LECTURE: TAX SALIENCE AND BEHAVIORAL PUBLIC FINANCE

HILARY HOYNES
UC DAVIS
EC230

Papers:

Chetty, Looney, and Kroft "Salience and Taxation: Theory and Evidence"
Amy Finkelstein "E-ZTax: Tax Salience and Tax Rates"

Motivation and Context

[a definition of salience….]

Tax Salience: Tax policy *a* is more salient than tax policy *b* if calculating the gross-of-tax price under *policy-a* requires less computation than calculating gross-of-tax price under *policy-b*.

Basic tax results relevant for this analysis:

- Incidence of a tax does not depend on whether tax is levied on consumer (e.g. added at point of sale) or levied on firm (e.g. posted price is inclusive of the tax)
- Behavior should respond to the net of tax price
- Behavioral response should be identical to prices and taxes
- (Optimal tax result-Ramsey Rule-inverse elasticity rule) Tax more heavily goods that have a lower elasticity

Bounded rationality:
Agents face a cost of processing information. They therefore (rationally) use heuristics to solve complex problems.

These papers show that if there are costs of processing information, then the salience of the tax can lead to the following results:

- Salience of the tax can affect measured elasticities:
    - If you are not aware that the tax changes, does your behavior respond?
    - If the costs of processing information are large, then you do not adjust in the same way
- Optimal tax result: Higher taxes being levied on the less visible or salient taxes (=lower elasticity)
- Political economy result: preference to raise less salient taxes
- Tax incidence: neutrality no longer holds
- Deadweight loss: taxes with small utility losses if ignored by individuals can still create large DWL overall. Ultimately, they still pay the tax and the DWL effects depend on HOW they adjust other spending and what the other spending is.

Connections to other literatures:

- I/O: There is empirical evidence on the differential responsiveness to different components of prices, costs, etc. Examples include: cost of appliance and energy costs, car purchases and manufacturer rebates.
- Taxes and the size of the government: the less visible the tax → tendency to have a larger government.
    - [To my surprise, this was a point made by the 2005 President's Advisory Panel on Federal Tax Reform in a concern to recommend a VAT (which was perceived as being less salient than an income tax)]
- Liebman and Zeckhauser "Schmeduling": labor supply responds more to average tax rates than marginal tax rates

These papers illustrate the kind of work being done by the best young people in empirical public finance

-- a move away from the pure “policy evaluation” of the identification-emphasized, reduced form (difference-in-difference, regression discontinuity) literature
-- instead use those methods to reveal something about behavior, theory

Much more connected to economics and economic theory.

Also part of emerging area of “behavioral public finance”
-- Individual faces cognitive constraints in achieving true optimum when faces with a complex tax system

Chetty et al "Salience and Taxation: Theory and Evidence"

What they test: whether a commodity tax has a larger effect on demand if it is included in the posted price (rather than added at point of sale).

Idea: tax included posted price is "more salient"

Examples of taxes that are included in price or not:

Included in price	*Not included in price*
Excise tax (gas, cigarette, alcohol)	income tax
Airline tax	sales tax

Their empirical evidence:

(1) Experiment in grocery store; augmenting posted price to advertise tax inclusive price

(2) Analysis of behavioral response of alcohol consumption to variation in excise (included in price) and sales taxes (added to price)

Organizing framework for paper:

$$U(x,y) = a\frac{x^{1-b}}{1-b} + y$$

2 goods: *x* and *y*
p=price of *x* (*y* is numeraire)
x is subject to ad valorem tax t^s so inclusive of tax price is $p_t = p(1+t^s)$

Economy consists of θ consumers who maximize subject to the gross of tax price p_t and $1-\theta$ who maximize subject to the pre-tax price of *p*

This results in aggregate demand for x (after approximation and logs) of:

$$\log \hat{x}(p,t,\theta) = \alpha + \beta \log p + \theta\beta \log(1+t^s)$$

Goal is to estimate θ, the fraction who take the sales tax into account. Under canonical neoclassical model, θ=1 and the elasticity of response to p and t is identical.

Empirical evidence 1: Grocery store experiment

Alter posted prices for 3 full categories of taxable items in grocery store in No. California over 3 week period.

