Do Taxes Impact Firm Location on the Margin? A Matched County-Pair Study

Kevin D. Duncan

Abstract

This paper uses a regression discontinuity approach to test the impacts of taxes on firm entry rates between neighboring states. We utilize matched county pairs as an approximate bandwidth around the sharp discontinuity in state policies imposed at their border. This allows us to control for unobserved or unaccounted for location specific drivers of firm entry. We test an array of seven top marginal tax rates, right to work status, and minimum wage rates as drivers of differences in firm entry between counties. Our results indicate that property and sales taxes have the largest distortionary effect on firm start up rates, and in more recent years income taxes as well.

1 Introduction

Each year states and counties try to incentivize new business start-ups in their locale over alternative choices. The most visible cases include governments offering firm's temporary reprieve from tax burdens, or deals to build infrastructure to support new entrants. There has been a growing literature addressing the efficacy of these exemptions both to state revenue and public welfare. States may also take a long run approach to incentive new firm start-ups by altering their tax and regulatory codes for all market participants.

This paper tests whether or not taxes impact firm entry rates between neighboring states. We also explore whether or not firms place preferences over government expenditures in their entry decision. This later situation would imply that increasing taxes to pay for certain services may not have distortionary effects on firm start up rates. The paper tests the impacts of seven top marginal tax rates, including property, income, corporate, capital gains, sales, workers compensation, as well as unemployment insurance, and state expenditures per capita on welfare, highways, and education.

We provide two methods to address these problems. We first estimate count data models of firm entry as a benchmark against existing literature in the field. We then outline our preferred estimator, which uses matched county pairs on either side of a state border. By matching counties, we get an approximate bandwidth around the sharp discontinuity in policies at state borders. The process may control for unobserved variables in count data models, as well as interaction terms between policy and location specific components of firm entry. This gives us the ability to identify just the policy effect.

This county-pair difference estimator is a generalization of the traditional difference in difference estimator extended over a larger panel of agents and time periods. It has been used to explore how different policy regimes affect local migration between geographic entities (McKinnish (2005) (2007)), as well as how differences in policies impact outcomes across borders (Holmes (1998), Rohlin (2011), Dube et al (2010), McPhail, Orazem, Singh (2010), Dhar and Ross (2012), Kahn and Mansur (2013)). Many of these papers focus on how firms or individuals respond to imposed opportunity costs in location from neighbor's policy action.

Broader economic theory provides mixed outcomes to if a higher number of firm start ups increases welfare. Cases where more firms are welfare decreasing appear when assuming fixed costs to entry, such that too many firms entering can exceed their marginal benefit. Comparably if consumers have preferences over differentiated products, increasing the number of firms ('varieties') can be welfare increasing. The theory we develop here works outside of these considerations. Instead our results can help governments achieve higher employment growth and government revenues, as well as extending existing determinants of firm sorting.

The paper proceeds in the following manner. First, we review literature relating to empirical firm sorting and entry and how regression discontinuity techniques have been used in related studies. We then provide a model to show why state borders may allow us to control for location specific terms of firm entry. Next, we explain our empirical design, which uses both count data models and a regression discontinuity technique using matched county pairs. Finally, we estimate reduced form models of the impact of various policies on firm start up rates. We then provide estimates for the impacts of taxes and government expenditures on firm entry into US counties. We conclude by showing which borders have the largest firm start up discrepancy, and talk about applications and limitations for our work.

2 Theory

Here we develop theory that provides justification for utilizing a regression discontinuity approach to estimate policy driven determinants of firm entry. We show that under some strong assumptions we get location specific determinants of firm entry to cancel out as we take differences close to regime borders. The results show that firms pick sides with the highest expected profit from state specific policies as we get approximately close to state discontinuities.

Assume there exists a spatial equilibrium where wages and capital costs are adjusted to local tax and location specific variables affecting firm level productivity. If markets are competitive firms will make zero economic profit in the long run, but in the short run demand or policy shocks can leave short run profits. We expect that if a regime changes its taxes over time, higher production costs and lower profits exist in that county, and that market will deter a relative amount of firms from entering. Since firms will bid up or down prices relative to taxes, those prices can be proxied by the tax rates directly. Firms make decisions based on information from the previous year, as governments might concurrently change policy along with market entry. Similarly, there might be time costs to set up a physical store in a chosen location.

Assumption 2.1. *Assume that a firms' profit can be expressed as a linear function,*

$$\pi_{i,j,t} = \gamma + \beta_i + \beta_j + X_{i,j,t-1}\beta + \epsilon_{i,j,t} \tag{1}$$

$$\epsilon_{ijt} \sim N(0, \sigma) \tag{2}$$

Where i indexes location, j a particular policy regime, and t time. X_{ijt} is a $1 \times K$ row vector.

