Models of Markets
(from empirical work to theory and back).

Ariel Pakes, Harvard University.

July 1, 2025

"The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier."

From: T. Haavelmo, in his 120 page Nobel Prize winning treatise, "The Probability Approach to Econometrics", *Econometrica*, 1944.

• Empirical work in I.O. Industrial Organization took Haavelmo quite seriously, and much of our profession seems to be coming along.

• This was facilitated by

- prior developments in theory (both economic and econometric),
- the increased availability of computer power (both for easing computational problems and for the generation of data),
- and the absence of a credible alternative way of analyzing counterfactual environments.

• Consider steps that might strengthen our ability to do counterfactuals.

Lessons from productivity analysis (Olley and Pakes,1996).

• Start with what I did & did not learn from this paper.

• The breakup of A.T.&T and productivity (price × quantity divided by an index of costs) in the telecommunications equipment industry.

- Empirically: the differential movement in the productivity measure was highly predictive of the growth, decline, & exit of firms.
- Partition changes in productivity into : i) reallocation of quantity to more productive plants and ii) changes in plant average productivity.

$$p_t = \sum_i s_{i,t} p_{i,t}, \text{ and } p_t \equiv \overline{p}_t + \sum_i s_{i,t}(p_{i,t} - \overline{p}_t).$$

- Productivity table.
  - Partial deregulation, 1975 (registration and certification program). Results in a shift in output to more productive plants
  - Consent decree (1982) to completion of breakup (1984).

TABLE 1

Decomposition of Productivity

Year	p_t	$\overline{p}_t$	$\Sigma_i \Delta s_{it} \Delta p_{it}$	$corr(p_t, k_t)$
1975	0.72	0.66	0.06	-0.11
1976	0.77	0.69	0.07	-0.12
1977	0.75	0.72	0.03	-0.09
1978	0.92	0.80	0.12	-0.05
1979	0.95	0.84	0.12	-0.05
1980	1.12	0.84	0.28	-0.02
1981	1.11	0.76	0.35	0.02
1982	1.08	0.77	0.31	-0.01
1983	0.84	0.76	0.08	-0.07
1984	0.90	0.83	0.07	-0.09
1985	0.99	0.72	0.26	0.02
1986	0.92	0.72	0.20	0.03
1987	0.97	0.66	0.32	0.10

Two other productivity facts.

• Prior to deregulation (in 1975) plant level capital was negatively correlated with productivity.

- A.T.& T : monopoly & purchased from their subsidiary, Western Electric, which had little no incentive to innovate.
- The breakup: Entry (Ericson, Northern Telecom & Hitachi) and the "baby bells" chose their more modern equipment.

• $\overline{p}_t$ declined rather sharply after the breakup.

- This is a period of rapid technological advance in the industry.
- Recall, this is revenue productivity and entry caused prices to fall.

• **Using productivity to understand the impact of events on firm growth & the intraindustry allocation of revenue seemed to work, but productivity was less helpful for understanding welfare.**

Why couldn't we dig deeper into welfare?

- There were two problems, and by now they are both widely recognized.
  - We had data on revenue, but not separately on price and quantity.
    - In some instances better data has become available (less aggregated measures of price and/or quantity), and has been helpful.
    - There was an attempt to go directly to markups, which gets to part of the problem. The jury is still out on this, but either way it lacks a basis for understanding why we have them, which is the core of the problem.
  - The dynamic models that were available were demanding. Even if demand and cost systems are known they
    - require cognitive abilities that the agents making the decisions are unlikely to have
    - lead researchers who ventured into the empirical analysis of dynamics to simplify in ways that, though perhaps the best available analysis, made it harder to believe the empirical results.

Welfare: Needs Demand and Pricing Equations.

• Differentiated products demand systems provided new techniques for handling old problems, and have been fairly effective, at least for analyzing a cross section of choices. This was kicked off by the two BLP articles (1995, 2004), and by now the basics are well known.

• How about prices? First cross sectional analysis of retail markets.