Sales tax is 7.375%

- Note: they chose categories that were relative high price (so sales taxes were nontrivial), high elasticity (so demand response would be detectable)
- Cosmetics, hair care accessories, deodorants
- All items in category were treated
- Data is scanner data (used at lot in IO) and records all transactions in the store.
- Data: (1) spans period before and after intervention, (2) covers two stores chosen as control stores, and (3) covers treated categories and control categories (toothpaste, skin care and shaving products)

Triple difference estimation:

	Treatment store			Control Stores		
	Before	During	Diff	Before	During	Diff
Treated products	y_{10}^T	y_{11}^T	Δ_1^T	y_{10}^C	y_{11}^C	Δ_1^C
Control products	y_{00}^T	y_{01}^T	Δ_0^T	y_{00}^C	y_{01}^C	Δ_0^C
	Diff-diff for T store		$\Delta_1^T - \Delta_0^T$	Diff-diff for C store		$\Delta_1^C - \Delta_0^C$
	Difference-in-difference-in-difference= $(\Delta_1^T - \Delta_0^T) - (\Delta_1^C - \Delta_0^C)$					

Why triple difference?
What does it capture?

-- what if you used difference for treated product in treated store?
-- what is additional value in adding difference for treated product in control store? what is left then?
-- what is value of then adding DD for control products?

Triple difference allows you to control for:
-- trends in treated category
-- trends in store
-- over time trends

Identifying assumption: no specific shock to treated category in treated store.

Back to framework model: what does this experiment capture?

In "control" world, we have the following aggregate demand:

$$\log \hat{x}(p,t,\theta) = \alpha + \beta \log p + \theta\beta \log(1+t^s)$$

In the "treated" world, everyone knows post-tax price costlessly so θ=1:

$$\log \hat{x}(p,t,1) = \alpha + \beta \log p + \beta \log(1+t^s)$$

So the change in log of x with the treatment is:

$(1-\theta)\beta \log(1+t^s)$, where β is the price elasticity of demand.

If everyone responds to post-tax prices, then the effect of the experiment should be zero. If everyone responds to pre-tax prices ("posted prices") then the response should be equivalent to the behavioral response of an increase in price of t.

TREATMENT STORE

Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	26.48 (0.22) [5510]	25.17 (0.37) [754]	-1.31 (0.43) [6264]
Experiment (2006: 8- 2006:10)	27.32 (0.87) [285]	23.87 (1.02) [39]	-3.45 (0.64) [324]
Difference over time	0.84 (0.75) [5795]	-1.30 (0.92) [793]	**DD_{TS} = -2.14** (0.64) [6588]

CONTROL STORES

Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	30.57 (0.24) [11020]	27.94 (0.30) [1508]	-2.63 (0.32) [12528]
Experiment (2006: 8- 2006:10)	30.76 (0.72) [570]	28.19 (1.06) [78]	-2.57 (1.09) [648]
Difference over time	0.19 (0.64) [11590]	0.25 (0.92) [1586]	**DD_{CS} = 0.06** (0.90) [13176]
		DDD Estimate	***-2.20*** (0.58) [19764]

Little change for the control store (good). So DDD≅DD

Using an elasticity of 1.59 (estimated in separate regression) you get an estimate of $\hat{\theta} = 0.35$

Conditional DDD

Note: to calculate a triple difference you need to control for
- Main effects: store, category, time
- 2-way interactions: storeXcategory, storeXtime, categoryXtime
- 3-way interaction: store X category X time (treatment effect)

Controls: price, time

Table 4
Table 5

TABLE 4

Effect of Posting Tax-Inclusive Prices: Regression Estimates

Dependent Variable:	Quantity per category	Quantity per category	Revenue per category ($)	Log quantity per category	Log revenue per category
	(1)	(2)	(3)	(4)	(5)
Treatment	**-2.20** (0.58)***	**-2.20** (0.59)***	**-13.12** (4.88)***	**-0.101** (0.03)***	**-0.123** (0.04)***
Average Price		-3.15 (0.26)***	-3.24 (1.74)*		
Average Price Squared		0.05 (0.00)***	0.06 (0.03)**		
Log Average Price				-1.59 (0.11)***	-0.39 (0.11)***
Category, Store, Week FEs		x	x	x	x
Sample size	19,764	19,764	19,764	18,827	18,827

Estimated elasticity from this model: well identified?
(Identified by exploiting the variation in average category-level prices across weeks within the stores.)