Broadly, we let location specific variables be any variable that is specific to a location, such as local agglomeration figures, education attainment, and other variables driven by the distribution of labor and productive factors in each regime. Variables at the regime level tend to include variables like taxes, regulatory policies, and government expenditures. Both sets of variables are allowed to evolve over time.

Now let us focus on two states be normalized on $[-1, 1]$, such that for $i \in [-1, 0)$ a firm is in state A, and for $i \in [0, 1]$, they are in state B. Therefore, if a firm has two choices, $y \in [-1, 0)$ and $\hat{y} \in [0, 1]$, then the firm chooses y over $\hat{y}$ if

$$E[\pi_{y,A,t} - \pi_{\hat{y},B,t} | X_{y,A,t-1}, X_{\hat{y},B,t-1}] = \beta_A - \beta_B + (X_{y,A,t-1} - X_{\hat{y},B,t-1})\beta > 0 \tag{3}$$

Assumption 2.2. *We can partition* $X_{i,j,t-1} = [X_{i,t-1} \quad X_{j,t-1}]$*, where* $X_{i,t-1}$ *is* $1 \times K_1$ *and* $X_{j,t-1}$ *is* $1 \times K_2$*. Equivalently,* $\beta = [\beta_1 \quad \beta_2]'$*, with* β_1 *being* $1 \times K_1$ *and* β_2 *being* $1 \times K_2$*.*

We let $X_{i,t-1}$ have all terms containing location specific coefficients, but this can include interaction terms with our regime specific variables $X_{j,t-1}$. Therefore this assumption simply states that our policy variables have to enter directly into the profit function. As a result, we can write the profit function.

$$\pi_{i,j,t} = \gamma + \beta_i + \beta_j + X_{i,t-1}\beta_1 + X_{j,t-1}\beta_2 + \epsilon_{i,j,t} \tag{4}$$

Now we make the final set of assumptions:

Assumption 2.3. β_i *and* $X_{i,t-1}$ *are continuous locally around* 0*, such that for* $|y - \hat{y}| < \psi, \quad \exists \delta$ *such that* $|\beta_y - \beta_{\hat{y}}| < \delta$ *and* $||X_{y,t-1,k} - X_{\hat{y},t-1,k})||_2 < \delta$

Theorem 2.4. $\lim_{y \to 0^-} \Pi_{y,A,t} - \lim_{\hat{y} \to 0^+} \pi_{\hat{y},B,t-1} \to \beta_A - \beta_B + (X_{A,t-1} - X_{B,t-1})\beta_2$

Proof. We know that we can rewrite each profit function as

$$\pi_{i,j,t} = \beta + \beta_i + \beta_j + X_{i,t-1}\beta_1 + X_{j,t-1}\beta_2 + \epsilon_{i,j,t}$$

Then, taking the difference we get

$$\pi_{y,A,t} - \pi_{\hat{y},B,t} = \beta_y - \beta_{\hat{y}} + \beta_A - \beta_B + (X_{y,t-1} - X_{\hat{y},t-1})\beta_1 \tag{5}$$

$$+(X_{A,t-1} - X_{B,t-1})\beta_2 + \epsilon_{y,A,t} - \epsilon_{\hat{y},B,t} \tag{6}$$

Next we take the limit, and apply expectations. Using Assumption (2.3), we get

$$E[\lim_{y\to 0^-} \pi_{y,A,t} - \lim_{\hat{y}\to 0^+} \pi_{\hat{y},B,t}] = \delta(1+\beta_1) + \beta_A - \beta_B + (X_{A,t-1} - X_{B,t-1})\beta_2$$

Then, as $|y - \hat{y}| \to 0$, $\delta \to 0$. Therefore,

$$E[\pi_{y,A,t} - \pi_{\hat{y},B,t} | X_{y,A,t-1}, X_{\hat{y},B,t-1}] = \beta_A - \beta_B + (X_{A,t-1} - X_{B,t-1})\beta_2 \tag{7}$$

□

The theorem states that as we get arbitrarily close to the border we see that which state a firm moves into is a function of state specific variables. As we move away from the border location characteristics might mitigate the policy effect, equivalently to Brulhart et al, especially where we expect policy effects to be small. This theory favors the use of regression discontinuity techniques for estimating policy treatment effects, especially when location specific drivers of firm entry might be unknown or unobserved.

3 Empirical Design

Our parameters of interest are the coefficients on our top marginal tax rates. We provide two approaches to identify these parameters. We first estimate traditional count data models using Poisson and Negative Binomial Maximum Likelihood Estimation. We then explain an alternative approach to identification using a regression discontinuity approach at state border.