- Example Trucks: Wollman's demand system for trucks generates markups Then price on characteristics determining cost + markup term (should get a coefficient of one). Instruments for markup.
- The good news is how well the static Nash in prices assumption tends to fit cross-sections of product prices in retail markets (in some cases you need to be careful about market definition).
- Hedonics predicted the importance of own characteristics but had no role for competitor characteristics; and its coefficient is close to the predicted value of one.

- Example Pakes (2021): Wollman's (2018) data/estimates/computations.

	Price	(S.E.)	Price	(S.E.)
Gross Weight	.36	(0.01)	.36	(.003)
Cab-over	.13	(0.01)	.13	(0.01)
Compact front	-.19	(0.04)	-0.21	(0.03)
long cab	-.01	(0.04)	-0.03	(0.03)
Wage	.08	(.003)	0.08	(.003)
$\widehat{Markup}$	.92	(0.31)	1.12	(0.22)
Time dummies?	No	n.r.	Yes	n.r.
R^2	0.86	n.r.	0.94	n.r.

- Time dummies pick up most of what the characteristics do not. Predictions for price changes almost entirely determined by markup changes (reflective of changes in the competitive environment). $R^2 \approx 50\%$. More room for improvement here.

- *Most, if not all, counterfactuals will require a pricing assumption. We know little about fit in*
  - *pricing models that emanate from vertical relationships*
  - *when firms realize that consumers reactions to price changes depend either on past choices or perceptions of future opportunities.*

Dynamics (goal: credible counterfactuals).

- I am going to divide my comments on dynamics into
  - single agent dynamics, mostly on consumer choice, and
  - the evolution of markets
- In both the focus will be on *better approximating observed behavior by limiting the cognitive requirements of our models.*
- Dynamic models of consumer behavior are
  - needed to analyze prices in any market where either (i) past choices or (ii) perceptions of future choices, are important, and
  - the increasing availability of panel data on discrete choices should enable us to do much more on this.

Single agent choices.

- I distinguish between two situations.

1. The perceptions of the future are integral to understanding behavior.
   - Prominent examples; investment decisions (firms or individuals).
   - Many advances here. Initial papers of Miller, 1984, Wolpin, 1984, Pakes, 1986, & Rust, 1987 (who developed an oft used framework).
2. Decisions which, because the implications of a bad choice can
   - easily be reversed if needed, &/or where repeated evaluations
   - have cognitive costs that rival implied utility gains.

- Inattention implies choice models need the impact of a policy *conditional* on *both*; (i) preferences and (ii) past choices.

• Inattention implies choice models need the impact of a policy *conditional* on *both*; (i) preferences and (ii) past choices.

• Some of the determinants of preferences are not observed and are relatively stable over time making them correlated with past choices.

• So if we put past choices into the model with no correction for unobserved preferences, its coefficient (say κ) will pick up their impact.

• Other coefficients will also be biased. E.g. for pricing responses is the fact that people stay with their last choice when prices change

- because they like their last choice, or because of inattention.

• If the unobserved preferences do not change much over short time intervals we can control for them by conditioning on individual-by-choice specific fixed effects.

• Similar to the the within-between analysis used in all the social sciences, but because of discrete choice we need moment inequalities.

Health insurance choices

- Question of policy interest: Is low response to price changes a concern?
  - The CommCare Data. Insurance for those with income less than 3$\times$ the federal poverty line (FPL=$10,380 in 2010).
  - We want switching costs that are not induced by a change in agent's physical environment, just by prices so we drop people who had changes in their choice set.
  - We need switches so we require data on sequences of choices (2 for upper bound, and 3 for lower); use shortest time interval for each.
  - Form cells: Cartesian product of a) couple of years, b) sequences of income (five income groups), d) region, and d) plan availability.
  - Only within-cell comparisons: fixed effects represent preference differences within these cells (partially linear utility function).

