased demand for ethanol by mandating an expansion in ethanol consumption. Thus, our results suggest that policies that increase the demand for ethanol have the potential for inducing investment in building ethanol plants.

Both the MTBE ban and the Renewable Fuel Standard can function as implicit blending mandates (De Gorter and Just, 2010; Anderson and Elzinga, 2012). However, whenever unpriced emissions are the sole market failure, a carbon tax or cap and trade program is more likely to achieve the first-best (Pigou, 1920; Coase, 1960), while fuel mandates are unable to replicate the first-best solution (Helfand, 1992; Holland, Knittel and Hughes, 2009; Lapan and Moschini, 2012). Lade and Lin Lawell (2015) show that when renewable fuel mandates are combined with a cost containment mechanism such as a credit window price, the efficiency of the mandate can increase substantially. Thus, while the MTBE ban and the Renewable Fuel Standard were effective in inducing investment in building ethanol plants, it is possible to increase their efficiency by combining them with cost containment mechanisms or by using a market-based instrument instead. We hope to explore these possibilities in future work.

References

Aguirregabiria, V., and P. Mira. (2007). Sequential estimation of dynamic discrete games. Econometrica 75(1): 1-53.

American Coalition for Ethanol. (2007). STATUS: A state by state handbook. Sioux Falls, SD.

Anderson, S., and A. Elzinga. (2012). A ban on one is a boon for the other: strict gasoline content rules and implicit ethanol blending mandates. Energy institute at Hass Working Paper, 2012.

Bajari, P., C.L. Benkard, and J. Levin. (2007). Estimating dynamic models of imperfect competition. Econometrica 75(5): 1331-1370.

Bartik, T.J. (1985). Business location decisions in the United States: Estimates of the effects of unionization, taxes, and other characteristics of states. Journal of Business and Economic Statistics 3(1): 14-22.

Biodiesel Magazine. (2008). Biodiesel Plant Lists. Last accessed online July, 2008. (biodieselmagazine.com).

Coase, R. (1960). The problem of social cost. Journal of Law and Economics 3: 1-44.

Cotti, C. and M. Skidmore. (2010). The impacts of state government subsidies and tax credits in an emerging industry: ethanol production 1980-2007. Southern Economic Journal 76(4): 1076-93.

De Gorter, H. and D.R. Just. (2010). The social costs and benefits of biofuels: The intersection of environmental, energy and agricultural policy. Applied Economics Perspectives and Policy 32(1): 4-32.

Dhuyvetter, K.C., T.L. Kastens, and M. Boland. (2005). The US ethanol industry: Where will it be located in the future? Agricultural Marketing Resource Center and Agricultural Issues Center, University of California.

Dixit, A.K., and R.S. Pindyck. (1994). Investment Under Uncertainty. Princeton, NJ: Princeton University Press.

DOE [Department of Energy]. (2008). Energy Time Lines: Ethanol. Revised June 2008. Washington, DC: DOE.

EIA [Energy Information Administration of the DOE]. (2009). Total Gasoline Rack Price, Total Industry Industrial Electrical Price, Natural Gas City Gate Price. Last accessed online July 2009. (http://tonto.eia.doe.gov/dnav).

EIA [Energy Information Administration of the DOE]. (2015). New EIA Monthly Data Track Crude Oil Movements by Rail. Last accessed online April 2015. (http://www.eia.gov/todayinenergy/detail.cfm?id=20592#tabs_Slider-5).

Ellison, G., and E.L. Glaeser (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review 89(2): 311-316.

Ethanol Producer Magazine. (2010). Plant lists. Accessed online December, 2010. (www.ethanolproducer.com).

Fatal, Y.S., S. Kotisir, H.A. Tejeda, and C. Zhan. (2012). Reducing GHG emissions and energy input in the U.S. supply chain of ethanol and gasoline. Working paper, North Carolina State University.

Gallagher, P. (2009). Roles for evolving markets, policies, and technology improvements in U.S. Corn Ethanol Industry Development. Regional Economic Development 5(1): 12-33.

Goetz, S. (1997). State- and county-level determinants of food manufacturing establishment growth: 1987-93. American Journal of Agricultural Economics 79: 838-850.

Haddad, M. A., G. Taylor, and F. Owusu. (2010). Locational choices in of the ethanol industry in the Midwest corn belt. Economic Development Quarterly 24(1): 74-86.

Helfand, G. (1992). Standards versus standards: The effects of different pollution restrictions. American Economic Review 81: 622-634.

Holland, S., C. Knittel, and J. Hughes. (2009). Greenhouse gas reductions under low carbon fuel standards? American Economic Journal: Economic Policy 1(1): 106-146.

Koplow, D. (2007). Biofuels- At what cost? Government support for ethanol and biodiesel in the United States: 2007 Update. Prepared for Global Subsidies Initiative (GSI) of the International Institute for Sustainable Development (IISD), Geneva.

Kwiatkowski, J.R., A.J. McAloon, F. Taylor, and D.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.

Lambert, D.M., M. Wilcox, A. English, and L. Stewart. (2008). Ethanol plant location determinants and county comparative advantage. Journal of Agricultural and Applied Economics 40: 117-135.

Lapan, H., and G. Moschini. (2012). Second-best biofuels policies and the welfare effects of quantity mandates and subsidies. Journal of Environmental Economics and Management 63: 224-241.

Lade, G.E., and C.-Y.C. Lin Lawell. (2015). Mandating green: On the design of renewable fuel policies and cost containment mechanisms. Working paper, University of California at Davis.

Lin, C.-Y.C. (2013). Strategic decision-making with information and extraction externalities: A structural model of the multi-stage investment timing game in offshore petroleum production. Review of Economics and Statistics 95(5): 1601-1621.

Maxwell, C., and M. Davison. (2013). Using real option analysis to quantify ethanol policy impact on the firm's entry into and optimal operation of corn ethanol facilities. Energy Economics 42: 140-151.

NASS [National Agricultural Statistics Service of the USDA]. (2010). Quickstats database. Last accessed December, 2010. (www.nass.usda.gov/QuickStats/).

NASS [National Agricultural Statistics Service of the USDA]. (2007). Ethanol co-products used for livestock feed. Washington, DC: NASS.

National Biodiesel Board. (2008). Plant Lists. Last accessed 2008. (www.biodiesel.org).

McNew, K., and D. Griffith. (2005). Measuring the impact of ethanol plants on local grain prices. Review of Agricultural Economics 27(2): 164-180.

Nebraska Energy Office. (2010). Ethanol Prices. Accessed online November, 2010. www.neo.ne.gov.statshtml/66.html

Pakes, A., M. Ostrovsky, and S. Berry. (2007). Simple estimators for the parameters of discrete dynamic games (with entry and exit examples). RAND Journal of Economics 38(2): 373-399.

Perrin, R.K., N.F. Fretes, and J.P. Sesmero. (2009). Efficiency in Midwest US corn ethanol plants: A plant survey. Energy Policy 37(4): 1309-1316.

Pigou, A. (1920). The Economics of Welfare. Macmillan and Co.

Renewable Fuels Association. (2010). Plant Lists. Last accessed December 2010. (www.ethanolrfa.org).

Ryan, S.P. (2012). The costs of environmental regulation in a concentrated industry. Econometrica 80 (3): 1019-1061.