The count data model works by showing that under a spatial equilibrium firm entry behaves like a Conditional Logit Model of entry, where there is a probability of a firm entering into particular locations. Other research has shown that by assuming structure on the profit function, we can show that estimating the conditional Logit model provides the same coefficients as estimating a Poisson regression. Our regression discontinuity approach follows directly from our theoretical section. This method takes the difference in variables from counties on either side of a state border. The border provides a sharp discontinuity in policies firms face, thus we treat counties as a closest bandwidth around the discontinuity.

Consistent with most panel data work, major obstacles include several levels of unobserved heterogeneity, and our reduced form estimates might have issues in proper accounting for the incentives to locate in a particular location. Our preferred estimator might solve some of these problems.

3.1 Count Data Models

Traditional firm location choice literature is motivated by firms entering across all possible locations in the market using the profit function described in (4). Our set up follows from McFadden (1974) and Wooldridge (2010, pp 619). This is done in the following fashion. First, we assume a profit function equivalent to equation 4, where we assume ϵ_{ijt} takes on an extreme-type-value-I distribution Weibuill distributed. Let us have $f = 1, ..., F$ number of firms trying to enter in a given time period. Each state in the US is denoted as a regime, j, and each county is a location i. Let us index them $ij = 1, ..., N$. Then the probability of a firm f locating at point ij in period t is;

$$p_{f,ij,t} = \frac{\exp(X_{i,j,t-1}\beta)}{\sum_{ij} \exp(i, j, t-1)} \tag{8}$$

Now let $d_{f,ij,t} = 1$ if a firm f enters at point ij in period t, and $d_{f,ij,t} = 0$ otherwise. Further, let us assume that there is no time dependence element, such that we can run this as TN independent events. Then the log likelihood becomes

$$\log L_{cl} = \sum_{ij=1}^{TN} \sum_{ij} d_{f,ij,t} \log p_{f,ij,t} \tag{9}$$

Here we are assuming a strong assumption that the vector of parameters $X_{i,j,t-1}$ is the same for all types of firms. Guimaraes, Figueiredo, and Woodward (2003) show that in the case where the profit function depends on the same characteristics across all firms, that (9) becomes a Poisson distribution consistent up to a constant as long as we believe that firm entry is directly a Poisson distribution as well. This happens as equation (9) becomes,

$$\log L_d = \sum_{ij} n_j \log p_{ij} \tag{10}$$

With n_{ij} being the number of firms that open up in location ij, let us assume;

$$E[n_{ij}] = \exp(X_{i,j,t-1}\beta) \tag{11}$$

They show that the log likelihood of the Poisson distribution becomes proportional up to a constant of the conditional Logit. Thus estimating a count data Poisson model is equivalent to estimating a full conditional log likelihood. This nice feature allows a fast and easy approach to identify the impacts of our tax variables. Taking the exponent of (11) gives us the Kernel to a Poisson distribution, with $E[n_{ijt}] = X_{i,j,t-1}\beta$. As a result, the Poisson likelihood takes on the form,

$$L(\theta|X, N) = \Pi_{i=1}^{m} \frac{e^{n_i \beta' n_i} e^{e^{-\beta' x_i}}}{n_i!} \tag{12}$$

We proceed by estimating a Poisson distribution in Table **??**.[1] We find that the the model is over dispersed by using Cameron & Trivedi (1990) regression based test for over

[1]Currently our estimates utilize just our matched county pair data so limits the firms' decision to just counties on the border of a state. However we plan to extend this estimation procedure over all counties. Due to time considerations leading to the existing draft of the paper this has not yet been done, but will be accomplished shortly.

dispersion. As a result, we run Negative Binomial regressions using the same expected value. This gives us space to relax the assumption that $E[n_{ij}] = Var[n_{ij}]$, and allow our errors to take on a more general shape.

3.2 Regression Discontinuity Approach

There are several issues with this approach. First, firm entry may be heavily dependent on terms such as population. Similarly, individuals may place preference in areas that have been experiencing large job growth. Finding instruments for these interactions can be difficult. Further, there may be other unobserved heterogeneity at the location level that is unobserved by the researcher. Regression discontinuity techniques are a way to possibly control for these variables.

By our theory we know that location specific terms, an terms shared across observations get canceled out as we take the difference while approaching the borer. Our data does not allow us to get a closer estimation to the true discontinuity than those provided by the borders of the county. The average county in our data set is 1260 square miles, or about 35 miles per side of believed to be approximately square. This distance is slightly longer than more refined approaches such as Rohlin (2011). In practice we match up counties on either side of a state border, let us denote them subject (sub) and neighbor (nbr), and their states stA and stB. Then, taking differences, we get by applying Theorem 2.4

$$\ln(n_{sub,stA,t}) - \ln(n_{nbr,stB,t}) = \beta_{stA} - \beta_{stB} + (X_{stA,t-1} - X_{stB,t-1})\beta_2 + \epsilon_{sub,stA,t} - \epsilon_{nbr,stB,t} \quad (13)$$