Logic of Non-Parametric Bounds

• Pakes, Porter, Shepard and CalderWang (2024) model the utility for consumer i of choice d at time t as

$$U_{i,d,t} = \underbrace{(-p_{d,i,t} - \kappa_0 \cdot 1\{y_{i,t-1} \neq d\}) \cdot \gamma_i \ + \ \lambda_{d,i}}_{\text{Structural Utility } (=SU_{d,i,t})} \ + \ \underbrace{\varepsilon_{d,i,t}}_{\text{Error}}$$

where $\lambda_{d,i}$ are product-by-individual fixed effects, γ_i allows each indivicual to have a different price coefficient, and κ_0 gives us the impact of past choices relative to the price coefficients.

• Central assumption. Once we allow for individual-by-product fixed effects & past choices, the remaining unobservables in the agent's utilities for the various choices are identically distributed over time.

• Then revealed preference makes the economic intuition underlying the bounds on parameters transparent.

- **Upper bound**. Take a group of agents who are at c in $t-2$
  - The relative price of c rises in $t-1$ and a portion switch out.
  - For each time t price change, there is a κ that is large enough to insure the fraction of the original group that switched out in t should be greater than that fraction in $t-1$.
  - If the fraction of the group who switched out in t is smaller than at $t-1$ then κ must be bounded from above. (Contrapositive)
- **Lower bound.** A group of agents who are at c in $t-3$
  - chose c at $t-2$.
  - the relative price of d goes up and some switch out.
  - then the price of c falls to a lower level then at $t-2$ yet less chose c in t than in $t-2$.

Health Insurance Choices: Income $\leq 3\times$ FPL ($\approx$ 10K).

- Point estimates from models without fixed effects on this data.
  - Detailed published model; \$1,164 (a different cut of CommCare data).
  - Our estimates; 250 rhs variables; \$1019 (as in other papers, the more rhs variables put in, the lower the estimate of κ).

Sample	**ID Set (κ)**		**95% CI**		**99% CI**	
	LB	UB	LB	UB	LB	UB
Min cell size 20	\$150	\$168	\$150	\$270	\$150	\$354
Min cell size 50 (main sample)	**\$102**	**\$186**	**\$78**	**\$294**	**\$78**	**\$450**
Min cell size 100	\$78	\$186	\$78	\$450	\$78	\$450

Note: Estimates from our nonparametrics estimator of the identified set and 95% and 99% confidence sets (or credible sets) for the switching cost κ in dollars per year.

- Recall that one κ CS [78,450].

Table: Switching Cost (κ) Estimates in Nonparametrics Heterogeneity Model

	Low-Income Group	**High-Income Group**
By Income-Only Model		
ID Set	**[$240, $257]**	**[$408, $432]**
95% CSet	*[$240, $343]*	*[$408, $480]*
99% CSet	*[$240, $369]*	*[$408, $516]*
By Income x Health Model		
Healthy ($s_{it} = 0$): ID Set	**[$292, $377]**	**[$432, $528]**
95% CSet	*[$240, $377]*	*[$408, $528]*
99% CSet	*[$240, $377]*	*[$408, $528]*
Sick ($s_{it} = 1$): ID Set	**[$75, $154]**	**[$120, $216]**
95% CSet	*[$0, $171]*	*[$0, $240]*
99% CSet	*[$0, $189]*	*[$0, $264]*

- Counterfactuals: We flattened out a price spike in the data & compared its impact with the different κ. Over two years they obtain different signs.

Back to the Complexity of Market Dynamics (assuming demand, cost, and pricing is known).

- The initial dynamic frameworks assumed[1]

1. state variables evolve as a Markov process (tractable and often verifiable, so we keep it)
2. and the equililbrium is some form of Markov Perfection (no agent has an incentive to deviate at any value of the state variables).

- The complexity of the perfectness assumption led to; (i) doubts about its ability to mimic behavior, and (ii) implementation problems for researchers.

- The ways one thought to circumvent the problems, like asymmetric information, just led to even more complex frameworks.

[1] Maskin and Tirole (1988) I.O. theory; Ericson and Pakes (1995) applied work.

