Field exam notes

Lucy Hackett

Contents

1 Meredith Fowlie **4**
1.1 Lecture 1: Intro and field experiments . . . 4
1.1.1 Field experiment fundamentals . . . 4
1.1.2 The potential outcomes framework . . . 5
1.2 Lecture 2: Experimental designs in EEE . . . 6
1.2.1 Experimental pitfalls . . . 6
1.2.2 EEE and the developing world . . . 8
1.3 Lecture 3: Power analysis . . . 8
1.3.1 Randomization inference . . . 8
1.4 Lecture 4: Instrumental variables and Hedonics . . . 9
1.4.1 Instrumental variables . . . 9
1.4.2 Hedonics overview . . . 10
1.5 Lecture 5: Electricity markets (and DiD) . . . 11
1.5.1 Electricity Supply . . . 11
1.5.2 Electricity Demand . . . 12
1.6 Lecture 6: Cap and Trade . . . 12
1.6.1 Leakage . . . 13
1.6.2 Equity . . . 13
1.7 Lecture 7: Intro to discrete choice . . . 13
1.7.1 Comparing logit models . . . 14
1.7.2 Random utility maximization . . . 14
1.7.3 Conditional Logit (CL) . . . 15
1.8 Lecture 8: Mixed Logit . . . 18
1.8.1 The Berry Transformation . . . 18
1.8.2 Mixed Logit . . . 18
1.9 Lecture 9: Logit applications . . . 20
1.9.1 Estimating WTP from logit models . . . 20
1.9.2 Estimating CS from logit models . . . 20
1.10 Lecture 10: BLP . . . 21
1.10.1 Preference heterogeneity . . . 22
1.10.2 Using market-level data . . . 22
1.10.3 Endogenous regressors . . . 23
1.10.4 Adding more structure via the supply side . . . 23
1.11 Lecture 11: BLP applications . . . 23

2 Joe Shapiro **28**
2.1 Lecture 1: Public Finance . . . 28
2.1.1 Sufficient statistics . . . 28
2.2 Lecture 2: Health demand . . . 29
2.3 Lecture 3: Climate damages . . . 30
2.4 Lectures 4+5: Hedonics and local public goods . . . 30
2.4.1 Rosen's model (1974) . . . 30
2.4.2 Estimation . . . 32
2.5 Lectures 6+7: Roback models . . . 34
2.6 Lectures 8+9: Water . . . 35
2.7 Lecture 10: Natural Resources . . . 37
2.8 Lectures 11+12: Trade, Leakage and GE land . . . 38
2.8.1 The Armington model . . . 39
2.8.2 Applications . . . 39
2.9 Lecture 14: Urban/ Spatial . . . 40

3 Jim Sallee **42**
3.1 Lecture 1: The Pigouvian prescription . . . 42
3.1.1 Basics of the Pigouvian prescription . . . 42
3.1.2 A model of abatement cost . . . 43
3.2 Lecture 2: Tradeable permits and instrument choice . . . 44
3.2.1 Quantity instruments . . . 44
3.2.2 Uncertainty . . . 44
3.2.3 Instrument choice . . . 44
3.3 Lecture 3: Instrument choice . . . 46
3.3.1 The Coase Theorem . . . 46
3.4 Lecture 4: Performance standards and rebound . . . 47
3.4.1 Rebound . . . 47
3.4.2 Performance standards . . . 47
3.4.3 The Gruenspecht effect (leakage) . . . 48
3.5 Lecture 6: Incidence and equity . . . 49
3.5.1 How regressive is environmental policy? . . . 49
3.6 Lecture 8: Second-Best policy I . . . 50
3.6.1 The Diamond model of heterogeneous consumers . . . 50
3.7 Lecture 9: Reviving the first best . . . 52
3.7.1 Two-part instruments . . . 52
3.7.2 Taxing ambient pollution . . . 52
3.7.3 Pigou in the presence of other markets . . . 52
3.7.4 Revenue recycling . . . 53
3.8 Lecture 12: Second best policy II . . . 53
3.8.1 More factors that complicate Pigou . . . 54
3.9 Lectures 13+14: Behavioral I and II . . . 55

4 Reed Walker **57**
4.1 Lectures 1+2: Intro and Measurement . . . 57
4.1.1 Measuring the Social Cost of Air Pollution . . . 57
4.1.2 Benefits of the CAA . . . 58
4.1.3 Non-Market valuation: Environmental Externalities I . . . 59
4.2 Lecture 3: Non-Market valuation of pollution . . . 60
4.3 Lectures 4+5: Non-Market valuation of climate change . . . 60
4.3.1 Identification and the construction of temperature variables . . . 61
4.3.2 Top–down approaches . . . 62
4.3.3 The Social Cost of Carbon . . . 63
4.4 Lecture 6: Non-Market valuation (hedonics) . . . 64
4.4.1 Hedonic valuation of climate impacts . . . 64
4.4.2 Non–market valuation . . . 66
4.5 Lectures 7+8: Cost benefit analysis . . . 67
4.5.1 IO models of a single sector . . . 67
4.6 Lectures 9+10: Production Functions . . . 72
4.6.1 Anatomy of a production function . . . 72
4.6.2 Data issues . . . 72
4.6.3 Estimation . . . 72
4.6.4 Index methods: Olley-Pakes (1996) . . . 73
4.6.5 High-level strategies when estimating production functions . . . 74
4.6.6 Estimating markups . . . 76
4.7 Lecture 13: Principle-Agent problems in EEE . . . 78
4.7.1 Rules versus discretion . . . 79

Part 1

Meredith Fowlie

Exam question style

2021 Conditional logit. Derive the linear estimating equation. Show how to use this to calculate Δ consumer surplus. Assumptions, both about CL and the empirical set-up.

2020 Power outages in developing countries. Derive CL and talk about assumptions both of interpretation as WTP and CL itself.

2019 CL; considerations when there is heterogeneity. Interpreting the parameter estimate. BLP/ IV with CL.

2018 Derive CL. Interpret parameter estimates. (Not) comparing parameters across estimates.

2017 Assumptions of experimental design. LATE interpretation. Encouragement designs. Mixed logit. Comparing coefficients.

Takeaways: Focus on mixed logit, mechanics, interpretation and application. Also spend time on experimental design and interpretation.

1.1 Lecture 1: Intro and field experiments

Definition 1. *A* ***structural model*** *is a model that explicitly combines theoretical structure/ assumptions with statistical assumptions in order to estimate or "recover" the parameters of a model or function from the joint density of economic data, in this case* $f(x, y)$.

In order to identify a parameter, we follow these 5 steps:

1. Write down a sensible theoretical model (e.g. a production function relating inputs and outputs).
2. Be explicit about what we can and cannot observe.
3. Think about the processes that give rise to "structural errors". Impose (and substantiate if possible) assumptions about the statistical properties of these errors.
4. Derive statistical objects/estimands from the model.
5. Identify the parameters that best match or rationalize the data we observe (conditional on the assumed structure).

1.1.1 Field experiment fundamentals

First, a critique of experiments. This rests on a couple of points:

- Experiments may be idiosyncratic (not externally valid) or unable to parse mechanisms, meaning that the knowledge learned from them does not accumulate. There also may be little economic interpretation to them.
- Experiments "play small ball", answering questions for which SUTVA assumptions hold or that are financially feasible, but cannot speak to big questions/changes.

Even with these critiques, and even if you will not do an experiment, it can be helpful to conceptualize your empirical strategy in terms of the ideal experiment you would run if you could.

Energy efficiency

- Energy efficiency is widely believed to be a low-cost way of reducing energy demand.
- Subsidizing the transition to more efficient technology can address market failures such as environmental externalities, behavioral biases, and asymmetric information.
- But ex-ante (engineering) estimates can oversell the gains due to rebound, variation in installation/quality, and overly optimistic assumptions about performance.

1.1.2 The potential outcomes framework

Units i receive treatment that can vary across but not within units.

Definition 2. *$Y_i(1)$ is the potential outcome that could happen for i if i is treated. $Y_i(0)$ is the potential outcomes for i if i is not treated (in the binary case).*

Definition 3. *$Y_i(1) - Y_i(0)$ is the unit–level causal effect of treatment.*

Definition 4. *$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ is the observed outcome for i.*

The fundamental challenge of causal inference is to construct our best estimate of the counterfactual outcomes we cannot observe! The assignment mechanism is the function that assigns probabilities to each possible *assignment* $D_{N\times 1}$. In general, assignment mechanisms can depend on some X or on the potential outcomes $Y(0), Y(1)$. There is already an assumption in the notation above:

Definition 5. *The **Stable Unit Treatment Value Assumption (SUTVA)** states that $Y_i(D) = Y_i(t) \forall\ D : D_i = t$.*

Classical assignment mechanisms all have the property that:

$$\sum_{D\in\{0,1\}^N} Pr(D|X, Y(0), Y(1)) = 1$$

They may also have the property of **unconfounded assignment:**

$$Pr(D|X, Y(0), Y(1)) = Pr(D|X)$$

Experiments generally have high internal validity. They may or may not have high external validity or construct validity.

Definition 6. ***Construct validity** is the extent to which the experimental treatment mimics the real-world intervention of interest.*

Definition 7. *The **Average treatment effect (ATE)** is:*

$$\frac{1}{N}\sum_i (Y_i(1) - Y_i(0))$$

Given randomization, this can be estimated:

$$\begin{aligned} E(Y_i(1) - Y_i(0)) = E(Y_i(1)) - E(Y_i(0)) &= E(Y_i(1)|D_i = 1) - E(Y_i(0)|D_i = 0) \\ \Rightarrow \widehat{\tau} &= \overline{Y}_T - \overline{Y}_C \end{aligned}$$

assuming:

1. **Ignorable assignment**: $(Y_i(1), Y_i(0)) \perp Di$
2. SUTVA

This can be estimated via OLS, using the regression:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

where ε_i captures treatment heterogeneity, sampling variation, and other determinants of Y_i. In general, we don't need controls, but may include them to increase precision, analyze heterogeneity, and if stratification was used. However, be aware that including covariates in the regression equation can introduce bias in finite samples because randomization does not guarantee zero correlation between covariates and treatment assignment in finite samples.

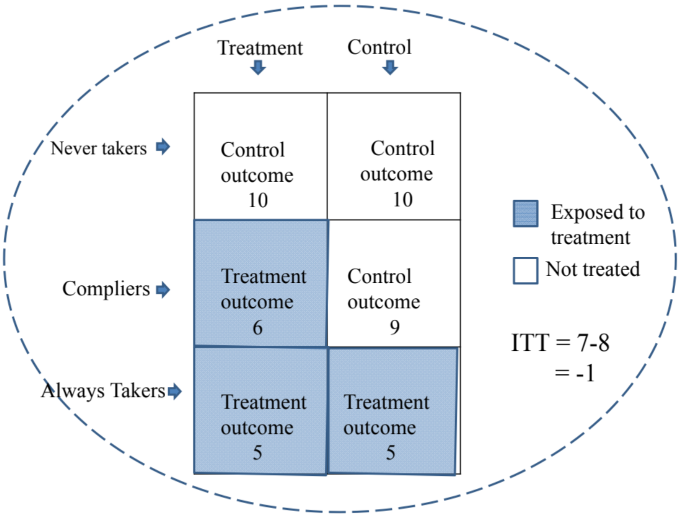


Figure 1.1: Estimating the ITT under imperfect compliance

1.2 Lecture 2: Experimental designs in EEE

1.2.1 Experimental pitfalls

Definition 8. ***Hawthorn effects:*** *when the knowledge that a subject is part of an experiment changes the way they behave.*

- This is important in energy settings because people normally don't think too much about their energy use, but they may more as part of the experiment.
- Schwartz et. al (2013): treatment consisted entirely of households being told they were part of an experiment. Energy consumption decreased 2.7% only during the experiment. Gosnell, List, and Metcalfe (2019) also find that control pilots dramatically change behavior.

Non–compliance can also be an issue, because it is usually not random, which changes the composition of treated and control groups. Solutions:

1. Estimate the ITT. Downside: power, note the ATE.
2. Manipulate the probability of treatment and estimate an IV. Downside: lower power, LATE $\neq$ ATE, requires some additional assumptions.
3. Focus on a sub–population whose treatment status you can manipulate (recruit and deny or recruit and delay). Downside: not representative (construct validity).

Encouragement designs

Note that encouragement designs assume no defiers! Mathematically:

$$\begin{aligned}
E[Y_i|Z_i=1] &= \pi_N E(Y_i(0)|N) + \pi_C E(Y_i(1)|C) + \pi_A E(Y_i(1)|A) \\
E[Y_i|Z_i=0] &= \pi_N E(Y_i(0)|N) + \pi_C E(Y_i(0)|C) + \pi_A E(Y_i(1)|A) \\
E[Y_i|Z_i=1] - E[Y_i|Z_i=0] &= \pi_C \left(E[Y_i(1)|C] - E[Y_i(0)|C]\right) \\
&= \tau_{ITT}
\end{aligned}$$

This depends on 2 additional assumptions:

1. Exclusion: $Y_i(D, Z_i = 1) = Y_i(D, Z_i = 0)$. The encouragement has no direct effect on the outcome.
2. Monotonicity: $D_i|Z_i = 1 \geq D_i|Z_i = 0 \; \forall \; i$

Finally, to get the ATE for compliers, just divide by the probability of being a complier, which you can back out from the sample moments in Table 1.1.

	$Z_i = 1$	$Z_i = 0$
Always takers	$D_i = 1$	$D_i = 1$
Never takers	$D_i = 0$	$D_i = 0$
Compliers	$D_i = 1$	$D_i = 0$
Defiers	$D_i = 0$	$D_i = 1$

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	$(\pi_N + \pi_C)(1 - P)N$	$(\pi_D + \pi_N)PN = \pi_N PN$
$D_i = 1$	$(\pi_D + \pi_A)(1 - P)N = \pi_A(1 - P)N$	$(\pi_C + \pi_A)PN$

Notes: P is the probability of being assigned to the treatment group. $\pi_D = 0$ under the assumption of no defiers (monotonicity).

Table 1.1: Experimental subgroups

Fowlie et al (2018): Weatherization Assistance program

This paper studies the causal effect of weatherization on energy consumption and expenditures, and seeks to compare realized reductions to engineering estimates as well as test for rebound. They use an encouragement design where they randomly encouraged and aided the application process for weatherization upgrades to hundreds of Michigan households.
Takeaways:

- Participation was a tough sale! Only 5% of encouraged HHs took up, and 1% of control HHs.
- Find a 20% reduction in energy consumption, which is large but $< 1/3$ the predicted. This is not necessarily evidence of rebound though, as indoor temps would have to rise 26° to justify the difference. Likely overly optimistic modelling of technology performance.
- Not cost effective, even using social cost.

Heterogeneous effects

One way to study heterogeneous effects is through interaction terms in a regression. But what if we don't know the salient attributes? Machine learning can help!

- In Random Forests, the split at each tree node is often performed by minimizing the mean squared error of the outcome variable Y.
- In Causal Forests, we cannot observe treatment effects for individuals!
- The prediction of a treatment effect is given by the difference in the average outcomes Y between the treated and the untreated observations in a leaf.
- Splitting criteria searches for a partitioning that maximizes differences in treatment effects across nodes subject to penalties for within-node variance in ATEs and treatment-control imbalance.

Equity and electricity tariffs

- The poor spend a greater share of their income on energy than the rich.
- Electricity utilities are natural monopolies → call for 2-part pricing (FC to cover fixed investments, marginal price = MC).
- Linear pricing would be more equitable, since lower–use users would pay less. But since consumption doesn't rise much with income, this would be a blunt instrument.
- Hahn and Metcalfe (2021) estimate elasticity of demand for natural gas using RED around signing up for the CARE subsidy.
- Find that CARE increases consumption of NG. Structural model → negative value of subsidy (no redistribution benefits). This is driven by the fact that the low income have higher elasticity to prices, and that they assume that increasing the CARE subsidy increases the price for non-CARE consumers. Therefore gas consumption overall rises, and if the social cost of gas is between CARE and non-CARE customers, then increasing the subsidy and the non-CARE price drives both prices away from marginal social cost.

1.2.2 EEE and the developing world

Carranza and Meeks (2021)

They estimates the impact of energy efficiency improvements on household electricity consumption and reliability in the Kyrgyz Republic. Improvements in energy efficiency can decrease overall stress on the grid during peak demand, improving reliability for all users.

- Implement randomized saturation design to test how adoption can reduce peak loads. Two-level randomization: transformers (C, L, H saturation) then HHs.
- Becker-de Groot-Marshack (BDM) to elicit WTP for subsidized bulbs.

They find:

1. Outages in high saturation areas see fewer blackouts.
2. CFLs decrease electricity consumption in low-blackout areas, but do not change it in high-blackout areas $\rightarrow$ reduction in blackout times lets them consume more (rebound).
3. Adoption spillovers are present.

1.3 Lecture 3: Power analysis

There are two ways of conducting power analysis: with closed form expressions or via simulations. These two methods are compared in Table 1.2.

	Randomization inference	**Simulations**
Drawbacks	Potentially strong assumptions	Need data
Description	• Randomness comes from the assignment mechanism, so shuffling the assignment generates a distribution of test statistics. • Expectation taken over the set of all assignments.	• Randomness comes from sampling variation, so potential outcomes are stochastic. • Expectation taken over sampling from the super population.

Table 1.2: Power calculation method characteristics

1.3.1 Randomization inference

We know that our test statistic will be $\widehat{\tau}/\sqrt{var(\widehat{\tau})}$, so we need an estimate of the variance of the estimator. Remembering that in this framework, D is fixed and $\varepsilon\ N(0, \sigma^2)$

$$\begin{aligned}
\text{Var}(\widehat{\tau}) &= \text{Var}((D'D)^{-1}D'y) = (D'D)^{-1}\text{Var}(D'y)(D'D)^{-1} \\
&= (D'D)^{-1}\text{Var}(D'(\alpha + D\tau + \varepsilon))(D'D)^{-1} \\
&= (D'D)^{-1}\text{Var}(D'\varepsilon)(D'D)^{-1} = (D'D)^{-1}D'\text{Var}(\varepsilon)D(D'D)^{-1} \\
&= (D'D)^{-1}\sigma^2 = \frac{\sigma^2}{\sum_i (D_i - P)^2} \\
&= \frac{\sigma^2}{PN(1-P)}
\end{aligned}$$

Note that this is minimized when $P = 1/2$. In order to reject the null hypothesis, we need:

$$\begin{aligned}
t_{1-\chi} + t_\alpha &< \frac{\widehat{\tau}}{\sqrt{\text{Var}(\widehat{\tau})}} \\
\widehat{\tau} &> (t_{1-\chi} + t_\alpha)\sqrt{\text{Var}(\widehat{\tau})} \\
MDE &= (t_{1-\chi} + t_\alpha)\frac{\sigma}{\sqrt{PN(1-P)}}
\end{aligned}$$

where α is the probability, under the null, that τ falls into the rejection region (type I error) and χ is the probability that a test will correctly reject the null (the power of the test, $1-$ type II error rate). This is illustrated in Figure 1.2.

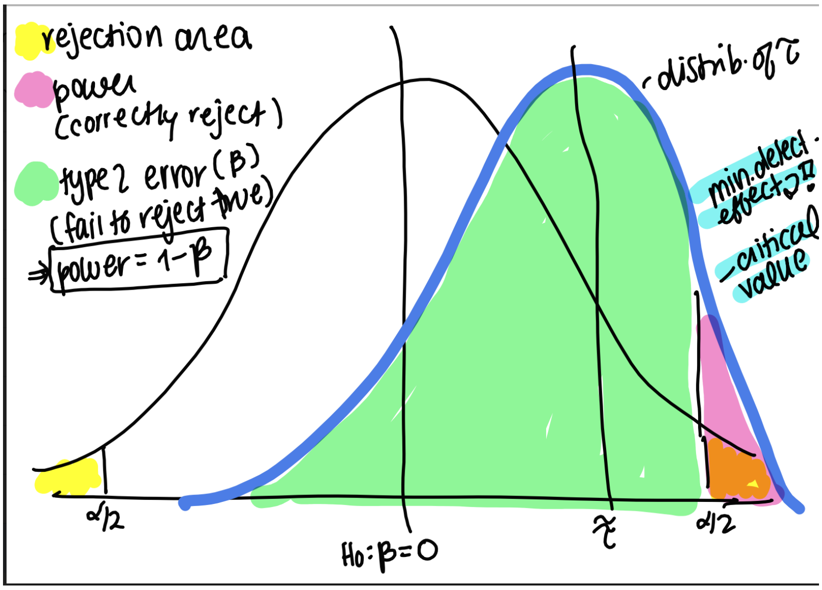


Figure 1.2: Power

Encouragement designs

Under encouragement designs, this becomes:

$$MDE = (t_{1-\chi} + t_\alpha)\frac{\sigma}{(c-s)\sqrt{PN(1-P)}}$$

where c is the share treated in the encouraged group, and s is the share treated in the control group.

Cluster randomization

Suppose the error is $\varepsilon_{it} = v_i + \epsilon_{it}$, so that v_i is the cluster variation distributed $\sim N(0, \sigma_v^2)$.

Definition 9. *The* ***Intracluster correlation coefficient*** *is given by*

$$\rho = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2}$$

Given this, the MDE is:

$$MDE = (t_{1-\chi} + t_\alpha)\frac{\sigma}{\sqrt{JP(1-P)}}\sqrt{\rho + \frac{1-\rho}{T}}$$

If $\rho \to 1$, then power is not increasing in T.

1.4 Lecture 4: Instrumental variables and Hedonics

1.4.1 Instrumental variables

Instruments must satisfy:

1. Exclusion.
2. Relevance.

Some characteristics of IV:

- β_{IV} can be biased in small samples since it is the ratio of 2 numbers.
- When the exclusion restriction fails, $\beta_{IV} = \beta + \frac{cov(z,v)}{cov(x,z)}$. When the instrument is weak, even small deviations from the exclusion restriction can result in serious bias.
- IV estimates are LATE estimates.
- If treatment effects are heterogeneous, then we need to assume away defiers.
- Heterogeneity and multiple instruments, you are capturing a mix of heterogeneous effects.
- Always report reduced form results! The exclusion restriction is weaker here and the ratio bias is not an issue.
- If you have multiple instruments, report the F stat using a just-identified specification with your best instrument.

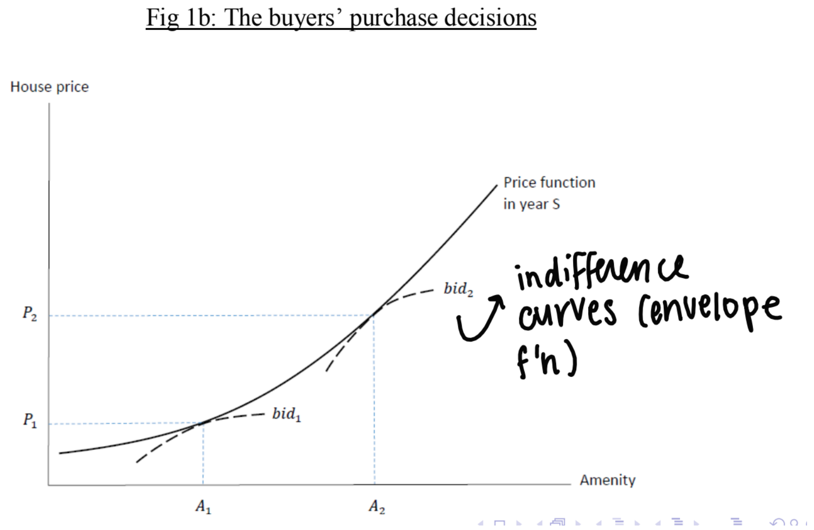


Figure 1.3: Hedonic price schedule

1.4.2 Hedonics overview

Why are real estate markets such good places to estimate WTP?

1. Real estate markets are usually competitive and feature a large volume of transactions.
2. Big ticket purchase → buyers invest a lot in learning about amenities.

The demand side:

- Buyers have heterogeneous preferences. Some forms of this heterogeneity are observable to us, others are not.
- In a competitive market, preferences for amenities are capitalized into housing prices.

The supply side: See Joe's half.

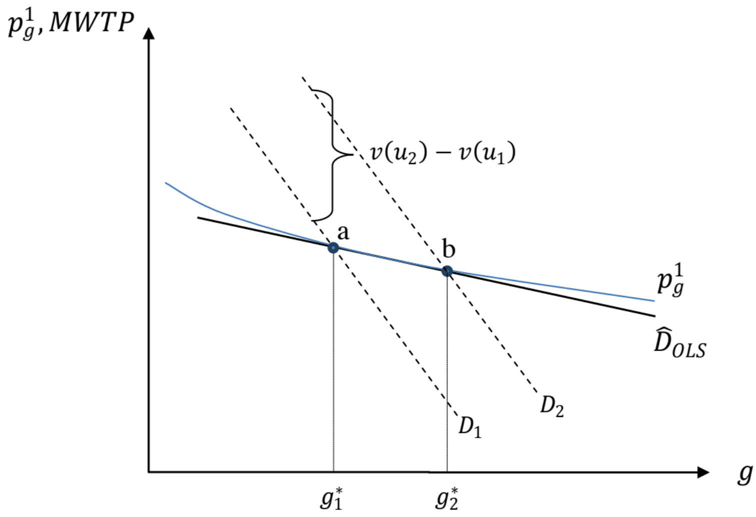
Consumer bid functions trace out their willingness to pay, and their tangency with offer curves traces out the price function for an amenity (see Figure 1.3). The slope of these bid functions at the intersection is the marginal willingness to pay for the amenity! But we do not observe bid functions, only the points of intersection...

Empirically, the steps are:

1. Specify and estimate $P(x, q)$.
   - Challenges: Defining the market spatially and temporally (you want the law of one price to hold without giving up variation in amenity, and for stable conditions to prevail over time), measuring variation in amenities, separating experienced utility from perceived utility, getting the functional form of P right.
   - OVB is a concern here! Separating the amenity from political power, wealth, other public goods, etc. may be difficult.
   - Quisiexperiments like DiD can be problematic because they cause the price function to change → SUTVA violation!
2. Trace out the demand function to back out marginal WTP.
   - At best, the first stage estimates a single point on an implicit inverse demand schedule. But to get MWTP for a discrete change, we need the whole schedule.
   - If implicit prices are non-constant, consumers can choose the marginal price they pay by sorting into a market segment with higher/lower levels of the attribute. This taste-based sorting complicates the identification of hedonic price schedules (except for the special case in which the hedonic price does not vary with the quantity consumed). See Figure 1.4:
     - Person 1's MWT for an increase in the amenity g from g_1 to g_2 is $D_1(g_2) - D_1(g_1)$. However, the researcher only observes points a and b, and b strongly prefers the amenity so they sorted into an area with worse other characteristics but higher g. Measuring WTP then as $D_1(a) - D_2(b)$ would underestimate the WTP.

Chay and Greenstone (2005)

Chay and Greenstone (2005) use an IV strategy to estimate marginal willingness to pay (MWTP) for cleaner air. They use 1975 non-attainment status as an instrument for changes in TSP concentrations. Their main



A. Endogeneity due to taste-based sorting within a market

Figure 1.4: Second stage sorting complicates identification

findings are:

1. Conventional hedonic methods produce unreliable and misleading estimates.
2. Mid-1970s TSP non-attainment designations increased housing values in non-attainment areas.
3. Provide evidence consistent with taste-based self-selection (LATE may underestimate ATE).

However, some issues include:

- Pool all of US– law of one price?
- Median prices can induce bias; coarse measure of air quality.

Deryugina et al. (2020)

Use wind direction as an instrument for PM 2.5 concentrations to measure the effect of short-term PM exposure on mortality and medical costs. They have data on the universe of medicare beneficiaries, and a new ML approach to estimating the life-years lost due to air pollution exposure. They find:

1. 1 μg/m3 increase in PM2.5 exposure for one day causes 0.61 deaths per million elderly individuals over three day window.
2. Increases in PM2.5 leads to more ER visits, hospitalizations, in-patient spending.
3. Accounting for rich medical history reduces life years lost estimate by 55% relative to standard approach that conditions on age and gender.

They use deviations from prevailing wind to avoid sorting bias. They also pool estimates across monitors to get around measurement error, because county pollution monitor reading is a crude proxy for air quality throughout the county.

More detail in the paper box from Reed's half.

1.5 Lecture 5: Electricity markets (and DiD)

1.5.1 Electricity Supply

Some electricity market basics:

- Under the old model, vertically integrated regulated monopolies recovered operating costs and earn a rate of return on capital investment under rate of return regulation → weak incentives to become more efficient.
- Restructuring has taken place over the last 3 decades, creating Regional Transmission Organizations that enable price arbitrage (wholesale market).

- This separated generation (competitive) from distribution (still natural monopoly).
- We have seen the exercise of market power in wholesale electricity markets during high-demand hours. Long-term contracts between wholesale buyers and sellers mitigate the exercise or market power. Real-time pricing of electricity can also mitigate market power distortions.
- Restructuring led to fewer plant shutdowns, increase in nuclear generation.