<u>Empirical Evidence 2: State excise taxes on alcohol</u>

Alcohol is subject to two state taxes:
Excise tax (imposed on wholesale price, fixed volume tax)
Sales tax (added at register)

Idea is to use variation across states and over time (in typical state- panel identification model) to see if consumers respond differentially to the more salient (excise) and less salient (sales) taxes.

Back to basic model

$$\log \hat{x}(t^E, t^S, \theta) = \alpha + \beta \log(1 + t^E) + \theta\beta \log(1 + t^s)$$

β = elasticity on excise tax related price changes
θβ = elasticity on sales tax related price changes

Model is estimated in changes (since autocorrelated taxes), state j, period t

$$\Delta \log \hat{x}_{jt} = \alpha_0 + \beta\Delta \log(1 + t_{jt}^E) + \theta\beta\Delta \log(1 + t_{jt}^S) + X_{jt}\rho + \varepsilon_{jt}$$

Data: 1970-2003

Excise taxes are higher (6.4%) than sales taxes (4.3%). They are also more variable, but also change in nominal terms less frequently.

Not a super great identification strategy—lot's of trending in both RHS and LHS variables. Secular year fixed effects included but no state specific time trends. Little attention to exogeneity of tax changes.

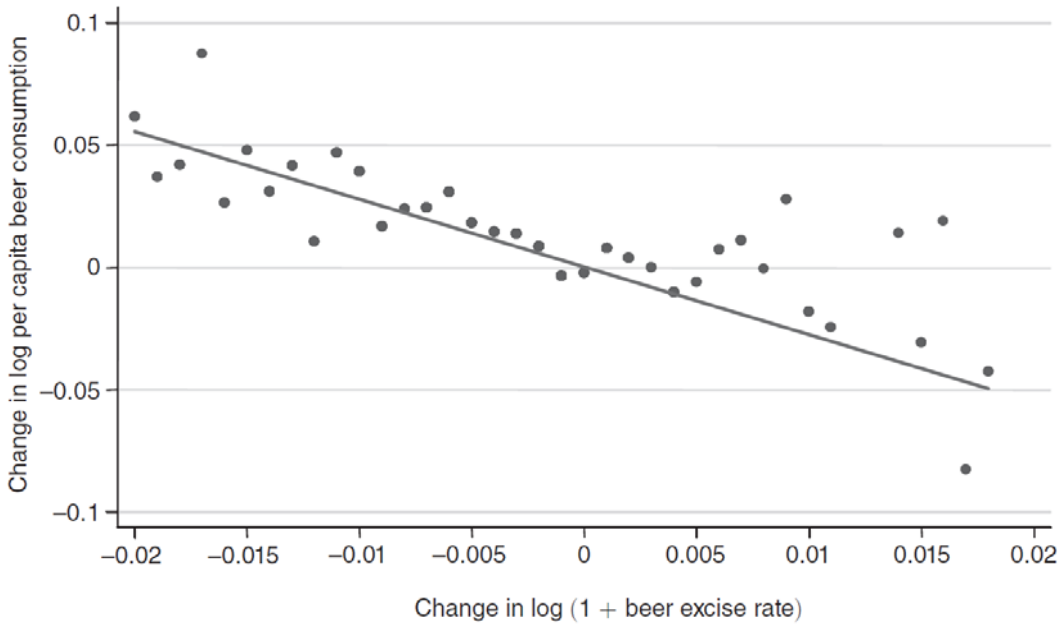


Figure 2A. Per Capita Beer Consumption and State Beer Excise Taxes

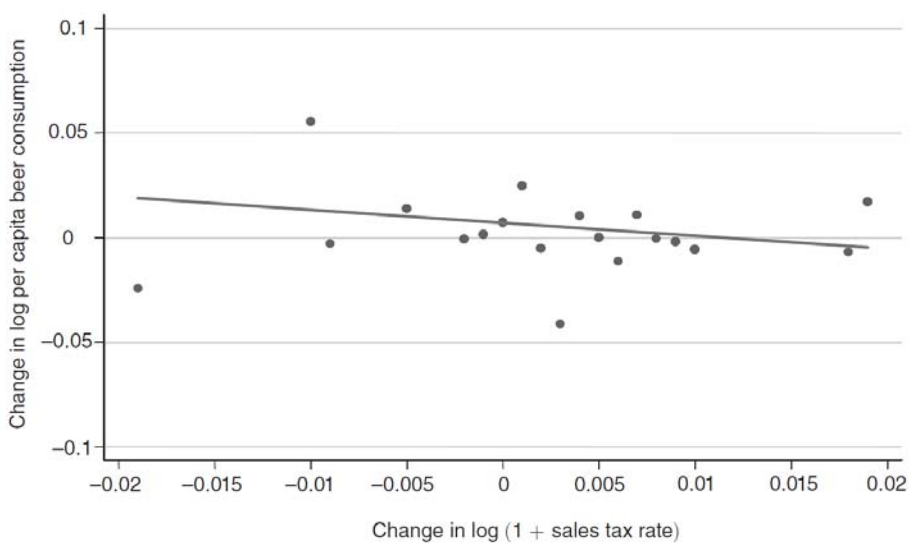


Figure 2B. Per Capita Beer Consumption and State Sales Taxes

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Change in Log(per capita beer consumption)

	Baseline	Bus Cycle	Bus Cycle Lags	Alc Regulations
	(1)	(2)	(3)	(4)
ΔLog(1+Excise Tax Rate)	**-0.87**	**-0.91**	**-0.86**	**-0.89**
	(0.17)***	(0.17)***	(0.17)***	(0.17)***
ΔLog(1+Sales Tax Rate)	**-0.20**	**-0.00**	**0.03**	**-0.02**
	(0.30)	(0.30)	(0.30)	(0.30)
ΔLog(Population)	0.03	-0.07	0.05	-0.07
	(0.06)	(0.07)	(0.19)	(0.07)
ΔLog(Income per Capita)		0.22	0.18	0.22
		(0.05)***	(0.05)***	(0.05)***
ΔLog(Unemployment Rate)		-0.01	-0.01	-0.01
		(0.01)**	(0.01)	(0.01)**
Lag Bus. Cycle Controls			x	
Alcohol Regulation Controls				x
Year Fixed Effects	x	x	x	x
F-Test for Equality of Tax Variables (Prob>F)	0.05	0.01	0.01	0.01