- Both theory & empirical work explored ways forward. Theory
  - weaker notions of equilibrium that were less cognitively demanding, &
  - learning processes that mimic how firms respond to change.
- Examples
  - Fudenberg and Levine (1993); Self-confirming equilibrium.
  - Osborne and Rubinstein (1998); Procedurally rational players.
  - Esponda and Pouzo (2016); Berk-Nash equilibrium (improper priors)

• Done in context of repeated game where the focus is on learning how competitors play. Related empirical work; Hortascu and Puller, 2008, Doraszelski, Lewis,& Pakes,2018, on electricity markets.

• Most of what we are interested in concerns the determinants and implications of investments in the "payoff relevant" states. To use these ideas to analyze those issues we need to modify the theoretical framework.

Cognitively less demanding framework for dynamics.

- Experience based equilibrium (Fershtman and Pakes, 2012) assumes
  1. Maximizing behavior. Agents chose actions which maximize their perceptions of the EDV of future net cash flow conditional on the information at their disposal.
  2. Partial consistency of perceptions.Their perceptions need not be correct but they are consistent with what they observe.
     - If a state is visited repeatedly they can learn the distribution of profits at the state, and optimize against that.

• Different agents can have different information sets (asymmetric information) and the theory does not constrain what they contain.

• This can greatly reduce cognitive requirements, but still has a perfectness assumption (just on a reduced state space).

• **Note.** *Though the theory is silent on what the state variables are, empirical work must find a good approximation to them.*

Moving to Data (Dubois and Pakes, 2025?)

- Two models
  - Empirical Model: Assumes maximizing behavior but does not assume correct perceptions of the future.
  - Equilibrium Model: Imposes the constraints of an EBE.

What is the role of the empirical model?

- Determine the state variables which consist of
  - variables in the data that advertising responds to, and
  - properties of unobservables (determinants of advertising that the firm responds to but are not in our data).
  - clarify the aspects of the the model needs replicate in order to provide a credible model of behavior

- Empirical model only differs from previously used models (Benkard et. al. 2007; Pakes et. al. 2007) in that
  - we do not assume equilibrium perceptions.
  - we do allow for serially correlated unobserved state variables.

The Equilibrium Model.

- Characteristics of equilibrium:
  - *Asymmetric information, firms only condition on (& formulate policies for) the states in their own information set (eases cognitive burden)*.
  - generates a Markov process for the vector formed from combining the state variables of all firms (for market structures).

Comparison to prior work.

• Two types of prior empirical analysis: helpful but perhaps not adequate for counterfactual analysis.

- Focus on implications of perturbations in incentives for one or more firms' investments holding all else constant. Equilibrium responses to an environmental change might differ.
- Use approximations (Benkard et. al. 2008). The environment that was approximated differs from the counterfactual environment

• The extent to which equilibrium effects would differ depends on the particular policy.

• We focus on the likely effects of banning Direct to Consumer Advertising of prescription pharmaceuticals.

Motivation.

- US is the largest pharmaceutical market in the world in both revenue ($\approx$ \$600B per year) and promotional spending: $\approx$ \$7B on Direct To Consumer (DTC) Advertising (3/4 on TV), and $\approx$ \$20B on detailing (promotion to prescribers)
- Only two developed countries allow DTCA of prescription drugs: New Zealand and the US (since 1985 in print, but for TV only in 1997).
- Clear that innovation reacts to pharmaceutical profits (Acemoglu and Linn, 2004, Blume-Kohout and Sood, 2013, Dubois et al., 2015); but the impact of advertising on profits is unclear.

• Prior economic empirical literature; implications of perturbations in firms' DTCA policy (notably Shapiro (2016,2018), and Sinkinson and Starc (2019)). Large health literature on the history & effects of DTCA.

Is DTCA Socially Beneficial?

- Arguments on direct welfare effects.
  - *Against*: (i) incentives for excessive use, (ii) returns largely a result of business stealing (no net benefit to society)
  - *For*: make (i) consumers aware that they can treat a condition before it becomes serious and/or improve adherence to regimen, (ii) providers aware of possible treatments
- Importance of impact on profits:
  - Back of envelope impact of pharmaceuticals on welfare (Ho and Pakes, 2025): $>>$ 3.25 trillion \$US a decade
  - R&D costs (private + public) $\approx$ 900 million and 80% is private.