Sarmiento, C., and W.W. Wilson. (2007). Spatial competition and ethanol plant location decisions. Annual Meeting of the American Agricultural Economics Association, July 27-29, 2008, Orlando, Florida.

Schmidt, T.M., J. Luo, and L.W. Tauer. (2009). Ethanol plant investment using net present value and real options analysis. Biomass and Bioenergy 33: 1442-1451.

USDA. (2007). Ethanol Transportation Backgrounder: Expansion of US corn-based ethanol from the agricultural transportation perspective. Washington DC: USDA AMS.

Urbanchuk, J.M. (2006). Economic Impacts on the Farm Community of Cooperative Ownership of Ethanol Production. National Corn Growers Association Report. Chesterfield, MO: NCGA.

Whittington, T. (2006). International Overview of Ethanol Production and Policies. Department of Agriculture and Food Western Australia Report.

Wooldridge, J.M. (2010). Econometric Analysis of Cross Section and Panel Data. Second Edition. Cambridge, MA: MIT Press.

Yi, F., and C.-Y.C. Lin Lawell. (2015a). Ethanol plant investment in Canada: A structural model. Working paper, University of California at Davis.

Yi, F., and C.-Y.C. Lin Lawell. (2015b). What factors affect the decision to invest in a fuel ethanol plant?: A structural model of the ethanol investment timing game. Working paper, University of California at Davis.

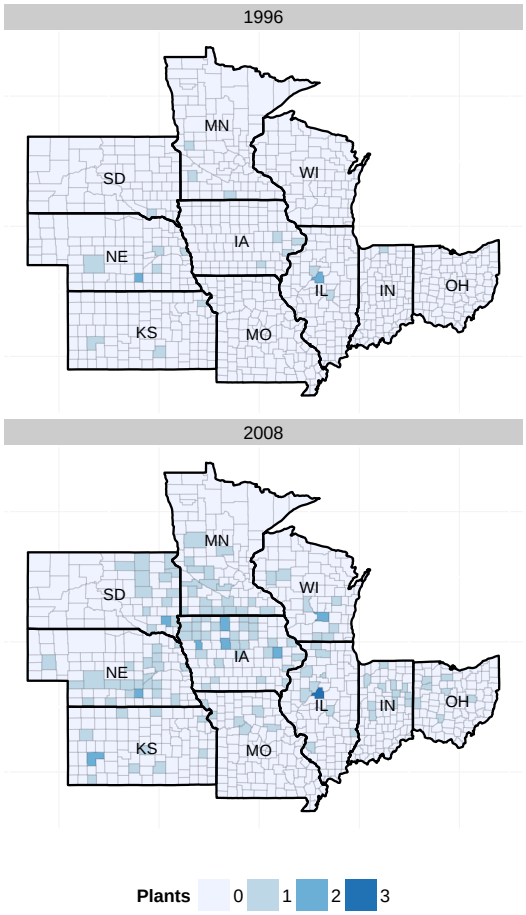

Figure 1: Number of operational ethanol plants by county in the Midwestern United States

Table 1: Summary statistics for variables used in reduced-form analysis

Variable	Counties with at least one new ethanol plant: 1996-2008 mean	Counties with at least one new ethanol plant: 1996-2008 std. dev.	Full Sample mean	Full Sample std. dev.	Spatial Resolution of Data
Investment in a New Plant [dependent variable: indicator]	0.034	0.182	0.004	0.066	
Existing Plants [count]	0.202	0.417	0.040	0.203	county
Existing Plants (spatial lag) [count]	0.632	1.069	0.395	0.819	region[a]
Existing Biodiesel Plant [indicator]	0.015	0.126	0.010	0.105	county
MTBE Ban [indicator]	0.582	0.493	0.487	0.500	state
Tax Credit [indicator]	0.370	0.483	0.346	0.476	state
RFS 1 [indicator]	0.167	0.373	0.163	0.369	national
RFS 2 [indicator]	0.146	0.354	0.150	0.357	national
Ethanol Price [$/gallon]	1.781	0.416	1.778	0.418	national
Gasoline Price [$/gallon]	1.349	0.553	1.341	0.562	state
Natural Gas Price [$/1000 ft3]	6.527	1.852	6.516	1.917	state
Electricity Price [cents/KwH]	5.107	0.495	5.232	0.536	state
Corn Price [$/bushel]	2.785	0.675	2.828	0.673	state
Soy Price [$/bushel]	7.106	1.690	7.160	1.693	state
Corn Intensity [acres planted/total acreage]	0.299	0.130	0.200	0.144	county
Corn Intensity (spatial lag) [acres planted/total acreage]	0.279	0.123	0.200	0.129	region[a]
Soy Intensity (county)[acres planted/total acreage]	0.245	0.127	0.183	0.131	county
Cow Density [head/acre]	0.103	0.055	0.085	0.051	district[b]
Observations	3687		28769		
Number of Counties	120		855		

[a] contiguous bordering counties [b] USDA definition

Table 2: Results from reduced-form fixed effects model of ethanol investment

	Dependent variable is probability of investment in a new plant					
	Base Model: Fixed Effects Logit		Linear Probability Models			
			FE Logit sample		Full sample	
	A	B	C	D	E	F
Existing Plants	-13.79***	-13.35***	-0.180***	-0.172***	-0.077***	-0.076***
	(1.722)	(1.585)	(0.011)	(0.011)	(0.004)	(0.004)
Existing Plants (spatial lag)	0.57	0.44	0.004	0.008	0.002*	0.003**
	(0.417)	(0.378)	(0.006)	(0.006)	(0.001)	(0.001)
Existing Biodiesel Plant	-0.51	-0.10	-0.037	-0.020	-0.004	-0.003
	(1.729)	(1.450)	(0.032)	(0.032)	(0.005)	(0.005)
MTBE Ban	-0.97	-0.98	0.003	-0.006	-0.003	-0.003
	(0.927)	(0.808)	(0.013)	(0.012)	(0.002)	(0.002)
Tax Credit	0.19	1.22*	0.006	0.015	-0.001	0.000
	(0.707)	(0.585)	(0.008)	(0.008)	(0.001)	(0.001)
RFS 1		0.72		-0.002		-0.002
		(1.433)		(0.020)		(0.003)
RFS 2		0.50		0.058		-0.000
		(3.067)		(0.041)		(0.005)
Gasoline Price	29.42*	5.40	1.120***	0.118*	0.105***	0.013*
	(12.462)	(3.230)	(0.246)	(0.048)	(0.031)	(0.006)
Ethanol Price		-0.54		-0.019		-0.004
		(2.379)		(0.033)		(0.004)
Natural Gas Price	0.94	-0.44	0.009	-0.009	0.002	-0.001
	(1.002)	(0.373)	(0.009)	(0.005)	(0.001)	(0.001)
Electricity Price	0.39	0.24	0.004	-0.003	0.003*	0.002
	(0.667)	(0.558)	(0.010)	(0.010)	(0.001)	(0.001)
Corn Price	-3.16	0.02	0.027	-0.022	0.009	-0.001
	(3.244)	(1.503)	(0.050)	(0.021)	(0.006)	(0.003)
Soy price	-2.70	0.13	-0.015	0.007	-0.002	0.000
	(1.417)	(0.380)	(0.025)	(0.006)	(0.003)	(0.001)
Corn Intensity	-6.27	-11.73	0.152	0.204	0.033	0.035
	(16.945)	(16.029)	(0.245)	(0.245)	(0.032)	(0.032)
Corn Intensity (spatial lag)	-5.54	-11.48	-0.265	-0.634	0.023	-0.016
	(22.536)	(18.690)	(0.348)	(0.332)	(0.047)	(0.045)
Soy Intensity	-15.91	-17.07	-0.171	-0.128	-0.034	-0.026
	(11.944)	(10.947)	(0.174)	(0.170)	(0.024)	(0.024)
Cow Density	7.61	10.78	-0.175	-0.057	0.294***	0.302***
	(26.297)	(25.524)	(0.409)	(0.404)	(0.077)	(0.077)
Year (trend)		0.75**		0.002		0.001
		(0.250)		(0.003)		(0.000)
Constant	NO	NO	YES	YES	YES	YES
County Fixed Effects	YES	YES	YES	YES	YES	YES
Year Effects	YES	NO	YES	NO	YES	NO
Observations	3,687	3,687	3,687	3,687	28,769	28,769
No. counties	120	120	120	120	855	855
Pseudo-R^2 or R^2	0.524	0.505	0.109	0.0989	0.0223	0.0213
Regression statistic (Chi^2 or F)	442.7	426.7	16.58	21.64	24.45	33.72
$Pr > Chi^2$ or $Pr > F$	0.000	0.000				

Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, * p<0.05

Table 3: Specification tests for reduced-form fixed effects model of ethanol investment

	Base Model: Fixed Effects Logit		Linear Probability Models			
			FE Logit sample		Full sample	
	A	B	C	D	E	F
Durbin-Wu-Hausman Test of Endogeneity of Corn Intensity (Ho: not endogenous)						
Coefficient on residual	15.75	5.243	0.327	-0.0892	-0.00702	-0.0245
$Pr > z$	(0.687)	(0.883)	(0.589)	(0.882)	(0.933)	(0.768)
Hausman test of Orthogonality of Random Effect (Ho: orthogonal)						
Hausman Chi^2	2674	2922	266.1	226.4	241.6	215.3
$Pr > Chi^2$	0.000	0.000	0.000	0.000	0.000	0.000
County Fixed Effects	YES	YES	YES	YES	YES	YES
Year Effects	YES	NO	YES	NO	YES	NO
Observations	3,687	3,687	3,687	3,687	28,769	28,769
No. counties	120	120	120	120	855	855

Table 4: Robustness of reduced-form fixed effects model to specification of corn and soy intensity and price

	Dependent variable is probability of investment in a new plant									
	Base Model: Fixed Effects Logit		Alternate Corn and Soy Intensity and Price Specifications							
	A	B	G	H	I	J	K	L	M	N
Existing Plants	-13.76***	-13.35***	-13.98***	-13.31***	-13.70***	-13.21***	-13.81***	-13.17***	-13.45***	-13.26***
	-1.72	(1.585)	-1.798	(1.581)	(1.701)	(1.565)	(1.749)	(1.561)	(1.643)	(1.560)
Existing Plants (spatial lag)	0.57	0.44	0.55	0.43	0.60	0.48	0.56	0.47	0.43	0.48
	(0.416)	(0.378)	(0.427)	(0.379)	(0.417)	(0.370)	(0.427)	(0.370)	(0.418)	(0.370)
MTBE Ban	-0.96	-0.98	-1.02	-0.95	-1.00	-0.96	-1.06	-0.92	-0.96	-1.03
	(0.927)	(0.808)	(0.930)	(0.806)	(0.927)	(0.808)	(0.927)	(0.808)	(0.924)	(0.784)
Tax Credit	0.20	1.22*	0.19	1.20*	0.06	1.06	0.10	1.03	0.45	1.12*
	(0.706)	(0.585)	(0.720)	(0.586)	(0.688)	(0.567)	(0.704)	(0.568)	(0.644)	(0.548)
RFS 1		0.72		0.84		0.58		0.72		0.38
		(1.433)		(1.417)		(1.428)		(1.409)		(1.208)
RFS 2		0.50		0.72		0.19		0.46		-0.20
		(3.067)		(3.003)		(3.092)		(3.022)		(2.619)
Corn Price	-3.24	0.02	-9.10	0.33	-2.88	0.31	-8.01	0.53	-3.18	0.63
	(3.241)	(1.503)	(5.201)	(0.727)	(3.276)	(1.506)	(5.064)	(0.726)	(3.279)	(0.667)
Soy Price	-2.69	0.13			-2.56	0.10				
	(1.415)	(0.380)			(1.411)	(0.380)				
Ratio of Corn to Soy Price			43.94	-3.59			37.71	-3.23		
			(32.010)	(7.048)			(30.863)	(7.032)		
Corn Intensity	-6.33	-11.73	-4.16	-11.94	7.80	3.18	9.46	2.82	6.02	4.33
	(16.927)	(16.029)	(16.761)	(16.005)	(14.556)	(13.489)	(14.741)	(13.492)	(13.375)	(12.874)
Corn Intensity (spatial lag)	-5.66	-11.48	-9.50	-12.15	-6.36	-12.14	-9.66	-12.73	-4.43	-12.68
	(22.525)	(18.690)	(22.878)	(18.733)	(22.630)	(18.671)	(22.946)	(18.714)	(22.396)	(18.468)
Soy Intensity	-15.88	-17.07	-16.06	-17.06						
	(11.918)	(10.947)	(11.867)	(10.922)						
Ratio of Corn to Soy Intensity					0.01	0.02	0.00	0.02		
					(0.097)	(0.090)	(0.098)	(0.089)		
Time Specification	Year Effect	Trend	Year Effect	Trend	Year Effect	Trend	Year Effect	Trend	Year Effect	Trend
Observations	3,687	3,687	3,687	3,687	3,687	3,687	3,687	3,687	3,736	3,736
number of counties	120	120	120	120	120	120	120	120	121	121
Pseudo-R^2	0.524	0.505	0.522	0.505	0.522	0.502	0.520	0.502	0.519	0.503

All estimates include county fixed effects. Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, * p<0.05

Table 5: Robustness of reduced-form fixed effects model to the specification of ethanol and gasoline prices

	Dependent variable is probability of investment in a new plant						
	Base Model: Fixed Effects Logit		Alternate Gasoline and Ethanol Price Specifications				
	A	B	O	P	Q	R	S
Existing Plants	-13.79***	-13.35***	-13.70***	-13.08***	-13.53***	-13.48***	-13.47***
	(1.722)	(1.585)	(1.717)	(1.508)	(1.705)	(1.612)	(1.620)
Existing Plants (spatial lag)	0.57	0.44	0.57	0.57	0.55	0.42	0.45
	(0.417)	(0.378)	(0.439)	(0.358)	(0.443)	(0.383)	(0.381)
MTBE Ban	-0.97	-0.98	-0.94	-1.45	-0.84	-0.95	-0.94
	(0.927)	(0.808)	(0.957)	(0.840)	(0.976)	(0.804)	(0.812)
Tax Credit	0.19	1.22*	-0.02	1.60**	0.00	1.18*	1.27*
	(0.707)	(0.585)	(0.719)	(0.614)	(0.722)	(0.589)	(0.599)
RFS 1		0.72		2.41*		1.11	2.18*
		(1.433)		(1.026)		(1.089)	(1.039)
RFS 2		0.50		4.03**		1.37	4.39**
		(3.067)		(1.562)		(1.818)	(1.645)
Gasoline Price	29.42*	5.40			-23.71	4.82**	
	(12.462)	(3.230)			(25.361)	(1.612)	
Ethanol Price		-0.54					3.29**
		(2.379)					(1.274)
Ratio of Ethanol to Gasoline Price			-47.42**	0.11	-73.36*	0.72	-1.74
			(15.537)	(1.874)	(32.735)	(1.993)	(2.170)
Time Specification	Year Effect	Trend	Year Effect	Trend	Year Effect	Trend	Trend
Observations	3,687	3,687	3,687	3,687	3,687	3,687	3,687
number of counties	120	120	120	120	120	120	120
Pseudo-R^2	0.524	0.505	0.530	0.494	0.531	0.505	0.502