First, let us index each state-pairs (stA, stB) by g. Next let us assume that $\beta_{stA} - \beta_{stB} = \beta_0$ for all sub, nbr pairs. Since we assign sub and nbr arbitrarily, this implies that

$\beta_{stA} = \beta_{stB}$. Then we make the following definitions.

$$\ddot{\ln}(n_{i,g,t}) = \ln(n_{sub,stA,t}) - \ln(n_{nbr,stB,t}) \tag{14}$$

$$\ddot{X}_{g,t-1} = \beta_0 + (X_{stA,t-1} - X_{stB,t-1}) \tag{15}$$

$$\ddot{\epsilon}_{i,g,t} = \epsilon_{sub,stA,t} - \epsilon_{nbr,stB,t} \tag{16}$$

Assume $\ddot{\epsilon}_{i,g,t}$ be an i.i.d white noise draw, then let $\ddot{X}_g = (\ddot{X}'_{g,0}, ..., \ddot{X}'_{g,T-1})'$ be a $T \times (1+K_j)$ matrix, and $\ddot{\epsilon}_{ig} = (\ddot{\epsilon}_{i,g,1}, ..., \ddot{\epsilon}_{i,g,T})'$ be a $T \times 1$ vector. Next we assume the traditional OLS moment conditions.

Assumption 3.1. *Let $\ddot{X}_g = (\ddot{X}'_{g,0}, ..., \ddot{X}'_{g,T-1})'$ be a $T \times (1+K_j)$, and $\ddot{\epsilon}_{i,g} = (\ddot{\epsilon}_{i,j,1}, ..., \ddot{\epsilon}_{i,j,T})'$ a $T \times 1$ vector. Then*

$$E[\ddot{X}'\ddot{\epsilon}] = 0, \quad \forall i, g \tag{17}$$

Assumption 3.2.

$$E[\ddot{X}'_g \ddot{X}_g] = 1 + K_j : \quad \forall g \tag{18}$$

We can estimate a pooled OLS estimator using Assumption's 3.1 and 3.2. This gives us the POLS estimator;

$$\hat{\beta}_2 = \left(\frac{1}{N^*} \sum_{k=1}^{T} \sum_{i=1}^{G} \sum_{j=1}^{N_G} \ddot{X}'_{g,t-1} \ddot{X}_{g,t-1} \right)^{-1} \left(\frac{1}{N^*} \sum_{k=1}^{T} \sum_{i=1}^{G} \sum_{j=1}^{N_G} \ddot{X}'_{g,t-1} \ddot{\ln}(n_{igt}) \right) \tag{19}$$

$$N^* = T(\sum_{g}^{G} N_g) \tag{20}$$

Donald and Lang (2007) show that increasing individual observations for each group doesn't provide better inference. They use a two stage estimator where they first calculate the mean for each side to show asymotitics with respect to the number of groups.Our estimator is a mean weighted version of their two stage estimator. We can rewrite 19 as:

$$\hat{\beta}_2 = \left(\frac{1}{TG} \sum_{t=1}^{T} \sum_{g=1}^{G} \frac{\sum_{i=1}^{N_g} \ddot{X}'_{g,t-1} \ddot{X}_{g,t-1}}{E[N_g]} \right)^{-1} \left(\frac{1}{TG} \sum_{t=1}^{T} \sum_{g=1}^{G} X'_{g,t-1} \frac{\sum_{i=1}^{N_G} \ddot{\ln}(n_{igt})}{E[N_g]} \right) \tag{21}$$

$$E[N_g] = \frac{\sum_{g=1}^{G} N_g}{G} \tag{22}$$

Compared to Donald and Lang's two stage estimator we underweight observations we observe only a few times compared to their true mean, and overweight observations we see many times compared to their true mean. Increasing N_g for some g doesn't improve our estimator, and only increase $E[N_g]$. Trying to keep $E[N_g]$ static requires making our asymptotics with respect to the number of group-pairings we have, G.

When doing inference there may be shocks to the state-pair border, for example the Mississippi river flooding along a border pair, but not shared with all other pairs in the sample. Therefore we use clustered errors on the state pair. Let $\ddot{X}$ be the $(\sum_g^G N_G \times T) \times (1 + K_j)$ regressor matrix. Thus our variance co-variance matrix takes on the form

$$\hat{V} = \frac{1}{G-2} \frac{\sum_{g=1}^{G} N_g - 1}{\sum_{g=1}^{G} N_g - 2} (\ddot{X}'\ddot{X})^{-1} (\sum_{t}^{T} \sum_{g}^{G} u_s u_s') (\ddot{X}'\ddot{X})^{-1} \tag{23}$$

$$u_s = \sum_i \ddot{\hat{\epsilon}}_{i,j,t-1} \ddot{X}_{g,t-1} \tag{24}$$

We assume this lag structure to indicate that firms respond to variables from the previous time period, and as they are starting up government's may choose to alter policies for the current year. In practice though most of our variables are heavily inter-temporally correlated, so no major difference occurs in sign, significance, or fit appears from using different lag structures.