What do we want to learn?

- DTC induces doctor visits where patients request a particular drug hoping the doctor (& formulary) are compliant. Do we need a model of detailings' response to get credible answers to the policy change?
- Both DTC and detailing are dominated by the highest selling products. Would banning DTC change market structure?
- How would banning DTC impact profits and hence the incentives to do R&D? New policies make this particularly worrisome.
- What would be the effect on demand for drugs? Can not distinguish who stops.

Differences across Markets.

• If DTC is particularly impactful in markets where the usefulness of the medication is least self-evident then we would expect differences across the four markets we study as

- Cholesterol. The need for a cholesterol drug is typically a result of an interaction with a doctor and a Lipid panel test.
- Depression. Lack of mental health awareness is an oft cited reason for depressed people not seeking health.

On the other hand

- Asthma. Seek help for Asthma after experiencing breathing problems.
- Ulcers. Seek help after experiencing digestive problems. Ulcers is disappearing in developed countries (H. pylori, and Anti-acids).

Steps in the analysis

- Estimate a BLP demand system; recover the "quality" (ξ) terms for each drug in each period.
- Estimate a controlled Markov process for how ξ_t;, the controls are the two types of advertising.
- Formulate & estimate an empirical advertising model.
- Introduce the equilibrium framework and a novel (and easy to compute) algorithm for equilibrium policies.
- Compare equilibrium policies; (i) to the policy functions from the empirical model, and (ii) to the data.
- Compute & analyze equilibrium after shutting down DTCA .

Empirical Properties of Markets.

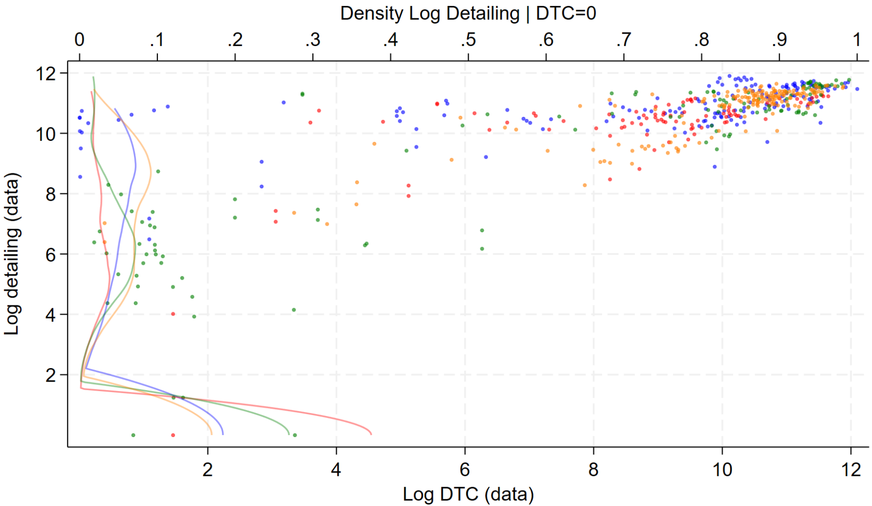
Detailing and DTC per product in thousands U.S. dollars per quarter.

	Det.	Top 4 Det.	DTC	Top 4 DTC	C4 Rev (Q)
AntiCholesterol	7,543	71,334	3,210	56,326	.79 (.70)
AntiAsthma	5,860	58,137	2,562	44,582	.65 (.66)
AntiDepression	4,603	71,066	1,698	47,906	.76 (.45)
AntiUlcer	2,508	44,991	671	53,016	.81 (.60)

- Markets are concentrated & largest selling products dominate DTC, and to a lesser extent detailing, in all markets.
- In markets where DTC per drug is high, detailing per drug is high (rank order perfectly correlated).
- Within markets DTC & detailing are also strongly related (picture).

• **Implication**: counterfactual requires counterfactual detailing.

Detailing versus DTC distribution



Blue: Anticholesterol, Red: Antiulcer, Green: Antidepressants, Orange: Antiasthma

Summary of Results for Empirical Model

• **Demand system:** random coefficients on price, molecule & time fixed effects, price interactions, patent status. 30-60 parameters.