All estimates include county fixed effects. Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, * p<0.05

Table 6: Robustness of reduced-form fixed effects model to the specification of tax credit and existing plant

	Dependent variable is probability of investment in a new plant					
	Base Model: Fixed Effects Logit		Tax Credit Specification		Existing Plant Specification	
	A	B	T	U	V	W
Existing Plants	-13.76***	-13.35***	-13.72***	-13.61***		
	(1.716)	(1.585)	(1.716)	(1.617)		
Existing Ethanol Capacity [gal per acre]					-4.53***	-4.63***
					(0.528)	(0.533)
Existing Biodiesel Plant	-0.50	-0.10	-0.49	-0.16		
	(1.722)	(1.450)	(1.704)	(1.454)		
Existing Biodiesel Capacity [gal per acre]					0.01	0.01
					(0.037)	(0.035)
MTBE Ban	-0.96	-0.98	-0.89	-1.08	-0.41	-0.75
	(0.927)	(0.808)	(0.926)	(0.821)	(0.837)	(0.710)
Tax Credit (indicator)	0.20	1.22*			1.04	1.45*
	(0.706)	(0.585)			(0.615)	(0.564)
Lifetime Tax Credit Benefit ($100,000)			0.04	0.10*		
			(0.047)	(0.040)		
RFS 1		0.72		0.85		0.16
		(1.433)		(1.438)		(1.265)
RFS 2		0.50		0.54		0.11
		(3.067)		(3.078)		(2.880)
Time Specification	Year Effect	Trend	Year Effect	Trend	Year Effect	Trend
Observations	3,687	3,687	3,687	3,687	3,687	3,687
number of counties	120	120	120	120	120	120
Pseudo-R^2	0.524	0.505	0.525	0.507	0.480	0.464

All estimates include county fixed effects. Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, * p<0.05

Table 7: Bin design of variables for structural estimation

	Base Bins		Alternate Bins		
Variables	Bin Design	Break	Bin Design	Break 1	Break 2
Cow Density (head/acre)	Bottom two thirds and top third	0.103	Middle Bin 1.5 Std. Dev. around Mean	0.048	0.124
Corn Intensity (acres planted / total acres)	Equal sizes	0.175	Middle Bin 1.5 Std. Dev. around Mean	0.078	0.191
Ethanol Price ($/gal)	Equal sizes	1.630	Middle Bin is middle 5 years	1.51	1.91
Gasoline Price ($/gal)	Equal sizes	1.110	Bottom third and top two thirds	1	
Output Price Indicator	High if both ethanol and gasoline prices are high				
Alternate Corn Price ($/bushel)	Equal sizes	3.010	Middle Bin 1.5 Std. Dev. around Mean	2.317	3.32
Corn Price ($/bushel)	Bottom third and top two thirds	2.340	Middle Bin 2 Std. Dev. around Mean	5.48	8.88
Natural Gas Price ($/1000ft3)	Equal sizes	6.810	Middle Bin 2 Std. Dev. around Mean	4.519	8.349
Electricity Price (cents/KwH)	Equal sizes	5.130	Middle Bin 2 Std. Dev. around Mean	4.702	5.741
Energy Input Price Indicator	High if both natural gasoline and electricity prices are high				

Table 8: Summary statistics for discretized variables used in structural estimation

	Base Bins				Alternate Bins		
	Full Sample		Sample of Entrants		Full Sample		Spatial Resolution
Variable:	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
New Plant	0.004	0.063	1	0	0.004	0.063	county
Existing Plant	0.036	0.185	0.083	0.277	0.037	0.188	county
Tax Credit	0.341	0.474	0.394	0.491	0.341	0.474	state policy
MTBE Ban	0.476	0.499	0.864	0.345	0.475	0.499	state policy
RFS I	0.153	0.360	0.182	0.387	0.153	0.360	national policy
RFS II	0.151	0.358	0.470	0.501	0.151	0.358	national policy
Cow Density	0.330	0.470	0.432	0.497	0.943	0.760	district[a]
Corn Intensity	0.494	0.500	0.803	0.399	0.917	0.669	county
Corn Price	0.677	0.468	0.742	0.439	0.918	0.712	state
Alternate Corn Price	0.513	0.500	0.644	0.481			state
Soy Price					1.093	0.596	state
Output Price Indicator	0.648	0.478	0.849	0.360			state
Ethanol Price	0.535	0.499	0.788	0.410	0.917	0.728	national
Gasoline Price	0.493	0.500	0.803	0.399	0.380	0.485	state
Energy Input Price Indicator	0.797	0.402	0.932	0.253			state
Natural Gas Price	0.492	0.500	0.811	0.393	0.945	0.649	state
Electricity Price	0.499	0.500	0.371	0.485	0.971	0.545	state
Metro Area	0.283	0.450	0.197	0.399	0.010	0.099	county
Biodiesel Plant	0.010	0.100	0.023	0.150	0.285	0.452	county
Observations	33307		132		33307		
Number of Counties	870				870		

[a] USDA definition

Table 9: Results from estimation of structural model

	Base Model	Alternate price specifications				
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Coefficients in the investment payoff on:						
Existing Plant	0.034	0.021	-0.237	0.042	-0.129	0.039
	(0.279)	(0.286)	(0.29)	(0.311)	(0.268)	(0.307)
Tax Credit	0.209	0.206	0.179	0.16	0.216	0.26
	(0.147)	(0.147)	(0.157)	(0.154)	(0.154)	(0.178)
MTBE Ban	0.814**	1.022***	0.837**	0.936**	0.907*	0.956*
	(0.293)	(0.303)	(0.305)	(0.299)	(0.372)	(0.323)
RFS 1	0.085	0.05	0.168	0.181	0.166	0.142
	(0.242)	(0.214)	(0.283)	(0.313)	(0.26)	(0.279)
RFS 2	0.727**	0.658**	0.786*	0.946**	0.816**	0.965***
	(0.256)	(0.231)	(0.32)	(0.338)	(0.309)	(0.27)
Cow Density	0.189	0.184	0.206	0.28‡	0.22	0.229
	(0.149)	(0.136)	(0.155)	(0.16)	(0.129)	(0.162)
Corn Intensity	1.012***	0.976***	0.962***	1.193***	0.986***	1.217***
	(0.181)	(0.163)	(0.201)	(0.198)	(0.213)	(0.22)
Energy Output Indicator	−0.423‡	−0.573‡			-0.542	-0.429
	(0.246)	(0.307)			(0.348)	(0.334)
Ethanol Price				-0.376		
				(0.364)		
Gasoline price			-0.289	-0.096		
			(0.286)	(0.245)		
Corn Price	-0.074	-0.071	-0.08	-0.085	-0.167	-0.183
	(0.265)	(0.197)	(0.239)	(0.205)	(0.259)	(0.231)
Energy Input Indicator		0.753*	0.517		0.67	
		(0.354)	(0.41)		(0.444)	
Natural Gas Price	0.374			0.436		0.383
	(0.275)			(0.404)		(0.349)
Electricity Price				0.036		
				(0.179)		
Constant	-4.97***	-5.164***	-5.087***	-6.108***	-5.042***	-5.962***
	(0.411)	(0.372)	(0.512)	(0.403)	(0.413)	(0.506)
Sigma	0.648***	0.612***	0.61***	0.786***	0.606***	0.776***
	(0.042)	(0.039)	(0.048)	(0.043)	(0.073)	(0.051)

Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, * p<0.05, ‡ p<0.01

Table 10: Results of estimation of structural model with alternate variable and bin specifications

	Base Model	Additional Covariates	Alternate (More) Bins	
	(i)	(vii)	(viii)	(ix)
Coefficients in the investment payoff on:				
Existing Plant	0.034	-0.123	0.135	-0.017
	(0.279)	(0.26)	(0.363)	(0.347)
Tax Credit	0.209	0.123	0.109	0.394
	(0.147)	(0.135)	(0.247)	(0.398)
MTBE Ban	0.814**	1.044***	0.502‡	1.014***
	(0.293)	(0.296)	(0.268)	(0.284)
RFS 1	0.085	0.044	1.287***	1.674***
	(0.242)	(0.209)	(0.295)	(0.403)
RFS 2	0.727**	0.651*	2.343***	1.869***
	(0.256)	(0.268)	(0.322)	(0.266)
Cow Density	0.189	0.225‡	0.708***	0.812***
	(0.149)	(0.131)	(0.159)	(0.13)
Corn Intensity	1.012***	0.965***	0.209	0.315*
	(0.181)	(0.168)	(0.173)	(0.131)
Energy Output Price Indicator	-0.423	-0.586*		
	(0.246)	(0.281)		
Ethanol Price			-0.518	-1.916**
			(0.646)	(0.636)
Gasoline Price			2.168***	2.546***
			(0.551)	(0.613)
Corn Price	-0.074	-0.071	-0.439	0.089
	(0.265)	(0.216)	(0.34)	(0.266)
Soy Price			-0.493	0.67
			(0.59)	(0.758)
Energy Input Price Indicator		0.792*		
		(0.382)		
Natural Gas Price	0.374		-1.549*	-1.104*
	(0.275)		(0.69)	(0.474)
Electricity Price			-0.179	
			(0.253)	
Metro Area		-0.244	-0.564	-0.369
		(0.2)	(0.569)	(0.589)
Existing Biodiesel Plant		-0.06	0.033	0.023
		(0.48)	(0.084)	(0.074)
Constant	-4.97***	-5.08***	-5.591***	-6.583***
	(0.411)	(0.287)	(0.607)	(0.587)
Sigma	0.648***	0.609***	0.997***	0.77***
	(0.042)	(0.046)	(0.083)	(0.092)

Standard errors in parentheses. Significance codes: *** p<0.001, ** p<0.01, *p<0.05, ‡ p<0.01

Table 11: Counterfactual Policy Scenarios

Counterfactual Policy	Description
Base	Replication with observed data
No RFS1	Remove RFS1 (set *RFS1*=0)
No RFS2	Remove RFS1 (set *RFS2*=0)
No Tax Credit	Remove state tax credit (set *Tax Credit*=0)
No MTBE Ban	Remove MTBE ban (set *MTBE ban*=0) [Pre-RFS (1996-2004) only]
No Policy	Remove all polices (set all G_{kt} variables=0)

Table 12: Replacement rules followed in simulations for missing entry probabilities $\tilde{g}(\cdot)$

		Replacement Rule Followed:		
Simulation	Number Missing	Set *other plant*=0	Match policy and significant state variables	Use annual mean $\tilde{g}$
Base	48.5	48.4	0.1	0.0
No RFS1	66.3	65.5	0.8	0.0
No RFS2	101.2	99.5	1.7	0.0
No Tax Credit	827.3	88.7	738.5	0.0
No MTBE Ban (1996-2004)	380.9	168.1	212.8	0.0
No Policy	4209.1	427.7	3781.2	0.2

The replacement rules are used to replace the missing value of $\tilde{g}(\cdot)$ for the simulated states of the world Ω^*_{kt} that we do not observe in the data.

Table 13: Number of entrants and welfare in data and Base Scenario

Full Period	Number of Entrants	Total Welfare of Entrants	Mean Welfare per Entrant	Std. Dev. of Welfare per Entrant
Data	132	273.28	2.07	0.64
Base	135.92 (14.97)	278.21 (31.62)	2.05 (0.15)	0.704 (0.04)
1996-2004	Number of Entrants	Total Welfare of Entrants	Mean Welfare per Entrant	Std. Dev. of Welfare per Entrant
Data	46	65.68	1.43	0.596
Base	47.60 (14.38)	64.01 (32.41)	1.3449 (0.154)	0.6958 (0.0649)

For the Base scenario, the reported statistics are averages over 50 simulations. Standard errors are in parentheses, and are calculated from using the parameter estimates from each of the 250 bootstrap samples. For each of the 250 bootstrap samples, 50 simulations are run using the parameter estimates from that bootstrap sample. Standard errors for a statistic is the standard deviation of the respective statistics over all 250 bootstrap samples.

Table 14: Number of entrants and welfare under simulated policy scenarios: Full Period

	Base Replication	No RFS1	No RFS2	No Tax Credit	No Policy
Number of Entrants	135.9	131.6	91.8	123.2	36.6
	(15.0)	(17.1)	(17.8)	(15.7)	(17.8)
Total Welfare of All Entrants	278.2	267.8	157.0	246.6	36.6
	(31.6)	(35.4)	(36.7)	(34.0)	(42.6)
Mean of Welfare per Entrant	2.05	2.03	1.71	2.00	1.00
	(0.15)	(0.14)	(0.14)	(0.15)	(0.18)
Std. Dev. of Welfare per Entrant	0.70	0.72	0.65	0.72	0.60
	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)
t-statistic from two-sample t-test of difference between this scenario and Base scenario					
Number of Entrants		2.98**	29.95***	9.26***	67.55***
Total Welfare		-1.58	-12.38***	-4.79***	-35.00***
Mean Welfare per Entrant		0.97	26.25***	3.45***	71.49***
Std. Dev. of Welfare per Entrant		-2.99**	13.76***	-3.11**	20.16***
t-statistic from two-smaple t-test of difference between this scenario and No Policy scenario					
Number of Entrants	-67.55***	-60.98***	-34.68***	-57.69***	
Total Welfare	35.00***	32.17***	22.34***	29.99***	
Mean Welfare per Entrant	-71.49***	-71.34***	-49.41***	-67.79***	
Std. Dev. of Welfare per Entrant	-20.16***	-22.74***	-10.42***	-22.42***	