Cicala (2021)

Evaluates the efficiency implications of moving to a regime in which a wholesale electricity market plays a larger role in coordinating dispatch. PCAs adopted market mechanisms in an abrupt and staggered way → DiD. He also must estimate counterfactual unit-level operations/market dispatch using a rich characterization of 'control' areas to control for changes that might otherwise bias a simple DID comparison, as PCAs trade between themselves and to avoid contamination from fuel price changes. He finds that:

1. 16% reduction in out of merit dispatch.
2. 55% gain from trade.

See more detail in Reed's half.

1.5.2 Electricity Demand

Electricity prices vary over space, time, and quantity due to block (step) pricing. This means that the marginal price you pay for electricity is endogenous!

- Using policy changes to instrument for electricity prices requires a parallel trends assumption, one that can be tricky given mean-reversion in HH electricity consumption. This mean reversion creates a negative correlation between current consumption and future consumption.

Ito (2014)

Ito (2014) studies whether HHs respond to the marginal or average price of electricity. He exploits the service boundary of an electric utility in Orange county to get variation in prices. He estimates:

$$\Delta \ln x_{it} = \beta_1 \Delta \ln MP_{it} + \beta_2 \Delta \ln AP_{it} + f_t(x_{itm}) + \gamma_{ct} + \delta_{bt} + u_{it}$$

where:

- Differences partial out HH-by-month FEs
- f_t controls for consumption percentiles → after this, the only variation left in prices is due to the service boundary.

He finds that:

- Consumers respond much more strongly to average pricing.
- In terms of welfare, this means that consumption increases under non-linear pricing since low consump. users have lower AC and high consump. users' AC doesn't change much.

1.6 Lecture 6: Cap and Trade

[Cap and trade theory omitted as this will be covered more thoroughly in Jim's section.]

Fowlie et al (2012)

They estimate the causal effect of a California-based emissions trading program on facility-level emissions vis a vis the command-and-control (CAC) regime it replaced, exploiting the fact that only a subset of industrial facilities in California were removed from CAC and required to participate in RECLAIM. They also investigate (crudely) how RECLAIM-induced changes in emissions are distributed across counties with different socio-economic characteristics.
RECLAIM: Caps NOx emissions from all point sources emitting more than 4 tons per year.

Empirical strategy: Semi-parametrically match treated and control facilities with similar pre-RECLAIM period emissions. They find that:

- NOx emissions fell by 20%
- No correlation of emissions changes with neighborhood SES.

1.6.1 Leakage

When markets are incomplete, leakage can occur through several channels:

1. SR shift of production and emissions to unregulated foreign firms.
2. LR relocation of firms to unregulated areas.
3. GE: If local demand for fuel is large, then the fall will decrease world carbon fuel prices, leading other unregulated areas to substitute towards these.
4. Negative leakage? Policy-induced reduction in green technology costs accelerates adoption elsewhere.

"Shuffling": CA imports clean energy from OR, who then backfills their energy supply with dirty imports from UT.

Lo Prete (2020)

To get around the SUTVA issue induced by shuffling, this paper uses matching to matches power plants in the western US with similar plants in other regions in order to assess the evidence on leakage/reshuffling. They find:

1. Evidence of massive leakage.
2. BUT: big parallel trends issue: The closure of of the SONGS nuclear plant in CA.

1.6.2 Equity

Hernandez Cortes and Meng (2021)

Investigate whether the pollution exposure gap between disadvantaged and other communities in California has widened or narrowed as a result of the GHG cap and trade program. Their strategy is:

1. Estimate a DiD approach where the treatment group are big emitters regulated under C&T $\rightarrow$ a single ATE.
2. use $\widehat{ATE}+$ plant FEs to generate heterogeneous C&T-driven abatement across regulated facilities.
3. Use HYSLPIT to measure changes in downwind concentrations.

They find that the EJ gap is narrowed under cap and trade. Questions remain about how different their treatment and control groups are. For example, refineries are in the treatment group but these had other policies affecting them in this time period such as the Low Carbon Fuel Standard (LCFS).

1.7 Lecture 7: Intro to discrete choice

The basic idea of discrete choice has the following characteristics:

- An agent (household/firm/consumer) makes a choice from among J possible actions or options.
- The dependent variable, Y, takes on non-negative, un-ordered integer values between zero and J.
- These choice sets must be **mutually exclusive**, **exhaustive**, and **finite**.
  - Note that these may not be as restrictive as they seem: choosing multiple options can be generalized by specifying categories as "Both A and C" for example. You can also add a "None" option to make the set exhaustive (Train, pg. 13).

- While discrete choice is generally conceptualized as answering "which", it can also answer "how much" if these can be binned discretely. For example, to predict how many cars one has we could say 0, 1, 2, or more than 2.

- Goal is to estimate/understand the conditional probability that agent n chooses option j conditional on observed choice characteristics X.
- Typically derived from an optimization problem, such as random utility maximization.

Definition 10. ***Nuisance parameters*** *are characteristics that the agent observes and cares about but that are unobserved to the econometrician.*

1.7.1 Comparing logit models

	Conditional logit	Mixed logit
$f(\varepsilon_{ni})$	iid EV1 $\forall i$, which implies that the unobserved factors are uncorrelated over alternatives, and have the same variance for all alternatives. Choices uncorrelated over time.	$\varepsilon_{ni} \sim$ iid EV, but $\beta_n \sim f(\beta\|\theta)$; the unobserved factors can be decomposed into a part that contains all the correlation and heteroskedasticity, and another part that is iid extreme value. The first part can follow any distribution, including non-normal distributions. We will show that mixed logit can approximate any discrete choice model and thus is fully general.
Example of distributional assumption failing	A person who dislikes travel by bus because of the presence of other riders might have a similar reaction to rail travel.	Fully general: "by specifying the explanatory variables and density appropriately, the researcher can represent any utility-maximizing behavior by a mixed logit model, as well as many forms of non-utility-maximizing behavior."
Evaluating choice probabilities	Closed-form solution	Numerically via simulation.

Table 1.3: Comparing logit models

1.7.2 Random utility maximization

Let U_{nj} be the utility that agent n receives from choosing option j. Then n chooses i if:

$$U^*_{ni} > U_{nj} \; \forall \; j \in J$$

U^*_{ni} is the individual's **latent utility** from their choice and is unobserved. However, we assume it can be expressed as:

$$U^*_{ni} = U(X_{ni}; \beta) + \varepsilon_{ni}$$

The vector of residuals ε_n includes all of the J choice specific disturbances associated with individual n. More interpretations of this error are (Train, pg. 17):

- The distribution of the unobserved portion of utility within the population of people who face the same observed portion of utility. Then, the probability P_{ni} is the share of people who choose alternative i within the population of people who face the same observed utility for each alternative as n.
- The effect of factors that are quixotic to the decision maker himself (representing, e.g., aspects of bounded rationality), so that P_{ni} is the probability that these quixotic factors induce the person to choose alternative i given the observed, non-quixotic factors.

Given this, we can write:

$$\begin{aligned} P(Y_n = i|X_n = x) = P(U^*_{ni} > U^*_{nj} \; \forall \; i \neq j) &= P(U(X_{ni}; \beta) + \varepsilon_{ni} > U(X_{nj}; \beta) + \varepsilon_{nj} \; \forall j \neq i) \\ &= P(\varepsilon_{nj} - \varepsilon_{ni} < U(X_{ni}; \beta) - U(X_{nj}; \beta) \; \forall j \neq i) \\ &= \int_{\varepsilon_n} 1[\varepsilon_{nj} - \varepsilon_{ni} < U(X_{ni}; \beta) - U(X_{nj}; \beta)] f(\varepsilon_n) d\varepsilon_n \end{aligned}$$

This can be computed if we know the distribution of the disturbances! To make this model empirically tractable, we need:

1. A plausible economic model of the underlying choice/optimization problem.
2. Assumptions about the statistical properties of the error distribution.

1.7.3 Conditional Logit (CL)

The conditional logit model puts the following structure on the errors:

$$U^*_{nj} = U(X_{nj}; \beta) + \varepsilon_{nj}$$
$$\varepsilon_{nj} \sim \text{iid } EV1$$

This distribution is attractive because the difference between 2 EV1 random variables has a logit distribution. Let $\varepsilon^* = \varepsilon_i - \varepsilon_j$. Then the density of ε^* is given by:

$$f(\varepsilon^*) = \frac{e^{-(x-\mu)/\sigma}}{\sigma(1 + e^{-(x-\mu)/\sigma})}$$

Plugging this distribution into the probability formula, and normalizing $\mu = 0, \sigma = 1$, we get that:

$$P(Y_n = i|X_n = x) = P_{ni} = \frac{\exp(-\beta' X_{ni})}{\sum_j \exp(-\beta' X_{nj})}$$

These probabilities have the characteristic that the marginal effect of the characteristics is largest in the middle, when one is less "decided", as in Figure 1.5.

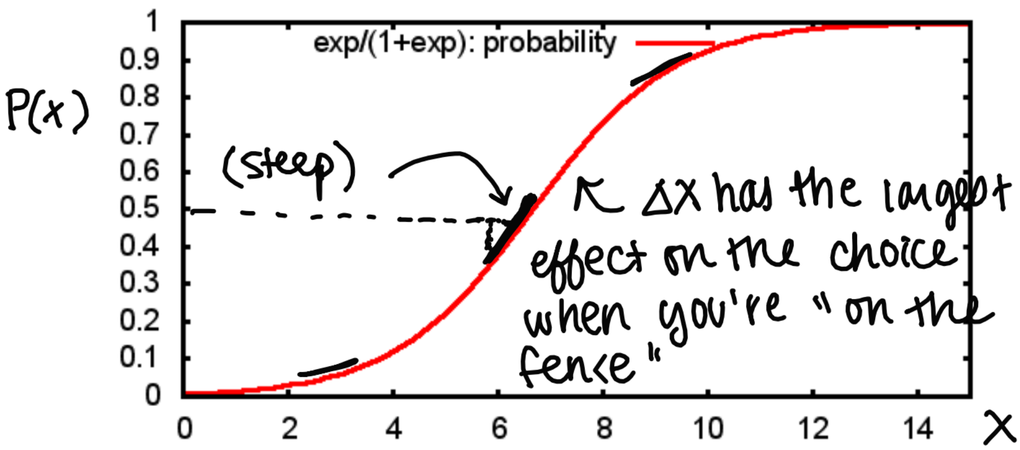


Figure 1.5: CL distribution

Estimating this object still requires more assumptions. We assume that:

1. All covariates are exogenous.
2. Our sample is either a cross section or a panel with zero serial correlation.
3. Each decisionmaker's choice is independent of the choices of other decision makers.
4. All choices are represented.
5. Disturbances are iid EV1.

Estimation

Because this is non-linear, it must be estimated using MLE.

$$P_n(y_{ni}) = \prod_i \left(\frac{\exp(-\beta' X_{ni})}{\sum_j \exp(-\beta' X_{nj})} \right)^{y_{ni}}$$

Then assuming independence across decision makers, we get:

$$
\begin{aligned}
L(\beta) &= \prod_n \prod_i \left(\frac{\exp(-\beta' X_{ni})}{\sum_j \exp(-\beta' X_{nj})} \right)^{y_{ni}} \\
LL(\beta) &= \sum_n \sum_i y_{ni} \ln(\frac{\exp(-\beta' X_{ni})}{\sum_j \exp(-\beta' X_{nj})}) \\
&= \sum_n \sum_i -y_{ni}(\beta' X_{ni}) - y_{ni} \ln(\sum_j \exp(-\beta' X_{nj}))
\end{aligned}
$$

Maximize by taking the derivative, solving.

What can you do with this?

Given MLE estimates of $\widehat{\beta}$, we can calculate a number of informative objects. These are treated in more detail in the section on Logit applications, but are summarized here to underscore the fact that these can be estimated for all logit–style models.

1. Willingness to pay $= -\beta/\alpha$, where α is the coefficient on the purchase price. This captures the amount you would need price to increase in order to leave you indifferent between increasing the amenity X and leaving it alone. For CL, there only one WTP number. In the Lucas paper, he estimates the "WTP to avoid an electricity mandate", which is more like CS, and not the WTP for a marginal change in some attribute.

2. Consumer surplus: $E[CS_n] = \frac{1}{\alpha_n} E(\max_j U_{nj} + \varepsilon_{nj}) = \frac{1}{\alpha_n} \ln \left(\sum_j \exp(U_{nj}) \right) + C$ when utility is linear in income; ie the marginal utility of income is independent of income OR constant over the implicit income changes implied by the policy.

Goodness of fit

We can assess goodness of fit by defining $s_n = 1(y_{nj} = i)$. Then our percent correctly predicted is:

$$
\frac{1}{N} \sum_n s_n
$$

Similarly, the likelihood ratio gives us the percent improvement over a no information model:

$$
LR = \frac{LL(0) - LL(\beta)}{LL(0)}
$$

Notes on Interpretation

1. Everything is relative

- Adding a constant to all utility terms does not change the relative choice.
- Therefore, you cannot directly estimate coefficients on attributes/characteristics that do not vary across choices (e.g. gender). However, you can make these vary across choices by setting the coefficient equal to 0 for some choice (so that the rest capture the relative effect of, say, income on the choice), or through interactions with choice-varying characteristics. For example, in Davis (2021), we makes household characteristics enter latent utility by normalizing their effect to be 0 in the case of gas:

$$
\begin{aligned}
u_{ie} &= \alpha_{0e} + \alpha_1 x_{ie} + \alpha_{2e} z_i + \epsilon_{ie} \\
u_{ig} &= \alpha_1 x_{ig} + \epsilon_{ig}
\end{aligned}
$$

- We can think about adding a constant to some of the terms: $U_{n1} = U(\beta' X_{n1}) + \varepsilon_{n1}$, $U_{n2} = \delta_2 + U(\beta' X_{n2}) + \varepsilon_{n2}$ (note that I cannot add deltas to both these as they would not be identified). Then δ_2 captures the average effect of unobserved attributes of choice 2 relative to the unobserved attributes of the omitted choice.
- Similarly, multiplying by some constant would not affect choices, so the βs themselves are only identified up to scale and we need to normalize. A common normalization is to divide everything by the unknown variance σ. This pins the variance of the EV1 errors to be $\pi/6$, identifying the model but leaving us estimating β/σ with no way to back out β.
- When aggregating, calculating predictions for the "average unit" may give seriously misleading estimates of marginal effects because of varying slopes. Therefore we are better off averaging over the predicted value of each decisionmaker. This average can we weighted/reweighted however one desires.

2. CL coefficients $\neq$ OLS coefficients

   - β measures the partial effect of an incremental change in the corresponding covariate x on the latent dependent variable scaled by the unknown/unidentified σ parameter.
   - β tells us how important X is relative to unobserved factors.
   - If the estimated β is large (small), this suggests that variation in costs explain much (little) of the choice variation relative to unobserved factors.
   - One option: calculate marginal effects:

   $$\frac{dP_{in}}{dx_{ik}} = \beta_k P_{ni}(1 - P_{ni}) \rightarrow \epsilon_{iik} = \beta_k(1 - P_{ni})x_{ni}$$

   or analyze how the choice of i changes with an attribute of j:

   $$\frac{dP_{in}}{dx_{jk}} = -P_{ni}P_{nj}\beta_k \rightarrow \epsilon_{ijk} = -P_{nj}\beta_k x_{jk}$$

3. Interactions are not what you think.

   - Interaction effects on choice probabilities depend on the values of all covariates, so estimated interactions effect can vary significantly across units in the data.
   - Interaction effects cannot be evaluated simply by looking at the sign, magnitude, or statistical significance of the coefficient on the interaction term when the model is nonlinear.

4. You cannot directly compare across groups.

   - It might be tempting to run CL for different sub groups and compare coefficients. But if the variance of disturbances σ is not homogeneous across groups, then these coefficients are not directly comparable!
   - A workaround: σ cancels when you take the ratio of two coefficients from the same model. Therefore if we have two groups A and B, we can compare the objects β_{1A}/β_{2A} and β_{1B}/β_{2B}.

The limitations of the CL assumptions

- **Preference heterogeneity:** CL accommodates systematic taste variation well, but not random taste variation.
  - That is, if we can formulate this variation as $\beta_{2n} = \alpha + \gamma SIZE_n$, we are good. If taste shocks are random across firms or correlated with unobservables, then we are not.
- **Panel data:** The CL model cannot accommodate correlation in the unobserved component of the latent utility/costs across choices made by the same agent, though it can allow for state-dependence through the inclusion of lagged variables. However, note that there still can't be serial correlation in errors.
- **IIA:** Wonkiness in this assumption takes a couple different forms.
  - Odds ratios are independent of other choices: $P_{in}/P_{jn} = \exp(\beta' X_{in})/\exp(\beta'_{jn})$
  - ϵ_{ijk} is the same expression for all the i not equal to j! This implies some very unrealistic substitution patterns. For example, in cars, this would assume that given a change in the price of a Tesla, the same percentage market share is lost by both the Chevy volt and the Toyota Tundra... likely not.
  - This gets at a deeper point: With CL estimation, we are not uncovering substitution patterns in the data, the structure of the model is imposing them!

* Mixed logit will relax many of these!
  - Models/accounts for random taste variation/heterogeneity.
  - Can accommodate correlation in unobserved factors across choices/time.
  - Substitution patterns uncovered versus imposed.
  - No more IIA

1.8 Lecture 8: Mixed Logit

1.8.1 The Berry Transformation

The Berry transformation starts from the Conditional Logit world. Let $\delta_j = X_j\beta + \alpha p_j$, and let utility from an outside option $J+1$ be 0 ($e^) = 1$). Then latent utility is given by:

$$U_{njt} = \delta_{jt} + \varepsilon_{njt}$$

and, given the CL assumption on disturbances:

$$Pr(U_{nit} \geq U_{njt}\ \forall j) = \frac{\exp(\delta_{it})}{1 + \sum_j \exp(\delta_{jt})}$$

Then the Berry transformation relies on the fact that the market share in a region of product j should be its choice probability:

$$\widehat{s_{it}}(\delta_{it}) = \frac{\exp(\delta_{it})}{1 + \sum_j \exp(\delta_{jt})}$$

Taking logs:

$$\log \widehat{s_{it}}(\delta_{it}) = \delta_{it} - \log\left(1 + \sum_j \exp(\delta_{jt})\right)\ \forall\, i = 0, \dots J$$

Taking differences with the outside option gives:

$$\log \widehat{s_{it}} - \log \widehat{s_{0t}} = \delta_{it} = X_{jt}\beta + \alpha p_{jt}$$

This equation is estimable with OLS! Our same CL concerns still persist, but the linearity allows us to apply IV methods and we can use aggregate shares instead of disaggregated data. For example, IIA still holds:

$$\frac{ds_{it}}{dp_{jt}} = \alpha s_i s_j$$

1.8.2 Mixed Logit

With mixed logit, our model becomes:

$$\begin{aligned} U_{ni} &= \beta_n X'_{ni} + \varepsilon_{ni} \\ \varepsilon_{ni} &\sim \text{iid } EV \\ \beta_n &\sim f(\beta|\theta) \end{aligned}$$

The difference is that β can vary randomly across agents. This reflects the fact that different decision-makers have different tastes/preferences. We no longer impose that taste parameters are fixed across agents.

- The $f(\theta)$ describes the density of these taste coefficients, where θ is a vector containing the parameters of the distribution of taste parameters.
- This is sometimes called the "mixing" distribution as it defines the weights in this mix of alternative logit functions.

The probability *conditional on a draw of* β_n is given by:

$$L_{ni}(\beta_n) = \frac{\exp(\beta'_n X_{ni})}{\sum_j \exp(\beta'_n X_{nj})}$$

In order to solve for the unconditional probability (that doesn't depend on the β_n because these are unknown), we need to integrate over the density of β:

$$P_{ni} = \int \frac{\exp(\beta' X_{ni})}{\sum_j \exp(\beta' X_{nj})} f(\beta) d\beta$$

This expression highlights the fact that mixed logit can be seen as a weighted average of the logit probability formula evaluated at different βs, where the weights are given by the density $f(\beta|\theta)$. Additionally, the β are integrated out, so that the only parameters to be estimated are in θ.

Estimation

To solve, integrate numerically via simulation using an assumed distribution of the βs:

1. Draw R times from the assumed distribution $f(beta_n|\theta)$.
2. Construct R values of the conditional choice probability evaluated using these R values.
3. Take an average. This gives you your simulated choice probability:

$$P_{ni}^R = \frac{1}{R}\sum_r \frac{\exp(\beta'_{nr}X_{ni})}{\sum_j \exp(\beta'_{nr}X_{nj})}$$

4. Insert these simulated probabilities into your simulated log-likelihood function:

$$SLL(\theta) = \sum_n \sum_j y_{nj}\log(P_{nj}^R)$$

5. The ML estimates of θ are those that maximize $SLL(\theta)$.

The distributional assumption depends on context; could be normal, or maybe for prices log normal makes more sense.

Panel data and ML

With panel data, let the choices $y_n = (y_{n1}, \ldots, y_{nT})$ be the sequence of choices in time. Then the probability of observing y_n is:

$$Prob(y|\beta) = \prod_t \frac{\exp(\beta'_n X_{nit})}{\sum_j \exp(\beta'_{nr}X_{njt})}$$

This ends up incorporating serial correlation in an individual's choices made by the same agent:

$$U_{ni} = \beta_n X_{ni} + \varepsilon_{ni} = (b + s\eta_n)X_{ni} + \varepsilon_{ni} = bX_{ni} + e_{ni}$$

where $s = cov(e_{nit}, e_{nit-1})$.

Substitution patterns

- With heterogeneous preferences, the agent who chose a more fuel efficient car has a β vector that weights fuel efficiency more heavily than average. In other words, error components are correlated across choices with similar choice attributes:
$$U_{ni} = \beta_n X_{ni} + \varepsilon_{ni} = (b + v_n)X_{ni} + \varepsilon_{ni} = bX_{ni} + e_{ni}$$
- This generates stronger substitution between more similar choices.
- The ratio of choice probabilities now depends on all of the data, including attributes of all choice alternatives.

The individual– specific parameters

Let $h(\beta|y, X, \theta)$ be the distribution of consumers who, given a choice environment characterized by X, would choose y. Applying Bayes' rule, we know that:

$$h(\beta|X_n, y, \theta)P(y|\theta, X_n) = P(y|\beta, X_n)f(\beta|X_n, \theta)$$

$$h(\beta|X_n, y, \theta) = \frac{P(y|\beta, X_n)f(\beta|X_n, \theta)}{P(y|\theta, X_n)} = \frac{P(y|\beta, X_n)f(\beta|X_n, \theta)}{\int P(y_n|X_n, \beta)f(\beta|\theta)d\beta}$$

Intuitively– move through the distribution of β in the population and weight each value by the probability that this agent would have made the choices he made had he had that β. Each coefficient β_k will have its own parameterization, so we can back out the distribution of the β_{nk} for each k and speak about, for example, the share of individuals who have a positive valuation for an attribute (See the example at the end of Chapter 6 in Ken Train's book).

1.9 Lecture 9: Logit applications

1.9.1 Estimating WTP from logit models

Supposing quasilinear utility, the consumer's problem is:

$$\max_j u_j + y - p_j$$

If latent utility takes the form:

$$u_j = \beta X_j - \alpha p_j$$

Then:

$$du_j = \beta dX_j + \alpha dp_j = 0 \Rightarrow \frac{dp_j}{dX_j} = \frac{\beta}{\alpha} = WTP$$

The above is for the CL case, since β is known. For mixed logit, we would need to simulate the β_n and report a distribution of WTP.

1.9.2 Estimating CS from logit models

Consumer surplus is the utility (monetized) that a person receives in a choice situation. Therefore, if we assume that a consumer's marginal utility of income α_n is constant over the price or choice set change, then we can translate utility into CS:

$$CS_n = \frac{1}{\alpha_n} \max_j U^*_{nj}$$

This assumption is reasonable if changing the choice set does not dramatically change your purchasing power. A price or cost variable enters the representative (indirect) utility in a consistent linear additive fashion. The negative of its coefficient is the marginal utility of income by definition.

Given the conditions:

1. Utility is linear in income.
2. $\varepsilon_{ni} \sim EV1$

then consumer surplus has a closed-form solution:

$$CS_n = \frac{1}{\alpha_n} \ln\left(\sum_j \exp(\beta_n' X_{nj})\right) + C$$

C is not identified because neither is the level of utility. But we can still evaluate changes! Then total consumer surplus in the population can be calculated as a weighted sum of logsums over a sample of decision-makers with the weights reflecting the number of people in the population who face the same representative utilities as the sampled person. In conditional logit, we know β so we can calculate CS_n for each n or for an alternative distribution of Xs, population weights, etc. In multinomial logit, we would need to simulate draws of β_n.

Davis (2021)

Lucas wants to analyze this discrete choice of heating fuel using household-level data. He first estimates a linear probability model which helps him identify key determinants of this choice. He then uses a conditional logit RUM framework to assess the welfare implications of an electrification mandate for new homes. He finds:

1. The key driver of HH heating electrification are energy prices.
2. Households in cold states prefer natural gas, and would be made worse off by \$2000 (or more) annually if forced to use electric.

Burgess et al. (2020)

Study household demand for electrification in Bihar, India using a revealed-preference measure of the value of electricity. They estimate how households value both grid and off-grid alternatives within a single demand system (so as to study substitution between sources), and use experimental variation in electricity price to address endogeneity concerns. The large-scale experiment randomly varied the offered tariff for solar microgrids across villages.

Experimentally estimated demand for this one energy source provides an incomplete welfare picture because there are other close substitutes available. The elasticity of demand for microgrids cannot be used to measure the STP for electricity access. When microgrid price rises, households may substitute to other options. Critiques of this paper:

1. Welfare analysis assumes estimated IV price coefficient estimates equally applicable to all energy supply alternatives. There are many reasons to think that an incremental change in the microgrid price will be perceived differently as compared to an incremental change in the village monthly average cost of grid connections or PV systems. LATE $\neq$ ATE.

2. CL not ML; some strong assumptions, especially considering it is a panel.

1.10 Lecture 10: BLP

Consider a market with J differentiated products. If we were to try and estimate the demand system, we would need to estimate a large space of $J \times J$ parameters. The solution: project these products onto a finite product attribute space. This method was developed by Berry, Levinsohn, and Pakes, and has 3 important groups of innovations:

1. Estimating models of consumer-level demand using aggregate (i.e. market level) versus disaggregated (i.e. household or customer level) data.

2. Demonstrating a tractable means of dealing with endogenous covariates (this is an extension of the Berry transformation.

3. Nesting the random utility model within a larger structural model of imperfect competition in order to obtain consistent estimation of underlying structural parameters.

We will need the following assumptions:

1. Firms sell directly to consumers.

2. Firms do not price discriminate. Consumers know all prices and attributes.

3. There are no dynamic considerations for either firms or consumers.

4. All non-price product attributes are assumed to be exogenous.

5. Consumers purchase at most one product per household

When defining a market, we should keep conceptualize product attributes (including prices) as varying across markets, but not across consumers within a given market. As before, the choice set contains an outside good, which allows us to use these models to study aggregate demand because we do not condition on purchasing a new car.

The base of BLP is Mixed logit. We define utility as:

$$u_{njt} = \alpha \ln(y_n - p_{jt}) + \sum_k x_{jtk}\beta_{nk} + \xi_{jt} + \varepsilon_{njt}$$

where t indexes markets and j indexes goods. Income enters as $\alpha \ln(y_n - p_{jt})$ because utility is modelled as Cobb–Douglass in BLP. In the quasi-linear case, it would be $\alpha(y_n - p_{jt})$. In both cases, α is the marginal utility of income.

- ξ_{jt} is the structural error term, and captures average valuation of attributes and quality that are observed by consumers but are not observable to the researcher (and thus not explicitly represented in the model).

- BLP assume that these are 'mean independent' of the observed attributes. That is, if your price coefficient is interacted with demographics and you think prices are endogenous then interaction terms will also be endogenous if you think the structural error term is product and demographic specific.

- Logit enters through ε_{njt}; Having included the ξ_{jt} to control the average effect of unobserved attributes, and allowing the preference coefficients to vary across consumers, there is presumably little variation left to be captured by this residual. This error term could capture measurement error, optimization error, etc.

- It is a standard normalization to set $\beta_{i0} = \xi_{0t} = 0$. The $\alpha \ln(y_n)$ drops out because it is common to all choices.

1.10.1 Preference heterogeneity

- In the context of demand models for differentiated products, it's important to accommodate preference heterogeneity (if preferences were homogeneous, why are product offerings differentiated?).
- Allowing preference parameters to vary systematically with observables reduces our reliance on the (often arbitrary) parametric assumptions we impose on the unobservable random components.