Sample Size	1607	1487	1440	1487

Much larger response to more salient tax.

$\hat{\theta} = 0.06$

Bottom line:

Both approaches show that there is a larger response under salience.
They conducted a small survey (at same grocery store) showing that this is not due to people simply not knowing about the sales tax (either what it applies to or how large it is).

Finkelstein "E-ZTax: Tax Salience and Tax Rates"

- Paper is motivated by the view that less salient taxes lead to larger government: slip in lots of new taxes without people noticing.
- The setting is the introduction of electronic toll collections on U.S. roads, tunnels, and bridges.
- She collects data on tolls, traffic, and the timing of introduction of toll collections in 123 of the 183 sites with tolls in place in 1985.
- She examines how the introduction of ETC affects
  - Tolls (analogy to size of government, or tax rates)
  - Elasticity of road use to toll price
- Evidence is compelling that tax rates rise when less salient tax is created (tolls rise with ETC introduction)
- Idea: once you use electronic payment you no longer pay attention to the toll amount.

Government Objective Function (max SWF of indi utils [j] by choosing taxes)

$$\max_{\vec{\tau}} \sum_j \upsilon_j W_j(\tau_j) + (1 - \sum \upsilon_j) R(\vec{\tau})$$

Leads to the inverse elasticity rule:

$$\frac{\tau_j^*}{p_j + \tau_j^*} = \frac{1}{\varepsilon_j} \left(\frac{1 - \left(\sum_j \upsilon_j \right) - \lambda_j \upsilon_j}{1 - \left(\sum_j \upsilon_j \right)} \right)$$

the tax rate on group j is decreasing in the demand elasticity, the MU of income (λ) and the social welfare weight

Prediction of the model for electronic toll introduction:
-- with electronic toll, the tax is less salient
-- therefore a change in the tax (toll) will lead to smaller changes in behavior (driving); elasticity falls
-- with less salient tax, tolls more likely to increase (=big government)

This analysis is predicated on the assumption that income effects of this change are small (which makes sense since tolls are small share of individual spending and government revenue)

Survey Evidence

What is the role of this?