For each scenario, the reported statistics are averages over 50 simulations. Standard errors are in parentheses, and are calculated from using the parameter estimates from each of the 250 bootstrap samples. For each of the 250 bootstrap samples, 50 simulations are run using the parameter estimates from that bootstrap sample. Standard errors for a statistic is the standard deviation of the respective statistics over all 250 bootstrap samples. Significance codes for t-tests: *** p<0.001, ** p<0.01, * p<0.05

Table 15: Number of entrants and welfare under simulated policy scenarios: Pre-RFS period (1996-2004)

	Base Replication (to 2005)	No Tax Credit	No MTBE Ban	No Policy
Number of Entrants	47.6	43.0	28.9	26.1
	(14.4)	(14.4)	(16.0)	(15.6)
Total Welfare of All Entrants	64.0	54.6	24.6	20.3
	(32.4)	(33.5)	(35.7)	(36.2)
Mean of Welfare per Entrant	1.34	1.27	0.85	0.78
	(0.15)	(0.15)	(0.21)	(0.19)
Std. Dev. of Welfare per Entrant	0.70	0.69	0.56	0.55
	(0.06)	(0.07)	(0.06)	(0.06)
t-statistic from two-sample t-test of difference between this scenario and Base scenario				
Number of Entrants		3.61***	13.71***	16.01***
Total Welfare		-2.45*	-9.53***	-11.22***
Mean Welfare per Entrant		5.47***	29.86***	36.4***
Std. Dev. of Welfare per Entrant		0.34	23.95***	24.74***
t-statistic from two-sample t-test of difference between this scenario and No Policy scenario				
Entrants	-16.01***	-12.56***	-1.97*	
Total Welfare	11.22***	8.74***	1.69	
Mean Welfare per Entrant	-36.40***	-31.88***	-4.00***	
Std. Dev. of Welfare per Entrant	-24.74***	-23.86***	-1.14	

For each scenario, the reported statistics are averages over 50 simulations. Standard errors are in parentheses, and are calculated from using the parameter estimates from each of the 250 bootstrap samples. For each of the 250 bootstrap samples, 50 simulations are run using the parameter estimates from that bootstrap sample. Standard errors for a statistic is the standard deviation of the respective statistics over all 250 bootstrap samples. Significance codes for t-tests: ***p<0.001, **p<0.01, *p<0.05

Table 16: Number of entrants and mean welfare per entrant by year

Number of Entrants	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Base	1.7	3.4	3.6	4.4	4.2	6.8	7.2	6.7	9.8	13.7	11.2	32.5	30.9
	(0.7)	(1.4)	(3.0)	(6.8)	(1.8)	(1.8)	(2.2)	(1.8)	(2.8)	(3.7)	(3.2)	(5.4)	(4.9)
No RFS1	1.7	3.4	3.6	4.4	4.2	6.8	7.3	6.7	9.8	10.9	9.4	32.6	31.0
	(0.7)	(1.4)	(3.1)	(6.8)	(1.8)	(1.8)	(2.1)	(1.7)	(2.8)	(3.1)	(3.9)	(5.4)	(5.0)
No RFS2	1.7	3.4	3.6	4.4	4.2	6.8	7.3	6.7	9.8	13.8	11.2	9.7	9.4
	(0.7)	(1.4)	(3.1)	(6.8)	(1.8)	(1.8)	(2.1)	(1.7)	(2.8)	(3.7)	(3.2)	(4.2)	(4.0)
No Tax Credit	1.6	3.1	3.4	4.1	3.8	5.8	6.7	5.8	8.7	12.6	10.0	29.5	28.1
	(0.7)	(1.4)	(2.9)	(6.6)	(1.6)	(1.5)	(2.2)	(1.6)	(2.6)	(3.4)	(2.9)	(5.2)	(4.9)
No MTBE	1.7	3.3	3.6	4.4	2.3	3.1	3.5	3.5	3.5				
	(0.7)	(1.4)	(3.1)	(6.8)	(1.9)	(1.3)	(1.3)	(1.1)	(1.5)				
No Policy	1.6	3.1	3.4	4.1	2.1	2.7	3.2	3.0	3.0	3.2	2.4	2.5	2.4
	(0.7)	(1.4)	(2.9)	(6.6)	(1.8)	(1.1)	(1.2)	(0.9)	(1.3)	(1.4)	(0.9)	(1.0)	(1.0)
Mean welfare per entrant	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Base	-0.29	0.37	0.57	0.77	1.10	1.52	1.59	1.66	1.86	2.04	2.04	2.51	2.56
	(0.33)	(0.24)	(0.26)	(0.30)	(0.22)	(0.19)	(0.22)	(0.17)	(0.16)	(0.19)	(0.23)	(0.21)	(0.20)
No RFS1	-0.28	0.36	0.56	0.78	1.11	1.52	1.60	1.66	1.86	1.95	1.96	2.51	2.56
	(0.33)	(0.24)	(0.27)	(0.30)	(0.22)	(0.19)	(0.21)	(0.17)	(0.16)	(0.16)	(0.24)	(0.21)	(0.21)
No RFS2	-0.28	0.36	0.56	0.78	1.11	1.52	1.60	1.66	1.86	2.04	2.04	2.05	2.12
	(0.33)	(0.24)	(0.27)	(0.30)	(0.22)	(0.19)	(0.21)	(0.17)	(0.16)	(0.19)	(0.23)	(0.23)	(0.22)
No Tax Credit	-0.35	0.30	0.50	0.72	1.02	1.44	1.55	1.60	1.81	2.01	2.00	2.48	2.53
	(0.31)	(0.23)	(0.25)	(0.29)	(0.19)	(0.19)	(0.22)	(0.18)	(0.16)	(0.19)	(0.22)	(0.21)	(0.21)
No MTBE Ban	-0.28	0.36	0.57	0.78	0.60	0.94	1.12	1.27	1.35				
	(0.32)	(0.25)	(0.26)	(0.30)	(0.26)	(0.25)	(0.19)	(0.18)	(0.21)				
No Policy	-0.35	0.30	0.50	0.72	0.55	0.87	1.08	1.22	1.29	1.43	1.44	1.56	1.66
	(0.31)	(0.23)	(0.25)	(0.29)	(0.24)	(0.23)	(0.18)	(0.18)	(0.19)	(0.19)	(0.23)	(0.22)	(0.21)

For each scenario, the reported statistics are averages over 50 simulations. Standard errors are in parentheses, and are calculated from using the parameter estimates from each of the 250 bootstrap samples. For each of the 250 bootstrap samples, 50 simulations are run using the parameter estimates from that bootstrap sample. Standard errors for a statistic is the standard deviation of the respective statistics over all 250 bootstrap samples.