We can decompose these preferences in the following way:

$$\begin{aligned} \alpha_n &= \alpha + \nu_{n\alpha} \\ \beta_{nk} &= \beta_k + \sum_r \mu_{kr} d_{nr} + \nu_{nk} \\ \nu_n | d_n &\sim N(0, \Sigma) \\ d_n &\sim f(d_n) \end{aligned}$$

where d_n is a vector of consumer demographics, typically normalized to 0 so that the β_k can be interpreted as averages. If we only have market-level data, then we do not observe the d_n and must make assumptions about the market-level distribution of these, $f(d_r)$. The ν_{ik} capture the effects of unobserved preference variation that we assume is randomly distributed in the population. Thus:

$$u_{njt} = \underbrace{\alpha p_{jt} + X_{jt}\beta + \xi_{jt}}_{\delta_{jt}, \text{Mean utility in } jt} + \underbrace{\left(\sum_r \mu_{kr} d_{nr} + \nu_{nk} \right)' X_{jt} - (\mu_\alpha + \nu_{i\alpha}) \frac{p_{jt}}{y_n} + \varepsilon_{njt}}_{v_{njt}, \text{Person-specific departure from mean, observ. and unobserv.}}$$

Now we divide the parameters into a linear set $\theta_1 = \{\alpha, \beta\}$ and a non-linear set $\theta_2 = \{\mu, \Sigma\}$. Once you have made your distributional assumptions about how the ν_i vary in the population, you can write down the expression for the unconditional choice probabilities:

$$P_{nit} = \int \frac{\exp(\delta_{it} + v_{nit})}{1 + \sum_j \exp(\delta_{jt} + v_{njt})} f(\theta_2)$$

From here, you use MLE or GMM in an iterative process in which the goal is to pin down the distributional parameters θ_2. In each iteration of θ_2 guesses, a fixed point algorithm works to pin down the δ that make observed market shares match estimated shares. This will give a value of the GMM objective function for that θ_2, and searches over these θ_2 for the minimizer.

1.10.2 Using market-level data

With market level data, identification comes from variation across markets. We additionally need to assume that the distribution of consumers' underlying tastes, conditional on an observed distribution of consumer characteristics, is invariant across markets.

To get some intuition, note that we can return to CL by setting $\nu = \Sigma = 0$ and inverting market shares as before. Estimation proceeds as follows:

1. Collect data on product attributes X and market shares y.
2. Specify latent utility function (which includes parametric assumptions about how random taste parameters are distributed in the population.
3. Simulate drawing from these assumed distributions (generate a set of R draws) using an initial guess for the non-liner parameters.
4. Estimate the implied market shares given these draws and an initial guess of the mean utilities δ (and an initial guess of the non-linear parameters θ_2)
5. Calibrate the simulated log likelihood function (SLL) or moment conditions and objective function (GMM).
6. Use numerical optimization algorithms to identify the estimates of non-linear parameters and the δ parameters.
7. Once converged, step outside and use estimated δ to recover the linear parameters.

1.10.3 Endogenous regressors

We can take advantage of the fact that the product-specific constants δ_{jt} are linear in the structural error to smuggle the endogeneity problem into a linear setting where the endogeneity issue can be addressed using standard methods (IV).

$$\delta_{jt} = X_{jt}\beta + \alpha p_{jt} + \xi_{jt}$$

But what are these δs? Within a simple conditional logit framework, the δ_{jt} can be estimated as $\ln(S_{jt}) - \ln(S_{0t})$. Outside of CL, this is more complicated. The key intuition, building on Berry (1994), is that for any value of θ_2, there exists a unique δ vector such that the predicted shares equal actual (observed) shares. Estimation therefore involves finding the δ that matches observed market shares given assumed θ_2, and BLP demonstrate an iterative contraction mapping procedure that finds the vector δ that equates the observed and predicted market shares for a given θ_2. The steps for this are:

1. Pick an initial guess of δ.
2. Conditional on this δ vector and assumed set of θ_2 parameters, compute (via simulation) the estimated shares.
3. Using these predicted shares, they then update the δ vector using:

$$\delta_j^n = \delta_j^{n-1} + \ln\left(\frac{s_j}{\widehat{s}_j(\delta_j^{n-1};\theta_2)}\right)$$

4. Recompute the market shares with this updated δ and repeat until the δ converges to the solution.

The empirical solution yields an empirical approximation to the non-linear market share functions.

1.10.4 Adding more structure via the supply side

BLP assume firms engage in Bertrand price competition. Because products are differentiated, Bertrand competition does not imply marginal cost pricing. The key assumption is that product lines (and non-price attributes) are fixed. Let Ω be the matrix of derivatives $\frac{dq_i}{dp_j}$ and $s(p)$ be the vector of market shares. Then the first order conditions give:

$$p - mc = \Omega(p)^{-1}s(p)$$

Adding in the supply side opens up a class of supply side instruments, such as:

- Own-product non-price attributes
- Sum of non-price attributes for other products offered by firm.
- Sum of non-price characteristics for products offered by rivals.

though note that non-price attributes can also be endogenous!

1.11 Lecture 11: BLP applications

Grigolon et al (2018)

Research question To what extent do car buyers undervalue future fuel costs, and what does this apply for the effectiveness and welfare of alternative policies?

Approach Use detailed data from a long panel of European countries and exploit variation in fuel costs by engine type. Their utility specification is:

$$u_{ijk} = x_{jk}\beta - \alpha_i(p_{jk} + \gamma G_{ijk}) + \xi_{jk} + \varepsilon_{ijk}$$

$$G_{ijk} = E[\sum_s (1+r)^{-s}\beta_i^m e_{jk}g_{ks}$$

Annual mileage β_i^m varies across consumers; estimated using the country-specific empirical distribution of mileage.
Contribution Underscore the importance of accounting for consumer mileage heterogeneity. Focus on mileage variation (across customers) and fuel cost variation (across engines).

Punchline Despite modest undervaluation of fuel costs, fuel taxes better target high mileage customers.

- When mileage heterogeneity is accounted for, the fuel tax is more effective than the vehicle tax at reducing total fuel usage because it targets high mileage customers.
- Using the RCL model, the authors show that the fuel tax welfare dominate an emissions equivalent vehicle tax (because a lower fuel tax required to achieve the same emissions reduction).

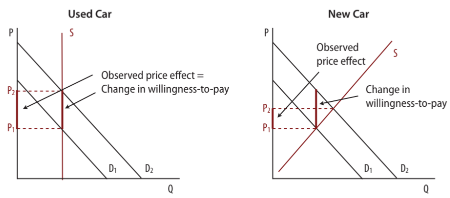
Busse, Knittel and Zettelmeyer (2013)

Use data on new and used car transactions to estimate a reduced form model of the short-run equilibrium effects of changes in gasoline prices on the transaction prices, market shares, and unit sales of cars of different fuel economics. Their main specification is:

$$P_{ijrt} = \lambda_0 + \lambda_1(G_{it} \times MPGQuartile_j) + \lambda_2 DEM_{it} + \delta_j + \tau_{rt} + \eta_{rt} + \varepsilon_{ijrt}$$

Having included region-year and region-month-of-year fixed effects, they argue they do not need any supply-side cost shifters. The effect varies with efficiency because the relative fuel cost advantage of an efficient MPG car rises with the price of gas

Figure 4: Effects of gasoline price change on hypothetical used and new cars



In the new car market, car dealers can to respond to changes in demand by altering prices, quantities, or both. The equilibrium price effect will be less than the change in the willingness-to-pay.
Authors assume a fixed supply curve in the used vehicle market (because the stock of used cars is predetermined). If true, the change in equilibrium prices is driven by the demand effect.

Overall they find:

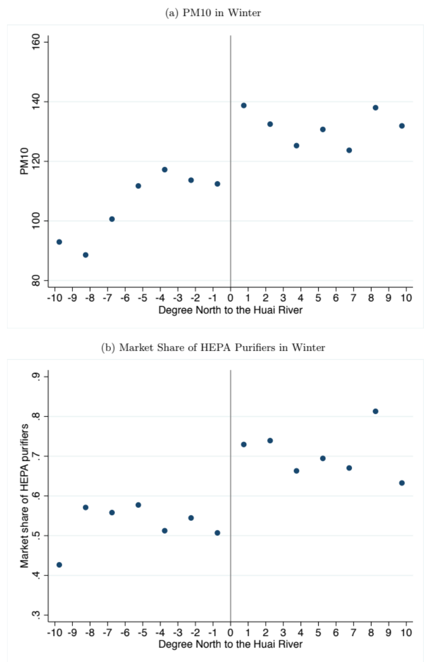
- SR changes in prices increase the price of the most fuel efficient cars, and decrease the price of the least efficient.
- Implicit discount rates are not large.

Ito and Zhang (2019)

Analyze WTP for defensive investments in air purifiers in China. Provide the first (?) revealed preference estimates of WTP for clean air in developing countries with an empirical framework based on a random utility model.

They exploit a discontinuity in heating policy along China's Huai River to isolate quasi-random variation in air quality and examine purchases of HEPA air filters north and south of river.

Figure 2: Regression Discontinuity Design at the Huai River Boundary



Conditional Logit
Applying the Berry transformation in the CL model, we can estimate the equation:

$$\ln s_{jc} = \beta x_c H_j + \alpha p_{jc} + \eta_j + \lambda_c + \xi_{jc}$$

where robust FE absorb observed/unobserved non-price attributes, City FE absorb city-level demand shocks, and x is instrumented with the Huai river RD, p is instrumented with distance to the producing plant.

Mixed Logit
They implement mixed logit with non–linear IV via GMM:

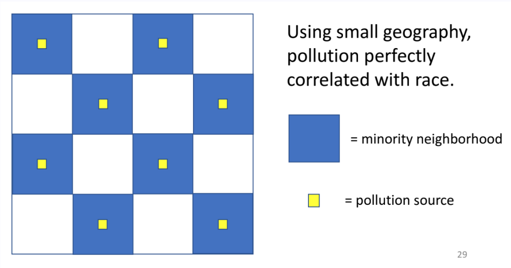
$$\widehat{\theta} = \text{argmin } \omega_{jc}(\theta)' Z_{jc} \Phi^{-1} Z'_{jc} \omega_{jc}(\theta)$$

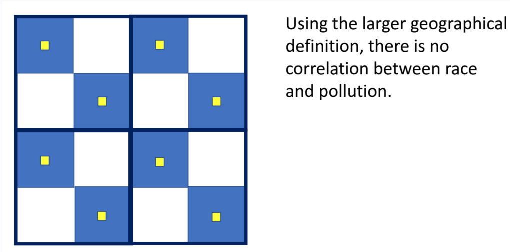
Note how the $Z'\omega(\theta)$ incorporates the IV moment condition, since $\omega(\theta) = \delta_{jc}(\theta) - (\beta_0 x_{jc} + \alpha_0 p_{jc} + \eta_j + \lambda_c)$ is the linear error term, and the θ containing the distribution parameters for α_n and β_n which are assumed to be log-normal. I think the second stage could be estimated via regular IV given the estimate of the δ, but this would almost be guaranteed to be inefficient (not homoskedastic). They find:

1. Estimates imply that a northern household is willing to pay USD \$32.70 per year to avoid the pollution increases induced by the Huai River policy.
2. Benefit/cost ratio of reforming heating is > 40.
3. These are likely underestimates: imperfect information, the omission of avoidance behavior and electricity costs, and the fact that they assume indoor pollution = outdoor pollution when in reality it is lower all understate WTP.

Ecological fallacy

1. When measuring the correlation between pollution and demographics, the ‘ecological fallacy’ can arise when inferring relationships between individual units (like households) from larger, more aggregated units (like counties) that contain those units.
2. Relationship estimated using aggregate data is only equal to correlations at the micro-level if there are no correlated group-level effects.
3. If peer preferences create segregation, positive correlations between race and exposure may be significant at local spatial scales but not at more aggregated scales.





Depro et al. (2015)

Argues that traditional DID empirical models are not actually identified. Without additional structure, individual sorting behavior cannot be identified using aggregate changes in population flows. This paper questions the strength of the prior evidence arguing that siting (versus residential mobility) explains inequalities in exposure.

The idea: Δpop_{it} does not identify the elements of a population dynamics matrix.

Example 1:

$$\begin{pmatrix} 0.00 & 0.60 & 0.60 \\ 0.50 & 0.25 & 0.20 \\ 0.50 & 0.15 & 0.20 \end{pmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 2.2 \\ 2.0 \end{bmatrix} \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.5 \\ 1.0 \end{bmatrix}.$$

Slope coefficient = −0.0583 [0.666] (coming to the nuisance).

Example 2:

$$\begin{pmatrix} 0.50 & 0.10 & 0.10 \\ 0.30 & 0.50 & 0.30 \\ 0.20 & 0.40 & 0.60 \end{pmatrix} \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 2.2 \\ 2.0 \end{bmatrix} \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.5 \\ 1.0 \end{bmatrix}.$$

Slope coefficient = 0.0833 [0.276] (fleeing the nuisance).

Empirically, they use BLP to estimate a model separately for each ethnic group. For each group, their model is:

$$s_{j,k} = \frac{\exp(\delta_j - \delta_k - \mu MC_{j,k})}{\sum_l \exp(\delta_l - \delta_k + \mu MC_{l,k})}$$

So for a given μ, there is a contraction mapping that makes the delta's perfectly match the shares. Then, μ is updated so that the share of stayers matches:

$$\%stay = \frac{\sum_k s_{k,k} pop_k^{2000}}{\sum_k pop_k^{2010}}$$

Their "outside loop" seems to be looping over values of μ, but this is not person–specific. Help?
Find:

1. Economically and statistically significant differences in Hispanic and white MWTP to avoid an incremental increase in cancer risk.
2. Relative to whites, Hispanics residential location choices reveal a smaller WTP for risk reduction.
3. These results are more consistent with (but not necessarily proof of) a residential mobility explanation for observed race and pollution correlations.

BUT... can we compare the marginal utility of income/ WTP between different socioeconomic groups?? WTP is a flawed way to compare valuations between people at different parts of the distribution.

Steering

Evidence from Christensen and Timmins (2019) suggests that realtors steer minority households to neighborhoods with higher concentrations of toxic contamination and pollution than their white counterparts. This implies:

- utility maximization assumptions underlying hedonic and residential sorting models may not hold, as
- different consumers have different choice sets

Part 2

Joe Shapiro

2.1 Lecture 1: Public Finance

2.1.1 Sufficient statistics

In some cases, model primitives are not necessary in order to make statements about welfare/ changes in welfare. This is where sufficient statistics enter: they are theoretically derived from a model, but can be estimated with data. This is a powerful tool, but it has its limitations:

1. Assumes away existing distortions.
2. Good for marginal, not discrete changes in policy.
3. New problem (model), new statistic.
4. Not always possible without making strong simplifications.

Harberger triangles example

For example, consider the welfare loss associated with distortionary taxation. The decentralized problem is given by:

$$\begin{aligned} \text{Consumer's problem: } & \max_{X,y} U(X_{J\times 1}) + y \qquad s.t.\, px + tx_1 \leq y \\ \text{Firm's problem: } & \max_{X} px - c(x) \\ \text{Market clearing: } & x^D(p) = x^S(p) \end{aligned}$$

whereas the planner's problem is:

$$\{\max_X U(X) + Z - tx_1 - c(X)\} + tx_1$$

In a fully structural approach, we need information on U, c. However, taking the first order condition of the planners problem with respect to the tax shows that:

$$\frac{dU}{dt} = -t\frac{\partial x_1}{\partial t} - x_1 + x_1 = -t\frac{\partial x_1}{\partial t}$$

For a discrete change in taxes:

$$\Delta W = \int_{t_1}^{t_2} = t\frac{\partial x_1}{\partial t}(t)dt$$

We can ignore the partials of other goods with respect to x_1 because of envelope conditions. Thus the change in consumption of x_1 is a sufficient statistic for the welfare change!

A more general setup

The agent maximizes utility under taxation and transfers, subject to M constraints $G_m(x, t, T)$. The Lagrangian is:

$$W = \max_x U(x) + \sum_m \lambda_m G_m(x, t, T)$$

Applying envelope conditions, the first order condition is:

$$\frac{dW}{dt} = \left(\sum_j \frac{\partial u}{\partial x_j}\frac{\partial x_j}{\partial t} + \sum_m \sum_j \lambda_m \frac{\partial G_m}{\partial x_j}\frac{\partial x_j}{\partial t} \right) + \sum_m \lambda_m \left(\frac{\partial G_m}{\partial T}\frac{\partial T}{\partial t} + \frac{\partial G_m}{\partial t} \right)$$
$$= \sum_m \lambda_m \left(\frac{\partial G_m}{\partial T}\frac{\partial T}{\partial t} + \frac{\partial G_m}{\partial t} \right)$$

We know the multipliers λ_m are the marginal utilities, which further simplifies this.

2.2 Lecture 2: Health demand

Health is an important channel of environmental benefits, and with some manipulations can be transformed into welfare.

Deschenes, Greenstone and Shapiro (2018)

Idea: Looking at health outcomes alone understates the health costs of air pollution. Full adaptation → no health impacts. They study the effect of the NOx Budget Trading Program (NBP) on pollution, medication purposes, hospital visits and mortality. Theoretically, this is conceptualized as a stock of negative health:

$$S(C, A, M) = S(D, M) = S(D(C, A), M)$$

The consumer solves:

$$\max_M S(C, M) \qquad s.t. \qquad I + p_W(T - L - S) = X + p_M M$$

Marginal willingness to pay for lower pollution concentrations is given by $WTP = p_w \frac{\partial s}{\partial c}$. This captures the effect of lower concentrations on sick days, holding abatement fixed. However, most studies are only able to identify the total derivative $\frac{ds}{dc}$, as mitigation and avoidance adjusts endogenously. Further simplifying, we get:

$$WTP = \underbrace{p_w \frac{ds}{dc}}_{\text{Econ. cost of sickness}} + \underbrace{p_a \frac{\partial a^*}{\partial c}}_{\text{Cost of mitigation}} - \underbrace{\frac{\partial U}{\partial s}\frac{ds}{dc}\frac{1}{\lambda}}_{\text{Monetized disutility of sickness/death}}$$

Thus the goal of this study is to identify the terms in this expression. Economic cost of sickness comes from hospital expenditures, mitigation costs from medication, and the monetized disutility of death from the VSL+mortality. Complication: medications market not perfectly competitive ⇒ price > MC; patients payment ≠ price.

- Short run: maybe drug expenditures are just a transfer from consumers to firms. But in the long run, if there were less disease burden these firms wouldn't even exist and these expenditures could be spent on other things, so this is social cost.

Empirical strategy: DDD using time, space, season to estimate the effect of NBP on outcomes:

$$y_{cst} = \gamma_1 NBP_c \times SUMMER_s \times POST_t + \eta_{ct} + \nu_{ts} + \mu_{st} + W_{cst}\beta + \varepsilon_{cst}$$

Note that regulation is designed to address non–attainment in ozone (NOx+stuff becomes Ozone in atmosphere). They also run an IV specification with $X =$ NOx, Ozone. However, note that the exclusion restriction for Ozone is tricky, since Ozone is affected by other factors that may have been impacted by the treatment. They find:

1. NPP led to 35% reduction in NOx emissions.
2. 1.5% Reduction in medication purchases: saved $800 million in defensive investments annually
3. 0.5% Reduction in mortality rate – prevented 2,500 summertime deaths each year, primarily age 75+ population.
4. Find effects for cardiovascular and respiratory but also other non–accidental deaths.
5. No detectable effects on hospitalization.
6. Underestimates: other defensive investments, pay full price for meds, etc.

2.3 Lecture 3: Climate damages

Definition 11. ***Climate adaptation*** *is "Adjustment in natural or human systems in response to actual or expected climatic stimuli or their effects, which moderates harm or exploits beneficial opportunities" (IPCC)*

The relationship of outcomes to weather shocks, net average climate, tells you how much adaptation has happened. For example, a hot day in Dallas is less lethal than a hot day in Quebec because they are less adapted.

Barreca, Clay, Deschenes, Greenstone and Shapiro (2016)

Idea: "modifiers" such as healthcare, air conditioning and residential electricity may have modified the temperature–mortality relationship over the 20th century.

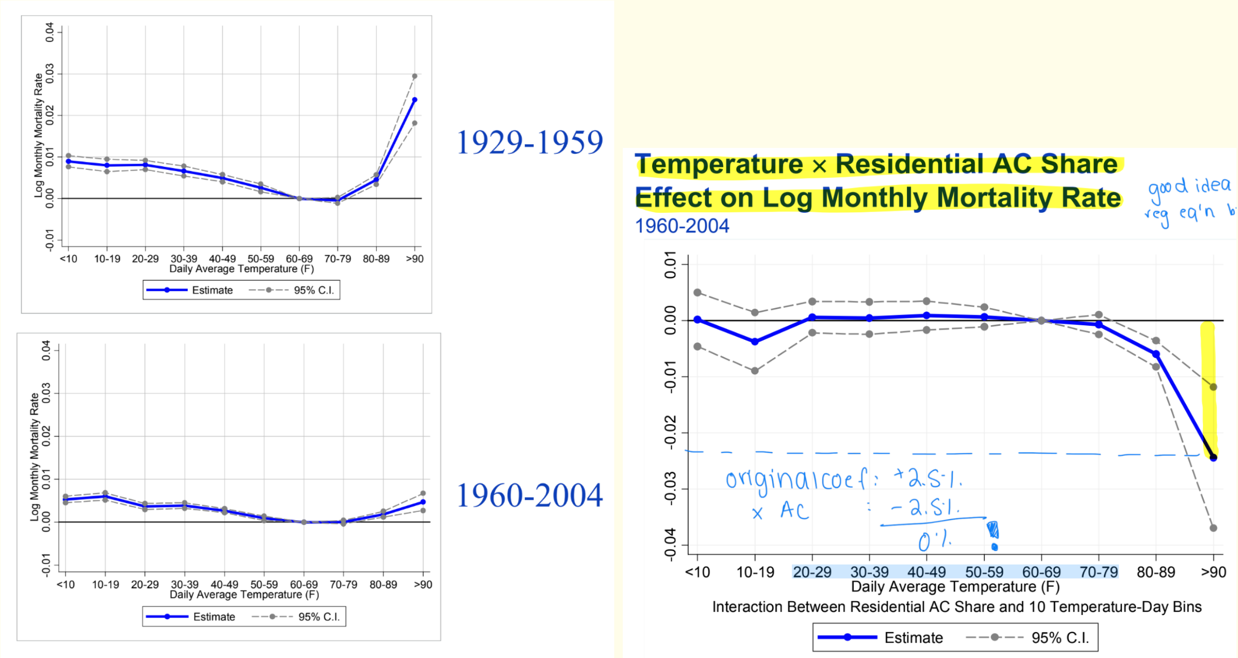
They compile annual mortality rate, state physicians per capita and share HHS with electricity, air conditioning for most of the 20th century, as well as daily temperature maximums. Their estimating equation is:

$$\log(Y_{sym}) = \sum_j \theta_j TMEAN_{symj} + \pi_L LOWP_{sym} + \pi_H HIGHP_{sym} + X_{sym}\beta + \alpha_{sm} + \gamma_{ym} + \varepsilon_{sym}$$

for j temperature bins. Data is at the state×year×month level. They also control for state–month trends. They find:

1. The temperature–mortality relationship flattened significantly over the 20th century: Impact of very hot days (> 90° F) on log mortality rate declined by factor of 5-6 after 1960.
2. This effect is driven by adoption of AC.
3. Protective effect of AC largest for youngest, oldest; blacks more than whites; significant for cardiovascular+respiratory disease, not infectious.

The AC paradox: AC protects us against climate change, but it is also causing it (↑ energy demand)



2.4 Lectures 4+5: Hedonics and local public goods

Hedonics is about measuring the value of *attributes* of goods.

2.4.1 Rosen's model (1974)

Consider a good that has n attributes. Let $z = (z_1, z_2, \dots, z_n)$ be the attribute vector, which maps to equilibrium prices through the hedonic price function $p(z)$. We assume that:

1. Exhaustive description of attributes
2. Consumers accurately perceive attributes

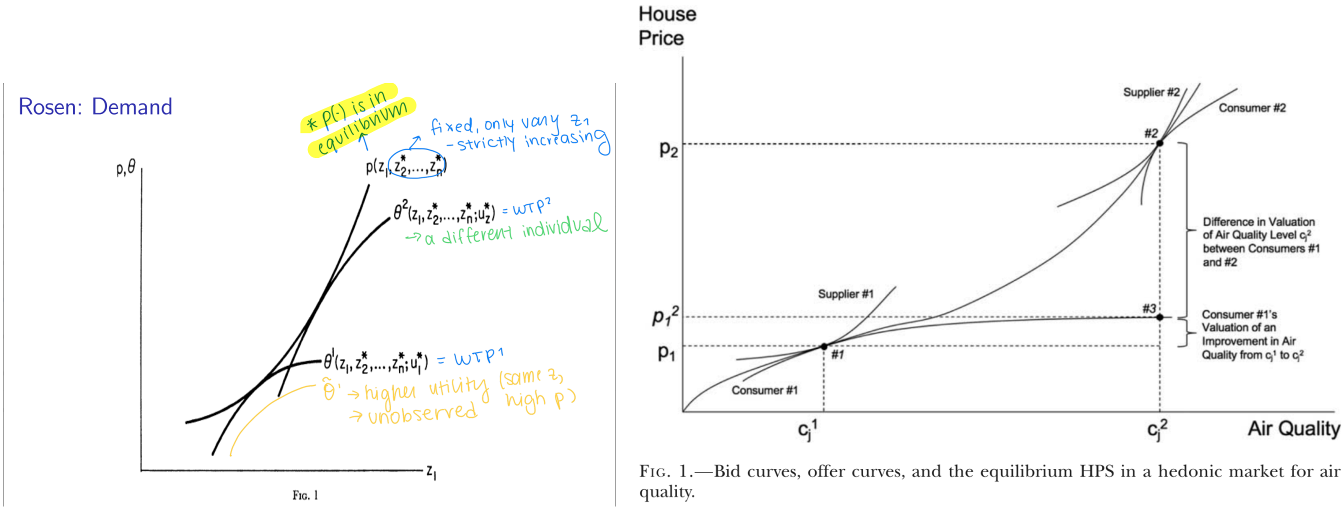


Figure 2.1: Demand in Rosen (1974)

3. z is a continuous space
4. Consumers purchase a single good
5. Prices are parametric (no market power for firms or consumers)

Demand

Demand is derived from consumer utility $U(x, z)$, where x is an outside composite good (numeraire). Utility maximization yields first order conditions:

$$\frac{\partial p}{\partial z_i} = p_i = \frac{\partial U}{\partial z_i} \Big/ \frac{\partial U}{\partial x}$$

Definition 12. *The* ***bid function*** $\theta(z; y, u)$ *solves:*

$$U(y - \theta, z) = u$$

and is the willingness to pay function as it represents the combinations of (y, z) *that maintain utility fixed.*

Deriving this with respect to an attribute z_i gives:

$$-\frac{\partial U}{\partial \theta}\frac{\partial \theta}{\partial z_i} + \frac{\partial U}{\partial z_i} = 0 \Rightarrow \frac{\partial \theta}{\partial z_i} = \frac{\partial U}{\partial z_i} \Big/ \frac{\partial U}{\partial \theta}$$
$$\Rightarrow \frac{\partial \theta}{\partial z_i} = p_i = WTP_{z_j}$$

This condition represents tangency of the bid functions with the price schedule, as depicted in Figure 2.1. $\frac{\partial \theta}{\partial z_i}$ is the **marginal WTP** for attribute z_i. Additionally, we require market clearing:

$$\theta(z; y, u) = p(z)$$

which imposes that the bid functions touch the price schedule. Note (as in the figure) that different consumers may have different bid functions. **In particular, I think these differences can only come from differences in preferences.**

Supply

Firms supplying the good produce $M(z)$ units of design z at cost $C(M; z, \beta)$. Plants choose M and z to maximize profit:

$$\max_{z,M} \pi = M(z)p(z) - C(M, z)$$

First order conditions give:

$$M\frac{\partial p(z)}{\partial z_i} = \frac{\partial C(M, z)}{\partial z_i}$$
$$p(z) = \frac{\partial C(M, z)}{\partial M}$$

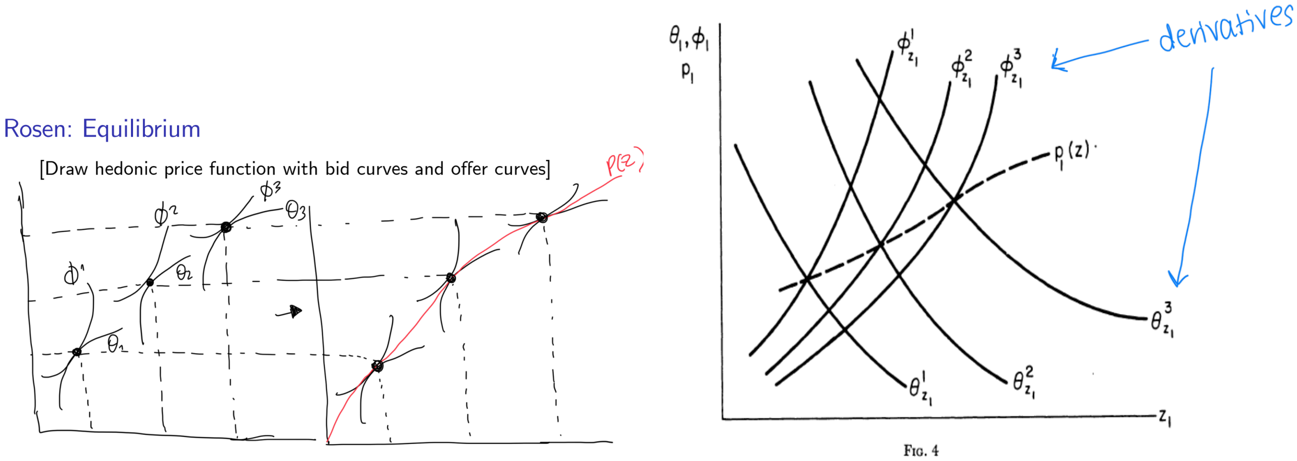


Figure 2.2: Equilibrium in the Rosen model

Definition 13. *The* ***offer function*** $\phi(z;\pi,\beta)$ *represents the unit prices the firm is willing to accept for a given z, so*

$$\phi(z;\pi,\beta) = \frac{\partial C(M,z)}{\partial M}$$

Equilibrium is given by tangency in the offer and bid functions. If all firms are identical (no variance in β), then $p(z) = \theta$. Similarly, if all consumers are the same, then $p(z) = \phi$

2.4.2 Estimation

Knowledge of the hedonic price function is not sufficient for making welfare statements when changes are non–marginal. This is because bid functions are unobserved, so we only see one point on each bid function for each consumer; other points come from other individuals who by definition have different preferences. For example, in Figure 2.1, Consumer 1's WTP for an improvement in air quality is $p_1^2 - p_1$. However, looking at the Hedonic price schedule, we would infer that it is $p_2 - p_1$, which is way larger. Other comments on non–marginal changes are:

- The stability of the price function itself can be an issue– if the demand/supply of the amenity change dramatically, the curve may shift!
- Greenstone and Gallagher (2008) [Superfund]: given no moving costs, movers have $\Delta U = 0$, but landowners near the improvement get a pure gain in the form of higher rents.