[Her own survey of Mass Pike drivers; also a commuter survey in NY/NJ]

Shows that those using ETC are:

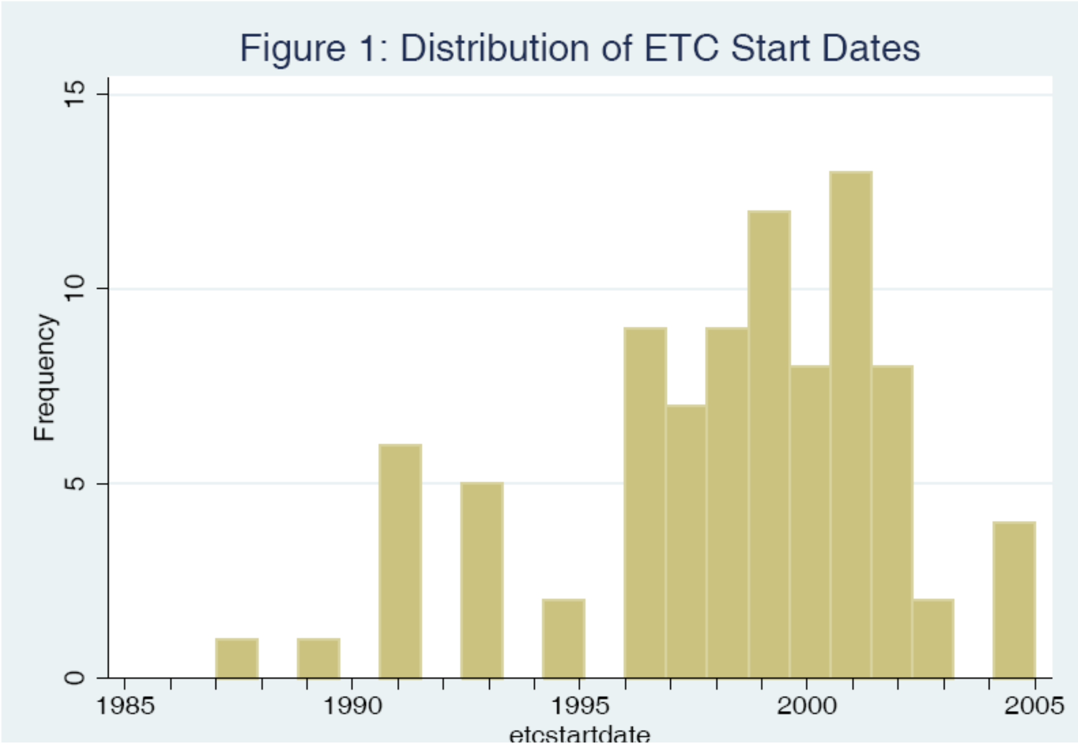
- Less likely to know amount of toll they pay (could this just be higher income folks are more likely to have ETC and are not paying attention?)

Table 1: Survey Evidence on Driver Awareness of Tolls, by Payment Method

	MA Survey				NYNJ Survey		
			Difference between ETC and Cash Drivers				
	ETC Drivers (1)	Cash Drivers (2)	No Covariates (3)	Covariate Adjusted (4)	ETC Drivers (5)	Cash Drivers (6)	Difference between ETC and Cash Drivers (No covariates) (7)
Fraction report "don't know"	0.618 (0.490)	0.021 (0.142)	0.597*** (0.060)	0.579*** (0.060)	0.381 (0.486)	0.200 (0.400)	0.18*** *(0.05)*
Fraction who incorrectly estimate toll	0.851 (0.359)	0.308 (0.463)	0.543*** (0.058)	0.512*** (0.067)	0.826 (0.379)	0.395 (0.489)	0.43*** *(0.06)*
Mean error, conditional on misreporting	$1.334 (1.850)	$0.162 (0.828)	1.172*** (0.275)	1.01*** (0.303)	$0.40	-$0.10	$0.50
N	68	146			*271*	*91*	

