Product Differentiation, Consumer Learning, and The Value of Me-too Drugs

Neha Bairoliya[1]

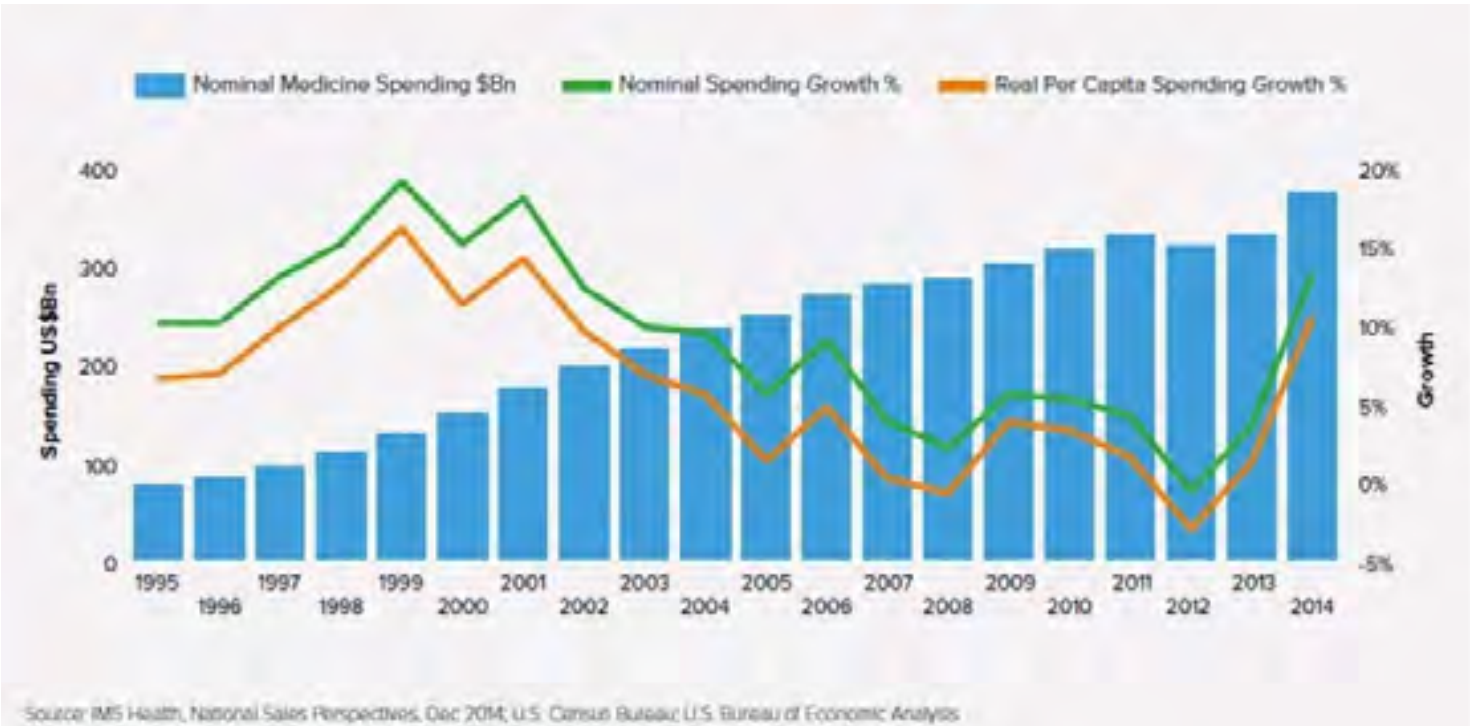
Pinar Karaca-Mandic[2] Jeffery McCullough[2] Amil Petrin[3]

December 12, 2015

[1]Harvard Center for Population and Development Studies
[2]School of Public Health, University of Minnesota,
[3]Economics Department, University of Minnesota

Introduction

- Pharmaceuticals are a large share of medical expenditures
  - $374bn in 2014 (IMS, 2015)
  - Part D: $62 billion in 2010, 12% of federal Medicare spending.



Patent Protection

- New drugs costs an average of $2.5 bn
  - e.g., DiMasi et al., 2014
- R&D financed through 20-year patents
  - Solves a public good investment problem
  - Leads to monopoly pricing
- Patents may lead to inefficient R&D allocation
  - May provide too-little incentive
    - Orphan drugs
  - May have "too-much" entry for large markets

Me-too Drugs

- Products are similar to other existing drugs
  - Each new chemical entity (NCE) is a unique molecule
  - Many use the same mechanism of action as other drugs
- Benefits
  - Efficacy and side effects may differ across patients and drugs
  - Price competition
- Inefficiency of free entry (Spence, 1976)
- Formularies may be too broad due to
  - Government policies
  - Uninformed consumers

Empirical Challenges

- Preference heterogeneity
  - Side effects and efficacy differ across drugs and patients
  - Idiosyncratic match values important for measuring welfare
  - Switchers have lower match values than those who stay
- Consumers are uncertain of their match values
  - They learn about their preferences over time
- Identification
  - Consumers with high demand may pick plans with lower copays
  - Prices and average product quality may be correlated
  - Working on identification issues now

What We Do

- Estimate structural model of R_x demand
  - Allow heterogeneous matches between consumers and drugs
  - Consumer learning about the stochastic match quality
  - Follows Crawford and Shum (2005), Shin et al., (2012)
- Recover covariance in match values
  - Essential for measuring the value of these products
  - The off-diagonal terms in the variance-covariance matrix
- Counterfactuals
  - Consumer welfare under alternative formulary designs
  - Value of formulary breadth is decreasing in the covariance
- Potential/alternative counterfactuals
  - What if line extension coverage were eliminated?
  - What if consumers were better informed
  - Consumer welfare loss from narrow formulary

Bayesian Learning Model

Utility

$$U_{ijt} = \underbrace{\beta_{ij,t-1}}_{\text{Drug Match Quality}} + \alpha X_{ij,t} + \epsilon_{ij,t}$$

where

$$\epsilon_{ij,t} \sim \text{iid Type 1 Extreme Value}$$
$$j = 1, ..., J \text{ drug choices}$$
$$i = 1, ..., N \text{ individuals}$$
$$t = 1, ..., T \text{ treatment periods}$$