For marginal changes, the coefficient on the amenity level in a well–identified hedonic regression (step 1) gives MWTP.

Estimation proceeds in 2 steps:

1. Estimate the hedonic price function.
    - Regress prices p on characteristics z. Consistency relies on exogeneity of z_j.
    - Issues: OVB; consumers observe things the econometrician doesn't; misconception of amenities; discrete set not continuous.
2. Estimate bid and offer curves.
    - Define the implicit marginal price $p(z)_{z_i} = \widehat{p_i}(z)$. Then regress implicit prices on attributes, using taste and cost shifters as instruments.
    - This step has been largely infeasible!

In the normal case, there is a distribution of both buyers and sellers, so you have a system of equations to identify. Other cases:

- If no variance in taste shifters, then all consumers are identical and $p(z)$ reflects bid function.
- If no variance in cost shifters, then all producers are identical and $p(z)$ reflects the offer function.

Black (1999)

Applies regression discontinuity across walk zone boundaries to estimate how much people value school quality. Her estimating equation is:

$$\ln(pricE_{iab}) = \alpha + X'_{iab}\beta + K'_b\phi + \gamma test_a + \varepsilon_{iab}$$

where K_b are boundary dummies. Different specifications limit estimation to smaller distances from the boundary.
Findings:

1. Homes closer to the boundary are more comparable (Table 3).
2. 5% increase in elementary school test scores (1 s.d.) increases home prices by 2.1 percent, or $4,000 given mean home price of $188,000. This effect is smaller than previous estimates.

Sorting may confound these effects: do people value their neighbors, are schools?

Bayer, Ferreira and McMillan (2007)

This paper builds on Black, and embeds her boundary RD in a mixed logit (BLP). They study this at the census-tract level for size counties in SF. Their empirical set up is similar to Black (1999). The mixed logit specification is:

$$V_{ih} = \alpha_{Xi} X_h - \alpha_{ip} p_h - \alpha_{id} d_{ih} + \theta_{bh} + \xi_h + \varepsilon_{ih}$$

we can partition this into household–specific terms and terms common to all HHs:

$$V_{ih} = \delta_h + \lambda_{hi} + \varepsilon_{ih} \tag{2.1}$$

$$\delta_h = \alpha_{0X} X_h - \alpha_{0p} p_h + \theta_{bh} + \xi_h \tag{2.2}$$

$$\lambda_{ih} = \left(\sum_k \alpha_{kX} z_{ik} \right) X_h - \left(\sum_k \alpha_{kp} z_{ik} \right) p_h - \left(\sum_k \alpha_{kd} z_{ik} \right) d_h \tag{2.3}$$

Notes:

- Distance does not have a shared component since it is the distance to i's place of work.
- They say: "Consequently, in the presence of heterogeneous preferences, the mean indirect utility δ_h estimated in the first stage of the estimation procedure provides an adjustment to the hedonic price equation so that the price regression accurately returns mean preferences. This is because with homogeneous preferences, then the bid curve is the price function so even though there are different points of intersection due to cost functions, the bid function represents the same level of utility $\rightarrow \delta_h$ (mean utility) is the same for everyone. [Equation 2.2]" Joe says (slide 64): " If households have homogeneous preferences, then δ_h same for all houses, equal to constant K... this implies our hedonic model" . What do these mean? Same preferences doesn't mean $\lambda_h^i = 0$?

They instrument for school quality using the border design, and instrument for price using the housing stock in rings > 3 miles away. They do not specify how their coefficients are distributed– Kendra says the standard is probably normal distribution.
Findings:

1. Prices, school quality jump at the border. Housing characteristics do not, but demogr. do! (Table 1)
2. The estimate for the value of school test scores falls by about half when sociodemographics are included, which "highlights the fact that the inclusion of boundary fixed effects in a hedonic price regression is not fully effective in controlling for all aspects of neighborhood quality"
3. Controlling for border FEs reduces the coefficient on race substantially $\rightarrow$ not controlling for unobserved neighborhood quality leads to overstating how much race is capitalized into housing values.
4. Hedonic regressions show people value education, income of their neighbors.

Big picture: Discrete choice modelling helps us some overcome the marginal change issue in hedonic models. More broadly,

Davis (2004)

"This paper measures the effect of health risk on housing values by exploiting a natural experiment that mitigates both econometric problems. The analysis focuses on an isolated county in Nevada where residents have recently experienced a severe increase in pediatric leukemia. Housing prices are compared before and after the increase with a nearby county acting as a control group." The estimating equation is:

$$PRICE_{jct} = \beta_1 X_{jct} + \beta_2 RISK_{ct} + \eta_{ct} + \varepsilon_{jct}$$

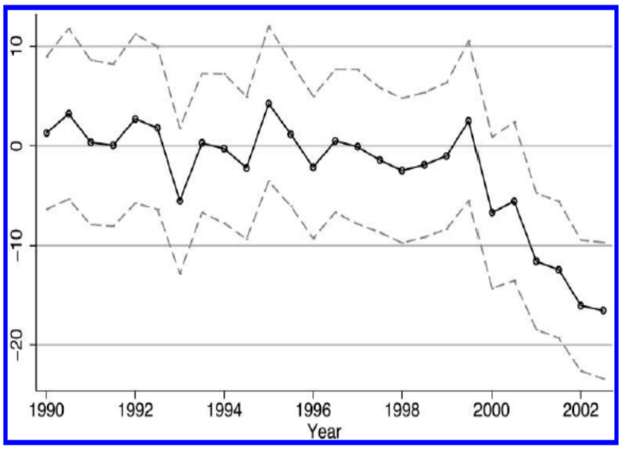


FIGURE 3. FALL IN HOUSING PRICES AFTER 1999: PERCENTAGE DIFFERENCE BETWEEN CHURCHILL COUNTY HPI AND NEVADA HPI

Find: 7.7% decrease in housing prices due to increased risk perception.

2.5 Lectures 6+7: Roback models

Roback models focuses on the role of wages and rents in allocating workers to locations with various quantities of amenities. The model has amenities that vary, with mobile labor and capital. Consumers:

$$\max U(x, l^c; s)$$

In equilibrium, utility must be equal everywhere because of labor mobility.

Firms maximize profits with CES production. Like consumers, firms must get the same profits everywhere. Consumer and firm indifference conditions pin down equilibrium wages. Some characteristics:

- Better amenities → more consumers, and fewer firms since better amenities implies stricter regulation (in this case). Both these lead to lower wages and an ambiguous effect on land rents.

The value of the amenity is given by:

$$\frac{p_s}{w} = \underbrace{s_l \frac{d\log r}{ds}}_{\text{land sacrificed}} - \underbrace{\frac{d\log w}{ds}}_{\text{numeraire sacrificed}}$$

where p_s is the income required to compensate for a small change in s. When goods are non–traded, then the numeraire's price is not pegged at 1:

$$G(w, r; s) = p(s)$$

Market clearing for the non-traded good y implies:

$$Ny(p) = y^D(p)$$

which leads to the amenity value:

$$p_s^* = y\frac{dp}{ds} - \frac{dw}{ds}$$

Kine and Moretti (2013) – TVA

They estimate the long-term consequences of a place-based policy: The Tennessee Valley Authority, which spent a lot of money on infrastructure and electricity in the TN valley throughout the 20th century.

Empirical strategy: Difference in differences using proposed but not implemented authorities as the control group. They do a matching exercise where they select the authority that minimizes the Euclidean distance between the authority and TVA on a vector of covariates. Oaxaca-Blinder regressions estimate:

$$y_{it}^{Non-TVA} - y_{it-1}^{Non-TVA} = \alpha + \beta X_i^{Non-TVA} + \Delta\varepsilon_{it}^{Non-TVA}$$

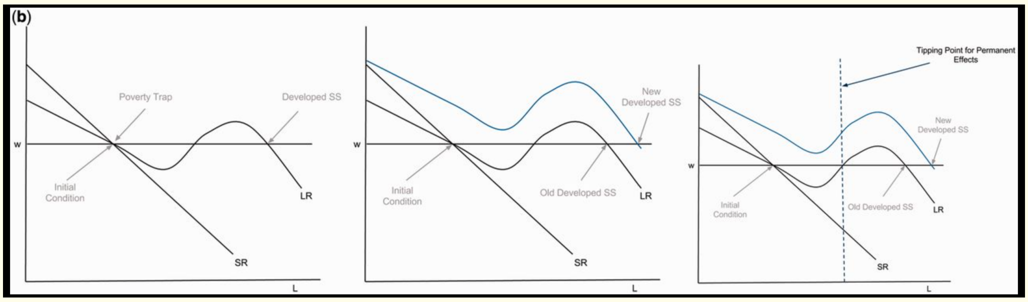
then use $\widehat{\beta}$ to predict counterfactual mean for TVA counties.
In the Roback model, housing in inelastic. But here elastic → wages, rents understate welfare effects.

Model: Counties are SOE's; labor is mobile; agglomeration economies. Assume productivity takes the form:

$$\ln A_{it} = g\left(\frac{L_{it-1}}{R_i}\right) + \delta_t D_i + \eta_i + \gamma_t + \varepsilon_{it}$$

This implies that the TVA has 2 effects:

- Directly raises output (+)
- Increases manufacturing workers → productivity spillovers. In general, this could be +/- depending on where workers are more valuable. But in this case, equal amenities → wages, productivity are equal everywhere → moving people has no effect.



I don't understand these graphs! They find:

1. LR increase in manuf., decrease in ag. but no changes in wages or housing values.
2. TVA raised the productivity of the U.S. manufacturing sector by roughly 0.3% between 1940 and 1960.
3. Roughly constant agglomeration elasticity → indirect effects limited

2.6 Lectures 8+9: Water

Keiser and Shapiro (2019)

Asks: how has water pollution changed since before 1972? Did the Clean Water Act cause these changes? How do residents value changes?

Strategy: Focuses on municipal grants for wastewater treatment plants and uses a DDD upstream/downstream design. The county would be a very noisy measure, so the unit of observation is the river.

Pollution trends estimated with:

$$Q_{icy} = \sum_{\tau=1963}^{\tau=2001} \alpha_\tau [Y_t = \tau] + X'_{icy}\beta + \delta_i + epsilon_{icy}$$

(cluster by watershed). The effect of grants on pollution estimated with:

$$Q_{pdy} = \gamma G_{py} d + X'_{pdy}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$

for plant p and downstream $\in \{0, 1\}$. This is a DDD: downstream × year × grant. Finally, the housing estimates are a DD estimated from:

$$Y_{pt} = \gamma G_{pt} + X'_{pt}\beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$

Finds:

- Pollution has declined substantially: share not fishable ↓ from 35% to 12% over 40 years.
- The clean water act grants caused some of this. The annual cost to make a river fishable was \$1.5 million.
- Treatment plants show almost complete pass–through of funds.
- Δ housing values $< \Delta$ costs; housing benefits/costs ~ 0.25

These are likely lower bounds: incomplete information, excluding non-use value, GE channels (ecology, changes in hedonic price f'n, etc.), and incomplete pass through (→ smaller costs than we think) all bias downward.

More generally, why are these benefits so much smaller than those estimated for the CAA?

- No market instruments
- Ignores a huge polluter: ag.
- Easier to substitute away from pollution
- limited health consequences?

Kremer et. al (2011) – Spring Cleaning

This paper achieves 3 goals:

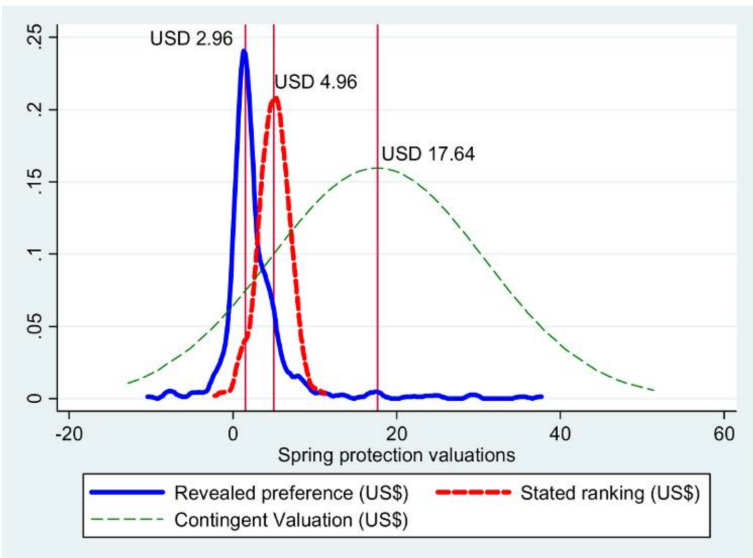
1. Effect of water source improvement on health
2. Structurally estimate WTP for water improvement using travel behavior
3. Property rights simulation

Experiment: Randomly treat springs by encasing them in concrete → decreases opportunity for contamination. Some springs later found unsuitable/defy T/C classification → estimate ITT and IV. Then, to estimate WTP, propose discrete choice model:

$$u_{ijt} = \beta_i T_{jt} + Z_j - C_i D_{ij} + e_{ijt}$$

where $e_{ijt} \sim$ iid EV1, D_{ij} is HH distance to source j, and C_i is the HH cost of time per minute. They can then express the WTP for spring cleaning in terms of walking time (in minutes) as β/C
Find:

1. Fecal contamination ↓ 66% at the source, but only 24% at home → re-contamination.
2. Cleaning reduced incidence of diarrhea, but did not improve BMI or weight.
3. From ML, the mean value of spring protection for a household is 32.4 work days, with considerable dispersion in valuations. This is almost half the stated preference (contingent valuation) value.
4. Implies a tiny WTP for health: VSL of 35 work years. Informational constraints?



Property rights: Propose a game where spring owners choose whether or not to protect and set water prices, then households choose what source to draw from. They analyze fully privatized, Lockean (charge if improved), Modified Lockean (charge if protected), and public funding regimes. Find:

1. Modified Lockean and public investment regimes the only ones with positive social welfare.
2. Private leads to under protection (can't capture all CS) and over protection (don't internalize externality on other spring owners).
3. Under higher incomes, private regime dominates others

Gallagher (2014); Learning about an Infrequent Event: Evidence from Flood Insurance Take-Up in the US

The idea: Insurance policies per capita measure beliefs; since insurance prices don't change after a flood, if we assume community-level flood probability is constant over 1958-2007, then changes in take-up reflect learning. Event study design:

$$\ln(takeup_{ct)} = \sum_{\tau=-T}^{T} \beta_\tau W_{c\tau} + \alpha_c + \gamma_{st} + \varepsilon_{ct}$$

Allow for effects on nearby communities:

$$\ln(takeup_{ct)} = \sum_{\tau=-T}^{T} \beta_\tau W_{c\tau} + \sum_{\tau=-T}^{T} \lambda_\tau N_{c\tau} + \alpha_c + \gamma_{st} + \varepsilon_{ct}$$

Learning model with discounting:

$$E(p|S_t', t') = \frac{S_t' + \alpha}{t' + \alpha + \beta} = \frac{\sum_s y_s \delta^{t-s} + \alpha}{\sum_s \delta^{t-s} + \alpha + \beta}$$

$\delta = 1 \rightarrow$ full information.
Find:

1. Insurance jumps after a flood.
2. Bayesian learning model matches take-up but not decline$\rightarrow$ need learning model with weight on more recent experience; Behavioral economics explanation: Availability Bias
3. Maybe HHs forget, don't know the past, or incorrectly think the odds are changing.
4. Nearby communities also respond a little bit
5. Population declining communities show more persistence in uptake

Lipscomb and Mobarak (2017) Decentralization and Pollution Spillovers... Redrawing County Borders in BRA

Study how regulatory decentralization leads to externalities. Their model predicts that:

1. pollution $\uparrow$ towards borders and at an increasing rate
2. there will be a break in the pollution function at the border
3. total pollution on a river increases in number of borders.

They find evidence of all of these.

2.7 Lecture 10: Natural Resources

Natural resource economics classically are defined by a concern for the dynamics of the problem. However, many natural resource papers ignore these dynamics and speak to the short run or spatial effects (environmental).

Ryan and Sudasrshan (2021) Rationing the Commons

Electricity rationing is India's de facto groundwater regulation. They study the welfare consequences of this regime by:

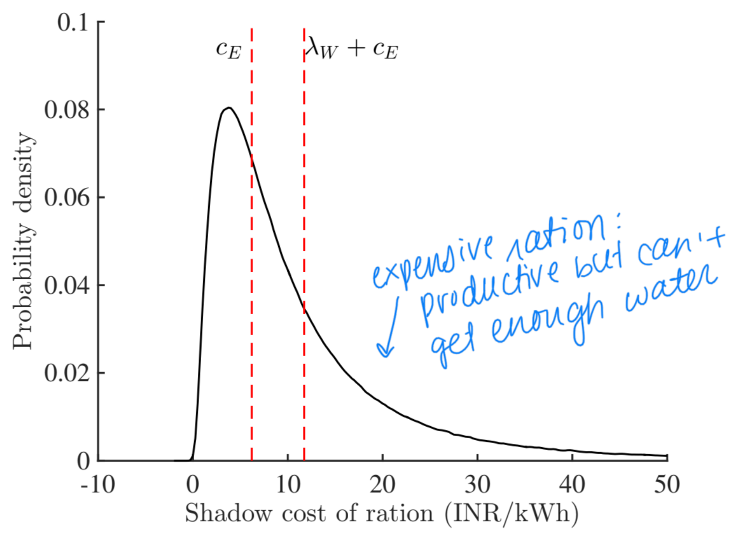
1. Estimate marginal return to increasing the ration by using variation in depth (instrumented with geology/topography) to get at non–existent variation in ration.

   → "Hedonic profit" regressions

2. Estimate agric. production function → study the distributional impacts of efficient, Pigouvian reform

   → Instrument for inputs using geography (water), no. adult men (labor), parcel size (land), seed prices (capital)

Figure: Shadow cost of the status quo ration



They find:

1. Depth has large effects on profits, yield, output.
2. The level of the ration is right on average (between private and social cost)
3. Rationing makes implicit transfers of 26% of HH income, relative to a Pigouvian regime → transfer from most to least productive.
4. Rationing lowers welfare by 16% because of reallocation
5. Pigouvian prescription is not a Pareto improvement.
6. Pigou → more water extraction with less electricity.

2.8 Lectures 11+12: Trade, Leakage and GE land

Some stylized facts about trade and the environment:

1. Climate change involves large and global risks and uncertainty.
2. Global problems need global solutions.
3. Environmental externalities ignore political boundaries
4. The costs of environmental regulation are concentrated in a few industries.
5. Resource extraction is not optimal.
6. Global CO_2 per dollar of GDP has been declining linearly over time.

2.8.1 The Armington model

Based on ACR, "New Trade models, same old gains?". We start with CES preferences:

$$U_j = \left[\sum_i q_{ij}^{(\sigma-1)/\sigma}\right]^{\sigma/(\sigma-1)}$$

which implies a price index:

$$P_j = \left[\sum_i (w_i \tau_{ij})^{1-\sigma}\right]^{1/(1-\sigma)}$$

and the famous gravity equation for trade flows:

$$X_{ij} = \left(\frac{w_i \tau_{ij}}{P_j}\right)^{1-\sigma} Y_j$$

2.8.2 Applications

Shapiro (2020) The Environmental Bias of Trade Policy

Do countries already have high barriers to trade on dirty industries? We might think this is true because these industries tend to be politically influential and could lobby for protectionism. If so, these could act like carbon tariffs.

Measure emissions by noting:

$$x = Ax + d \Rightarrow x = (1-A)^{-1}d$$
$$E = e(1-A)^{-1}ds$$

where A is the input–output matrix, e is CO_2 emitted per dollar of fuel and E is CO_2 emitted per dollar of output.

Upstreamness is measured by:

$$U_i^S = \sum_j \Omega_{ij} y_j / y_i \qquad \text{share sold to firms}$$

Empirically, he estimates:

$$t_{js} = \alpha E_{js} + \xi_j + \varepsilon_{js}$$

Find:

1. Countries implicitly have carbon border adjustments, but they are subsidies → trade policy subsidizes climate change.
2. Controlling for upstreamness makes the slope of tariffs wrt the carbon rate 0 → Political economy explanation: energy–intensive industries are upstream, and everyone downstream lobbies for low protectionism on upstream industries (inputs).
3. Counterfactual where all tariffs are set to the mean (harmonizing the tariffs on clean and dirty goods) decreases CO_2 emissions by 3.59% and increases world real income by 0.65%.

Shapiro and Walker (2018) Why is Pollution from U.S. Manufacturing Declining?

This paper first seeks to statistically decompose the decreases in pollution in US manufacturing into different channels. These are:

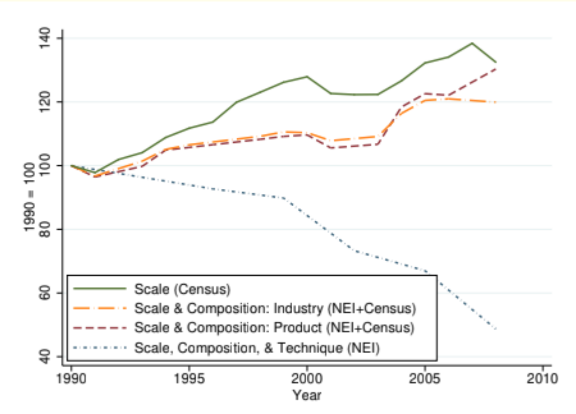
1. Scale
2. Composition
3. Technique

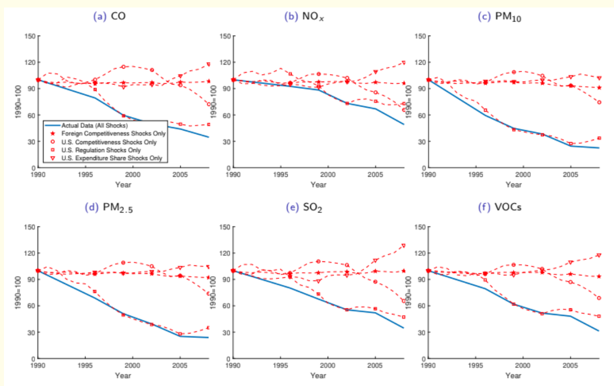
$$Z = \sum_s x_s e_s = X \sum_s \kappa_s e_s = X\kappa' e$$

$$dZ = \underbrace{\kappa' e(dX)}_{\text{Scale}} + \underbrace{Xe'(d\kappa)}_{\text{Composition}} + \underbrace{X\kappa'(de)}_{\text{Technique}}$$

Then they build a trade–environment model with CES preferences and monopolistic competition $\to$ gravity equation, and use this top simulate counterfactuals such as holding foreign competitiveness at historical levels, holding US preferences fixed, or holding regulation shocks fixed.
Find:

1. Essentially all the fall is attributable to technique.
2. Simulations reveal that regulatory shocks are the only shocks capable of explaining the fall in emissions.





2.9 Lecture 14: Urban/ Spatial

Greenstone, Hornbeck and Moretti (2010) Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings

The paper's goal is to quantify agglomeration spillovers by estimating how openings of large plants affect TFP of incumbent plants. They compare runner–up cities for the location of a million dollar plant to establish quasi–exogenous siting. Possible reasons suggested by Roback for agglomeration include:

- Dense labor market $\rightarrow$ lower search frictions.
- Denser areas have lower transportation costs
- Knowledge spillovers
- Local amenities
- Productive amenities

Agglomeration implies $\delta A/\delta N > 0$. We expect:

$$\frac{d\Pi}{dN} = \left(\frac{\partial f}{\partial A}\frac{\partial A}{\partial N}\right) - \left(\frac{\partial w}{\partial N}L^* + \frac{\partial q}{\partial N}T^*\right)$$

Criticisms:

1. CES? If there are increasing returns, then productivity would seem to increase even though there are no spillovers.
2. Price indices? Changing markups?

They find:

1. After million dollar plant opens, incumbent plants in winning counties experience 12% TFP gain after 5 years.
2. Mechanisms: larger spillovers in industries that share labor pools with the million dollar plant, or share technological linkages.
3. Large heterogeneity in gains, and several places where TFP decreases.

Saiz (2010) The Geographic Determinants of Housing Supply

This paper asks what is the role of geography in determining the elasticity of housing. Saiz uses water and slopes over 15% as exogenous limits on where housing can be built. He models:

- Circular city with radius Φ_k, with share Λ_k develop-able $\rightarrow$ mean rent is $r_k = r(2/3\Phi_k)$
- More land–constrained cities see housing value grow more given a positive demand shock $\rightarrow$ less land availability implies lower housing supply elasticities. (slide 54)

Part 3

Jim Sallee

3.1 Lecture 1: The Pigouvian prescription

3.1.1 Basics of the Pigouvian prescription

Definition 14. *An **externality** affects costs or utility directly, not through prices. The presence of externalities imply that the market outcome is not Pareto efficient.*

Notably, things that are NOT externalities are:

- Market power
- Learning
- Network effects

Definition 15. *The **Pigouvian prescription** is to set a tax equal to marginal damages at the optimum. If the tax is set correctly, then outcomes under a Pigouvian tax are Pareto efficient (a Caldor-Hicks improvement), though note that the transition may not necessarily (or even frequently) be a Pareto improvement.*

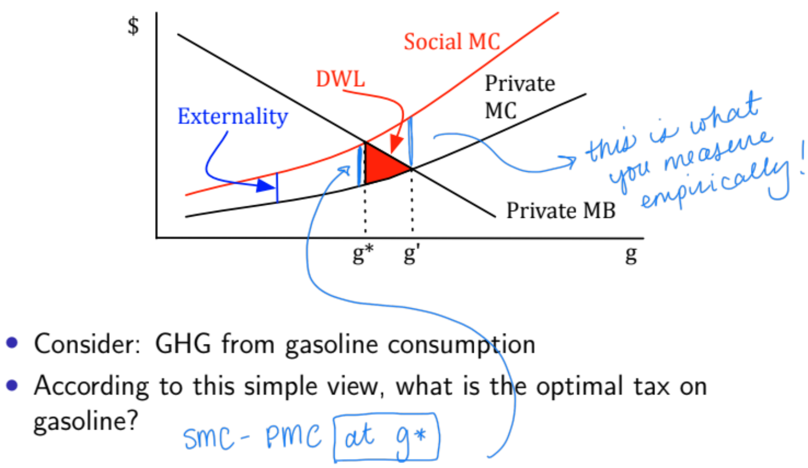


Figure 3.1: Undergrad version of externalities

Analytically, the Planner's problem is:

$$\max_{X_1,X_2} SWF = U(X_1,X_2) - \phi X_2$$
$$\text{s.t. } p_1X_1 + p_2X_2 \leq Y$$

The consumer's problem is:

$$\max_{X_1,X_2} U = U(X_1,X_2)$$
$$\text{s.t. } (p_1+t_1)X_1 + (p_2+t_2)X_2 \leq Y + T$$

These problems match when:

$$t_1 = 0 \qquad t_2 = \phi/\mu$$

where dividing ϕ by μ transforms utils to dollars.