3.3 Sensitivity Tests

We subject our estimator to a series of robustness checks. For all of our regressions, we test models with and without amenities. We want to check whether or not our tax and regulatory variables become statistically insignificant once we account for these additions, and in our second model check whether or not they properly become indistinguishable from zero. Next, we test a version of this model where we do not impose the coefficients are the same across borders.

$$\ddot{\ln}(n_{g,t}) = X_{stA,t-1}\beta_{sub} + X_{stB,t-1}\beta_{nbr} + e_{igt} \tag{25}$$

Instead we let coefficients take on their own value in the difference, and do a set of F-tests on whether or not our assumption that $\beta_{i,A} = -\beta_{i,B}$ holds in the difference as assumed. The results of this regression are reported in Table **??**. Corresponding F tests are presented in Table **??**. Next we run our regression discontinuity estimator while forcing the coefficients to be the same. We present results for this model in **??**.

In Table **??** we test a set of regressions where we estimate period-specific coefficients and compare them to our pooled estimator to try and estimate of whether or not it is safe to assume that profit parameters are roughly stable over time.

$$\ddot{\ln}(n_{g,t}) = X_{stA,t-1}\beta_{stA} + X_{stB,t-1}\beta_{stB} + e_{i,g,t} : \quad t = 1999, ..., 2008 \tag{26}$$

Which leads to the POLS coefficient;

$$\hat{\beta}_2 = \left(\frac{1}{G} \sum_{i=1}^{G} \frac{\ddot{X}'_{g,t-1}\ddot{X}_{g,t-1}}{E[N_g]} \right)^{-1} \left(\frac{1}{G} \sum_{i=1}^{G} \ddot{X}'_{g,t-1} \frac{\sum_{j=1}^{N_G} \ddot{\ln}(n_{igt})}{E[N_g]}) \right) \tag{27}$$

Next we test a version of our model that includes a dummy variable for each state-pair in our sample. By construction of our estimator, we are claiming that any county level fixed effects take the form of location specific terms, which have to cancel out when we take the difference but state specific fixed effects may remain. We favor using dummy variables over Fixed Effect or First Difference transformations because our policy variables are incredibly stable over time. States very rarely change tax policies, and correlations with current tax rates with each of five periods of lags shows that taxes even at their weakest are still more than 85% correlated with each other. Equivalently, right to work status changes once in our sample, and minimum wages rarely alter at the state level as well. Therefore these transformations do not provide enough variation to get valid inference.

Finally, we do not test for general endogeneity where states change taxes in response to the difference in firm entry rates. This is because the aforementioned stability of all of our policy parameters, it seems unlikely that they are responding to comparatively more volatile firm start up rates. Further, there is no reason to assume counties favor one set of borders over any other, unless counties find themselves systemically at a loss compared to neighbors, a corner solution we do not check for.

4 Variables and Data

Our primary variable of interest is county level firm start up rates for all firms in a year. This data set is generously provided by the Census Bureau's Business Dynamic Statistics program.[2] The data also includes how many firms expanded or contracted employment, a variety of broad NAICS coded industries, and the number of firms that shut down in a given year, however we omit these variables in practice. Since many counties in our sample do not have all the sub coded, for completeness our results rely on the summation across different industries.

For our matched county pair estimation, our data set includes 105 state-pairs, with

[2] http://www.census.gov/ces/dataproducts/bds/overview.html

1213 matched counties over 10 years. Additional borders could be added, namely New Mexico - Oklahoma, and Arizona - Oklahoma. However, due to the number of regressors we exclude the pairs for degrees of freedom.

As part of our empirical design we drop counties not on the border of a state to ensure that as we linearly approach the border from either side we have a single limit. We match counties that share borders by arbitrarily assigning subject (*sub*) and neighbor (*nbr*) classification. As a result, our main variable of interest, *births_ratio* takes on the form,

$$births\_ratio_t = \ln(n_{sub,t}) - \ln(n_{nbr,t}) \tag{28}$$

We also include estimations for the raw number of implied firm difference,

$$births\_diff = n_{sub,t} - n_{nbr,t} \tag{29}$$

We also include a variable *allstarts*, that is a single column vector of all observations for the *sub* and *nbr* counties stacked on top of each other. This later variable is what we use to estimate our count data models. In the future we will extend our count data model data set to include all counties in the US between 1998 and 2009.