<u>Adoption of ETC</u>

Across states



Basic Model

$$y_{it} = \gamma_t + \beta_1 \text{ETCAdopt}_{it} + \beta_2 \text{ETC}_{it} + \varepsilon_{it}$$

where ETC=1 if electronic in place in year t
ETCAdopt=1 if adopted this year
Year controls

$$y = \Delta \log(\text{minimum toll})$$

minimum toll: when ETC implemented sometimes there is a discount offered in the early years. To get around endogeneity of this, her preferred sample excludes the areas where the discount is used.

Since the model is in differences, then the gammas capture average growth rates by year, and the betas capture deviations from those growth rates.

Why identify an impact of the year of adoption? (why include ETCAdopt)
Source of identification in the model?

ETCAdopt is used to capture the initial impact of the introduction, maybe with a discount

The identification is a simple DD model, comparing changes across areas with and without ETC.

Identifying assumption:

ETC are not endogenous (places with ETC implemented are on same trend line and placed w/o ETC). ETC implementation is not correlated with changes in toll setting relative to its norm.

Details in estimation:
-- weighting: she weights 49 operating authority equally. Why does that make sense? Why not weight by usage?
-- she clusters on state (why?)

Basic approach amounts to two steps:

1) Establish that ETC is associated with an increase in tolls.

2) Present evidence in support of the hypothesized mechanism, namely that ETC increases the equilibrium toll rate by decreasing its salience.

- -- short run elasticity of driving with respect to toll declines with ETC
- -- toll setting behavior becomes less sensitive to the local election calendar

Table 3: Impact of ETC on Toll Rates.

	Δ Log Min. Toll	Δ Log Manual Toll	Δ Log Toll	Δ Log Toll	Δ log Min. Toll	Δ log Min. Toll
	(1)	(2)	(3)	(4)	(5)	(6)
ETC_{it}	0.015 (0.006) [0.018]	0.020 (0.006) [0.004]	0.024 (0.012) [0.061]			
$\Delta ETC\_Penetration_{it}$				0.623 (0.285) [0.044]	0.557 (0.262) [0.045]	0.501 (0.261) [0.067]
$ETCAdopt_{it}$	-0.051 (0.035) [0.158]	0.016 (0.032) [0.622]	-0.033 (0.019) [0.097]	-0.051 [0.035] [0.166]	-0.105 (0.109) [0.348]	-0.097 (0.108) [0.380]
Mean dep. var	0.020	0.022	0.017	0.017	0.020	0.020
# of states	24	24	17	17	24	24
# op. author	49	49	31	31	49	49
# facilities	123	123	70	70	123	123
N	5,079	5,079	2,875	2,751	4,815	4,815
Estimation	OLS	OLS	OLS	OLS	IV	IV
Sample restriction			No ETC discount	No ETC discount		

Installing ETC leads to 75% more increase in tolls (=1.5/2.0). [2 is in the denominator because it is the average increase in the period].

Effect in the first yr is the sum of the two coefs.

IV: instrument ETC penetration with dummy for ETC introduction.

Overall, diffusion of ETC to steady state (60%) leads to an increase of 20-40 percent increase in rates.
[ETC_penetration=fraction of toll transactions or revenue collected by ETC]

Table 5 adds facility FE. Why isn't this in the main model?