Observations:

- In a first best scenario, the Pigouvian prescription does not directly depend on elasticities!!
- The size of the welfare gain does, however.
- When marginal damages are non-constant, you need an estimate of elasticity to locate optimal quantity, which is needed to determine marginal damages at optimum.
  - Example: If damages are ϕX_2^2, then the optimal tax is $2\phi X_2^*$, and you need the elasticity to know X_2^*.
- The optimum is where all choices are made efficiently– a lot of bodies buried here.
  - If optimal defensive expenditures are much less than the observed, then marginal damages at optimum may be much higher than observed. i.e., at (optimally) lower levels of damages, defensive investments would be lower implying higher observed marginal damages.
- Cost effectiveness does not depends on getting the tax right!

Definition 16. *A policy is* ***cost effective*** *if it gives everyone the same marginal incentive to reduce emissions across all relevant margins. This condition is known as the* ***equimarginal principle****.*

What reasons could there be for modifying the prescription?

1. Spatial complexity of the shape of the damages function.
2. Other approaches may require less information.
3. Double dividends from cap and trade?
4. Uncertainty (Weitzman)

3.1.2 A model of abatement cost

The planner's problem is:

$$\max_{a_j, x_j} SWF = U(X) - \sum_j \left[c_j(x_j) + g_j(a_j) - \phi(e_j(x_j) - a_j)\right]$$

$$\begin{aligned} \text{FOCs:} \quad & U_{x_j} = c_j'(x_j) - \phi e_j'(x_j) \\ & g_j'(a_j) = -\phi \end{aligned}$$

This problem complies with:

- The equimarginal principle: the marginal cost of abatement is equal across all agents, $g_j' = g \,\forall\, j$
- Both channels (scale, abatement) have the same marginal cost:

$$\underbrace{g_j'}_{\text{MC abatement}} = \underbrace{\frac{U_{x_j} - c_j'}{e_j'}}_{\text{MCS reducing scale}}$$

The decentralized Pigouvian problem is:

$$\text{Consumer:} \quad \max_X U(X) \quad \text{s.t. } PX \leq Y + T$$
$$\text{FOCs: } U_{x_j} = p_j \lambda \,\forall\, j$$

$$\text{Firm}_j\text{:} \quad \max_{X,A} p_j x_j - (c_j(x_j) + g_j(a_j) + \tau(e_j(x_j) - a_j))$$
$$\begin{aligned} \text{FOCs: } & p_j = c_j' + \tau e_j' \\ & g_j' = \tau \end{aligned}$$

When I try and put together the firm and the consumer's FOC's, I get a λ hanging around... how to deal with this?
Notes:

- Given any (non–optimal) τ, the firm's FOCs hold $\rightarrow$ equimarginal principle $\rightarrow$ cost effective!

We can further show that this is cost effective by comparing it to a Planner's emissions cap, which is by definition cost–effective as maximizes welfare given a level of emissions reduction:

$$\max_{a_j, x_j} U(X) - \sum_j (c_j(x_j) + g_j(a_j)) - \phi E$$
$$\text{s.t. } \sum_j e_j(x_j) - a_j \leq E$$

This implies:

$$\Lambda = U(X) - \sum_j (c_j(x_j) + g_j(a_j)) - \phi E + \lambda(E - \sum_j e_j(x_j) - a_j)$$
$$\text{FOCs:} \quad U_{x_j} - c_j' - \lambda e_j' = 0$$
$$- g_j' + \lambda = 0$$

This matches the Pigouvian set-up given $\lambda = \tau$! I feel there is an interpretation of the marginal social cost of another unit of emissions being τ, but I'm not quite sure what it is.

3.2 Lecture 2: Tradeable permits and instrument choice

3.2.1 Quantity instruments

Quantity instruments are also "market–based instruments" that mimic the cost–effectiveness of a tax. Their allocation determines incidence but not efficiency! The general problem set up for firms is:

$$\max_{a_j, x_j, w_j} \pi = px_j - (c_j(x_j) + g_j(a_j)) + p_w(d_j - w_j)$$
$$\text{s.t. } e_j(x_j) - a_j \leq d_j - w_j$$

w_j are net permits purchased, and d_j are endowed permits. The first order conditions are:

$$p_j = c_j' + \mu_j e_j'$$
$$g_j' = \mu_j$$
$$p_w = \mu_j$$

which matches the planner's problem when $p_w = \phi$. Note:

- Permit prices reflect marginal cost of compliance!

3.2.2 Uncertainty

Under perfect certainty, permits and taxes are essentially equivalent (ignoring incidence). But when there is uncertainty about costs or benefits, this equivalence breaks. As the graphs indicate:

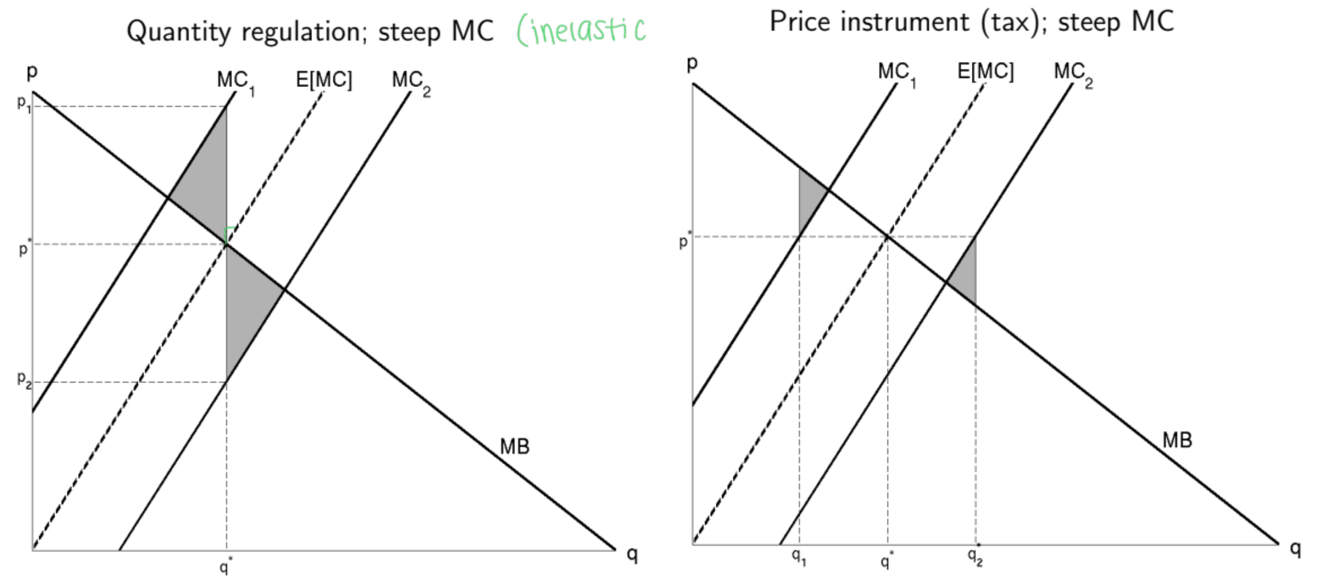
- If supply is relatively elastic (flat) and there is uncertainly over MCs, you prefer quantity regulation (costly to get Q wrong). Conversely, if MC is inelastic and there is uncertainty over MCs, then you prefer price regulation (costly to get price wrong).
- If uncertainty is over benefits, then price and quantity regulation are equivalent.

Pizer and Prest write, "One intuition is that government policy is attempting to replicate society's expected marginal benefits of abatement in the form of a demand schedule in the regulated market. A flat schedule (prices) works better when marginal benefits are relatively flat; a vertical schedule (quantities) works better when marginal benefits are relatively steep".

3.2.3 Instrument choice

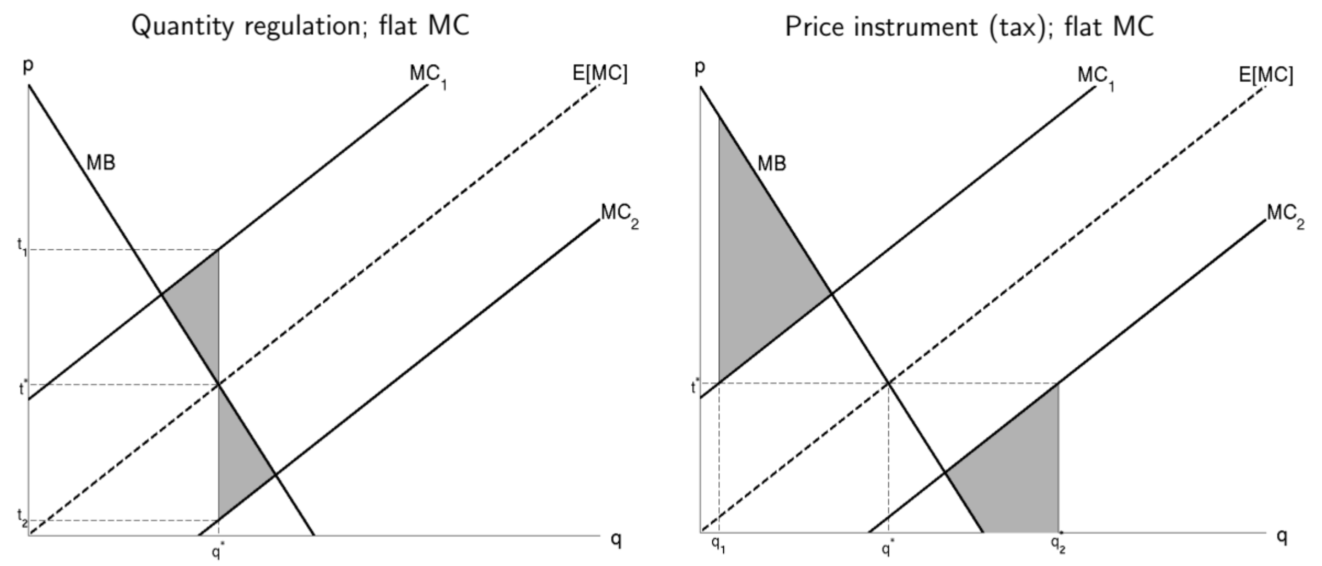
Goulder and Parry (2008) outline criteria for choosing a policy instrument:

- Efficiency
- Cost effectiveness



Notes: DWL in quantity regulation given by the distance between MC and MB from q^* to the q you would want if you knew the realized MC. In the case of price regulation, DWL is the distance between MC and MB from realized $MC(p^*)$ the the equilibrium you would want if you knew the realized MC.

Figure 3.2: Inelastic supply



Notes: DWL in quantity regulation given by the distance between MC and MB from q^* to the q you would want if you knew the realized MC. In the case of price regulation, DWL is the distance between MC and MB from realized $MC(p^*)$ the the equilibrium you would want if you knew the realized MC.

Figure 3.3: Elastic supply

- Equity/distribution
- Robustness to uncertainty
- Political feasibility
- Political flexibility/robustness
- Admin costs
- Enforcement

* Information, spillovers

3.3 Lecture 3: Instrument choice

Observations:

1. Information makes Pigou less attractive, but remember: command and control requires a LOT of information $\rightarrow$ less information is required for market mechanisms to be cost effective, but maybe not optimal.
   - Direct regulatory mechanisms "less inefficient" when heterogeneity limited
2. Equity may not always need to be part of the design
   - **Optimistic separability** suggests that equity and efficiency issues can simply be severed and treated separately.

Pizer and Prest (2019) Prices versus Quantities with Policy Updating

Does the criteria for choosing between price and quantity instruments change when policies change over time? They explore a setting where:

1. Policy changes over time
2. Firms can anticipate these policy changes and react accordingly, and
3. Quantity instruments are tradeable over time

They write the gain of a price instrument over a quantity instrument as:

$$\Delta^0 = 2\frac{\sigma^2}{2c_2^2}(c_2 - b_2) \qquad \text{Weitzman 2 periods}$$

$$\Delta^U = \frac{-(\sigma_\eta^2 + (b_2/c_2)^2\sigma_\theta^2)}{2(b_2+c_2)} < 0 \qquad \text{Pizer and Prest}$$

where:

- σ_θ^2 is the variance of the MC shift parameter
- σ_η^2 is the variance in the MB shift parameter

Intuitively, "if firms can anticipate second-period policy updates before they make choices in the first period, the potential to trade across periods leads them to equalize current and anticipated future prices driven by the anticipated policy update".
They find:

1. "Inter-temporally tradeable quantity regulation can achieve a first-best outcome if firms can correctly predict future updates to government policy and if the government updates policy in order to maximize benefits."

3.3.1 The Coase Theorem

Coase's Theorem forces us to contemplate transaction costs and institutions. Ellickson (1991): "The essence of Coase's argument...is that transaction costs are large and that economic actors arrange their institutions with an eye to these costs"

- Government is one way of overcoming transaction costs, but so is direct bargaining, and so are various other institutions.

Usher (1998) heavily criticizes Coase, arguing that if transaction costs are zero, then there is no need for property rights or even markets. But Coase's Theorem has some implications that we should wrestle with:

1. If bargaining is feasible and is taking place, the Pigouvian tax creates an inefficiency (taxing a non-existent problem is a problem)
2. It is wrong to compensate victims, except in a lump sum manner.

3.4 Lecture 4: Performance standards and rebound

3.4.1 Rebound

Definition 17. ***Jevons paradox:*** *increasing efficiency leads to an increase in energy consumption.*

Rebound is an example of the substitution effect; increasing efficiency decreases the cost of using an appliance. Example, cost of driving a mile is:

$$c = \frac{p_g}{mpg} \rightarrow \frac{\partial c}{\partial mpg} = \frac{-p_g}{mpg^2}$$

Borenstein (2015) decomposes changes in energy consumption given an efficiency upgrade:

Effect	Δ Energy consump.	Notes
Static efficiency	(−)	Direct gain in efficiency
Indirect rebound	(+)	↑ income → buy more goods that require energy
Direct rebound (income)	(+)	↑ income → consume more energy directly
Direct rebound (substitution)	(+)	Consume more energy because cheaper
Cross effects	(−)?	Substitution to energy → ↑ consump. of other goods.

Table 3.1: Energy accounting

Other characteristics matter too!

- If nudged into getting a more efficient car that is also smaller, more cramped → affects how much you drive.
- Another example: strict regulation may force downgrading (goes against rebound) or may increase overall technology of the vehicle (amplify rebound).
- If technology makes a tool more efficient and more productive, then use more.

3.4.2 Performance standards

Performance standards are not effective because they fail to get reductions on the intensive margin (utilization), whereas the Pigouvian tax would achieve reductions on both the extensive (greater efficiency) and intensive (drive less) margin. Rebound → potential increases on intensive margin.

CAFE in particular:

- Fails to reduce mileage.
- Does not encourage savings among used cars (only affects new cars/sales).
- Rebound exacerbates other externalities (mortality, congestion).

Comparing a performance standard to a Pigouvian tax:

Instrument	Δ price efficient option	Δ price inefficient option
Pigouvian tax	(+)	(++)
Performance standard	(−)	(+)

Table 3.2: Market size effects

Holland, Hughes and Knittle (2009) Greenhouse Gas Reductions under Low Carbon Fuel Standards?

The LCFS requires that the carbon content of fuels be, on average, below a certain threshold → performance standard!
They model two highly–substitutable fuels H and L with perfectly competitive supply and separable costs. The centralized problem is:

$$\max_{q_H, q_L} U(q_H, q_L) - C_H(q_H) - C_L(q_L) - \tau(\beta_L q_L + \beta_H q_H)$$
$$\text{FOC:} \qquad U_{q_i} = MC_i(q_i) + \tau\beta_i$$

1st best Pigouvian tax: $\tau\beta_i$.

In contrast, a performance standard sets the constraint:

$$\frac{\beta_H q_H + \beta_L q_L}{q_H + q_L} \leq \sigma$$

New constrained planner's problem:

$$\max_{q_H, q_L} U(q_H, q_L) - C_H(q_H) - C_L(q_L) + \lambda(\sigma(q_H + q_L) - \beta_H q_H - \beta_L q_L)$$

with $\lambda = \tau$, we *almost* get the planner's problem, except with an extra term: $\lambda\sigma(q_H + q_L)$ → **this is a subsidy on output!!**

3.4.3 The Gruenspecht effect (leakage)

Suppose vehicles are scrapped when their value (price) falls below a threshold. Then:

- Pigouvian tax on gasoline → ↓ value, ↑ chance of scrapping
- Performance standard → ↑ price of new cars → ↑ price of used cars → ↓ scrapping
- This gets worse when thinking about vehicle type; if CAFE means new trucks are expensive, might buy used truck (worse) and not a new prius.

Ito and Sallee (2018) The Economics of Attribute-Based Regulation: Evidence from Fuel Economy Standards

Attribute–based regulation is regulation that targets some characteristic of a product or firm, but which takes some secondary attribute into consideration when determining compliance. The idea of this paper: Fuel economy standards are a function of car size; might this distort this secondary attribute, making cars larger?

Theory of a positive externality from efficiency e:

- Non–attribute subsidy: se
- Attribute–based subsidy: $s(e - \sigma(a))$, $\sigma'(a) < 0$

The attribute–based subsidy creates 2 effects:

- $S_e = s$ Pigouvian subsidy of externality
- $S_a = s\sigma'(a) < 0$ unnecessary distortion of a.

Why would we ever want ABR?

- Equalize MC (but why not trade, which would be better?)
- Incidence: you want to transfer wealth
- Safety (but making cars bigger makes them less safe on the road)
- Technology forcing: spillovers
- More stable shadow price than a performance standard.

3.5 Lecture 6: Incidence and equity

PSA: Pigouvian taxes restore efficiency, but that does not imply that they are socially optimal! For that we need to know the Social Welfare Function.

If a policy is regressive, should we abandon it or modify it?

- **Conventional view:** Modify according to equity–efficiency trade-offs.
- **Optimistic separability:** (Kaplow) equity and efficiency are separable. In particular, if a policy is regressive, we can adjust other policies to preserve desired distribution. This is essentially based on the second welfare theorem.

Example: pricing carbon. LIHH spend more of their income on raw energy → energy taxes would be regressive.

Takeaways from Kaplow's view:

- Think about how you will use revenue.
- Doesn't make sense to think about environmental policy in isolation.

Kaplow's prescription has issues. Namely:

- Assumes preference homogeneity.
- Targeting is difficult since heterogeneity is poorly predicted by the transfer function.

However, the conventional view also has to grapple with some issues:

- Arrow's Impossibility Theorem: Determining social preferences is hard.
- Pareto efficiency is only a partial ordering, and tends to bias towards the status quo. How to compare among Pareto efficient outcomes?
  - Other ideas of efficiency exist: Kaldor–Hicks
- Heterogeneity implies it is hard to target losers, implying that we can't really get Pareto improvements from Pigouvian taxes.

Definition 18. ***Kaldor–Hicks compensation criteria:*** *A change from allocation A to B is preferred if a set of hypothetical transfers exists that could be made that would ensure no one was worse off under A and someone is made better off (making the pie bigger).*

Fullerton (2011) Six distributional channels

1. Higher prices for goods
2. Changes in factor returns
3. Allocation of permit rents
4. Distribution of the benefits of environmental quality
5. Transition effects (unemploy)
6. Capitalization of effects into home, stock, land values.

Remember: In GE land, anything can happen!

3.5.1 How regressive is environmental policy?

Evidence is mixed; many studies find that energy/fuel taxes would be regressive. But when using consumption as the base, this evidence is attenuated, and even more if you use lifetime consumption as the base. Cronin, Fullerton and Sexton (JAERE 2019) even find that carbon taxes would be progressive, since social transfers in the US are indexed to prices.

3.6 Lecture 8: Second-Best policy I

The Pigouvian prescription may need to be modified if:

1. You can't tax the externality directly.
   - For example, damages vary by source but the tax rate must be uniform.
   - Tailpipe emissions of co-pollutants. Pigou: tax emissions. But we can only tax gasoline.
2. What if there is another market failure?
3. The market already fixed it (Coase)
4. Equity concerns.

3.6.1 The Diamond model of heterogeneous consumers

Consumption of a good x causes an externality, and each consumer has a different externality:

$$\begin{aligned} &\max_{x_i} U^i(x_1, \ldots, X_N) + \mu_i \\ &\text{s.t.} \qquad \partial^2 U^i / \partial x_i \partial x_j = 0 \,\forall\, i \neq j \\ &\qquad\qquad (p+t)x_i + \mu_i = Y_i \end{aligned}$$

The planner solves:

$$\begin{aligned} &\max_t SWF = \sum_i U^i(x_1(p+t), \ldots, x_N(p+t)) + \mu_i \\ &\text{s.t. } p \sum_i x_i + \sum_i \mu_i = Y_i \end{aligned}$$

which implies FOC:

$$\sum_i \sum_j \frac{\partial U^i}{\partial x_j} \frac{\partial x_j}{\partial p} - p \sum_i \frac{\partial x_i}{\partial p} = 0$$

From the consumer's problem, we know $\frac{\partial U^i}{\partial x_i} = p + t$:

$$\begin{aligned} \sum_i \sum_{j \neq i} \frac{\partial U^i}{\partial x_j} \frac{\partial x_j}{\partial p} + \sum_i \frac{\partial U^i}{\partial x_i} \frac{\partial x_i}{\partial p} - p \sum_i \frac{\partial x_i}{\partial p} &= \sum_i \sum_{j \neq i} \frac{\partial U^i}{\partial x_j} \frac{\partial x_j}{\partial p} + (p-t) \sum_i x_i \frac{\partial x_i}{\partial p} - p \sum_i \frac{\partial x_i}{\partial p} \\ &= \sum_i \sum_{j \neq i} \frac{\partial U^i}{\partial x_j} \frac{\partial x_j}{\partial p} + t \sum_i \frac{\partial x_i}{\partial p} = 0 \end{aligned}$$

Solving for t:

$$t^* = \frac{-\sum_i \sum_{j \neq i} \frac{\partial U^i}{\partial x_j} \frac{\partial x_j}{\partial p}}{\sum_i \frac{\partial x_i}{\partial p}} = \sum_j \left(\frac{\frac{\partial x_j}{\partial p}}{\sum_k \frac{\partial x_k}{\partial p}} \sum_{i \neq j} \frac{\partial U^i}{\partial x_j} \right)$$

Intuitively, the optimal tax is a weighted average of the Pigouvian taxes that would apply to each type, where the weights are the demand derivatives; the tax should be "most like" the tax that would be applied to the most responsive consumers.

Knittle and Sandler (2018) The Welfare Impact of 2nd Best Uniform-Pigou Taxes: Evidence from Transport.

How effective is a gas tax at reducing local air pollution? They show that gasoline price elasticity is positively correlated with emissions (older, dirtier cars have bigger mileage response to gasoline price) $\rightarrow$ 2nd best gas tax $>$ average damages/gallon. Note that this depends on marginal damages being heterogeneous OR non-constant with different preferences! If damages were homogeneous and constant, then this is not necessary. For example, if $\partial U^i / \partial x_j = \alpha \,\forall\, i, j$, the average tax would be right.

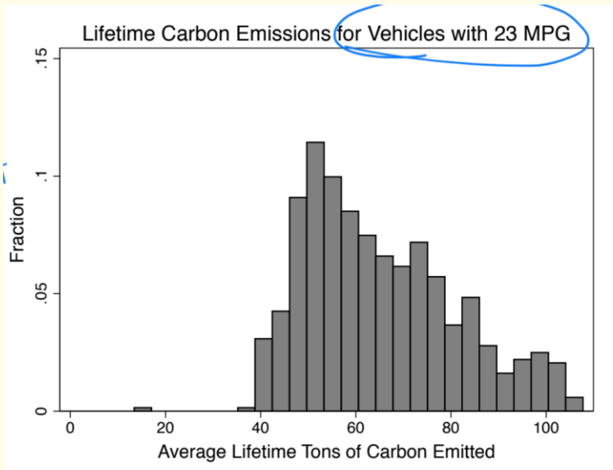
In order to estimate welfare gains, they need additional assumptions on:

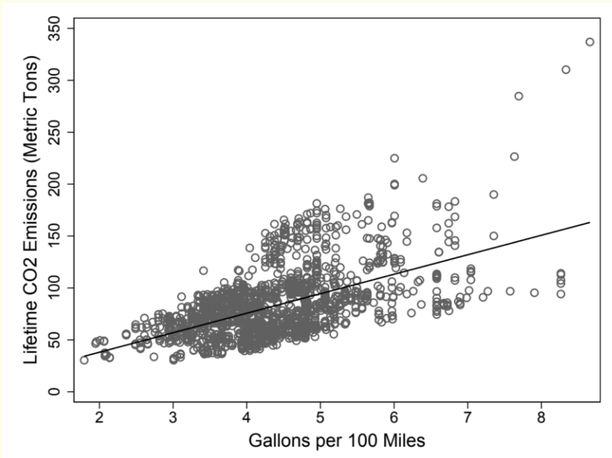
- Mileage distribution
- Distribution of the externality.

Jacobson, Knittle, Sallee and van Bentham (2020) Using Regression Stats to Analyze Imperfect Pricing Policies

This paper asks, what is the welfare cost of basing a policy on a subset of the determinants of an externality? In math, if the externality is $e(A, B)$, what is the welfare loss of basing policy on only A?

Example: lifetime emissions depend on fuel economy AND mileage, but we only regulate on mileage. No other variation in emissions for a given MPG is regulated. CAFE sets a tax like $t_j = \alpha + \beta gpm_j$. If no variation in emissions conditional on GPM, then this is the Pigouvian tax. Variance not captured measures the failure of the policy.





Punchline: R^2 measures how close the policy gets to the Pigouvian tax.

If we are stuck with the second-best, is a different instrument better than a tax (tradeable permits)?

Consider a model with J goods, and each has an externality ϕ_j. Suppose supply is perfectly competitive and utility is quasilinear.

- No intensive margin $\rightarrow$ no utilization effect
- Model can be recast as including random failure rates for durable goods, so long as all consumers share ex ante distribution
- Fixed products (short run, no technology change)

Let $e_j = \tau_j - \phi_j$. The planner's problem is:

$$\max_{t_j} U(x_1(p_1 + t_1), \dots x_J(p_J - t_J)) + \mu + \sum_j \phi_j x_j - C(x_1, \dots x_J)$$

which gives first order condition for each j:

$$\sum_i \left(\frac{\partial U}{\partial x_i} - \frac{\partial C}{\partial x_i} - \phi_i \right) \frac{\partial x_i}{\partial t_j} = 0$$

Given that the consumer's FOC satisfies $\frac{\partial U}{\partial x_i} = p_i + t_i$ and the firm sets $p_i = \frac{\partial C}{\partial x_i}$ this becomes $\sum_i (t_i - \phi_i) \frac{\partial x_i}{\partial t_j} = 0$ and DWL becomes:

$$DWL = -\frac{1}{2} \left[\sum_j e_j^2 \frac{\partial x_j}{\partial t_j} + \sum_j \sum_{k \neq j} e_j e_k \frac{\partial x_j}{\partial t_k} \right]$$

DWL then depends on cross effects and own–effects. The paper's result is that conditional on cross effects being zero (differences in externalities are uncorrelated with substitutability), the regression R^2 is a sufficient statistic for efficiency, as:

$$R^2 = \frac{\text{DWL(OLS) - DWL(constant)}}{\text{DWL(Pigouvian) - DWL(constant)}}$$

the fraction of the possible welfare gain achieved by the linear policy.

3.7 Lecture 9: Reviving the first best

3.7.1 Two-part instruments

Multi-part tariffs can sometimes mimic an unavailable Pigouvian tax! The tax may be unavailable because you do not observe the action, for example, throwing away trash vs. recycling, but two-part tariffs come to the rescue. Idea:

- Tax the action up front (ie, tax buying a bottle), then refund the tax if recycled.
- Intuition: Just subsidizing recycling would result in too many bottle purchases.