Bayesian Learning Model

Learning Process

- Prior Beliefs

$$\beta_{ij,0} \sim N(\mu_{\beta_{ij,0}}, \sigma^2_{\beta_{ij,0}})$$

- Drug Experience

$$d_{ij,t} = \begin{cases} 1 & \text{if drug } j \text{ taken in period } t, \\ 0 & \text{otherwise} \end{cases}$$

- Quality Signal

$$q_{ij,t} \sim N(\beta_{ij}, \sigma^2_{\beta_{ij}})$$

- Posterior Beliefs

$$\beta_{ijt} \sim N(\mu_{\beta_{ijt}}, \sigma^2_{\beta_{ijt}})$$

Bayesian Learning Model

Learning Process

- Posterior mean of match quality

$$\begin{aligned}\mu_{\beta_{ij,t}} &= \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij,t-1}}}\mu_{\beta_{ij,t-1}} + d_{ij,t}\frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij}}}q_{ij,t} \\ &= \frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij,0}}}\mu_{\beta_{ij,0}} + \sum_{\tau=1}^{t} d_{ij,\tau}\frac{\sigma^2_{\beta_{ij,t}}}{\sigma^2_{\beta_{ij}}}q_{ij,\tau}\end{aligned}$$

- Posterior variance of match quality

$$\begin{aligned}\sigma^2_{\beta_{ij,t}} &= \frac{1}{\frac{1}{\sigma^2_{\beta_{ij,t-1}}} + \frac{d_{ij,t}}{\sigma^2_{\beta_{ij}}}} \\ &= \frac{1}{\frac{1}{\sigma^2_{\beta_{ij,0}}} + \frac{\sum_{\tau=1}^{t} d_{ij,\tau}}{\sigma^2_{\beta_{ij}}}}\end{aligned}$$

Bayesian Learning Model

Learning Process

- Perception Bias

$$\nu_{ijt} = \mu_{\beta_{ijt}} - \beta_{ij}$$

- Signal Noise

$$\eta_{ij,t} = q_{ij,t} - \beta_{ij}$$

$$\mu_{\beta_{ij,t}} = \beta_{ij} + \frac{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}\nu_{ij,0} + \sum_{\tau=1}^{t} d_{ij,\tau}\eta_{ij,\tau}}{\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}} + \sum_{\tau=1}^{t} d_{ij,\tau}}$$

Osteoporosis

- Characterized by low bone mass
  - Especially prevalent in post-menopausal women
  - May result in fractures
- Approximately 8 million women and 2 million men in the US

Treatment

- Lifestyle
  - Diet and exercise
  - Avoid smoking and alcohol
- Bisphosphonates
  - Most common treatment
- Estrogen antagonists
- Biologics

Products and Market Share

Product	Active Ingredient	Brand/Generic	Market Share	
			2007	2008
Alendronate	Alendronate	Generic	-	0.47
Fosamax	Alendronate	Brand	0.40	0.05
Fosamax +D	Alendronate+ Cholecalciferol	Brand	0.09	0.06
Boniva	Ibandronate	Brand	0.18	0.15
Evista	Raloxifene	Brand	0.06	0.05
Actonel	Risedronate	Brand	0.26	0.19

Out Of Pocket Costs

Product	2007		2008	
	Mean	Std. Dev.	Mean	Std. Dev.
Actonel	51.72	28.33	56.14	30.65
Alendronate	-	-	17.51	15.32
Boniva	57.66	24.97	62.46	32.13
Evista	51.74	31.48	56.48	34.22
Fosamax	48.64	27.78	60.06	25.76
Fosamax Plus D	47.65	26.30	48.38	24.89

Switches

Drug	Switched (from)	Switched (to)
Fosamax	773	74
Fosamax Plus D	106	38
Actonel	176	90
Alendronate	68	934
Boniva	152	118
Evista	36	57
Total	1311	1311

Estimation
Identification

- Can't identify all learning parameters $\{\beta_{ij}, \sigma^2_{\beta_{ij}}, \nu_{ij,0}, \sigma^2_{\beta_{ij,0}}\}$
  - $\beta_{ij} \sim MVN(\bar{\beta}, \Omega_{\bar{\beta}})$
  - Only $\frac{\sigma^2_{\beta_{ij}}}{\sigma^2_{\beta_{ij,0}}}$ can be identified ($\sigma^2_{\beta_{ij}} = 1$)
  - Normalization is analagous to an outside good

Estimation

- θ: Parameter of the data generating model.
- $k(\theta)$: Researcher's prior beliefs about θ.
- $Y = \{y_1, ..., y_N\}$: Observed choices.
- $L(Y|\theta)$: Probability of observing Y.
- $K(\theta|Y)$: Updated beliefs re. θ given Y.

Then according to **Bayes' rule**:

$$K(\theta|Y) = \frac{L(Y|\theta)k(\theta)}{L(Y)}$$

$$K(\theta|Y) \propto L(Y|\theta)k(\theta)$$

- Estimate via MCMC

◂ Details

Parameter Estimates

Table: Parameter Estimates from the Bayesian Learning Model

	Full Model		Restricted Model	
Product	Posterior Mean	Heterogeneity	Posterior Mean	Heterogeneity
Price	-0.01		-0.02	
True Mean Quality				
Generic (β_1)	16.67	15.52	7.44	6.08
Boniva (β_2)	11.60	14.87	-4.02	11.19
Evista (β_3)	9.26	15.93	-16.85	15.26
Fosamax (β_4)	12.31	15.30	-0.98	5.76
Fosamax Plus D (β_5)	12.58	16.38	-12.36	12.12
Initial Perception Bias				
Generic (ν_1)	-5.45		-6.25	
Boniva (ν_2)	-4.07		-3.26	
Evista (ν_3)	-5.73		-4.48	
Fosamax (ν_4)	-2.91		-1.45	
Fosamax Plus D (ν_5)	-5.60		-3.09	
Initial Precision ($\frac{1}{\sigma^2_{\beta_{ij,0}}}$)	0.02		0.02	

Note: Unobserved heterogeneity is measured by the posterior mean of the square root of the diagonal element of the covariance matrix $\Omega_{\tilde{\beta}}$.