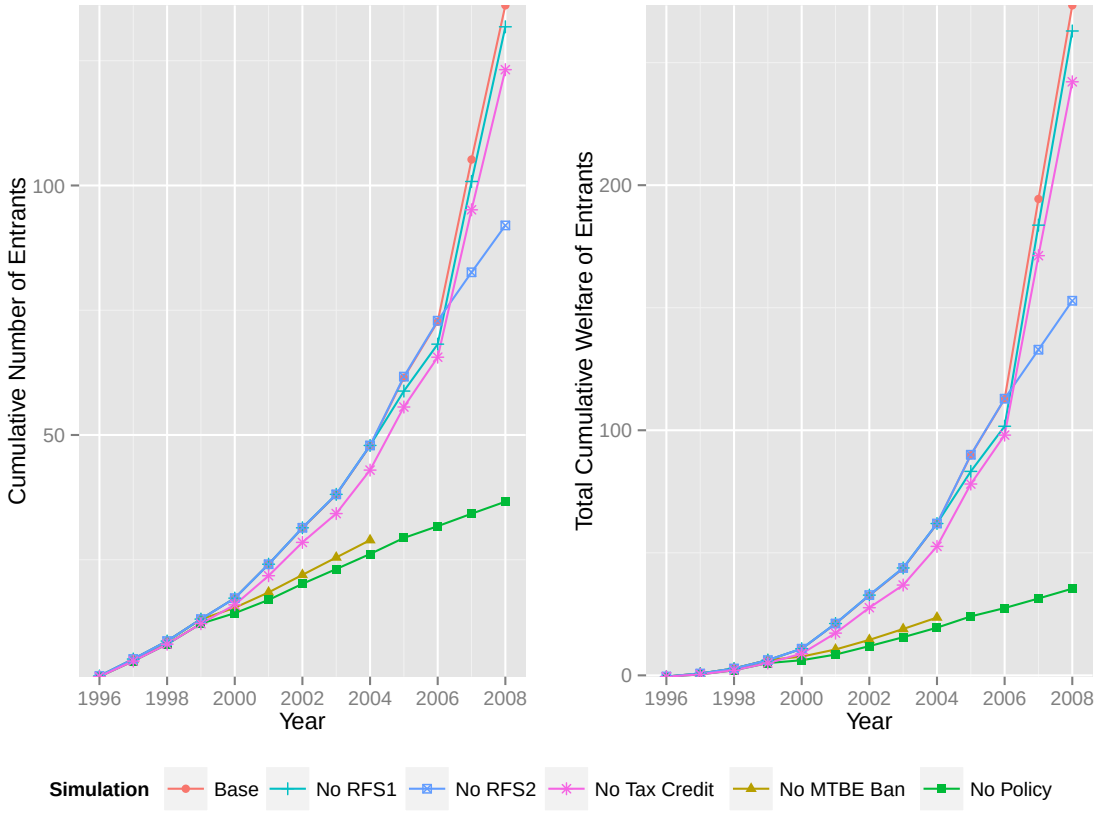

Figure 2: Cumulative number of entrants and total cumulative welfare of entrants under different policy scenarios over time

Figure 3: Mean welfare per entrant by year under different policy scenarios

Table 17: Number of entrants and mean welfare per entrant by state in full and pre-RFS periods

Number of Entrants		IL	IN	IA	KS	MN	MO	NE	OH	SD	WI
Full Period	Base	16.2	16.8	23.6	8.7	14.3	8.1	19.9	8.4	12.2	7.6
		(2.7)	(3.3)	(4.2)	(1.8)	(2.8)	(1.6)	(3.4)	(1.6)	(2.9)	(1.7)
	No RFS1	15.6	16.2	22.9	8.4	14.0	7.8	19.5	8.1	11.9	7.2
		(2.9)	(3.2)	(4.5)	(1.9)	(2.9)	(1.5)	(3.7)	(1.6)	(3.0)	(1.7)
	No RFS2	10.2	10.3	16.7	5.6	10.5	5.0	15.0	5.3	8.0	5.2
		(2.4)	(2.5)	(3.9)	(1.7)	(2.8)	(1.4)	(3.2)	(1.3)	(2.9)	(1.5)
	No Tax Credit	16.2	14.3	23.6	6.9	13.5	6.4	17.7	8.4	9.5	6.7
		(2.7)	(2.6)	(4.2)	(1.8)	(2.6)	(1.5)	(3.4)	(1.6)	(2.7)	(1.6)
	No Policy	5.1	4.9	6.1	2.2	3.8	2.2	4.7	3.1	2.5	2.1
		(1.8)	(2.3)	(2.6)	(1.7)	(2.6)	(1.4)	(2.0)	(1.1)	(2.7)	(1.2)
Pre-RFS (1996-2004)	Base	4.2	4.3	9.7	2.4	6.6	1.8	9.9	2.2	4.1	2.3
		(1.3)	(2.1)	(2.6)	(1.3)	(2.5)	(1.0)	(2.5)	(0.8)	(2.4)	(1.1)
	No Tax Credit	4.2	4.3	9.7	2.0	5.7	1.6	7.7	2.2	3.7	1.9
		(1.4)	(2.0)	(2.6)	(1.4)	(2.3)	(1.2)	(2.0)	(0.8)	(2.5)	(1.0)
	No MTBE Ban	3.3	3.5	4.3	1.9	3.4	1.8	4.7	2.2	2.0	1.9
		(1.3)	(2.1)	(2.2)	(1.4)	(2.7)	(1.0)	(2.1)	(0.9)	(2.6)	(1.1)
	No Policy	3.4	3.5	4.3	1.6	2.8	1.6	3.5	2.2	1.8	1.5
		(1.3)	(2.1)	(2.1)	(1.4)	(2.4)	(1.2)	(1.6)	(0.8)	(2.6)	(1.1)
Mean Welfare per Entrant		IL	IN	IA	KS	MN	MO	NE	OH	SD	WI
Full Period	Base	2.05	2.15	2.11	1.80	1.99	1.89	2.05	1.89	2.19	2.02
		(0.19)	(0.18)	(0.14)	(0.17)	(0.13)	(0.21)	(0.15)	(0.21)	(0.17)	(0.17)
	No RFS1	2.03	2.14	2.10	1.79	1.97	1.88	2.05	1.87	2.18	1.99
		(0.19)	(0.17)	(0.14)	(0.16)	(0.13)	(0.21)	(0.14)	(0.20)	(0.17)	(0.16)
	No RFS2	1.64	1.74	1.81	1.41	1.70	1.45	1.81	1.46	1.84	1.69
		(0.18)	(0.16)	(0.13)	(0.17)	(0.13)	(0.20)	(0.15)	(0.18)	(0.17)	(0.19)
	No Tax Credit	2.05	2.03	2.11	1.67	2.00	1.75	2.04	1.89	2.03	1.99
		(0.20)	(0.15)	(0.14)	(0.17)	(0.13)	(0.20)	(0.16)	(0.21)	(0.15)	(0.18)
	No Policy	1.01	1.09	1.12	0.60	1.04	0.70	1.08	0.94	0.98	0.93
		(0.18)	(0.17)	(0.18)	(0.17)	(0.20)	(0.19)	(0.20)	(0.19)	(0.23)	(0.21)
Pre-RFS (1996-2004)	Base	1.02	1.14	1.53	0.81	1.46	0.62	1.64	0.74	1.50	1.12
		(0.15)	(0.16)	(0.15)	(0.19)	(0.16)	(0.24)	(0.18)	(0.20)	(0.19)	(0.21)
	No Tax Credit	1.01	1.14	1.53	0.66	1.40	0.46	1.49	0.74	1.40	0.97
		(0.15)	(0.16)	(0.15)	(0.18)	(0.15)	(0.19)	(0.18)	(0.19)	(0.18)	(0.19)
	No MTBE Ban	0.72	0.88	0.91	0.48	0.92	0.62	1.04	0.74	0.84	0.85
		(0.17)	(0.20)	(0.20)	(0.22)	(0.26)	(0.24)	(0.26)	(0.21)	(0.28)	(0.26)
	No Policy	0.71	0.88	0.90	0.35	0.82	0.46	0.89	0.74	0.73	0.70
		(0.17)	(0.20)	(0.20)	(0.20)	(0.23)	(0.19)	(0.22)	(0.19)	(0.25)	(0.22)

For each scenario, the reported statistics are averages over 50 simulations. Standard errors are in parentheses, and are calculated from using the parameter estimates from each of the 250 bootstrap samples. For each of the 250 bootstrap samples, 50 simulations are run using the parameter estimates from that bootstrap sample. Standard errors for a statistic is the standard deviation of the respective statistics over all 250 bootstrap samples.