Equivalently to Orazem, McPhail, and Singh (2010) our tax variables are provided by the following sources. The National Bureau of Economic Research estimates of state marginal income tax and long-term capital gains tax rates. When applicable, we pull from the highest marginal tax rates available, as this is the rate most applied to small business and S corporations, or calculate the highest implied tax rate.[3]

Corporate and sales tax rates were compiled from The Council of State Governments

[3] `http://users.nber.org/~taxsim/allyup/` `http://users.nber.org/~taxsim/marginal-tax-rates/` `http://users.nber.org/~taxsim/state-marginal/`

Book of States, where marginal rates are the highest state tax rates on business corporations. Where rates differ between banks and non-banks, we use the non-bank rate, and we restrict to sales tax rates levied on general merchandise, and rather than food, clothing, and medicine. Property taxes are calculated from household level data provided by the Minnesota Population Center's Integrated Public Use Micro-data Series (IPUMS). The top marginal unemployment insurance tax rates were provided to Orazem, McPhail, and Singh by the US Department of Labor. To calculate, they multiply the top marginal tax rate, $\tau_{u,it}^{max}$, by the maximum wage level to which the rate is applied, W_{it}^{max}. They then normalize by the average wage in a state in a state in a current year, $\bar{W}_{it}^{max}$. Then the unemployment insurance tax is calculated as;

$$\tau_{u,it} = \frac{\tau_{u,it}^{max} W_{it}^{max}}{\bar{W}_{it}^{max}} \tag{30}$$

Workers compensation is provided between Thomason et al (2001) for between 1977 and 1995, with data afterwards provided by the Oregon Department of Consumer and Business Services. In all cases we use the lagged difference in their top marginal values, such that for a tax rate τ_i we get that for each pair of states sub, nbr and time t the tax ratio is calculated

$$tax\_ratio_{i,t} = \tau_{i,sub,t} - \tau_{i,nbr,t}$$

All of our tax variables are scaled to be between 0 and 100, where a 100% tax rate would be 100. This creates an intuitive interpretation of our estimated coefficients later.

We include two major non-tax variables that we expect to have a major impact on firm entry, right to work status and minimum wage rates. The former is the difference in a states right to work status coded as a binary variable with 1 if the state has right to work laws, and 0 otherwise. Right to work is a law that enables workers to exempt themselves from joining unions. For minimum wages we simply take the difference between the minimum wage for each state in a given year,

$$min\_wage\_ratio_t = min\_wage_{sub,t} - min\_wage_{nbr,t}$$

The Tax Policy Center also provides historical state minimum wage, education spending per capita, highway expenditure per capita, and welfare spending per capita data from 1983-2014.[4] Right to work status was compiled from the National Conference of State Legislators.[5] s

Lastly, amenity data was acquired from the USDA.[6] Since we care less about the coefficients of these variables, we use the normalized values of hours of sunlight in January, temperature in July, humidity in July, topology coefficient, and percent of county that is water. In all cases, the amenity variables are coded to be normal with mean zero and standard deviation 1. As a result, interpretation of these terms should be done in terms of deviations from the mean.

As a final series of controls, we also include percent of workforce unionized, log real fuel prices, population density, percent of industry manufacturing, and percent of population with high school education. This provides a robust set of location modifiers to explain a lot of non-economic preference for geographic amenities. Table **??** provides a series of summary statistics for each of our differenced terms.

It should be noted that for our data set we see some state-pairs more often than others. As a result, the range of observations range between 20 for Delaware and New Jersey, Delaware and Pennsylvania, and Arizona and Nevada, and 350 between Oklahoma and Texas, out of 12130 total observations available before any additional transformations are

[4] `http://www.taxpolicycenter.org/taxfacts/displayafact.cfm?Docid=603`

[5] `http://www.ncsl.org/research/labor-and-employment/right-to-work-laws-and-bills.aspx`

[6] `http://www.ers.usda.gov/data-products/natural-amenities-scale.aspx`

taken. We graph all the borders in Figure 1, where red represents the subject county, and blue is its matched neighbor.

Firm start ups look like exponential decay with an incredibly fat tail. From zero there is a large spike up, followed by decreasing frequency that exhibits extreme values in its tail. We can see this initial spiking behavior much more clearly as we truncate the data set to only show data points where the number of new firm start ups was less than 1000.

Taking the difference between our subject and neighbor counties we can see that the difference exhibits the exponential decay split across sides, and the log form looks like a slightly skewed normal distribution. This is expected from how subject and neighbor status of counties is arbitrarily assigned.

We plot the aggregate un-weighted tax differential by summing up the difference between tax rates for each matched county pair. The difference almost looks like two normal distributions with opposite skews. Further, similar to the firm start up data, we see that the right hump appears to have higher variance toward it's tail .

Finally we produce correlative graphs between all the variables. Along the top row we see the correlation between the difference in log number of firm start up rates with each of our tax variables. At first glance there does not appear to be any correlation between firm start up rates and any of the individual tax rates. As a result, most of the correlations appear to be "black boxes." However in my sample income tax and corporate tax rates are highly positively correlated (.64). Using the data set of all taxes from 1977-2008 the correlation is (.55), thus it appears as if over time states have generally streamlined their tax and corporate tax rates to be related to each other on purpose. Further, corporate tax rates are also correlated with income tax and capital gains tax rates.