Parameter Estimates

Table: Correlation Coefficient Estimates

Product	Generic	Boniva	Evista	Fosamax	Fosamax Plus D
Generic	1.00	0.94	0.92	0.99	0.98
Boniva		1.00	0.96	0.94	0.96
Evista			1.00	0.92	0.97
Fosamax				1.00	0.99
Fosamax Plus D					1.00

Out of Sample Performance

Table: Market Share

Product	2007		2008	
	Actual	Predicted	Actual	Predicted
Actonel	0.26	0.23	0.18	0.21
Generic	0.00	0.00	0.46	0.29
Boniva	0.17	0.15	0.16	0.15
Evista	0.06	0.05	0.06	0.04
Fosamax	0.40	0.50	0.05	0.25
Fosamax+D	0.09	0.04	0.06	0.02

Price Elasticity: Learning Model

Table: Price Elasticity

Product	Mean
Actonel	-0.30
Generic	-0.08
Boniva	-0.36
Evista	-0.39
Fosamax	-0.25
Fosamax Plus D	-0.35

Robustness

Table: Mean Drug Prices in 2008 by Cohort

Product	2007 Cohort	2008 Cohort
Actonel	55.87	56.42
Generic	17.95	17.04
Boniva	62.49	62.41
Evista	56.69	56.25
Fosamax	60.47	59.63
Fosamax Plus D	48.50	48.24

Counterfactual

Table: Welfare (CV)

	Full Model		Restricted Model
Product Eliminated	Mean	Std. dev	Mean
Generic	-270.71	403.91	-50.54
Boniva	-125.89	301.19	-66.14
Evista	-48.95	208.10	-34.21
Fosamax	-108.11	242.25	-31.42
Fosamax Plus D	-27.20	131.82	-25.26

Empirical Concerns

- We are good at whether, but not which, switches occur
- Can't disentangle product preferences from automatic substitution
- Estimate lower elasticity for generics than for brands

Estimation
Bayesian Inference

- Parameter vector to be estimated
  $\Theta = \{\beta_{i1}, ..., \beta_{iJ-1}, \{\eta_{ij,t}\}_{\tau=1}^{T_i-1}, \bar{\nu}_1, ..., \bar{\nu}_{J-1}, \bar{\sigma}^2_{\beta_0}, \alpha\}$
- Liklihood function

$$L_i(y_i|X_i;\Theta) = \prod_{t=1}^{T_i}\prod_{J=1}^{J}\left(\frac{\exp(\bar{U}^E_{ij,t})}{\sum_{j'}\exp(\bar{U}^E_{ij',t})}\right)^{d_{ij,t}}$$

- Joint posterior distribution

$$K(\Theta|\{d_i, X_i\}_{i=}^{N}) \propto \prod_{i=1}^{N} L_i(d_i|X_i;\Phi_i,\Psi)k(\Theta)$$

- Use a Gibbs Sampler to draw from this joint posterior density

Back

1. Update $\beta_i = \{\beta_{i1}, ..., \beta_{iJ-1}\}$ using a Metropolis-Hastings sampler.
2. Update $\bar{\beta}$ and $\Omega_{\bar{\beta}}$ using a Gibbs sampler
3. Update α by a Metropolis-Hastings sampler
4. Update $\{\bar{\nu}_1, ..., \bar{\nu}_{J-1}, \bar{\sigma}^2_{\beta_0}\}$ using a Metropolis-Hastings sampler
5. Update $\{\eta_{ij,\tau}\}_{\tau=1}^{t-1}$ by a Metropolis-Hastings sampler

◂ Back

Product Differentiation, Consumer Learning and the Value of Me-too Drugs

Neha Bairoliya (Presenter), Harvard University
Pinar Karaca-Mandic, Jeffery McCullough, Amil Petrin,
University of Minnesota

Dec, 12 2015 AHEW & AWEHE
Discussant: Daniel Miller, Clemson University

Overview

Value of Drugs Entering Market

- Generics...benefit of low prices
- "Me-too" Brands...benefit of product differentiation

Estimate Demand System for Osteoporosis Drugs

1. Heterogenous "patient-drug" match quality
   - benefit of product differentiation
   - do me-too drugs add value? Are generic equivalents really perfect substitutes?
2. Learning
   - Do learning frictions diminish value of generics, differentiation?

Application to Medicare Part D

- 2007-2008 individual-level claims data
- Policy counterfactual: Restricting choice through narrow formularies

Osteoporosis Drugs

Very large market: 30% Medicare Beneficiaries, $16 billion

- Fosamax mkt leader (54% mkt share in 2006)
- "Me-Too" brands
  - Fosamax Plus D
  - Actonel,Boniva, Evista
- Generic Fosamax enters in 2008 (70% mkt share in 2012)

30-day supply prices	2006	2012
Fosamax (Generic)	-	9
Fosamax (Branded)	58	127
Fosamax Plus D	96	109
Actonel	78	130
Boniva	67	129
Evista	80	149

Given high drug prices (inflation). Is there value in brands? Does a generic contain spending?

Learning Model

Utility

$$U_{ijt} = \mu\beta_{ijt-1} + \alpha P_{ijt} + \epsilon_{ijt}$$

Quality Signal

$$\mu\beta_{ijt} \sim N(\beta_{ij}, \sigma^2_{q_{ij}})$$

Bayesian updating about match quality

$$\mu\beta_{ijt} = \frac{\sigma^2_{q_{ijt}}}{\sigma^2_{q_{ijt-1}}}\mu\beta_{ijt-1} + d_{ijt}\frac{\sigma^2_{\beta_{ijt}}}{\sigma^2_{\beta_{ijt-1}}}$$

Learn β_{ij} overtime based on choice history $d_{ij0}, d_{ij1} \ldots d_{ijt}$

at pace depending on signal noise $\sigma^2_{q_{ij}}$

Key Model Features/Assumptions

- β_{ij} heterogeneity in patient-drug match quality
- myopic consumers (not forward looking), dynamics generated through state dependence
- Identification: β_{ij} steady state choice (T large), price variation (benefit design). Learning parameters (switching frequencies)

Results

Very Preliminary Results

- β_{ij}: significant heterogeneity in match quality
- highly correlated match qualities across drugs
- Generic Fosamax very high mean utility
- High signal noise, initial perception bias
- $\alpha \approx 0$: very inelastic demand

Comments

Majority of comments address issues of modeling, identification, data that may be presenting estimation challenges

Substitution Patterns

Are generics perfect substitutes?