• **Advertising's contribution to demand.**

- The two types of advertising are strategic complements. This implies that detailing will go down when we shut down DTC.
- Strong serial correlation in unobservable determinants of demand.

• **Serial correlation in disturbances.** Annual correlation coefficients between .82 and .89.

- **Reduced form results: Observables that advertising responds to**
  - The derivative of log profits w.r.t. advertising. The value function is an iterate of the profit function, but far too complex to compute. So using the impact of advertising on profits as a proxy is logical.
  - Time to loss of patent exclusivity. Advertising goes down as we approach the end of the patent life.
  - Advertising of competitors in the same therapeutic (ATC4) class has a negative impact. The profit function imposes a negative cross partial, so this accentuates the fact that competitors' advertising are strategic substitutes.

Parameters estimated for each market from advertising equation.

- marginal cost for each product.
- separate demand system and dynamic parameters (12) per market.

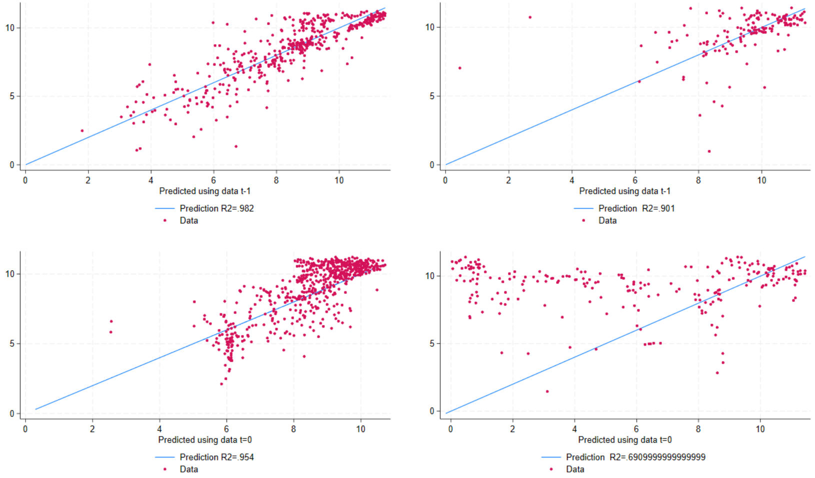
Assumptions used in counterfactual comparisons.

- The market is in equilibrium (an alternative would be to use a learning algorithm; e.g. in Fershtman and Pakes, 2012).
- Underlying processes for unobservables do not change.
- Both equilibria are selected
  - by using the states of the firm in the initial period as an initial condition (we are looking at what would have happened in the in-sample period, were the counterfactual institutions in place) and
  - to be boundary consistent.

• Caveat we are working on: Prices are held constant (we are trying different approaches here). Also still doing robustness.

• **Next**. Some pictures to illustrate fit.

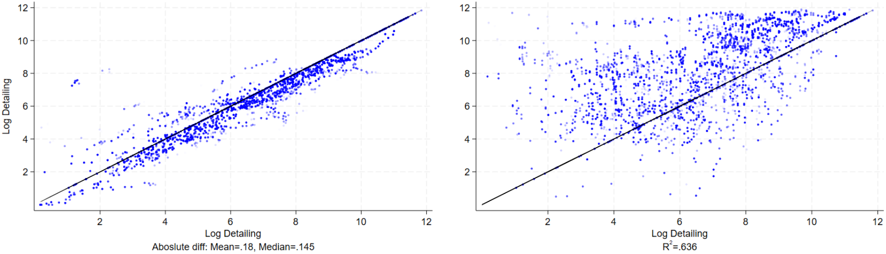
Results: Fit of the Empirical Model.



Note: Detailing on the left, DTC on the right. Using data t-1 for top graphs and only data at t=0 for bottom graphs.