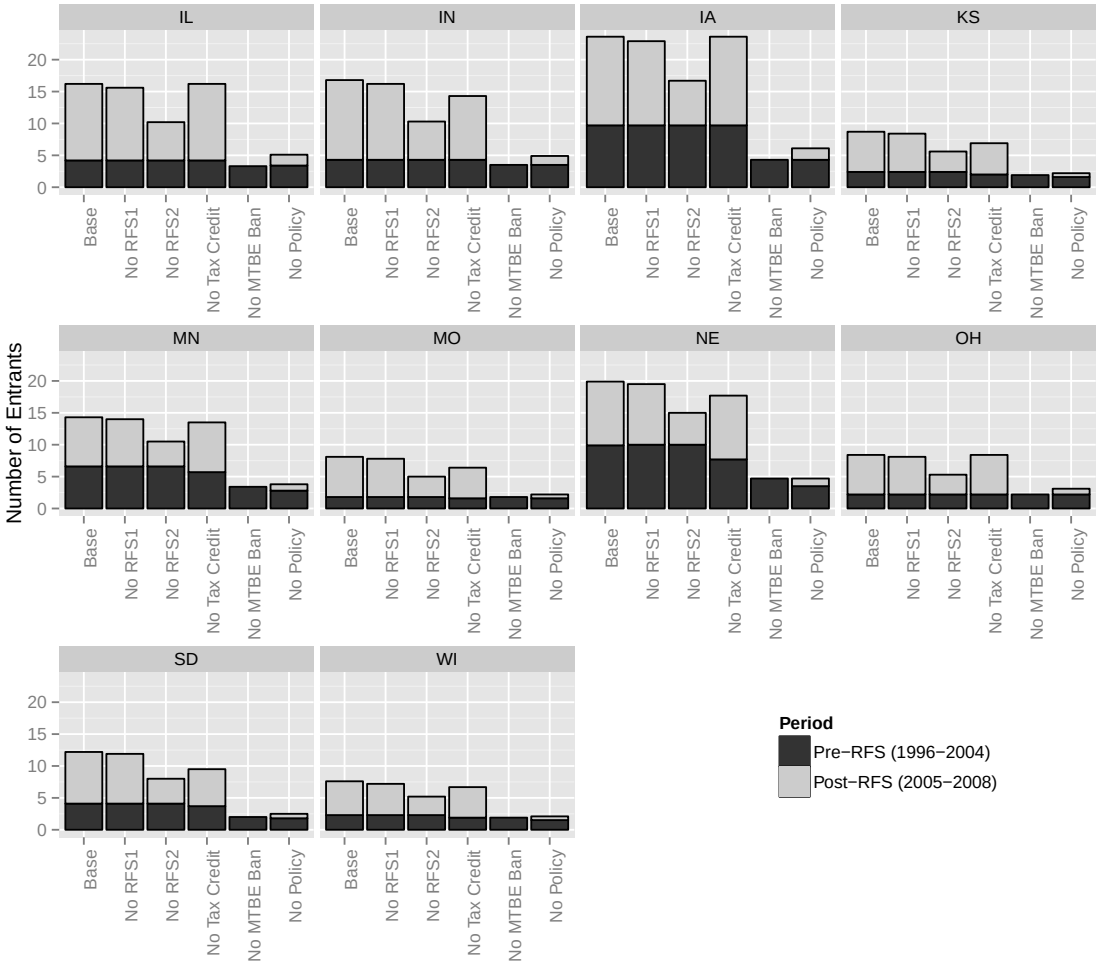

Figure 4: Number of entrants by state under different policy scenarios

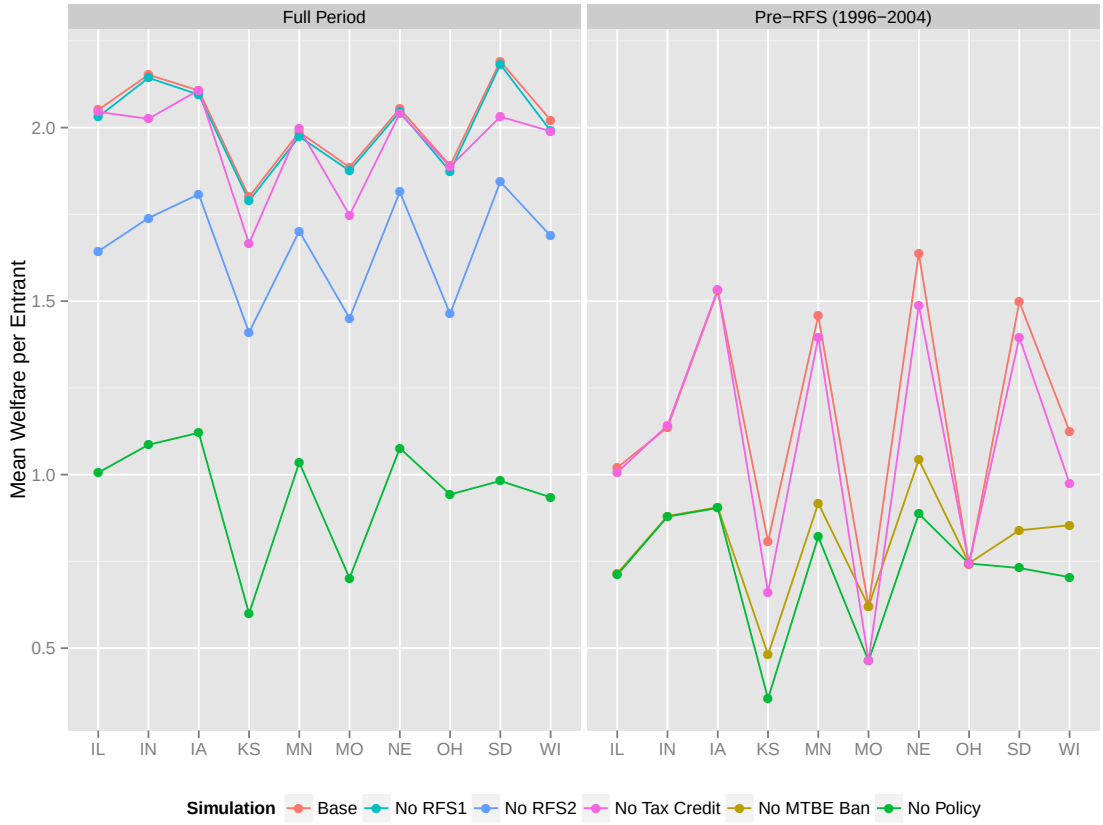

Figure 5: Mean welfare per entrant by state under different policy scenarios