- Yes: high correlation in β_{ij} w/ branded
- No: generic much higher mean utility

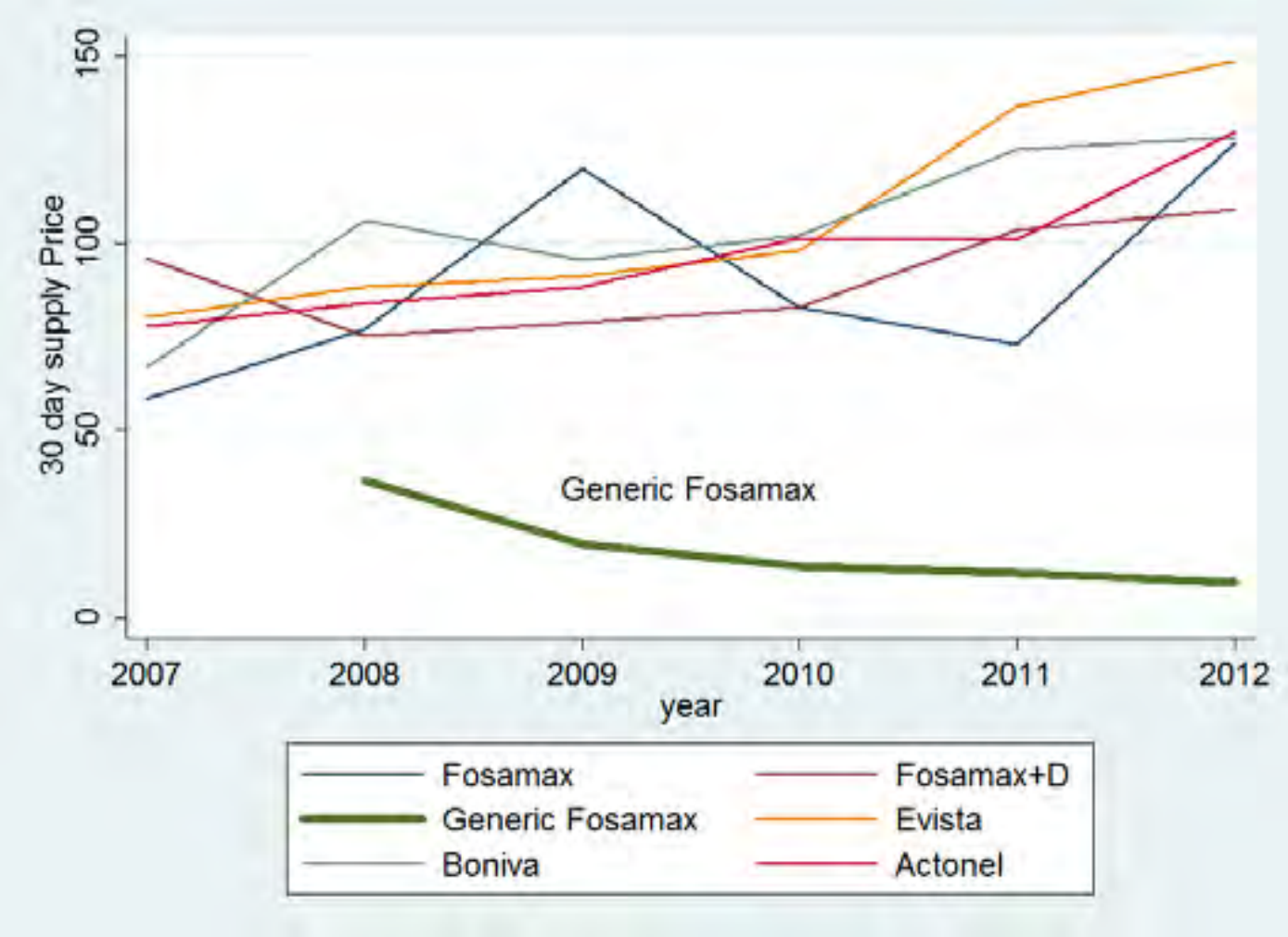
Modeling unrestricted $corr(\beta_{ij}, \beta_{ij'})$

- Allows for rich substitution patterns
- Information assumption? no knowledge of corr: current model. consumer has prior info on corr: learn match value of j' even if $d_{ij't} = 0$. need additional learning (signal noise) parameters about correlation
- mixing discrete types (severe diagnosis, not severe)
- generic/branded equivalents as special case $corr = 1$

price coefficient $\alpha \approx 0$

- Generic preference driven by price not quality
- price endogeneity...supply-side, detailing ('08 marketing blitz?)
- "pay-to-delay" tactics to retain market power (price series graph)