3.7.2 Taxing ambient pollution

When individual contributions to an ambient pollution level are unobserved, taxing all polluters with respect to total pollution can induce desired behavior (Segerson, 1988)! Consider the following tax scheme:

$$T(x) = \begin{cases} t(x-\overline{x})+k, & x > \overline{x} \\ t(x-\overline{x}), & x < \overline{x} \end{cases}$$

where $\overline{x}$ is the pollution target. The polluter's problem is:

$$\max_{a_i, y} \pi_i = py - E[T(x(a,e))] - C(y,a)$$

giving FOCs:

$$p = \frac{\partial C}{\partial q_i}$$
$$tE[\frac{\partial x}{\partial a}] - kF_a - \frac{\partial C}{\partial a} = 0$$

and the planner's problem is:

$$\max_{a,y} W = py + E[B(x(a,e))] - C(y,a)$$

giving FOCs:

$$p = \frac{\partial C}{\partial y}$$
$$E[\frac{dB}{dx}\frac{\partial x}{\partial a}] - \frac{\partial C}{\partial a} = 0$$

If $k = 0$, then these are equivalent when:

$$t^* = \frac{E[\frac{dB}{dx}\frac{\partial x}{\partial a}]}{E[\frac{\partial x}{\partial a}]}$$

Note:

1. The indeterminacy in t and and k masks the fact that in the LR, these combinations imply different industry sizes!

3.7.3 Pigou in the presence of other markets

Sandmo (1975) shows that that in general equilibrium, but with no distortions other than the externality, the Pigouvian prescription holds. The simple model assumes linear production, and goes:

$$\Lambda = nU(1 - x_0, x_1, \ldots, x_J, nx_J) + \lambda(X_0 - \sum_j a_j x_j)$$
$$\frac{u_j}{u_0} = a_j, \qquad j = 1, \ldots J-1$$
$$\frac{u_J + nu_{J+1}}{u_0} = a_J$$

implying that the optimal tax vector is:

$$t^* = (0, \ldots, 0, -n\frac{u_{J+1}}{u_0})$$

However, other distortions such as a revenue requirement complicate this picture. A revenue requirement $T \leq \sum_j t_j x_j$ where lump-sum taxes are not permitted changes the tax on the dirty good, as:

1. Ramsey taxation: We want to tax the less elastic good more.
2. Corlett-Hague Rule (1953) model: put higher taxes on goods that are more complementary to leisure

The result:

$$t_j = \left(1 - \frac{-\mu}{\lambda}\right)\left(\frac{\sum_k x_k D_{jk}}{D}\right) \qquad j = 1, \ldots, J-1$$
$$t_J = t_j + \frac{-\mu}{\lambda}\left(-n\frac{u_{J+1}}{u_0}\right)$$

- λ: shadow price on revenue constraint
- μ MU income
- μ/λ inverse marginal cost of public funds = inverse ratio of value of \$1 to government over \$1 to consumer

This displays the **Additivity property:** the dirty good has a Ramsey tax plus the Pigouvian tax. This is first best-y:

1. The tax on the dirty good moves with externality
2. The tax on clean goods is independent of their relation to the dirty good
   - Dixit (1985)'s **Principle of Targeting**;you want to correct an externality at its source; target directly and do not worry about correlated margins of choice.

Marginal cost of public funds

The MCPF is the welfare cost of raising \$1 of public funds. We usually assume MCPF > 1, that is, the government needs more than a dollar to spend a dollar.

3.7.4 Revenue recycling

Definition 19. *The **Revenue recycling effect** is a benefit from using a Pigouvian tax to offset a distortionary tax. In partial equilibrium, this yields a **double dividend!***

However, framed in general equilibrium, the incidence of a tax may be born by the factors of production. $\uparrow$ price $\rightarrow$ fall in the real wage. This is the **tax interaction effect**. The tax interaction effect empirically can be divided into 2 cases:

1. **Strong double dividend** revenue recycling $>$ tax interaction $\Rightarrow$ optimal tax $>$ marginal damages
2. **Weak double dividend** revenue recycling is beneficial (better to cut taxes than use lump-sum rebates)

Punchline: We very likely want to tax below Pigouvian prescription is this because of the MCPF? (strong double dividend fails), but it is very valuable to use revenue to cut distortionary taxes (weak double dividend holds).

3.8 Lecture 12: Second best policy II

Definition 20. *The **Marginal value of public funds** is the cost-benefit ratio of government expenditure:*

$$MVPF = \frac{\textit{Beneficiaries' WTP}}{\textit{Net cost to gov't}}$$

The MVPF is "infinite" when it decreases costs and has a net benefit. Otherwise, use when expenditures benefit the same group (different groups require welfare weights).

Key advantage is that Net Cost takes into account all of the ways that behavior changes in response to the program. You need the net effect, but not a decomposition of the channels. Claim is that this is more empirically relevant. Is this an advantage of cost or benefits? Why isn't it true in both? SLIDE 8

The MVPF has an advantage over the MCPF in that it can remain agnostic about how the expenditure is financed.

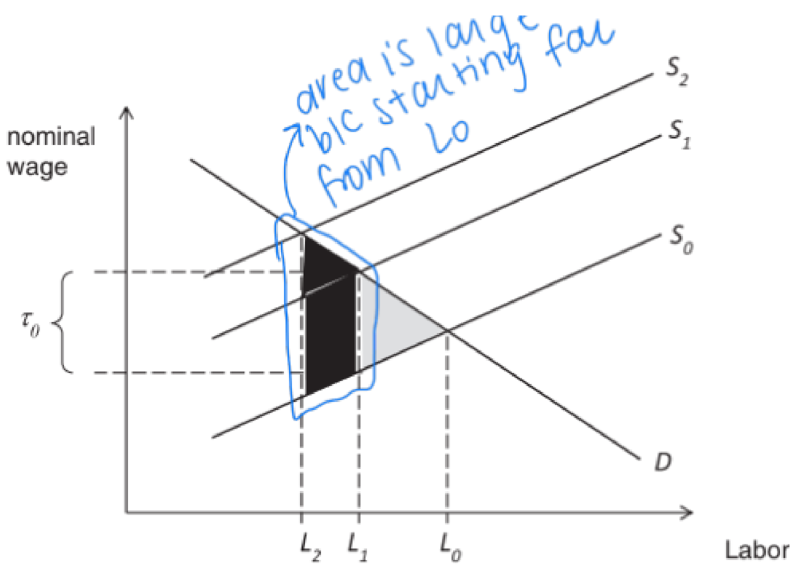


Fig. 2. The tax-interaction effect.

Source: Goulder (2013)

Figure 3.4: Tax interaction effect (trapezoid)

Willingness to pay

Remember (again!) that WTP depends on income! It is not always the appropriate measure, and does not give you a pass on taking a stand on the SWF.

It is the right measure when:

1. Individuals are swapping a resource.
2. Costs, benefits born by the same people.

Goulder, Hafstead and Williams (2016) General Equilibrium Impacts of a Federal Clean Energy Standard

They propose a GE model that compares the efficiency of energy intensity standards and a carbon tax when there are other distortions in the tax code. They find:

1. Efficiency standards can achieve a given reduction in emissions at a lower cost than a carbon tax under plausible conditions.
2. This is because the tax raises the cost of electricity more, which interacts with other tax distortions to worsen the distortion of real factor returns.

3.8.1 More factors that complicate Pigou

More than one problem usually requires more than one solution!

1. Market power
   (a) The Pigouvian tax can lower welfare when a pre-existing distortion exists → monopolist lowers welfare in the direction of the tax; whether or not they overshot is an empirical question.
   (b) BUT market power is endogenous! If the intervention changes market power, this can matter (Buchanan)
2. Non-marginal cost pricing
3. Co-benefits
   (a) With co-benefits, may want to increase the Pigouvian tax.
4. Behavioral frictions

5. Leakage
   (a) Holland (JEEM 2012) shows performance standard can dominate Pigouvian tax with leakage
   (b) The idea: implicit output subsidy mitigates leakage.

Parry and Small (2005)

The gas tax in theory addresses externalities related to climate change, local pollution and externalities. However, for the latter two, the idea, tax would rather be a vehicle miles traveled tax; these are different because of endogenous choice of fuel efficiency, making a fuel tax a poor targeter of local pollution and congestion. In general we might think the optimal fuel tax, compared to the Pigouvian tax, is:

- Lower: because environmental taxes are less efficient at raising revenue than are labor taxes.
- Higher: because gas can be a complement to leisure and its consumption is relatively inelastic (Ramsey)
- Higher: because it lowers congestion which raises labor supply, which alleviates distortions.

They find:

1. Gas taxes are too low in the US and too high in the UK.
2. The biggest externalities from driving are congestion then local pollution, but since the gas tax poorly targets these they are not reasons to raise the tax too much.

Boomhower and Davis (2020) Do Energy Efficiency Investments Deliver at the Right Time?

This paper asks how do fiscal distortions affect the welfare impact of a green subsidy?

Suppose we want to subsidize adoption q, where q has a public benefit τ and fiscal interactions mean funds have a net efficiency of η:

$$W = \underbrace{U(q) - C(q)}_{\text{priv. benefits, costs}} + \underbrace{\tau q}_{\text{public benefit}} + \underbrace{qs - \eta qs}_{\text{Net transfer}}$$

$$\frac{dW}{ds} = \underbrace{[U' - C' + \tau + (1-\eta)s]\frac{dq}{ds}}_{\text{Marginal particip.}} + \underbrace{q(1-\eta)}_{\text{Infra-marginal transfers}}$$

$$\rightarrow \eta = 1 \Rightarrow \tau^* = s$$

$$\rightarrow \eta < 1 \Rightarrow \tau^* < s$$

Intuition: when dq/ds is large, then the first term dominates. When q is large, the second term does. Infra-marginal transfers matter when revenue is distortionary.

3.9 Lectures 13+14: Behavioral I and II

Definition 21. *A **nudge** is a situation in which the choice set/prices are not changed, but the environment or framing is changed. A nudge that has an effect is proof against homo economicus.*

- *homo economicus makes mistakes when he lacks information, homo sapien makes mistakes even with full information.*

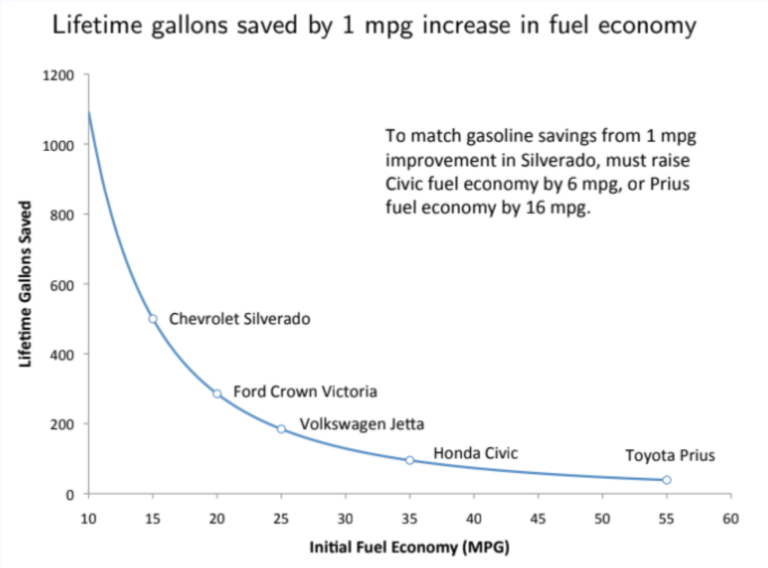
Note that the welfare effects of nudges are generally ambiguous! Can nudges hurt people? And lots more questions:

1. Are nudges persistent or fleeting?
   - Ito, Ida and Tanaka (2018) find prices are much more persistent than moral persuasion.
   - Alcott and Rogers (2014): Consumers rebound after OPower nudges but not fully.
2. Welfare impacts?
3. What is an optimal nudge?
4. Do nudges provide information or operate around biases?

- Ito (2014): consumers respond to average price → information is hard to understand.
- Providing coarse information can help some consumers parse the difference but can crowd out more accurate information for some consumers (Energy star categories)

The MPG illusion

MPG is non-linearly related to gasoline savings, so gas savings is a decreasing function of fuel economy!



The energy efficiency paradox

The observation that apparently negative cost energy efficiency technologies often enjoy low take up in the market. May be due to:

- Split incentives
- Consumer undervaluation
- Engineers are wrong.

Part 4

Reed Walker

4.1 Lectures 1+2: Intro and Measurement

Where are we today? Pollution levels in carbon monoxide, sulfur dioxide, etc. have fallen by about 85% since 1980. The main cause in the US: The Clean Air Act, which sets National Ambient Air Quality Standards for certain criteria pollutants, administered at the county–pollutant–year level.

4.1.1 Measuring the Social Cost of Air Pollution

There are 5 main challenges in measuring the social cost of air pollution:

1. Causal inference: Non–attainment correlated with urban activity, economic growth, etc.
2. Pollutants are correlated between themselves → instruments for pollution affect multiple pollutants.
3. SR vs. LS: SR is more exogenous, but less policy-relevant.
4. Adaptation and heterogeneity in the dose–response function:
   - Adaptive measures mean we understand marginal damages.
   - Non–linearity in damages means we need to estimate the entire damage function → need useful (exogenous) variation in pollution of the entire support of the pollution distribution.
   - Heterogeneity → need exogenous variation in pollutant AND in explanatory factor.
5. Monetization: hospital costs ≠ welfare.
   - WTP may be > hospital costs. ie, how much would you pay to avoid having your leg broken?
   - There are a LOT of social costs associated with pollution. How to monetize each is a separate question.

Climate change impacts are *somewhat* easier to study, as exogenous changes in temperature along the entire support are readily available, not so much of a multiple pollutant problem. LR/SR still difficult, but this is why there has been a lot of recent progress on the social cost of carbon.

Deryugina et. al (2019) The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction

Asks what is the effect of acute air pollution exposure on mortality, life-years lost, and health care utilization among the US elderly. The key innovations are:

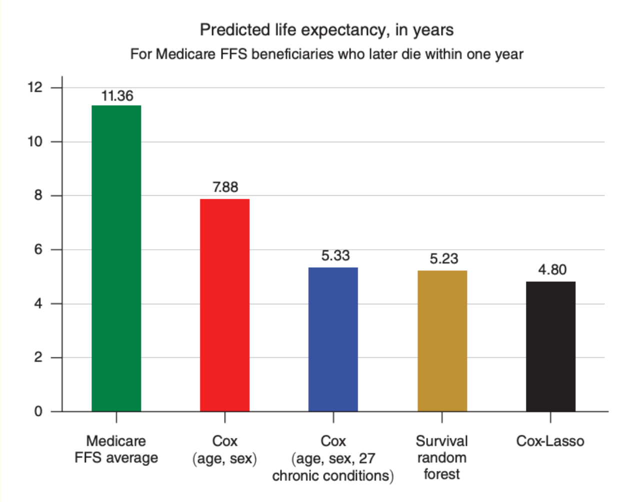
1. Full medicare claims data (full geographic coverage and detailed outcomes)
2. Predict life expectancy with LASSO to better measure life years lost (treatment effect heterogeneity). This minimizes u_{it} (below).
3. Instrument: daily wind variation.

Let $\widehat{L}_{it}$ be predicted life expectancy, and define $u_{it} = \widehat{L}_{it} - L_{it}$. Then we can frame the regression of interest as:

$$\widehat{L}_{it} = \alpha + \overline{\gamma} PM2.5_{it} + (\gamma_i - \overline{\gamma}) PM2.5_{it} - u_{it} + e_{it}$$

it is easy to imagine that $\gamma_i PM2.5$ is corelated with $PM2.5$ → self-selection bias slide 46 lecture 3.
Find:

1. 3.65 $\mu g/m^3$ reduction in PM2.5 from 1999-2011 implies 55,000 reduction in elderly deaths and 150,000 life years saved per year.
2. With a \$100,000 VSL, this gives a benefit of \$15 billion.
3. This estimate would be 3x larger if you just used average life expectancy to calculate life years lost $\rightarrow$ those that die because of higher PM2.5 exposure had lower life expectancy.



4.1.2 Benefits of the CAA

Note that the pollution damages function $\neq$ CAA benefits– in order to estimate these benefits, we need to calculate the causal effect of the CAA on air quality and then use the damage function to estimate out the benefits. In terms of costs, there are a lot of channels through which these may be manifested:

1. Structure/level of prices and quantities
2. Market structure.
3. Change static costs of production
4. Dynamics: rate of innovation, productivity, etc.
5. Product quality and variety.
6. Transition costs.
7. Distributional concerns.

We have a couple options in terms of methods for measuring these, each with its own pros and cons:

1. Program evaluation
    - Cons: Ignores GE; ignores effects most closely related to welfare (prices, costs); Non-attainment is a blunt instrument.
2. CGE
    - Cons: Black box; What is the right amount of complexity?; some strong assumptions.
    - Trade/geography models making progress here.
3. Ground-up IO models of a single sector
    - Cons: Not always clear how to generalize; partial equilibrium; non-transparent empirics.

4.1.3 Non-Market valuation: Environmental Externalities I

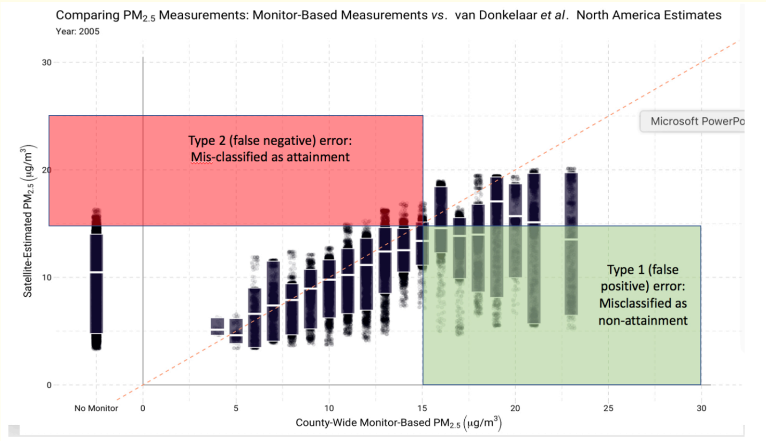
In general, to estimate Δ welfare, incidence, we need to know:

1. Supply curve/ marginal costs
2. Demand curve/ preferences
3. Market structure
4. Marginal damage function/ externality

Fowlie, Rubin and Walker (2019) Bringing Satellite-Based Air Quality Estimates Down to Earth

They want to use spatially-fine, satellite estimates of PM2.5 concentrations to estimate the extent to which limited EPA monitoring has led to the over/under detection of violations of PM2.5 standards.

Methodology: Use satellites to re-calculate the design values, and see how non-attainment status compares with the EPAs designations using the sparse monitoring network.



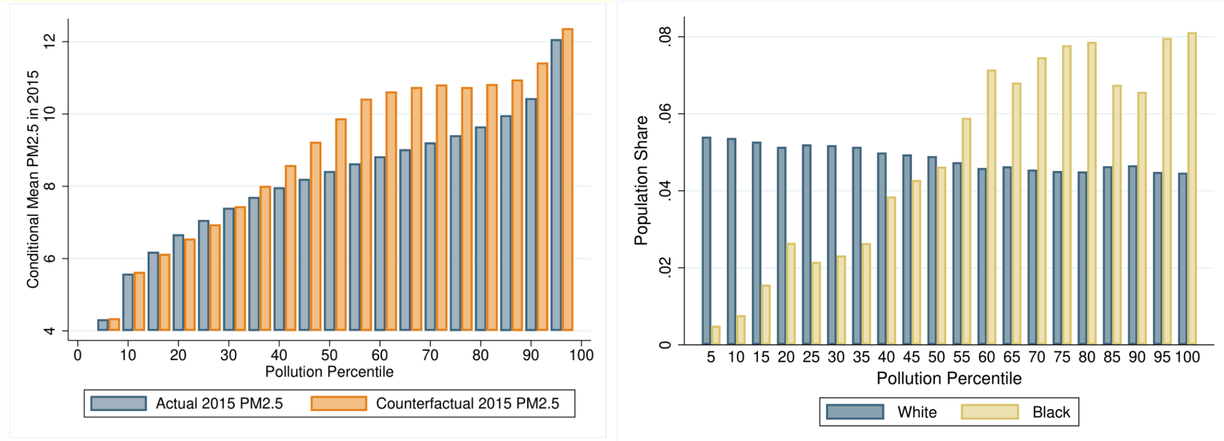
Find:

1. Type 1 errors (misclassify as non-attainment) more common than Type 2.
2. Type 1 errors are more urban, minority
3. Mortality implications of Type 1 errors (i.e. avoided deaths) could be much more consequential than the foregone mortality benefits associated with Type 2 errors. This is partly because many more people live in Type 1 error areas than Type 2 error areas (11% vs. 2%).

Currie, Voorheis and Walker (2021) Environmental Inequality

This paper combines spatially continuous pollution data with large-scale demographic data at a fine scale (ACS: census block) to provide new facts on environmental disparities. Find:

1. Blacks exposed to greater PM2.5 than whites, but the gap has shrunk significantly over the last 15 years.
2. This gap does not disappear when controlling for socioeconomic variables, nor when removing migration patterns by holding individuals fixed in their baseline place of residence.
3. Black neighborhoods may be getting relatively cleaner because the CAA cleans up the worst areas where they are disproportionately represented: 62% of the change in the B/W gap is attributable to the CAA.



4.2 Lecture 3: Non-Market valuation of pollution

Chay and Greenstone (2003) The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession

The idea of this paper is to use the recession of 1980-1982 as an instrument for decreased output $\rightarrow$ decreased pollution, and measure its effect on infant mortality.

While there was already a lit linking infant mortality to pollution, it was largely cross-sectional/OVB issues. Their FEs method uses the identifying assumption:

$$E[\Delta TSP_{ct}, \Delta\varepsilon_{ct}|\Delta I_{ct}] = 0$$

and they implement an IV method where they instrument changes in income and TSP with lagged levels. Find:

1. A 1μg reduction in TSPs would result in an additional 200 additional infant surviving past 1 year.

Overall, we can think of a taxonomy for probing the internal validity of a health paper:

1. Pre-trends/event studies $\rightarrow$ balance.
2. Post trends in other observables that are correlated with health $\rightarrow$ gives you an idea of composition.
3. Placebo-like tests on things not affected by the intervention (eg, accidental deaths).

Chen et al (2013)

Exploits the fact that north of the Huai river heating is done by coal, which leads to higher particulate levels to estimate an RD of the effect of long-run PM exposure on life expectancy. He finds:

1. life expectancies are about 5.5 y lower in north stemming from increased cardio-respiratory mortality.
2. long-term exposure to an additional 100 μg/m^3 of TSPs is associated with a reduction in life expectancy at birth of about 3.0 years.
3. BUT... results are a bit sensitive to polynomial order. Other covariates at the border?

4.3 Lectures 4+5: Non-Market valuation of climate change

How can we make inferences about the costs of climate change? Past experiences can be a benchmark; ask how sensitive society is to temperature, and what evidence there is about the ability for society to adapt. Some takeaways from this literature are:

- Non-linearities in the climate dose–response function matter.
- You need to be careful when aggregating data.

	Top–down approach	Bottom–up approach
Benefits	Comprehensive, captures adaptation, easier	Clearer mechanisms, captures distributional effects, more credible welfare calculations
Challenges	Less credible welfare estimates, non-market and growth effects not captured	Distant from welfare, sometimes need to model adaptation explicitly, requires prices

Table 4.1: Comparing approaches to estimating costs of climate change

4.3.1 Identification and the construction of temperature variables

The idea: compare a geographic area to itself across years (FEs), and measure the effect of small changes in the annual distribution of daily temperature. In order to capture non-linearities, this should be as flexible as possible. For example, using temperature bins.

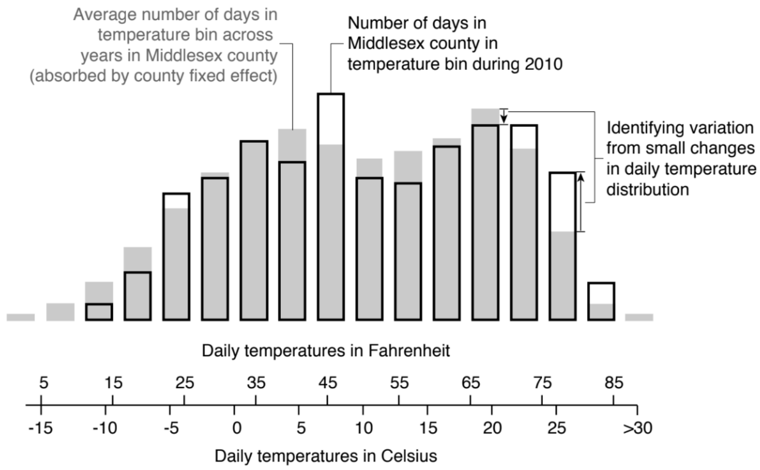


Figure 4.1: Identifying variation using binned temperature variation

Barreca et al (2017) Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century [round 2

] "Paper provides the first large-scale empirical evidence on long-run adaptation opportunities through changes in the use of currently existing technologies."

They use state–month level data with state, time, other FEs to measure the effect of binned temperature on mortality:

$$\log(Y_{sym}) = \sum_j \theta_j TMEAN_{symj} + \pi_L LOWP_{sym} + \pi_H HIGHP_{sym} + X_{sym}\beta + \alpha_{sm} + \rho_{ym} + \varepsilon_{sym}$$

where $TMEAN_{symj}$ is the number of days in a state–month where the mean was in temperature bin j. In order to see the effect of air conditioning, add the following terms to the specification:

$$\phi MOD_{sy} + \sum_j \delta_j TMEAN_{symj} \times MOD_{sy}$$

In order to monetize the benefits of AC, we can either sum across a wide array of benefits (mortality, morbidity, hospitalization costs, etc.), or directly estimate consumer surplus. They do the latter by estimating demand for electricity used for AC with demand function:

$$q_{is} = \beta_0 + \beta_1 AC_{is} + \beta_2 p_{is} + X_{is}\gamma + \varepsilon_{is}$$

They make assumptions about supply, and argue that price isn't endogenous since it is regionally set (can IV or control for this).

Burgess, Deschenes, Donaldson and Greenstone (2017) Weather, Climate Change and Death in India

Observe that the there are large differences in the heat–mortality relationship between urban and rural areas in India. Why the difference? The relationship they explore: formal banking as a mitigating technology (smooths consumption).

Empirical strategy uses the quasi–random rollout of a banking system to test this effect. Find:

1. In hot years, agricultural yields/ wages fall, reflecting the inability to smooth consumption.
2. Bank branch expansion largely offsets the effect of hot days on mortality: Areas with the median number of banks (0.44) saw a 0.3% increase in mortality for every day over 90°F vs. a 1.2% increase in areas with no bank branches.

4.3.2 Top–down approaches

Dell, Jones and Olken (2012) Temperature Shocks and Economic Growth: Evidence from the Last Half Century

This paper reopens a looooong standing debate as to whether or not temperature is central to economic development. Historically, this literature is crappy because largely cross sectional.

The strategy in this paper: country–level panel regressions:

$$g_{it} = \theta_i + \theta_{rt} + \sum_j \rho_j T_{it-j} + \varepsilon_{it}$$

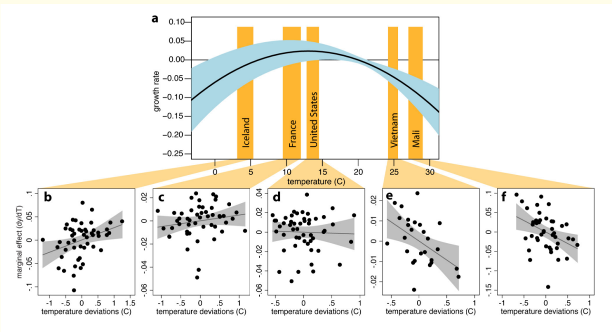
Find:

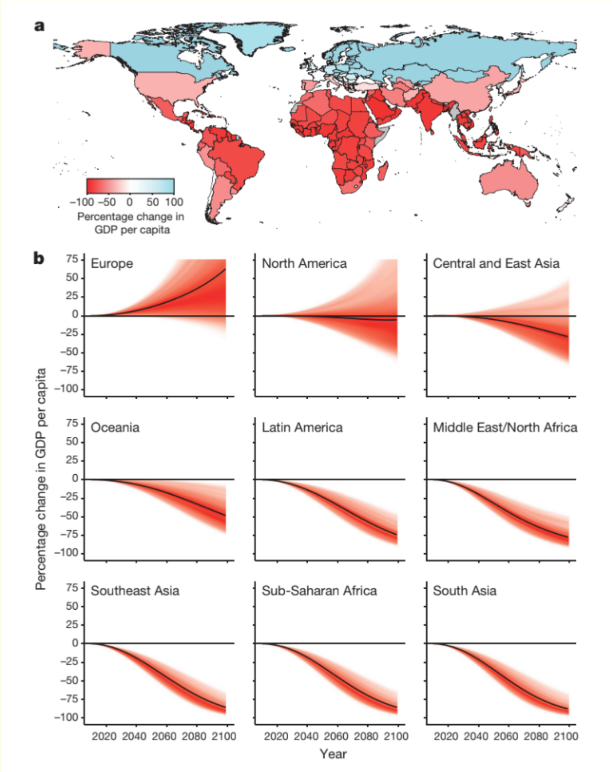
1. Temperature matters for economic growth, but only for poor countries. Is this evidence of adaptation?
2. There are some issues with this paper; non-linearities are misspecified.

Burke, Hsiang and Miguel (2015)

This paper's goal is to reconcile disparate findings: temperatures seem to be important for economic output in some countries but not in others. They explore whether this is because of non-linearities and aggregation issues.