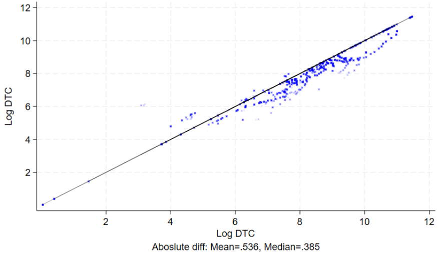
Results: EBE vs (i) Empirical and (ii) vs Data

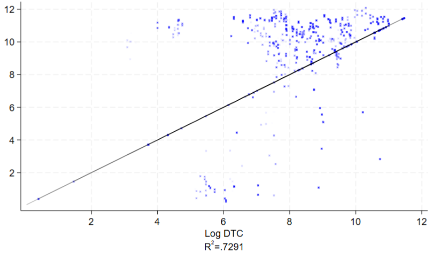
Detailing



- Left panel: empirical model vs EBE. Right panel: Data vs EBE.
- EBE is noticeably different from empirical model, but mimics the data about as well.

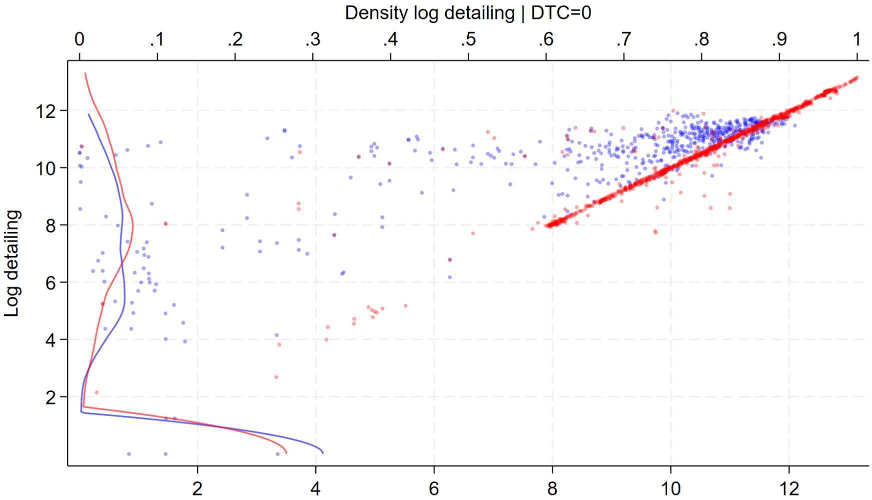
DTC.





- Left panel: Empirical model vs EBE. Right panel: Data vs EBE.
- EBE is noticeably different from empirical model, but mimics the data about as well.

Detailing vs. DTC in EBE (red) and data (blue)



Counterfactuals: Effects of Banning DTC

• Detailing falls dramatically; by 75% for Cholesterol, 64% for Asthma, 59% for Depressants, and 24% for Ulcers.

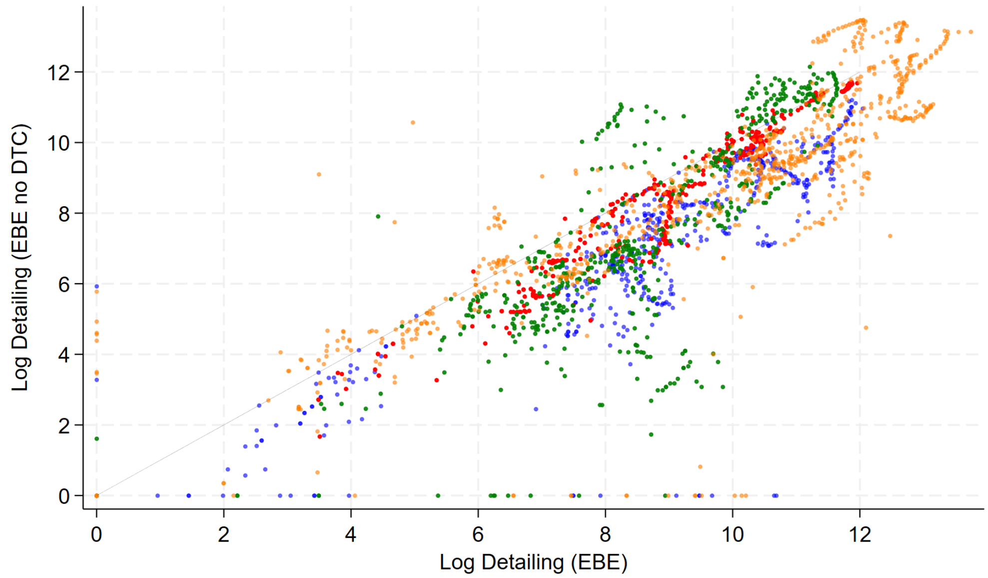
- The falls are concentrated in the firms that do DTC.