Drug Prices



Generic Drug price high in 2008, but falling

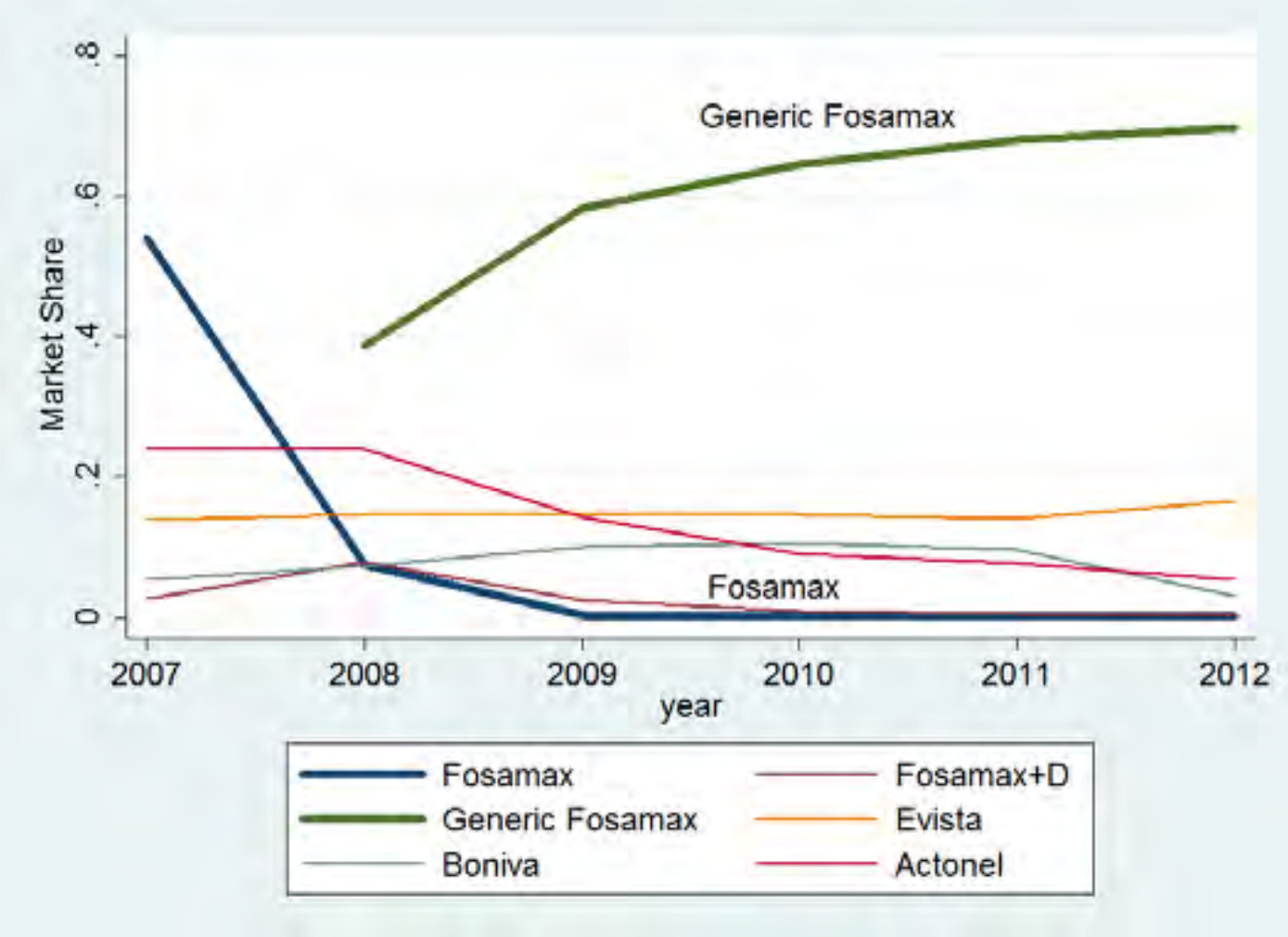
Source: 5% Part D Event Files

Time Horizon

Short Run vs. Long Run Demand

- learning choice friction creates differences in short vs long run demand
- Data from (relatively) short period 2007-2008, when market evolving with generic entry
- Long run: **branded Fosamax market share** ≈ 0 which might imply generic is perfect substitute (market share graph)
- introduction of generic, 2008 mkt share > 0: raises question of whether choice friction learning or other switching friction (waiting for next doc visit to switch over to generic)
- distinguish learning: early vs late switching intensity. differ in data for branded to generic switches? If not, non-learning switching parameter for generic, learning amongst branded

Market Shares



Branded Fosamax market share drops to zero, but positive in 2008

Source: 5% Part D Event Files

Dynamics: Forward Looking Expectations Curative Effects

Forward looking expectations about drug effectiveness: **cured**

- Crawford Shum (Anti-ulcer), Dickstein (Depression), Chorniy (ADHD)
- Experiment to learn curative effect and tradeoff w/ side effects
- Identification: length of treatment, ends at $t = T$
- Include curative effect: pharmacology osteoporosis, long enough panel to observe $t = T$, treatment stoppage observed in data

Dynamics: Forward Looking Expectations Prices

p_{ijt}: **OOP Price** varies across consumers and time

- Source of variation: plan choice, benefit phase (deductible, IC, donut hole, catastrophic)
- Large price changes from crossing attachment points (deductible \$80, to IC \$20) won't induce switch if consumers form expectations about annual drug spending
- could explain $\alpha \approx 0$

Incorporating price expectations

- Specify fully dynamic model with complicated non-linear demand features (gaps, bunching)
- myopic model using avg(Price) for year based on patient's annual spending and progression through benefit phases
- check data for more intensive rate of drug switching at attachment points
- Include retail price, not just OOP price to reflect physician preferences

Data

Strengths

- claims, price variation, year introducing generic
- Any formulary exclusions? useful for identifying random coefficient

Low Income Subsidy Beneficiaries (excluded from sample)

- interesting population >50% Part D drug spending
- useful for estimation: $p_{ijt} \approx 0$ as model restriction
- pin down detailing effects & physician preferences over retail prices
- Address endogeneity of plan choice...Random reassignment of LIS

Formulary Restrictions/drug formulations

- quantity limits, prior authorization
- step therapies interesting b/c structures learning process
- other formulation/packaging (daily, weekly, monthly), strength, tablet/injectable

Counterfactual

Welfare effects of restricting formulary choice

- Policy lever in Part D to steer patients towards cost-effective drugs
- narrow formulary reduces learning friction, but also reduces product differentiation

Other interesting counterfactuals

- Benefit design (closing donut hole, copay caps)
- Price controls
- 30 vs 90 day supply (slow down learning)
- Nudging: step therapies
- Interaction of drug choice and plan choice switching frictions
  - Nudging: Default plan assignments based on patient drug history