Empirics: estimate a FEs (country, year, country time trends) growth model with non–linear (quadratic) temperature response.





There is still a lot we don't know about the drivers of adaptation. Why do some populations adapt and others do not? Incentives, credit, costs, biases, political economy, collective action?

4.3.3 The Social Cost of Carbon

The SCC

"The social cost of carbon is an economic metric intended to provide a comprehensive estimate of the net damage– that is, the monetized value of the net impacts, both negative and positive– from the global climate change that results from a small (1-metric ton) increase in carbon-dioxide (CO2) emissions." (NAS 2017)
It is meant to be a comprehensive measure of climate change damages, including changes in agricultural productivity, energy demand, health, floods, ecosystem services, and more.

Integrated Assessment Models (IAMs) attempt to measure the SCC in 4 steps:

1. Projecting future global and regional population, output, and emissions
2. Calculating the effect of emissions on temperature, sea level, and other climate variables
3. Estimating (explicitly or implicitly) the physical impacts of climate and, to the extent possible, monetizing those impacts on human welfare (i.e., estimating net climate damages)
4. Discounting monetary damages to the year of emission.

The IAM damage function includes all the damages listed as part of the SCC, converted to the present value using discounting. The inputs to the function should be plausibly causal and reflect adaptation and its costs. Also, the discount rate matters a lot here.

There is a lot of uncertainty in these models; we have a lot of predictions, but what should we do with them? One option: meta analysis. Meta-analysis can involve:

1. Taking the median of estimates (the "most common parameter")
2. Precision–weighting different estimates
3. Analyze the full distribution of estimates
4. Do Baysian analysis over estimates

Costinot, Donaldson and Smith (2016): Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World

They develop model of ag markets designed to capture where crops are produced, how shocks affect supply/prices, and how changes in productivity and prices map into consumption and welfare changes. Find:

1. Impact of climate change on agriculture $\sim 0.26\%$ reduction in global GDP (when trade and production patterns are allowed to adjust).

4.4 Lecture 6: Non-Market valuation (hedonics)

Some more notes about empirical challenges in hedonics:

- Estimating OLS, IV estimates *someone's or weighted average* of WTP on a specific part of the attribute's support $\rightarrow$ think about who's and where.
- You need to do the almost–impossible 2nd step to really look at the welfare implications of non–marginal changes. You can do this by making assumptions about preferences (discrete choice), which gives you WTP functions for amenities that allow you to identify non-marginal changes (think Lucas Davis' paper). Or, make other assumptions.
- Large changes in amenities, large changes in supply, time (long differences??) can cause the hedonic price schedule to move. This is because the schedule is an equilibrium object, NOT a model primitive.
- Benefits may be improperly valued in revealed preference methods if imperfect information, credit constraints, discrimination, etc. present.

Greenstone and Gallagher (2008) Superfund

This paper exploits the fact that the criteria for a Superfund site to be cleaned up was defined as a specific numerical cut off to estimate an RD model of the effect of cleanup on housing values. Remember, you need to look at both supply and demand (or prices AND quantity) to get at welfare effects, as the price change is a function of the elasticity of supply.

They use decadal median housing values at the census tract level to estimate the RD. They find:

- No effect of Superfund cleanup on housing prices, neighborhood composition or housing supply.
- NOTE: when Gamper-Rabindran and Timmins (2013) re-estimate this with geocoded housing data and looking at quartiles of housing prices, they find a 19% increase in housing prices in areas closer to superfund sites.

4.4.1 Hedonic valuation of climate impacts

First, why is agriculture important if it's such a small percentage of GDP? See Figure 4.2– GDP is not a welfare measure!

We can use different approaches to estimate the effect of weather on crop production:

1. Agronomic approach: Lab estimates of effect of temperature on crops, yields. Ex: Rosezweig and Parry (1994).
2. Statistical studies from the real world.

Rosenzweig and Parry (1994) Potential impact of climate change on world food supply

They asked scientists to predict yield changes at many field sites around the world given an increase in temperature.

This holds crop choice fixed! An improvement: model crop switching (see next paper).

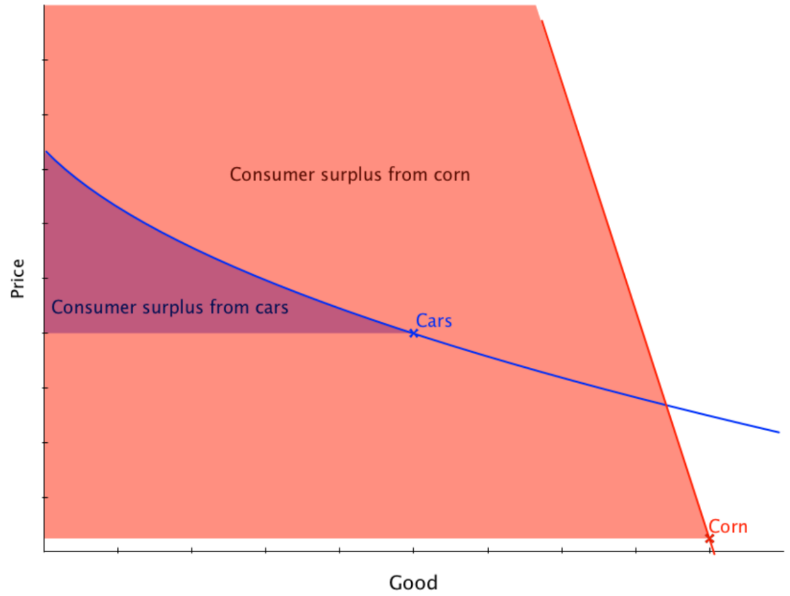
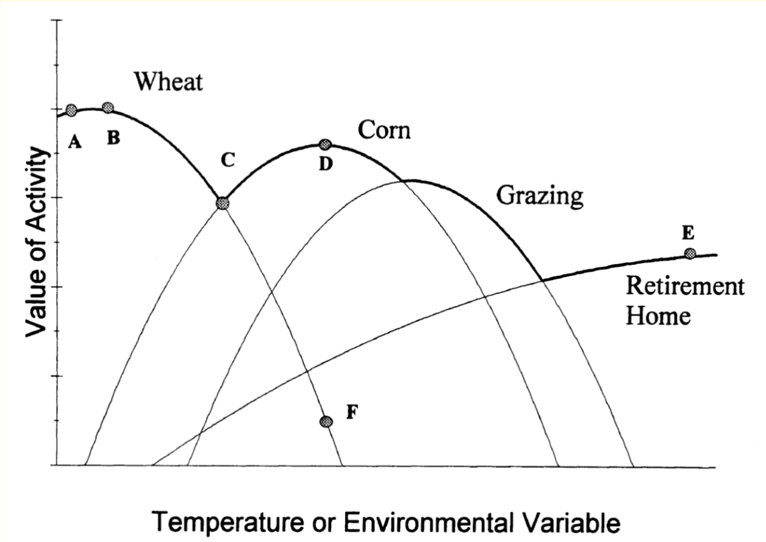


Figure 4.2: Consumer surplus of agriculture

Mendelsohn, Nordhaus and Shaw (1994)

They implement a Ricardian approach by studying land rents directly with hedonic methods. This is a cross–sectional study, but it takes into account crop-switching as a method of adaptation.

An improvement on this: allow the effects to vary over space, as a function of irrigation intensity (see Schlenker, Hanemann, Fisher 2005).



Deschenes and Greenstone (2007) The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather

They improve on the climate change hedonics literature by estimating county FE models of the relationship between annual weather variation and firm profits.

Their empirical strategy:

- Uses degree days
- Controls for county and year FEs

The find:

- Positive relationship between temperature and profits.

BUT: Fisher, Hanneman, Roberts and Schlenker reply to this, finding that correcting errors and filling in missing data leads to a *negative* relationship.

4.4.2 Non–market valuation

Hedonics and market valuations have their limits. Sometimes there is no revealed preference measure if the value is not capitalized into some other market value. There are also other types of valuations such as existence values that are not traded on a market. This is where non–market valuation comes in.

Contingent valuation

When people never use or visit a good, there is no way to estimate trade–offs for the existence good. When this is the case, we need to use contingent valuation methods. In broad strokes, these use surveys where they give people hypothetical choices.

When people do use a good, we could alternatively use travel cost methods to infer valuation. Here the trade-off is between the cost of traveling to a site and the number of visits.

In general, CVs are heavily criticized:

- Maybe surveyees don't take these seriously
  - Do people take any survey seriously? What about the ACS?
  - Better to do these in person.
- Maybe people answer the wrong question (overall valuation instead of individual)
  - Referendum format easier to answer than open–ended questions eliciting maximum WTP
- People don't have enough information, experience to know their answer
  - More information, specificity can help.
- Maybe people are motivated by warm glow.
  - Survey should continuously remind people of what it would cost them.
- Responses are frequently inconsistent and apparently irrational
- There is NO incentive for the surveyee to get things right, and they may in fact be incentivized to lie.

Despite this, they don't seem to do as horribly as you might think. Carson, Flores, Martin and Wright (1996) compare CV and RP estimates, and find that the density of the ratio CV/RP is not crazy (Figure 4.3).

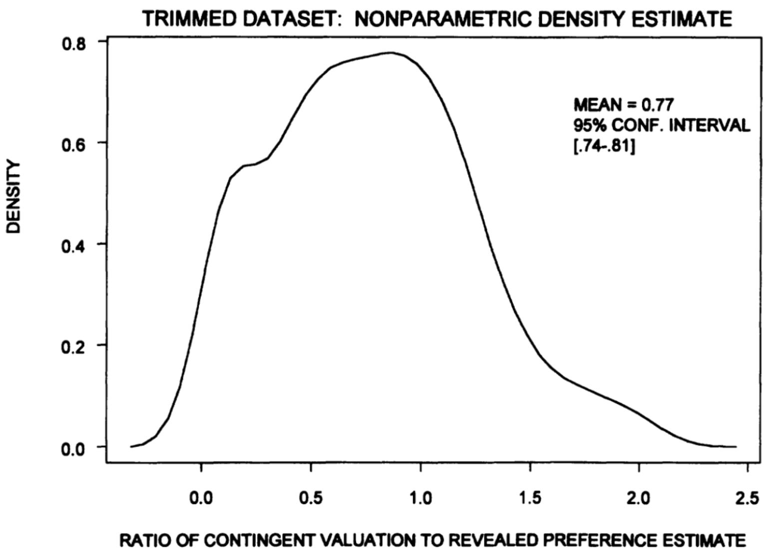


Figure 4.3: Density of the ratio of CV to RP

4.5 Lectures 7+8: Cost benefit analysis

When evaluating the cost of energy and environmental policy, there are a couple of issues to keep in mind:

1. Welfare losses demand on market structure.
2. Regulation might change the FC of entry → induce market power
3. Market spillovers
4. Pre-existing distortions and interaction effects

In broad strokes, protecting the environment lowers the capacity to meet other demands (moves PPF inwards), and may inhibit investment in productive capital. Additionally, reallocation (transition) costs may be high. As before, we can model these costs using:

1. CGE models
2. Empirical IO models (ground-up)
3. Program evaluation

4.5.1 IO models of a single sector

These models are motivated by the idea that individual industries are sufficiently distinct (and details sufficiently important), that cross-industry variation isn't so useful. They focus on a single industry/market, with careful attention for institutional details, measurement of key variables, and econometric identification issues, and hope to learn generalizable insights from a narrow focus. The ingredients to these models are:

1. Economic model with the key features of the industry.
   - Describes the economic environment: the market structure and institutions, actors, and information.
   - Lists primitives: technology, preferences, endowments/assets.
   - Need to know what is exogenous to the agents/the environment.
   - Specify decision variables, timing, objective functions.
   - Describe equilibrium.
2. Data: prices, quantities, number of firms, costs if possible.
3. Econometric specification of the model
   - Need to specify heterogeneity/shocks; what is observed/unobserved.
   - Optimization errors? Measurement errors?
   - Frequently requires us to restrict utility, production functions to a specific form, and/or restrict preference, productivity distributions to be of a known form → computational tractability.
   - Identification: Given a distribution of the observed data, is there a unique set of model parameters to match that distribution?

Ryan (2012) The Costs of Environmental Regulation in a Concentrated Industry

Most previous studies ignored dynamics in calculating welfare → when regulation changes costs of entry, entry/exit could be important. This paper explores this in a concentrated industry: Portland cement. The cement industry is characterized by:

1. Large (indivisible) capital investments (FC of entry) and economies of scale → market power
2. Highly polluting/energy intensity
3. Highly local competition → concentrated local markets

Model sketch: Oligopolists make optimal decisions over entry, exit, production, investment given the strategies of their competitors. In a dynamic game:

1. Estimate SR profit f'n, market demand → policy functions for entry, exist, investment as a function of the public state vector S_t (vector of capacities).
2. Find parameters that rationalize data as best responses. Parameters: capacity adjustment cost, distribution of entry costs (assumed iid normal), scrap values. Conditions: entry, exist, investment.

$$C_i(q_i,\delta) = \underbrace{\delta_0}_{FC} + \underbrace{\delta_1}_{MC} q_i + \underbrace{\delta_2 1(q_i > \nu s_i)(q_i - \nu s_i)^2}_{\text{Capacity constraint cost}}$$

In each period:

1. Incumbents draw from scrap value distribution → decide whether or not to exit
2. Entrants draw from distribution of investment and entry costs, remaining incumbents draw from distribution of costs of investment. All firms simultaneously make entry and investment decisions.
3. Capacity-constrained incumbents compete over quantities.
4. Firms enter and exit. Investments mature.

The empirical context: The 1990 CAA amendments which required cement plants to undergo an environmental certification process → increase sunk costs.

- Allow structural cost parameters to change after the amendments.
- Uses market-level data on prices, quantities, N firms, energy and labor prices + kiln-level data on capacity, production, investment.
- Needs to estimate:
  1. Profits (depends on consumer demand) → Instrument for prices in constant-elasticity market demand with goal, gas, electricity and wage prices.
  $$\log Q^D_{mt} = \alpha_0 + \alpha_1 \log P_{mt} + \alpha_{2m} + \alpha_{3t} X_{mt} + \epsilon_{mt}$$
  2. Numerically find δ_1, δ_2, ν ($\delta_0 := 0$) that recover Cournot quantities
  3. Investment policy function using (S, s) model.
  4. Entry/exit policy functions: probit model as a function of competitor capacities (exit depends also on own capacity); can vary before/after regulation.

Counterfactual: MPNE post–1990 using the cost distribution pre–1990.
Finds:

1. Amendments doubled sunk costs of entry
2. Regulation → reduced exit and entry, high investment by incumbents, lower aggregate market capacity.
3. Consumer welfare ↓ 2.5% due to increased market power.

Fowlie (2010) Emissions Trading, Electricity Restructuring and Investment in Pollution Abatement

She studies whether heterogeneity in electricity market regulation affect how coal plants choose to comply with NOx emissions trading program. The idea: existing distortions in product markets subject to CAT regulation may interfere with the emissions permit market.

The idea:

- Regulated firms are guaranteed to recoup fixed investments, making it more likely they will invest in high-FC, high-abating technology.
- Deregulated plants have greater uncertainty over their ability to recoup such investments:
  – Marginal unit of power is not usually coal but rather gas, which has a lower compliance cost.
  – → electricity prices would not reflect their high environmental compliances costs, further increasing uncertainty.

Empirical setting: Variation in electricity market regulation comes from interstate variation in industry restructuring. Plants choose abatement technology in order to comply with the NOx Budget Program. She then:

- Estimates a random coefficients logit model to estimate demand for abatement choice as a function of regulated status, the plant's cost of capital and cost recovery parameters.

- Uses this model to estimate 2 counterfactuals: All coal plants in the emissions trading program are subject to electricity rate regulation or are restructured.

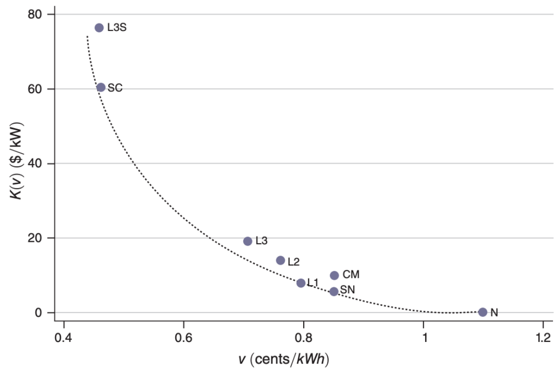
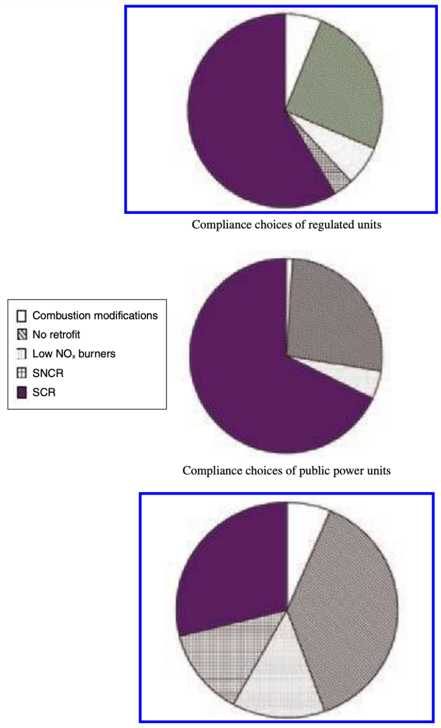


FIGURE 1. ESTIMATED NO_x CONTROL COSTS FOR A 512 MW T-FIRED BOILER

Notes: In generating this figure, I assume that the unit will achieve perfect compliance. This assumption finds empirical support (US EPA 2005). I further assume that compliance will not be achieved through reductions in output. Support for this assumption is provided in the Web Appendix. Section III includes a detailed discussion of how these cost estimates are generated.

Strategy code	Technology	lbs NO_x/mmBtu
N	No retrofit	0.42
SN	Selective Non-Catalytic Reduction (SNCR)	0.34
CM	Combustion modification	0.33
L1	Low NO_x burners with overfire air option 1	0.31
L2	Low NO_x burners with overfire air option 2	0.28
L3	Low NO_x burners with overfire air options 1&2	0.26
SC	Selective Catalytic Reduction (SCR)	0.13
L3S	L3 + SCR	0.11



Compliance choices of regulated units

Compliance choices of public power units

Compliance choices of unregulated units

She finds:

1. Restructured plants are less likely to adopt more capital intensive environmental compliance options as compared to regulated or publicly owned plants.
2. Larger share of the permitted pollution is emitted in states where air quality problems more severe (i.e. due to heterogeneity in electricity market regulations)

Natural monopoly regulation

Top-down regulation of natural monopolies attempts to solve the DWL issue of natural monopoly pricing by setting price and giving firms a nominal rate of return. However, this results in poor incentives (all costs passed to consumers) and distortions (asymmetric information). So which are better? Regulated or unregulated?

- Stigler and Friedland (1962): Hedonic regression; find that regulation did little to reduce prices below unregulated monopoly levels → take this as evidence against regulation. This had a coding error!! Actual effect: 20% lower prices.
- However, deregulation can increase costs by getting rid of vertical generation, coordination.

In practice, segments viewed as competitive (e.g. electricity generation), have been increasingly deregulated while network segments (transmission) still operate under natural monopoly cost. Now generating units are scheduled according to bids in day-ahead, uniform price auctions instead of engineering estimates and real-time auctions determine who actually produces electricity. Now: around 60% of electricity is generated by markets.

Cicala (2015) When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation

He studies the shift from cost-of-service regulation (i.e. cost-plus) to market based dispatch. Asks: What happened to input costs, specifically coal procurement? Why?

The idea: regulation may (1) be biased toward capital investment; (2) suffer from asymmetric information; (3) be plagued by interest groups. He tests these by looking for (1) capital intensity of abatement tec; (2) look at markets with observable input prices (NG); (3) look at fraction coal purchased in-state.

Empirical strategy: Use DID combined with a matching estimator to estimate the effect of deregulation on coal procurement.

Finds:

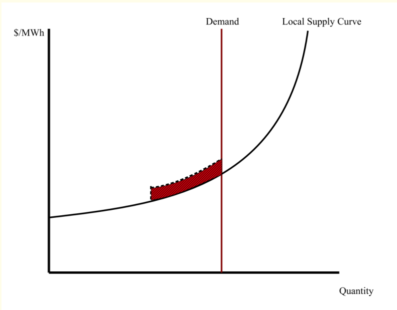
1. Deregulated plants substantially reduce the price paid for coal
2. Deregulated plants employ less capital-intensive production techniques
3. Deregulated plants expand coal imports from out of state by 25%
4. Reallocation of coal purchases toward more productive mines

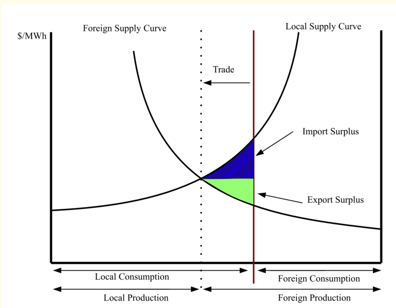
Cicala (2022) Imperfect Markets versus Imperfect Regulation in US Electricity Generation

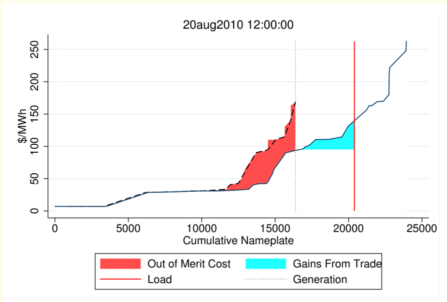
Asks: What are the welfare gains from moving to a deregulated, wholesale electricity market? Do markets (including all of their flaws) outperform other methods for electricity dispatch?

Empirical strategy: Uses a DID of regulated vs. deregulated markets around the staggered creation of wholesale electricity markets, and calculates the impact on merit order and the gains from trade. There are two key welfare determinants:

- Out of merit costs; sourcing from a source that is not lowest-cost
- Gains from trading electricity







Finds:

1. Gains from trade increase 50%
2. Out of merit costs decrease 15%

Caveats:

- Static. But could investment incentives change capacity?
- Perfectly inelastic demand $\rightarrow$ no welfare changes on demand side.

Shapiro and Walker (2021) Is Air Pollution Reg. Too Stringent?

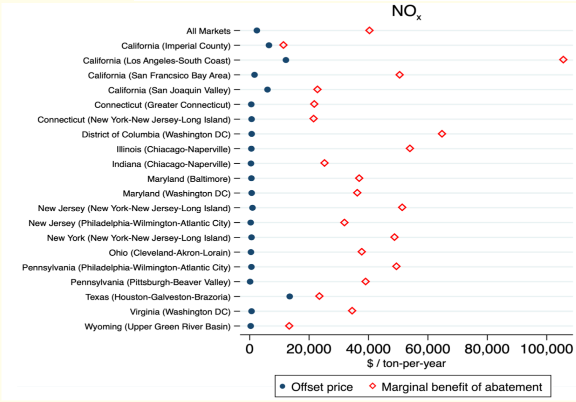
Air pollution levels have fallen dramatically; have we reached the point where costs equal or are above benefits?

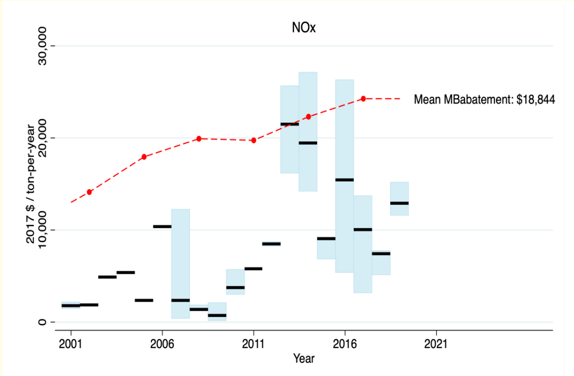
Empirical strategy: Use the fact that in non-attainment areas, entrants must pay for pollution permits from incumbents. These are decentralized markets $\rightarrow$ estimate MC of abatement. They want to estimate:

$$P = -\frac{\partial C_i(X_i)}{\partial X_i} - \frac{\partial T_i(X_i)}{\partial \partial X_i}$$

where X_i is abatement.

Table 3—Ratio of Offset Prices to Marginal Benefits of Abatement by Region, 2010-2019				
	NO_x		VOCs	
	(1)		(2)	
Panel A. National				
1. Offset prices / marginal benefits of abatement	0.06	0.08	0.14	0.11
2. Mean pollution offset prices (per ton of emissions)	$2,416	$4,058	$2,798	$2,213
3. Mean pollution marginal benefits of abatement	$40,501	$51,466	$20,620	$20,602
Weight	Tons	Population	Tons	Population





Find:

1. MBs of air pollution abatement are still far above MCs.
2. Offset prices are very volatile → evidence of steep marginal abatement cost curve? → planner may prefer price regulation (Weitzman)

Walker (2012) The Transitional Costs of Sectoral Reallocation: Evidence From the Clean Air Act and the Workforce

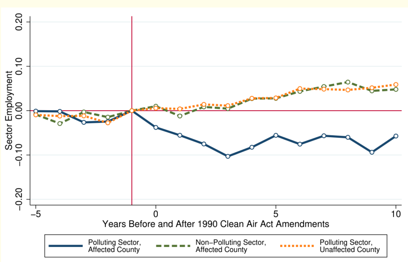
Asks: Shifting production from dirty to cleaner production necessitates reallocation costs for workers/firms. How large are these transitional costs born for workers in affected firms?

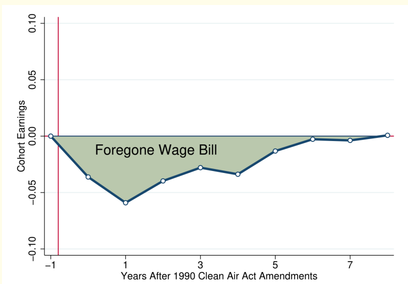
Empirical strategy: DDD: use county-industry-year variation in Clean Air Act stringency to estimate event studies with worker-firm linked data. The workers are divided into pre-post cohorts. Isolates:

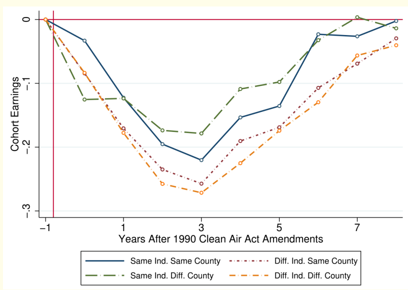
- Equilibrium wage changes within a firm/sector
- Costly worker transitions across firms/sectors

Identifying assumption: No unobserved shocks to polluting sectors of counties that switch into non-attainment in years after non-attainment went into place.

$$Y_{cst} = \eta_1 P_s \times N_c \times 1(\tau_t \leq 0) + \rho_{cs} + n_{ct} + p_{st} + \gamma_t + \epsilon_{cst}$$







Finds regulation leads to:

1. Employment losses
2. $9 billion foregone wage bill
3. Increased rate of job separations

4. Greatest earnings loss among those who change industries $\rightarrow$ losses are due to transitions, not lower wages.

* If workers are paid their MP, then these are social losses. But the losses are $<$ benefits of CAA.

4.6 Lectures 9+10: Production Functions

In order to study changes in competition, cost shocks, (de-)regulation, merger analysis, or the effects of policies on efficiency, we need to know/estimate productivity, costs and markups.

4.6.1 Anatomy of a production function

Consider output of the form:

$$Y_{it} = \exp(\omega_{it}) F(L_{it}, K_{it}, M_{it})$$

ω_{it}: Hicks neutral productivity. As Cobb–Douglas:

$$Y = AL^{\alpha} K^{\beta} M^{\gamma}$$

- Output elasticities: α, β, γ (CES+ no markups = cost shares)
- A: TFP
- $\alpha + \beta + \gamma =$ returns to scale

4.6.2 Data issues

Estimating even this simple form has its issues:

- Firm-level data sets may lack establishment-level info, less detail about inputs/outputs, and introduce sample selection issues.
- We want quantities of inputs and outputs, but data usually have revenue/expenditures.
- Capital stocks/investment rarely observed; when they are, reliability?
- Quality differentiation? Multi-product production?
- Transmission bias: $A \rightarrow L^*, K^* \rightarrow$ Simulteneity!!
- Selection: Productive firms less likely to exit $\rightarrow$ don't observe entire distribution of A.
- Multicollinearity among inputs.

4.6.3 Estimation

Taking logs of the Cobb-Douglas example, we can estimate A as the residual of:

$$\ln Y_y = \alpha \ln K_t + \beta \ln L_t + \ln A_t$$

Another option: FEs

$$A_{it} = \mu_i + \epsilon_{it}$$

Assumption: ϵ_{it} is uncorrelated with input decisions. But this comes with its own issues:

- Measurement error/attenuation: FEs kill the noise to signal ratio.
- What if productivity is time invariant?