Interestingly, we see the presence of clusters in the unemployment insurance tax rate with every other variables as immediate outliers. This is present in the right most column by by a few data points appearing in the right most part separate from the remaining states' unemployment insurance tax rates. This suggests there are some states that have an abnormally high unemployment insurance tax rate. Through analysis, it appears as is Delaware has a significantly higher unemployment insurance tax rate than any other state in our sample.

Broadly we infer the following; Taking the log number of firm start ups, especially by plotting their difference, makes the data look incredibly normal. Further, simple correlations between our dependent variable and any single of our independent variables does not immediately seem to indicate any expected results, though the independent variables themselves clearly have some minor dependence on each other. Thus, utilizing more powerful conditioned regression models will hopefully tell a better story than what is often done in many empirical tax works by only utilizing one or a small sample of tax variables.

4.1 Results

We report results from our count data models in Table **??**. The first and second columns report the results for our Poisson regression. Consistent with the literature, we confirm they are overdispered by running a Cameron & Trivedi (1990) test for over-dispersion. Despite this, we see that for most of our tax variables we find they are negative and significant. The exceptions for this are corporate and workers compensation taxes. Our coefficients for unemployment tax are incredibly large, which is given that due to the weighting process the calculated tax rates are relatively small, such that a 1% increase in the effective top marginal tax rate is almost a 10x increase in the current rates. Equivalent logic holds for why the property tax rate coefficient is particularly large.

Interesting, we find that higher percent of workforce with a high school education decreases firm start ups, higher minimum wage and right to work status increase firm start ups, and most government expenditures on highways and education decrease firm start up rates, while expenditures on welfare increases firm start up rates.

Columns 3 and 4 provide results for a Negative Binomial regression, where Column 3 includes amenity variables and Column 4 omits them. By letting the error structure vary more, the sign and significance switch for income and unemployment insurance tax rates. Equivalently, now higher minimum wages decrease firm start up rates, and both corporate tax rates and right to work status become statistically insignificant. We continue to see that higher education spending and higher percent of the populace with a high school education reduces firm start up rates.

In Table **??** we report coefficients for our matched county pair design where we do not impose that coefficients are the same on either side of the border. Columns 1 and 2 take the raw difference in firm start ups, with the first column having amenity controls and the second excluding them. Columns 3 and 4 use the difference in log firm start ups, with the same structure on including amenity controls. Visually we see for the most part the coefficients are in fact the equal and opposite from our regressions. Same with the count data models, higher corporate tax rates imply a higher number of new firm start ups that is statistically significant. The difference in state minimum wage rates and right to work status are both indistinguishable from zero. Of note though is that many of the papers that have found positive sign on these restrict their start up data to a finer selection of predominately small firm industries, which we do not do in our sample. In contrast to the count data model estimates, all the government expenditure figures are not statistically significant below the 5% threshold, and further as expected, all the amenity figures are not statistically significant.

Table ?? provides F tests for the assumption of the coefficients being the same across borders. We test for each variable that $\beta_{i,sub} = -\beta_{i,nbr}$. The results verify our belief that coefficients are the same and opposite in our design is a valid assumption. The exception is sales tax rates, for the subject county they are strongly and negatively significant, but for the neighbor they not significant at all. However, given that the rest of them pass, this might be a spurious result due to the number of regressors. We see an equivalent note in the workers compensation figures in our F tests, where for the neighboring county it appears to be significant, but not for the subject county.

Table ?? presents our main results. While in previous regressions by not imposing equality or taking differences our estimations have presented some positive and significant estimates for taxes, by imposing equality in the difference, we see that only tax variables that are negative and significant remain. This includes property taxes, income taxes, and sales taxes as the biggest driver of the tax differential. Weirdly, we see that a higher difference in the fuel price also imposes higher firm start up rates. In our differenced count regressions, only property taxes and right to work status seem to impact the number of firm start ups. And finally, as we saw in Table ??, all the amenity variables become statistically insignificant in the difference as is expected by construction of our design.

Table ?? shows regression results for *births_ratio* for the years 1999, 2002, 2006, and 2008 to be relatively equally spaced. Property taxes continue to be strongly negatively significant for all 4 time periods, with it getting slightly weaker near the end. Conversely, income taxes go from largely irrelevant to strongly significant in the early 2000's. Sales tax rates remain relatively constant in sign, magnitude, and significance across our sample as well. Alternatively, 1999 seems to have a variety of variables that appear to be significant, including right to work laws, highway spending per capita, welfare spending per capita, and workers compensation, all of which are strongly positively significant. In later years these terms become statistically insignificant of the regression, which is a feature shared in Table **??**'s results as well. Thus, while our tax variables seem to be the most stable among our terms in sign and significance, the assumption that terms do not change over time does not seem to hold exactly.