PARALLEL TRADE OF PHARMACEUTICALS: THE DANISH MARKET FOR STATINS

Susan J. Mendez
University of Melbourne

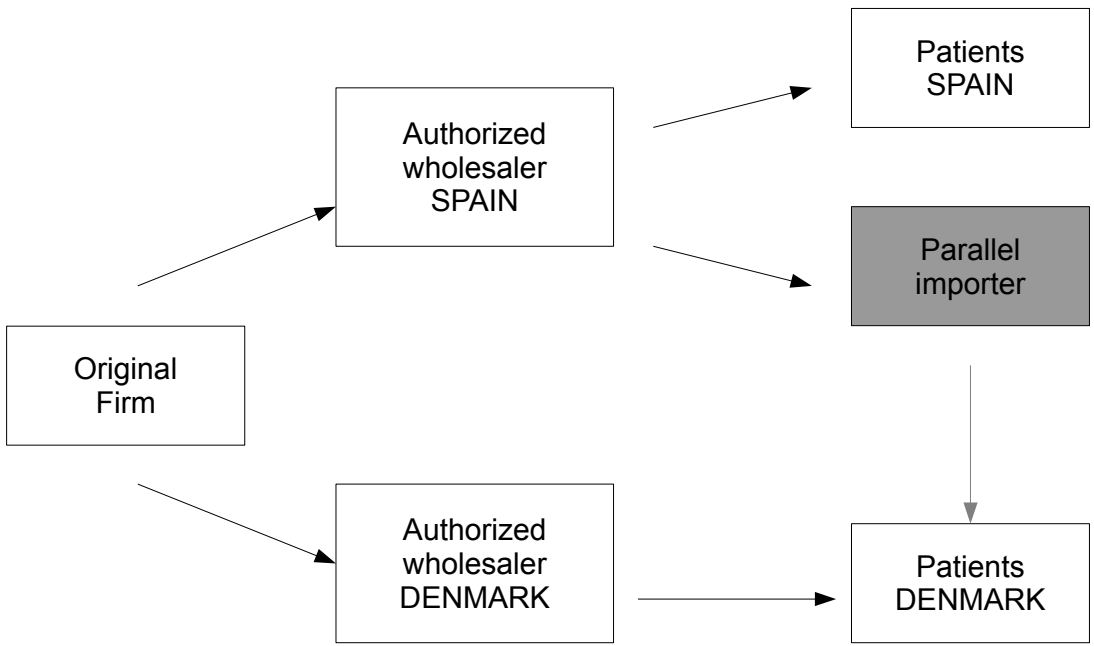
Joint 2015 AHEW and 7th AWEHE
University of Hawaii, December 11 – 14, 2015

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

What is a parallel imported pharmaceutical product?

Parallel traded pharmaceuticals are legally marketed in one country but distributed in another country without authorization of the property right holder. PI are legal in EU and EEA. PI are illegal outside this area.



Motivation

- There is significant price dispersion in the European market for pharmaceuticals, this induces arbitrage opportunities and a profitable market for parallel imports

\+ PI should promote (price) competition in import country generating savings to patients and insurers [Ganslandt and Maskus, 2004; Kanavos and Cost-Font, 2005]

– PI can harm export country by restraining supplies or delaying entry [Kyle, 2010]

– Long run: PI weaken intellectual property rights protection - original producers have less incentives to innovate [Szymanski and Valletti, 2005; Grossman and Lai, 2008]

Research question and results

Research Question:
What are the effects of PI? – welfare

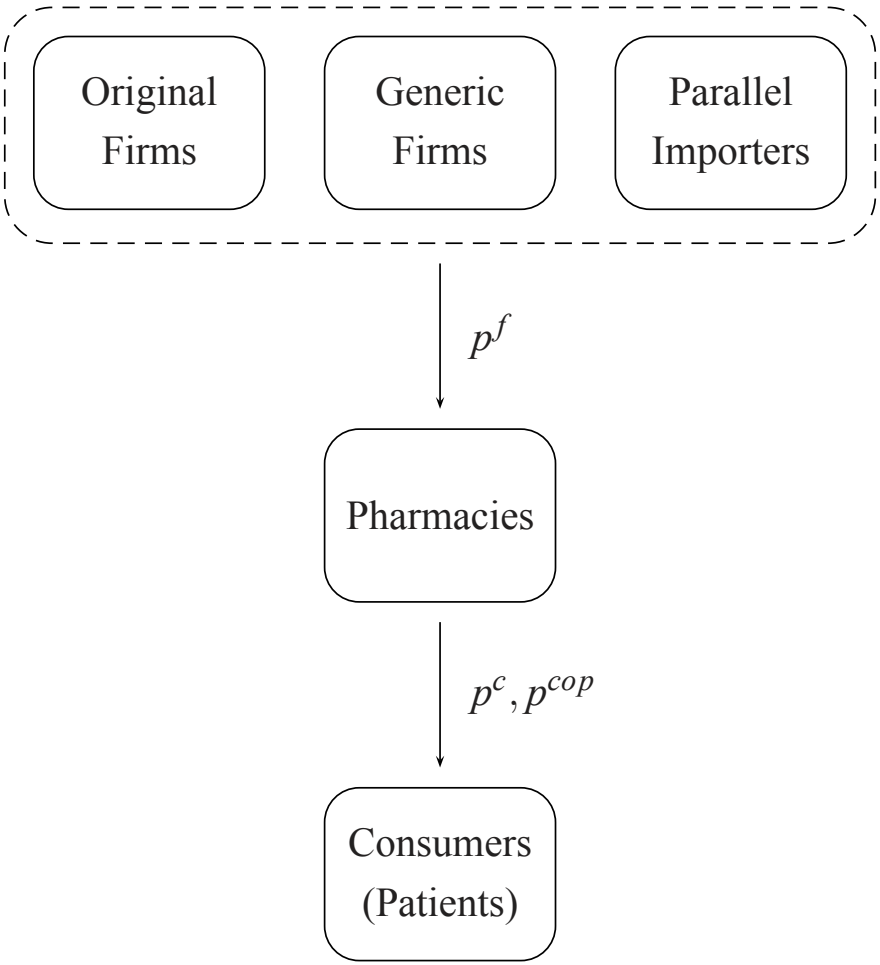
1. Set up the framework: Structural model of demand and supply
2. Use parameters to calculate counterfactual market equilibrium under prohibition of parallel imports

Results:
Using data on statins in Denmark, a prohibition of parallel trade:

- results in higher prices for original firms and generics firms [3%]
- leads to substitution from parallel imported products towards original products
- results in an increase in consumer expenditures [$12 mio./year], government expenditures [$30 mio./year], and firm profits [$9 mio./year]
- decreases consumer surplus, leading to an overall decrease in welfare [$9 mio./year]

The Danish pharmaceutical industry

Overview



- Firms are free to set their prices p^f which are publicly available
- Pharmacies face generic substitution and their markups are regulated: $p^c = \mu p^f + k$
- Consumers are entitled to free and equal access to health care services: reimbursement is based on reference pricing:

$$p^{cop} = p^c - 0.8 * p^r$$

  - p^r: reference price is pharmacy retail price up to the average price in EU-15 members, excluding Greece, Luxembourg, Spain, and Portugal – before April 2005

Data set

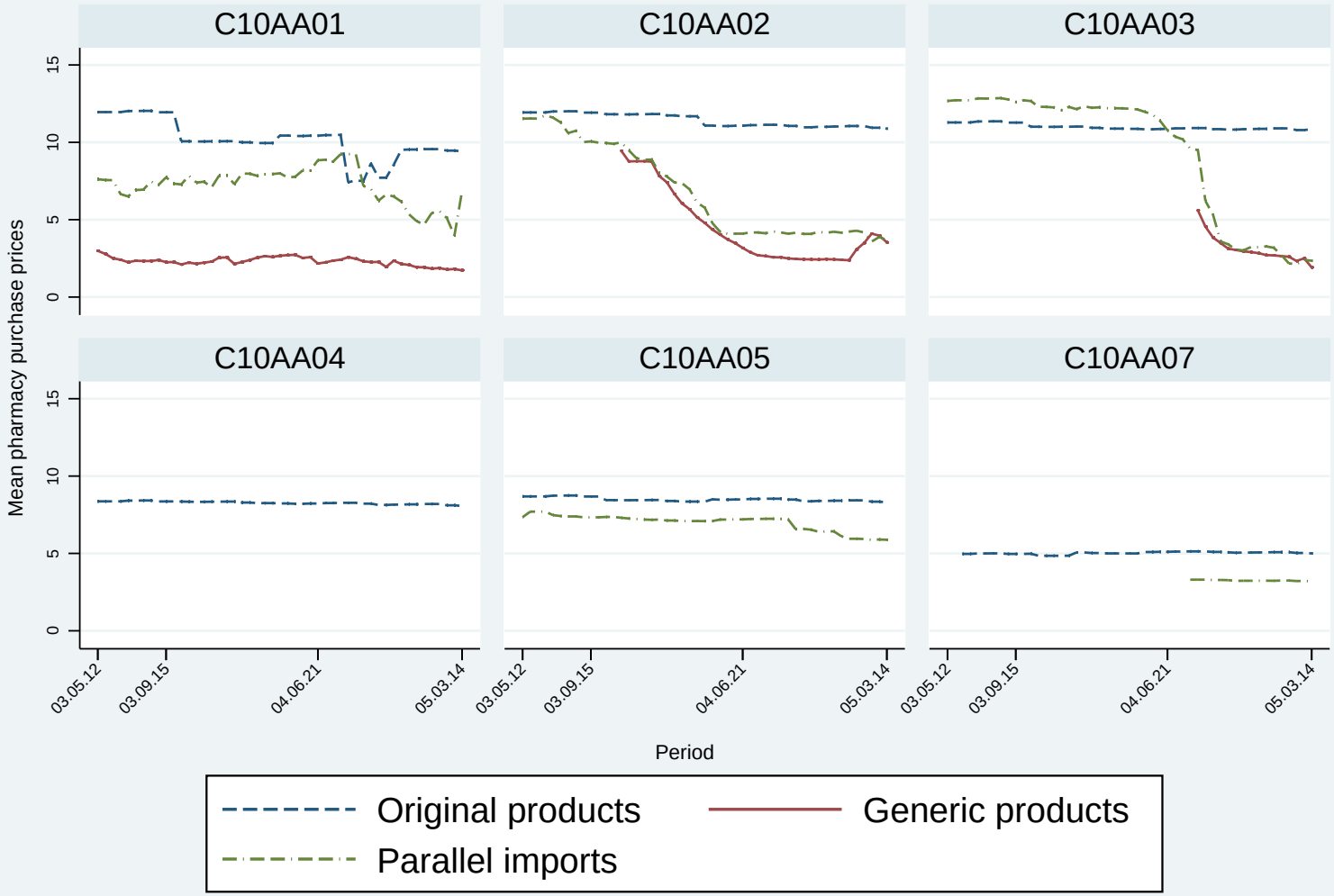
Fortnightly (2003 - 2005) data on group of anti-cholesterol drugs – HMG CoA reductase inhibitors (Statins) [In 2004: over $50 mio.]

- Product: active ingredient (ATC), name, package size, strength, firm (6 Molecules - 213 Products)

Simvastatin (C10AA01)	*Zocor*	*20 Tablets*	*80 mg.*	*Merck*
Atorvastatin (C10AA05)	*Lipitor*	*40 Tablets*	*20 mg.*	*Pfizer*

- Unit: WHO Defined Daily Dose (DDD)
- Substitute: DKMA – same active ingredient (ATC) and strength (similar package size)

Average pharmacy purchase price



Demand estimation

$$U_{ij} = \underbrace{X_j\beta - \alpha p_j^{cop} + \xi_j}_{\delta_j} + \sum_g [d_{jg}\zeta_{ig}] + (1-\sigma)\varepsilon_{ij}$$

- Mean utility $\delta_j \equiv X_j\beta - \alpha p_j^{cop} + \xi_j$
- X_j observed product characteristics, ξ_j unobserved product characteristics
- d_{jg} is equal to one if product j belongs to group J_g and zero otherwise
- ζ_{ig} is common to all products in group g
- σ is the nesting parameter and measures within group correlation
- ε_{ij} unobserved patient-specific heterogeneity and are Gumbel distributed