More options:

1. IV. But how to find exogenous input demand shifters?
2. Control function approaches: currently frontier, but requires strong assumptions about productivity DGP (See below: Index methods)
3. Dynamic panel methods: lagged (differenced) instruments $\rightarrow$ high data demands

4.6.4 Index methods: Olley-Pakes (1996)

Cobb-Douglas + CRS + no markups ⇒ Output elasticities = cost/revenue shares. ie:

$$\beta = \frac{\text{labor expenditures}}{\text{revenue}}$$

Analyzes effects of deregulation in telecommunications equipment industry. The main idea: model choice of inputs as a function of unobserved term. Cobb-Douglas implies:

$$y_{it} = \beta_k k_i t + \beta_l l_{it} + \omega_{it} + \epsilon_{it}$$

A1 ω_{it} follows Markov proces: $p(\omega_{it+1}|I_{it}) = p(\omega_{it+1}|\omega_{it})$

A2 KLM: $k_t = (1-\delta)k_{it-1} + i_{it-1}$.

A3 Investment is a f'n of observables and ω: $i_{it} = I_t(k_{it}, \omega_{it})$, increasing in ω.

A4 Labor variable and non-dynamic, i.e. chosen each t, current choice has no effect on future (can be relaxed).

Given $I_t()$ increasing in ω, it can be inverted for productivity: $\omega_{it} = h_t(i_{it}, k_{it})$. Estimation:

1. **Step 1:** Estimate production function semi-parametrically $\rightarrow \widehat{\beta}_l$:

$$\begin{aligned} y_{it} &= \beta_k k_i t + \beta_l l_{it} + \omega_{it} + \epsilon_{it} \\ &= \beta_l l_{it} + \beta_k k_{it} + h_t(i_{it}, k_{it}) + \epsilon_{it} \end{aligned}$$

   given $\widehat{\beta}_l$, can estimate $\widehat{\phi}_{it} := \beta_k k_{it} + h_t(i_{it}, k_{it}) = y_{it} - \widehat{\beta}_l l_{it}$

2. **Step 2:** Identify β_k. Let $g(\omega_{i,t-1}) = E[\omega_{i,t}|\omega_{i,t-1}] \Rightarrow \omega_{i,t} = g(\omega_{i,t-1}) + \xi_{i,t}$ (specify Markov transition probabilities). Then the second stage equation is:

$$\phi_{i,t} = \beta_k k_{i,t} + g(\omega_{i,t-1}) + \xi_{i,t} = \beta_k k_{i,t} + g(\phi_{i,t-1} - \beta_k k_{i,t-1}) + \xi_{i,t}$$

   with non-random entry/exit, we need to model this as it affects estimates of output elasticities.:

$$\omega_t = \underbrace{\overline{\omega}_t}_{\text{Mean prod.}} + \underbrace{\sum_i \Delta s_{it} \Delta\omega_{it}}_{\text{Selection: cov(prod., market share)}}$$

Issues:

- Relies strongly on single index restriction of productivity
- Measurement errors in investment creates problems
- Cannot accommodate zero investment or non-monotonic investment

Unobserved product quality or markups

Write the production function as:

$$y_{it} = x_{it}'\beta + \psi_{it} + \epsilon_{it}$$

But if we only observe revenue and expenditures, and sectoral price indices, right:

$$\begin{aligned} r_{it} &= y_{it} + p_{it} && \text{Revenue} \\ e_{it} &= x_{it} + w_{it} && \text{Expenditures} \\ p_{it} &= \overline{p}_t + \tilde{p}_{it} && \text{Prices} \\ w_{it} &= \overline{w}_t + \tilde{w}_{it} && \text{Input prices} \end{aligned}$$

This gives the above in terms of deflated revenues and expenditures:

$$r_{it} - \overline{p}_t = e_{it} - \overline{w}_t{}'\beta + \tilde{p}_{it} - \tilde{w}_{it}'\beta + \psi_{it} + \epsilon_{it}$$

This clarifies the potential for 2 types of biases:

1. $Cov(e_{it}, \tilde{p}_{it}) \neq 0 \rightarrow$ output price bias. Suppose no variation in input prices but a firm can somehow charge a higher price, lower quantity $\rightarrow$ lower input demand leads to negative correlation between prices, inputs $\rightarrow$ downward bias in coefficients.
2. $Cov(r_{it}, \tilde{w}_{it}) \neq 0 \rightarrow$ input price bias. Suppose there is no variation in output prices, but there is in input prices (eg, lower amenities, higher wages). High input prices and same quantity will lead us to observe firms with high expenditures and the same revenues $\rightarrow$ negative bias in coefficients.

Levinsohn and Petrin (2003) extension

Main idea: rather than use investment to control for unobserved productivity, use materials inputs. Benefits:

1. Less likely to have zero materials expenditure in the data.
2. Investments may be lumpy and not respond to productivity shocks; materials may be better behaved.

Identification issues (ACF)

If labor is also a function of materials, capital, then when we non-parametrically control for these in the regression, there shouldn't be extra variation left over to identify β_l! Enter: Ackerberg, Caves and Frazer (2015). The propose the first stage:

$$y_{it} = \theta(l_{it}, k_{it}, m_{it}) + \epsilon_{it}$$

Once you have predicted values $\widehat{\theta}_{it}$, you know productivity up to a vector of output elasticities:

$$\omega_{it}(\beta) = \widehat{\theta}_{it} - \beta_k k_{it} - \beta_l l_{it} - \beta_m m_{it}$$

The second stage solution proposes the moment condition:

$$E\left[\left(\omega_{it}(\beta) - g_t(\omega_{it-1}(\beta))\right)\begin{pmatrix} k_{it} \\ l_{it-1} \\ m_{it-1}\end{pmatrix}\right] = 0$$

Under a stronger assumption, you may also include l_{it} if you assume labor also chosen in advance. Interpretation: Current period productivity "innovation" should be orthogonal to lagged input choices for variable inputs (i.e. materials) and orthogonal to current period input choices for dynamic inputs (i.e. capital... because chosen in previous period).

4.6.5 High-level strategies when estimating production functions

One strategy is to ignore all these issues. Or, just focus on a specific industry with very good data and where you can write our production/costs in specific detail (ie, Ryan (2012)).

In order to really address input/output price biases, some tips are:

- Interpret TFP as a measure of profitability $\rightarrow$ can use sales/expenditures.
- Focus on homogeneous-goods industries.
- Model markups directly.
- In all cases: make sure conclusions aren't sensitive to the approach.

Greenstone, List and Syverson (2012) The Effects of Environmental Regulation on the Competitiveness of U.S. Manufacturing

Conceptual framework:

$$Q = A(\lambda_L L)^{\alpha}(\lambda_K K)^{1-\alpha}$$

$\lambda_L, \lambda_K \leq 1$ if inputs need to be used for environmental compliance. This gives:

$$TFP = \lambda_L^{\alpha}\lambda_K^{1-\alpha}A$$

They want to estimate if the CAA leads to lower TFP due to decreases in the λs. Empirically estimate:

$$TFP_{it} = \sum_p [\beta_p NOATT_{cpt} + \delta_p POLL_{ip} + \gamma_p NOATT_{cpt} \times POLL_{ip}] + X_{it}\Phi + \eta_i + \varepsilon_{it}$$

This is a DDD (polluting industry-county-year) estimated for each pollutant p. They find:

1. Non-attainment for any pollutant decreases TFPR by 2.6%.
2. On average, when industry TFP rises by 1%, indust price falls 0.35% $\Rightarrow \Delta TFP_q = \Delta TFP_{rev} - \Delta \ln p = 2.6 - 0.35\Delta TFP_q \Rightarrow \Delta TFP_q = 2.6/0.65 = 4\%$ fall in TFP_q.
3. If firms are price-takers, then $\Delta TFP_{rev} = \Delta W$. Counterfactual output = 412.5B $\rightarrow \Delta W = 412.5 \times 0.026 = 11B$

Some issues:

1. Would be really nice to know some more about price effects (and/or markups)
2. Mechanisms? FC/MC? Decomposition?
3. Endogeneity of non-attainment?

De Loecker and Collard Wexler (2015) Reallocation and Technology: Evidence from the US Steel Industry

To what extent are productivity gains in steel driven by technological innovation versus reallocation to more productive plants? This is interesting/important bc it gets at the underlying drivers of productivity, which we know little about.

Their strategy: Use product level input usage, output to construct plant-level price indices for inputs and outputs. Their production function model:

- Cobb-Douglas in labor, materials, capital
- Estimate separately for mini-mills, vertically integrated (VI) producers
- ACF estimation

	Input and Output Price Deflators		No Plant-Level Output Price Deflator		No Plant-Level Input Price Deflator	
	GMM	OLS	GMM	OLS	GMM	OLS
	I	II	III	IV	V	VI
Material	0.680	0.631	0.650	0.610	0.680	0.631
	[0.65 0.73]	*[0.58 0.69]*	*[0.62 0.70]*	*[0.52 0.67]*	*[0.64 0.73]*	*[0.58 0.69]*
Labor	0.274	0.327	0.282	0.332	0.273	0.327
	[0.24 0.31]	*[0.28 0.37]*	*[0.24 0.32]*	*[0.29 0.38]*	*[0.24 0.31]*	*[0.28 0.37]*
Capital	0.079	0.050	0.082	0.051	0.082	0.050
	[0.04 0.11]	*[-0.01 0.10]*	*[0.05 0.11]*	*[-0.01 0.10]*	*[0.05 0.11]*	*[-0.01 0.10]*
VI premium	-0.075	-0.018	-0.038	0.013	-0.076	-0.018
	[-0.12 -0.04]	*[-0.04 0.00]*	*[-0.08 0.00]*	*[-0.01 0.03]*	*[-0.12 -0.04]*	*[-0.04 0.00]*

Then they decompose changes in productivity into changes in:

1. Average productivity of MM, IV plants.
2. Covariance output, productivity (reallocation)
3. Reallocation across technologies

Table 7: Static Decompositions of Productivity Growth (Change 1963-2002)

Aggregate TFP $\Delta\Omega$	22.1%	
Olley-Pakes Decomposition:		
Unweighted Average: $\Delta\overline{\omega}$	15.7% (0.71)	
Covariance: $\Delta\Gamma^{OP}$	6.4% (0.29)	
Between Decomposition:		
Unweighted Average: $\Delta\overline{\Omega}$	17.0 % (0.77)	
Between Covariance: $\Delta\Gamma^{B}$	5.1 % (0.23)	
Within Decomposition:	Minimills	Integrated
Aggregate TFP: $\Delta\Omega(\psi)$	9.6%	24.3%
Unweighted Average: $\Delta\overline{\omega}(\psi)$	5.4% (0.55)	18.4% (0.83)
Within Covariance: $\Delta\Gamma^{OP}(\psi)$	4.4% (0.45)	3.7% (0.17)

Find:

1. Gains from average prod. > gains from reallocation (15.7 vs. 6.4)
2. Gains from average greater than reallocation across IV/MM (17 vs. 5.1)
3. VI saw greater average productivity increases.
4. 14-25% increase in CS depending on elasticity of demand, share of price fall due to minimills.

4.6.6 Estimating markups

At a high level, the deviation between output elasticity of input and revenue growth identifies markup. Take a cost-minimizing firm's Lagrangian:

$$L = \sum_v P^v V^v + rK + \lambda(Q - Q(V, K, \Omega))$$

From the first order condition, we have:

$$P^v = \lambda \frac{\partial Q}{\partial V^v} \Rightarrow \underbrace{\frac{\partial Q}{\partial V^v} \frac{V^v}{Q}}_{\text{output elast.}} = \underbrace{\frac{P}{\lambda}}_{\text{markup}} \times \underbrace{\frac{P^v V^v}{PQ}}_{\text{Input share of rev.}}$$

Note: $\lambda = MC$ since $dL/dQ = \lambda$. Thus the markup can be identified given an estimate for the output elasticity and the revenue share of the input. If we had prices, we could just directly estimate MC from estimated markups.

Estimating costs without cost data

... or, demand-side estimation. If we assume a specific utility function and pricing rule, these imply particular markups. Once we know their structure, we can estimate costs via:

$$\ln MC = \ln P - \ln MU$$

ex: BLP utility or CES preferences.

Advantages	Disadvantages
Full (partial equil.) modeling of the market. Permits counterfactual analysis	Result depends on assumptions. Most commonly applied to case studies where institutional setup can inform assumptions
Mechanisms clear	Emerging consensus: approach does well in cross-section.
Not good at explaining time series of price/markups	

Table 4.2: Benefits and drawbacks of demand-side cost estimation

In contrast, production function approaches to estimating markups:

- Arguably fewer assumptions
- Apply to many industries
- Can be implemented using manufacturing firm surveys, which are increasingly available

Ganapati, Shapiro and Walker (2017)

How do externality correcting taxes affect consumers and producers (and ultimately welfare) in imperfectly competitive product markets?

Their approach:

- Reduce dimensionality of problem to estimating a set of reduced-form parameters (e.g., cost pass-through).
- Derive sufficient statistics expressions for welfare/incidence of input taxes under imperfect competition.

Define: Incidence $= I = \frac{dCS/dt}{dPS/dt}$, pass-through rate $= \rho = \frac{dp}{dt} = \frac{1}{1+\epsilon_D/\epsilon_S}$. We know that:

$$I^{PC} = \frac{\rho}{1-\rho} \qquad I^M = \rho$$

under perfect competition, we can expect:

$$\Delta CS = \rho Q^* \qquad \Delta PS = (1-\rho)Q^*$$

Bottom line: Pass-through is a sufficient statistic for incidence under perfect comp. and monopoly. More generally, let:

$$\theta = \left(\frac{p - mc}{p}\right) \times \epsilon_D$$

then the general incidence formula when the tax is on inputs is:

$$I = \frac{\rho}{dmc/dt - (1 - L \times \epsilon_D)\rho} = \frac{\rho_{MC}}{1 - (1 - L \times \epsilon_D)\rho_{MC}}$$

Then they:

1. Estimate cost-pass through only to bound incidence between the perfect competition and monopoly cases.
2. Use plant-level data to estimate production functions for a subset of homogeneous goods.
   - Translog, gross-output production function with 3 factors (K, L, M)
   - Output is quantity (not revenue), avoiding output price bias

	Energy Cost Share (1)	Output Elasticities: Labor (2)	Materials (3)	Capital (4)	Returns to Scale (5)	Markup (6)	Observations (7)
Boxes	0.02	0.04	0.95	0.04	1.00	1.47	1414
Bread	0.02	0.28	0.63	0.09	1.13	1.20	248
Cement	0.33	0.91	1.08	0.19	2.46	2.30	229
Concrete	0.02	0.11	0.68	0.16	1.09	1.12	3369
Gasoline	0.88	0.01	0.99	0.03	1.02	1.11	284
Plywood	0.02	0.02	0.95	0.11	0.92	1.48	139
Mean	0.02	0.10	0.070	0.14	1.09	1.15	5683

Note: Translog, 3-factor (K, L, M), gross-output production function. Materials include electricity+fuels. This table shows mean values of energy cost shares, output elasticities, and markups. An observation is a plant-year.

3. Use optimal input choices and production to infer productivity and markups $\rightarrow$ estimate MC from prices, markups.
4. Examine how marginal costs, markups, and unit prices are affected by (plausibly exogenous) changes in energy costs
   - Exploit balkanized nature of U.S. electricity markets $\rightarrow$ Different regions use different fuel mixes to generate electricity.
5. Estimate ρ_{MC} by instrumenting MCs with the fuels share $\times$ fuel price instruments:

$$p_{it} = \rho_{MC} mc_{it} + X'_{it}\gamma + \eta_i + \epsilon_{it}$$

Find:

1. Fuel prices increase MCs: 1% increase in gas price $\rightarrow$ 0.29% increase in MC. Avg. gas share = 25% $\rightarrow 0.25 \times 0.29 = 0.07\%$ increase in MC.
2. Fuel prices decrease markups: 1% increase in gas price $\rightarrow 0.25 \times -0.069 = -0.02\%$ change in markups
3. Pass-through $\rho_{MC} = 0.715$
4. With ρ, and estimates of L, $\epsilon_D \rightarrow$ Estimate incidence for each industry, bounding the estimates with PC, monopoly cases. Find consumers bear most of the burden, but not as much as generally believed.

	Lag (t-0) (1)	Lag (t-2) (2)	Lag (t-5) (3)	Lag (t-0) (4)	Lag (t-2) (5)	Lag (t-5) (6)
Coal Price × Coal Share	0.092	0.156	-0.110	0.357	0.374	0.123
	(0.387)	(0.363)	(0.311)	(0.244)	(0.293)	(0.255)
Gas Price × Gas Share	0.779***	0.788***	0.866***	0.235***	0.225***	0.291***
	(0.175)	(0.140)	(0.191)	(0.086)	(0.084)	(0.061)
Oil Price × Oil Share	0.136	0.229	0.013	-0.070	0.047	-0.029
	(0.341)	(0.290)	(0.207)	(0.121)	(0.118)	(0.139)
Plant FE	X	X	X	X	X	X
Year FE	X	X	X			
State Trends	X	X	X	X	X	X
Region-Year FE				X	X	X
Product-Year FE				X	X	X

	Lag (t-0) (1)	Lag (t-2) (2)	Lag (t-5) (3)	Lag (t-0) (4)	Lag (t-2) (5)	Lag (t-5) (6)
Coal Price × Coal Share	-0.012	-0.092	0.049	-0.098	-0.215	-0.058
	(0.220)	(0.183)	(0.171)	(0.156)	(0.172)	(0.172)
Gas Price × Gas Share	-0.288***	-0.286***	-0.334***	-0.049	-0.021	-0.069*
	(0.080)	(0.066)	(0.096)	(0.045)	(0.053)	(0.038)
Oil Price × Oil Share	-0.035	-0.056	0.065	0.062	0.010	0.107
	(0.181)	(0.134)	(0.097)	(0.090)	(0.069)	(0.077)
Plant FE	X	X	X	X	X	X
Year FE	X	X	X			
State Trends	X	X	X	X	X	X
Region-Year FE				X	X	X
Product-Year FE				X	X	X

	(1) Boxes	(2) Bread	(3) Cement	(4) Concrete	(5) Gasoline	(6) Plywood
Panel A: Incidence Components						
MC Pass-Through (ρ_{MC})	1.43	0.69	1.81	0.74	0.31	1.08
Demand Elasticity (ϵ_D)	3.24	2.42	1.82	5.53	8.70	1.39
Mean Lerner Index (L)	0.33	0.18	0.57	0.13	0.12	0.41
Panel B: Consumer Share of Burden (by Market Structure)						
Symmetric Oligopoly	0.57	0.53	0.63	0.49	0.24	0.67
Asymmetric Oligopoly	0.60	0.66	0.69	0.63	0.25	0.87
Monopoly	0.59	0.41	0.64	0.43	0.24	0.52
Perfect Competition	1.43	0.69	1.81	0.74	0.31	1.08

Muehlegger and Sweeney (2017) Pass-Through of Own and Rival Cost Shocks: Evidence from the U.S. Fracking Boom

Fracking reduced costs for US oil producers. How do these costs reductions affect equilibrium outcomes? Idea: The magnitude of strategic response depends on nature of competition and whether the shock is market-wide or firm-specific.

Conceptual framework: Let MCs of frim i be $c_i = \alpha + \alpha_i$. Then pass-through of a cost shock is:

$$\rho_\alpha = \sum_i \frac{\partial p}{\partial \sigma_i} \left(\frac{\partial \sigma_i}{\partial \alpha_i} + \sum_{j \neq i} \frac{\partial \sigma_i}{\partial \sigma_j} \frac{\partial \sigma_j}{\partial \alpha} \right)$$

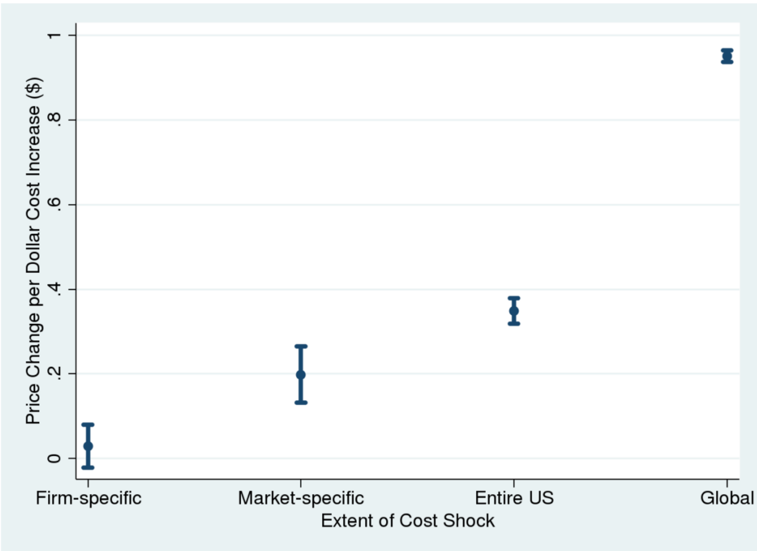
where σ is the firm's strategy variable. Example: in Cournot σ is the firm's quantity. If the shock is market-wide, then the pass-through is amplified through market-wide reductions in quantity. Takeaways:

- Pass-through depends on nature of cost shock. Firm-specific shocks lead to lower pass-through.
- Patterns of pass-through distinguish the nature of competition.

Empirical setting: Shale boom temporarily reduced input costs for refineries near deposits. Observe reductions in input costs for some refiners, while the costs for other firms (sometimes in similar areas + identical products) was unchanged.

Regressing prices on costs generally leads to bias: OVB from competitor's costs. They address this by controlling for observables in rival's costs. The idea: controlling for rival's costs (market-time FEs) and world oil prices leads you to estimate the effect of only firm-specific shocks. Coarsening the FEs $\rightarrow$ estimate pass-through for wider-reaching shocks. Finds:

1. Controls matter: with time$\times$geography dummies (ie, firm-specific shocks), little pass-through
2. Pass-through increases with scale of the shock.



4.7 Lecture 13: Principle-Agent problems in EEE

Two polar models:

- **Moral hazard:** hidden actions or imperfect information where the Principal cannot perfectly observe all of these actions: she observes only a noisy signal about these actions.
- **Screening/adverse selection:** Principal-agent models w/ hidden knowledge, or under incomplete info where Agent has private information on some relevant parameter for the transaction. The Principal does not know this parameter and has only prior beliefs on its value (Bayesian setting).

4.7.1 Rules versus discretion

Regulatory agencies enforce standards using imperfect information on the parties they regulate. Imperfect info means we should maybe be flexible (allow discretion), but discretion can lead to bad behavior among enforcers.

Duflo, Greenstone, Pandne and Ryan (2018) The Value of Regulatory Discretion: Estimates From Environmental Inspections in India

Study this by running a field experiment where they increased the frequency of inspection for a sample of industrial plants in the Indian state of Gujarat. Two treatment arms:

1. Provided resources to bring all plants up to minimum required frequency of inspections
2. Removed regulator's discretion: inspections allocated randomly across all treatment plants.

They then estimate a dynamic discrete game between the regulator and emitting firms using the data on chains of interaction → estimate conditional choice probabilities for regulatory actions depending on pollution, past actions and the implicit cost of compliance for plants.
Find:

1. Increased inspections lead to more violations being found, more citations, BUT
2. Tiny increase in compliance, no increase in penalization, no change in emissions.
3. Why? They think removal of discretion: the regulator only punishes egregious violators, and randomizing plants leads to no increase in the number of extreme violators found.
4. Game: costs of abatement much smaller than cost of penalties.
5. Uniform rule does poorly → discretion leads to better regulation.

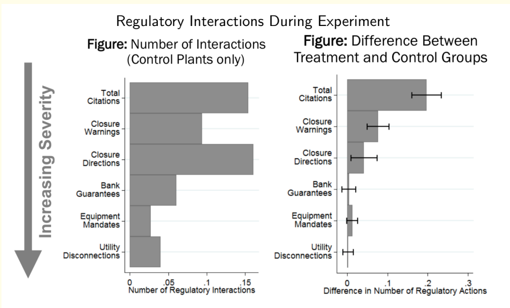
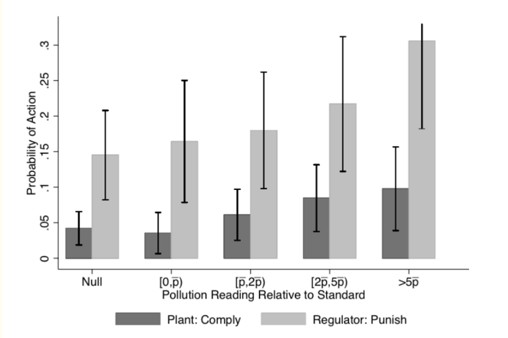


Figure 5: Predicted Probabilities of Compliance and Punishment

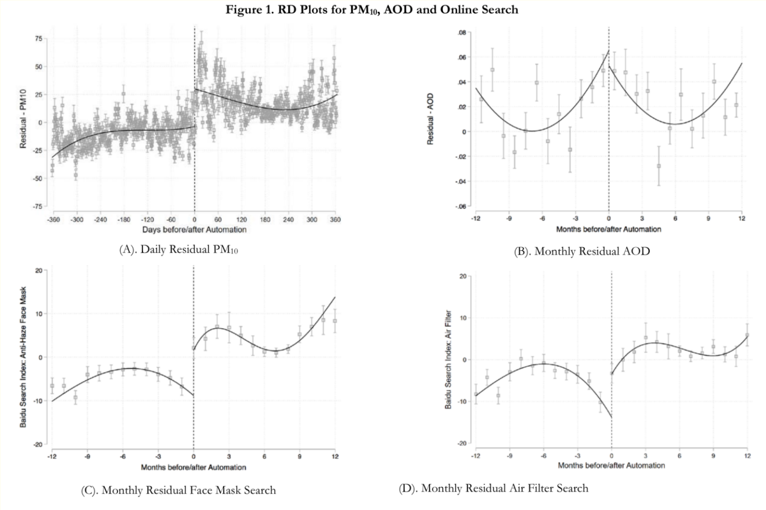


Greenstone, He, Jia and Li (2022) Can Technology Solve the Principal-Agent Problem? Evidence from China's War on Air Pollution

In China, local officials have high-powered incentives to achieve certain economic, social targets → incentive to cheat on pollution measures? Does under-reporting exist, and how does automated monitoring by central gov't alter these incentives?

Empirical setting: Introduction of centralized pollution monitoring from central govt → High frequency event study (or RD) + longer run event study. Find:

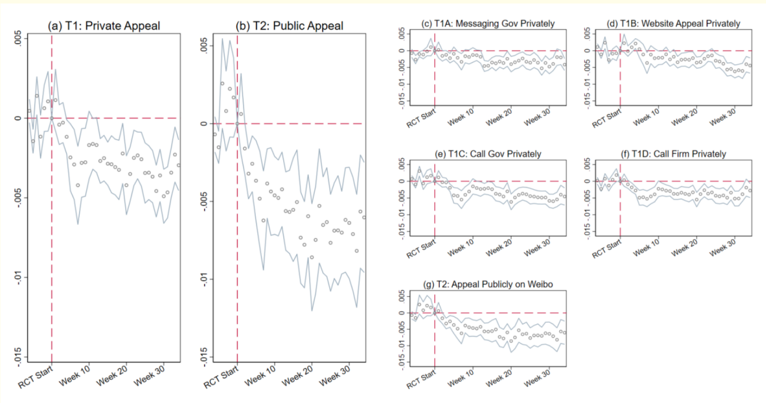
1. Evidence of the under-reporting of air pollution before automation, improvement in data quality after.
2. No discontinuity in Aerosol Optical Depth (AOD) around the automation date → no change in satellite-based measures
3. Lots of variation in treatment effects across cities. City-level under-reporting negatively corr. with GDP; positively corr. with PM10.
4. Automation →↑ defensive investments ⇒ welfare losses from biased info.



Buntaine et al. (2021) Citizen monitoring of waterways decreases pollution in China by supporting government action and oversight

Asks: When does citizen participation affect regulation and pollution? They run a field experiment where they randomly vary how citizen appeals about pollution standard violations are sent to regulators and violating firms. Treatments:

- Private appeal to gov't via message.
- Private appeal to gov't via website.
- Private appeal to gov't by phone.
- Private appeal to firm by phone.
- Generic appeal on social media.
- Boosted appeal on social media.



Cicala, Hemous and Olsen (2021) Adverse Selection as a Policy Instrument: Unraveling Climate Change

They study asymmetric information in the context of a Border Tax Adjustment to combat regulatory leakage. The idea: International sovereignty may restrict what governments can mandate of foreign firms, but does not foreclose the possibility of creating incentives to shape their behavior:

- Create certification mechanism as an incentive for relatively clean firms to separate themselves from more intensive polluters.
- Uncertified firms have an output-based fee that tracks average rate of damage among those choosing not to participate in certification program.
- This leads to an unraveling in favor of certification.

Find:

1. Optimally-implemented certification program would achieve nearly 75% of welfare gains of universal carbon tax.

2. Potential issue: **backfilling**, where dirty firms expand to serve the untaxed foreign market.