Finally, Table **??** reports a model with dummy variables for each state-pair interaction. When we make this transformation, we find that none of our variables are significant. However, this is somewhat to be expected. For certain state-pairs we only have 10 observations, and for others we have 350. Further, for each of our previous regressions, the R^2 is quite small, leading to a lot of additional variation that might be being soaked up by these terms. Despite this, we see that the coefficients with the highest t-test remain property, sales, and income, all of which are negative.

As a final output of our paper, we compare two different rankings. First we calculate the weighted tax differential by multiplying the tax coefficients from Table **??**, column 4 times each states marginal tax values. These are reported in Table **??**. We then calculate the mean number of firm start up differential over the entire state border by taking a pooled average across counties for each time period in our sample. In both cases we rank them from highest to smallest, but report only the top 10, where results are reported in Table **??**.

We see that for most states the weighted tax differential is very small, especially given that taxes are on a scale of 1 to 100. Thus the implied impact of taxes on relative firm start up rates is ultimately unsurprisingly small. However, for a few counties, this is not the case, and we see clear outliers where more than 1% of the differential is motivated by the difference in tax rates. Since the coefficient on unemployment insurance tax is so large, we also tested a version of the weighted tax differential with it removed, but nothing changes, and inspection on the data itself implies that the weighting process to calculate the term makes it on average very small.

Comparably, we can look at what the empirical difference is in the log number of firms. Thus we can get a comparably idea to measure the relative impacts of these tax policies on firm start up rates, as well as show which borders currently have the largest discrepancy. Below I present the top 10 states with the largest discrepancy in their relative weighted tax impact.

5 Conclusion

Our paper tests the impact of taxes on firm start up rates, and if firm entry seems to be dependent on government expenditures. We present a model illustrating when using regression discontinuity techniques around the border may provide identification for the impacts of government policies on firm start up rates. We then estimated both count data models and a model where we took the difference in county firm start up rates on opposite sides of a state border in a pseudo-regression discontinuity design.

In our empirical results, we included an array of state top marginal tax rates, right to work status, and minimum wage as costs, and counterbalance it with spending per capital on education, highways, and welfare. We also included a variety of controls, such as geographic amenities, population density, fuel prices, union rate, and percent of population with a high school degree.

Our Count Data Model estimates show that property, capital gains, and corporate taxes impose a burden on the number of firm start up rates. Surprisingly, both higher minimum wage, and a lack of right to work status, imply higher firm start ups. Also, education and highway spending per cap lower firm start up rates, but welfare spending does not. Thus it is not directly clear when cutting or raising taxes pays for itself in increased public expenditures. The inclusion of scaled amenity variables does not strongly impact the significance or sign of terms, but does tend to drive coefficients to be lower by some margin. These results may be biased by the lack of constricting our current choice to just counties on the border, rather than all counties.

In these models we see a high fit and significance for almost all variables. There may be issues both in endogeneity, as well as unobserved characteristics that entice firms to enter into one market over another. Our border discontinuity design may correct for some of these obstacles. In this model we take the difference between two counties on either side of a state border. In these estimates property taxes, income taxes, and sales taxes have the strongest determining factor on firm start up rates. This coincides with the observation that many companies are small S corporations, such that in the short run considerations such as capital gains or corporate tax rates shouldn't impact the decision choices of most firms.

In our specification tests, we find that it is reasonable to assume that coefficients are the same across counties for our pooled estimator. We also show that the sign, size, and significance of property and sales taxes remain consistent for each time period in our sample. Finally, we show that when we include an array of state-pair specific fixed effects all of our estimates become insignificant, but our tax variables remain the largest, keeping their sign and relative importance.

Comparably government expenditure variables do not seem to impact firm start up rates. This might be due to the fact that individuals can live in one county that has a preferred public expenditure bundle and still set up a businesses in a neighboring county that has a preferred regulatory policy. This allows for min-maxing of results for aspiring entrepreneurs. in comparison with other studies, our minimum wage and right to work variables do not seem to impact firm start up rates, but compared to studies focusing on those variables we do not restrict our analysis to restaurants, or other predominately low wage or manufacturing sectors.

Going forward, we would like to do more empirical tests for the impacts of taxes on different types of industries, as is common among related literature (Dube et al (2010), Rohlin (2011). However using our border discontinuity approach limits how many observations and makes identification harder. Further, we would like to provide greater robustness checks within our empirical framework. A simple extension would be to find out how our estimates vary as we test counties further away from the border, though the outline in our theoretical section might dissuade such regressions as able to properly identify a treatment effect over possible changes in location specific terms.