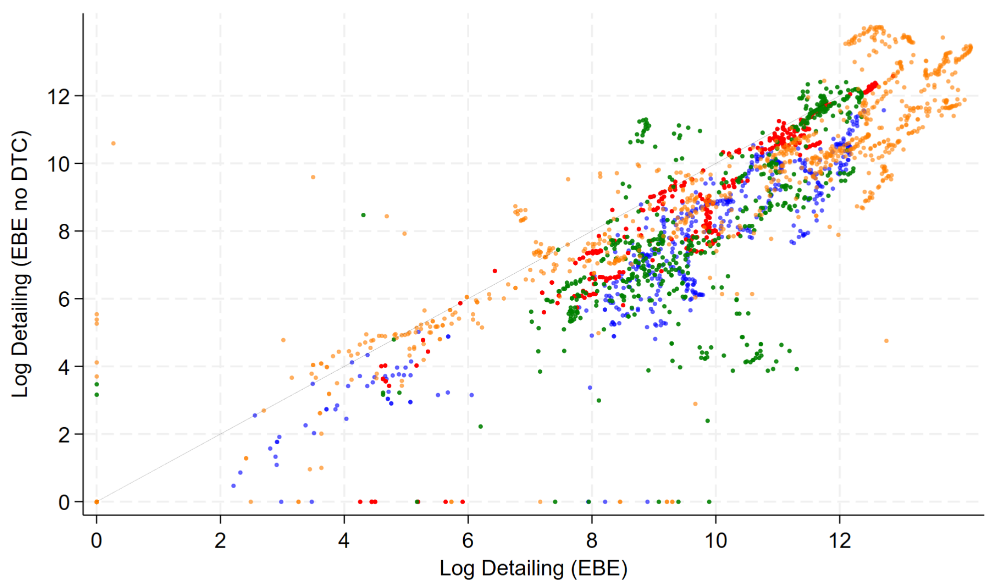
• Profit changes (they save detailing & DTC expenditures).

- Markets where DTC was needed to draw consumers: Cholesterol loses 7.6% and Depression loses 9.3%.
- Markets where disease is more self evident: Asthma increases profits by 5.4%, and Ulcers stays about the same (-..4%).

• In the markets where there were losses, they are all in patented drugs. Generics always increase their profits. Bodes poorly for R&D incentives.

Figure 18: Equilibrium Detailing if DTC banned versus Equilibrium Detailing when DTC allowed (all markets) (with price change)





Note: Detailing if no DTC on the vertical axis versus Detailing if DTC on horizontal. Colors for each market (Blue for Anticholesterol, Red for Antiulcer, Green for Antidepressants, Orange for Antiasthma). The density distribution of Detailing if no DTC when equilibrium Detailing with DTC is zero is plotted horizontally on the left vertical axis. Quarterly discount factor $\beta = .92$ for top graph and $\beta = .97$ for bottom graph.

Table 10 shows the EBE parameters when DTC is banned. Table **??** shows the results of the EBE for total quarterly market level detailing, DTC advertising and net profit with and without allowing DTC advertising. The results show that banning DTC advertising leads to a reduction of detailing expenditures,

- Sales become far less concentrated.
  - Those of the largest selling products fall dramatically.
  - The midsize and smaller firms often increase their sales; especially in Asthma which accounts for the increase in profits their, and this is also noticeable in Depression.
- Share of sick population that no longer uses the drug varies by between 1.1% and 3.1% on average over the years, but declines over time.

That is all we have for now.
Thanx for listening.

63