Figure 3A: Full Sample

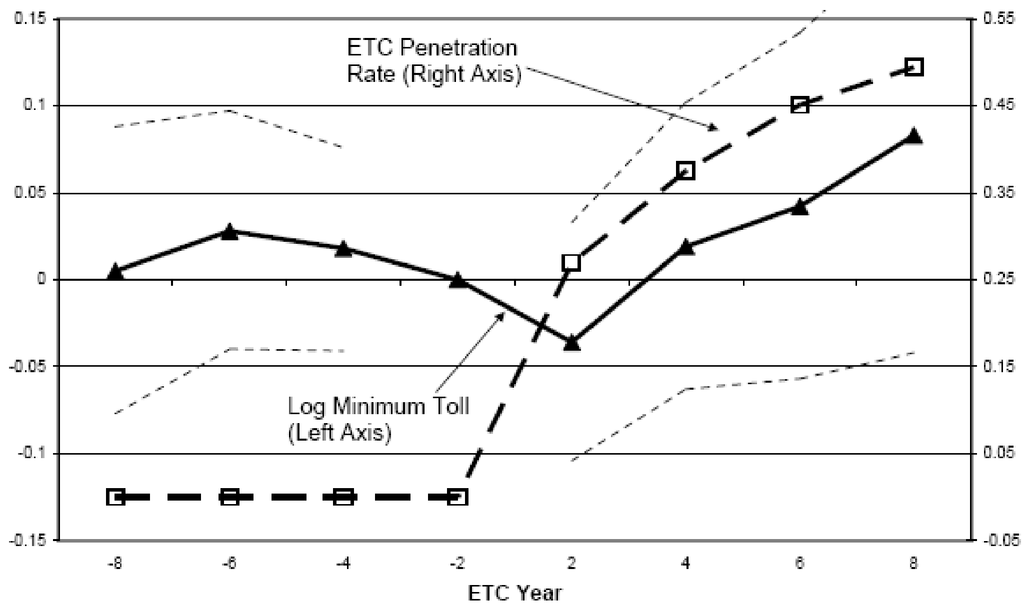
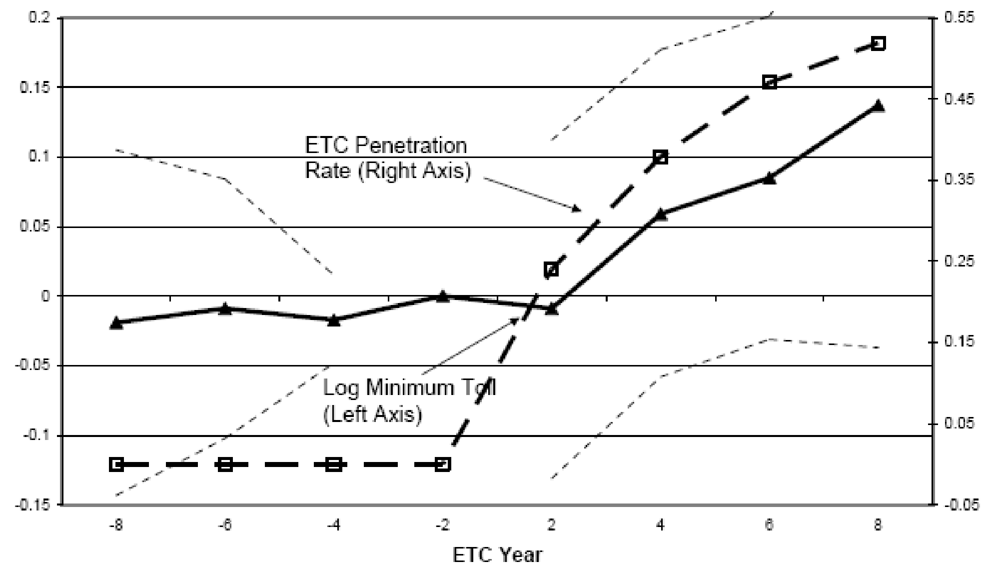


Figure 3B: Balanced Panel



Examine timing of ETC on toll increases using "event study"

- Look for pre-trend to be flat (note text around balanced panel)
- Examine how treatment effects change (increase) with time since treatment

Impacts on traffic:

$$\Delta\log(traffic) = \gamma_t + \beta_1 \Delta\log(\min toll_{it}) + \beta_2 \Delta\log(\min toll_{it}) * Never\_ETC +$$
$$\beta_3 \Delta\log(\min toll_{it}) * ETC\_penetration +$$
$$Never\_ETC + ETC\_penetration + \varepsilon_{it}$$

I think we know that the price elasticity of demand for gas (and driving?) has decreased over time. I think that means we want to control for mintoll*year in the regression (?)

again just year FE in model not facility FE
why no instrumenting for penetration? I would like to see the simple DD.

Table 5: The elasticity of traffic with respect to tolls

	(1)	(2)	(3)	(4)	(5)	(6)
Δ log min. toll $_{it}$	-0.049	-0.058	-0.061	-0.057	-0.062	-0.060
	(0.015)	(0.018)	(0.019)	(0.017)	(0.039)	(0.037)
	[0.004]	[0.008]	[0.009]	[0.006]	[0.145]	[0.135]
Δ log min. toll $_{it}$ * ETC_Penetration $_{it}$			0.134		0.141	
			(0.038)		(0.076)	
			[0.005]		[0.091]	
Δ log min. toll $_{it}$ * ETC_Year $_{it}$				0.006		0.006
				(0.001)		(0.003)
				[0.002]		[0.062]
Δ log min. toll $_{it}$ * Never_ETC $_{i}$			-0.071	-0.073	-0.009	-0.006
			(0.136)	(0.131)	(0.209)	(0.205)
			[0.611]	[0.588]	[0.966]	[0.976]
Mean dep. Var	0.049	0.042	0.043	0.042	0.040	0.039
# of states	21	12	12	12	12	12
# op authors	32	16	16	16	16	16
# of facilities	76	33	33	33	33	33
N	2,200	727	671	727	292	305
Sample restriction(s)		No ETC discounts	No ETC discounts	No ETC discounts	No ETC discounts +2/-2	No ETC discounts +2/-2

Base elasticity= -0.06

Larger if never ETC= -0.061-0.071

Smaller if ETC (makes sense since less salient)

Overall, though, really small elasticities.

Political Economy

-- baseline assumption is that legislators do not want to increase taxes in election years.
-- if less salient taxes do not change behavior (people are less aware of the tax changes) then there should be less of an election year effect with the less salient tax

$$y_{it} = \gamma_t + \beta_1 \text{ETCAdopt}_{it} + \beta_2 \text{ETC}_{it} + \beta_3 1(ElecYear_{st}) +$$
$$\beta_4 1(ElecYear_{st}) * \text{ETCAdopt}_{it} + \beta_5 1(ElecYear_{st}) * \text{ETC} + \varepsilon_{it}$$

Expect:

$\beta 3$=negative (fewer toll increases in election years)

$\beta 5$=positive (less negative due to less salient)

Table 6: The impact of ETC on the politics of toll setting

	Δ Log Min Toll (1)	Min Toll Raised? (2)	Δ Log Min. Toll (3)	Min Toll Raised? (4)	Δ Log Min. Toll (5)	Min Toll Raised? (6)	Δ Log Min. Toll (7)	Min Toll Raised? (8)	Δ Log Min. Toll (9)	Min Toll Raised? (10)
ETC_{it}	0.015 (0.006) [0.018]	0.073 (0.024) [0.006]	0.016 (0.006) [0.017]	0.074 (0.024) [0.006]	0.016 (0.006) [0.016]	0.074 (0.024) [0.005]	0.006 (0.009) [0.507]	0.044 (0.022) [0.042]	0.006 (0.009) [0.494]	0.044 (0.022) [0.042]
AnyElec $Year_{st}$			-0.015 (0.006) [0.017]	-0.026 (0.011) [0.023]			-0.016 (0.004) [0.000]	-0.029 (0.010) [0.003]		
GovElec $Year_{st}$					-0.017 (0.006) [0.011]	-0.036 (0.013) [0.010]			-0.016 (0.005) [0.001]	-0.036 (0.012) [0.002]
LegOnly $ElecYear_{st}$					-0.013 (0.007) [0.064]	-0.014 (0.013) [0.263]			-0.015 (0.005) [0.005]	-0.021 (0.012) [0.085]
AnyElec $Year_{st}$ $*ETC_{it}$							0.017 (0.012) [0.140]	0.055 (0.027) [0.041]		
GovElec $Year_{st}$ $*ETC_{it}$									0.004 (0.014) [0.791]	0.016 (0.033) [0.617]
LegOnly $ElecYear_{st}$ $*ETC_{it}$									0.030 (0.014) [0.038]	0.094 (0.033) [0.005]

Indeed, under ETC there is less of an election year effect. Less sensitive to electoral business cycle.