Berry (1994):

$$ln(s_j) - ln(s_0) = X_j\beta - \alpha p_j^{cop} + \sigma ln(s_{j|g}) + \xi_j$$

Elasticities:

$$\begin{aligned}
\eta_{jj} &= \frac{-\alpha}{(1-\sigma)} p_j^{cop}[1 - \sigma s_{j|g} - (1-\sigma)s_j] \\
\eta_{jk} &= \frac{\alpha}{(1-\sigma)} p_k^{cop}[\sigma s_{k|g} + (1-\sigma)s_k] \\
\eta_{jl} &= \alpha p_l^{cop} s_l
\end{aligned}$$

Demand estimation – results

	OLS		IV - Nested Logit	
	Coef.	Std. Error	Coef.	Std. Error
Copayment price	-0.0531	0.0041	-0.8315	0.0514
Conditional share	0.8803	0.0075	0.3147	0.1229
Strength in ddd	0.3466	0.0221	-0.8071	0.0673
Package size	0.0237	0.0004	0.0182	0.0011
On-Patent	0.9790	0.0639	1.6971	0.1189
No. prod. in nest	0.2387	0.0053	-0.2121	0.0514
Constant	-11.4166	0.6099	-10.6689	0.9525
Firm dummies	yes		yes	
Period dummies	yes		yes	
Mean own-price elasticity η_{jj}				-3.607
Mean cross-price elasticity - same nest η_{jk}				0.179
Mean cross-price elasticity - different nest η_{jl}				0.0014

Number of observations: 6,388. Instruments for the IV - nested logit are: number of products of rival firms, average price of products from the same firm in other groups, sum of characteristics of rival firms. F-test statistics: 36.40 (p-val:0.00) for p^{cop} and 26.57 (p-val:0.00) for $s_{j|g}$.

Supply estimation

Profit of firm f:

$$\Pi_f = \sum_{j \in \vartheta_f} (p_j^f - c_j) s_j M - F$$

The first order condition for product j:

$$\frac{\partial \pi_j}{\partial p_j^f} = M \left(s_j + \sum_{h \in \vartheta} (p_{jh}^f - c_{jh}) \frac{\partial s_{jh}}{\partial p_{jm}} \right) = 0$$

Set of J FOC in vector notation $S(p, x, \xi) - \Delta(p, x, \xi)(P - C) = 0$

$$C = P - \Delta(p, x, \xi)^{-1} S(p, x, \xi)$$

Supply estimation – results

	marginal cost	markups	markups in %
All Products			
	5.277	0.648	20.92
	(4.486)	(0.137)	(19.08)
By firm type			
Original	7.940	0.745	11.06
	(3.673)	(0.114)	(7.351)
Generic	2.035	0.584	31.93
	(2.003)	(0.125)	(21.10)
Parallel Imp.	7.014	0.631	15.96
	(4.992)	(0.114)	(16.91)
By patent status			
Off-patent	5.077	0.617	22.50
	(4.740)	(0.125)	(20.47)
On-patent	6.208	0.789	13.53
	(2.855)	(0.099)	(6.465)

Average marginal cost and markups. Std. Dev. in parentheses.

Counterfactual market equilibrium

Prohibition of PI

Calculate new shares and prices under new equilibrium:

- use expression for shares:

$$s_j(\delta) = \frac{e^{\delta_j/1-\sigma}}{(\sum_{j\in J_g} e^{\delta_j/1-\sigma})^{\sigma} \sum_g (\sum_{j\in J_g} e^{\delta_j/1-\sigma})^{(1-\sigma)}}$$

- use first order conditions

$$P^f = C + \Delta^{-1} S$$

- use regulation rules

$$p^{cop} = \mu p^f + k - 0.8 p^r$$

Consumer surplus is given by:

$$CS = \frac{1}{\alpha} M \log \left[1 + \sum_{g=1}^{G} (\sum_{j\in J_g} e^{\delta_j/1-\sigma})^{(1-\sigma)} \right]$$

Results

Average change in prices

Eliminating parallel trade reduces average prices but results in higher prices for both original and generic products

	Pharmacy purchase price (p^f)			Copayment price (p^{cop})		
	real	count.	change in%	real	count.	change in%
All products						
	5.92	5.33	-10.08	3.21	2.84	-11.50
	(4.53)	(4.17)		(4.42)	(3.93)	
By firm type						
Original	8.69	8.79	2.49	4.63	4.79	3.45
	(3.62)	(3.63)		(4.84)	(4.94)	
Generic	2.62	2.63	3.55	1.31	1.33	1.14
	(2.03)	(2.01)		(1.77)	(1.78)	
Parallel imp.	7.64			4.40		
	(5.04)			(5.50)		

Fortnightly average prices for a DDD in DKK.

Exchange rates in June 2005: DKK 1 = $ 0.1634 = €0.1343

Results

Average change in shares and markups

Eliminating parallel trade leads to substitution from parallel imported products towards original products

	Shares			Markups		
	real	count.	change in%	real	count.	change in%
All products						
	0.124	0.243	96.34	0.648	0.706	9.03
	(0.43)	(1.54)		(0.14)	(0.21)	
By firm type						
Original	0.113	0.481	324.5	0.745	0.852	14.34
	(0.21)	(2.23)		(0.11)	(0.25)	
Generic	0.184	0.058	-68.19	0.584	0.593	1.58
	(0.62)	(0.14)		(0.12)	(0.05)	
Parallel imp.	0.048			0.631		
	(0.16)			(0.11)		

Fortnightly average shares per product in percentage and average markups per DDD in DKK.
Exchange rates in June 2005: DKK 1 = $ 0.1634 = €0.1343

Results

Average welfare effects

Eliminating parallel trade:

- Increases firm profits, consumer expenditures as well as government expenditures
- reduces consumer surplus and increases firm profits, leading to an overall decrease in welfare

Yearly average	real	counterfactual	change	change in %
Government Expenditures	271.51	454.22	182.71	80.90
Patients Expenditures	87.29	162.29	75.01	123.06
Consumer surplus	232.35	119.78	-111.41	- 49.29
Profits	38.03	94.54	56.51	167.50
Total welfare	270.38	214.32	-54.90	-20.73

All figures in million DKK. Exchange rates in June 2005: DKK 1 = $ 0.1634 = € 0.1343

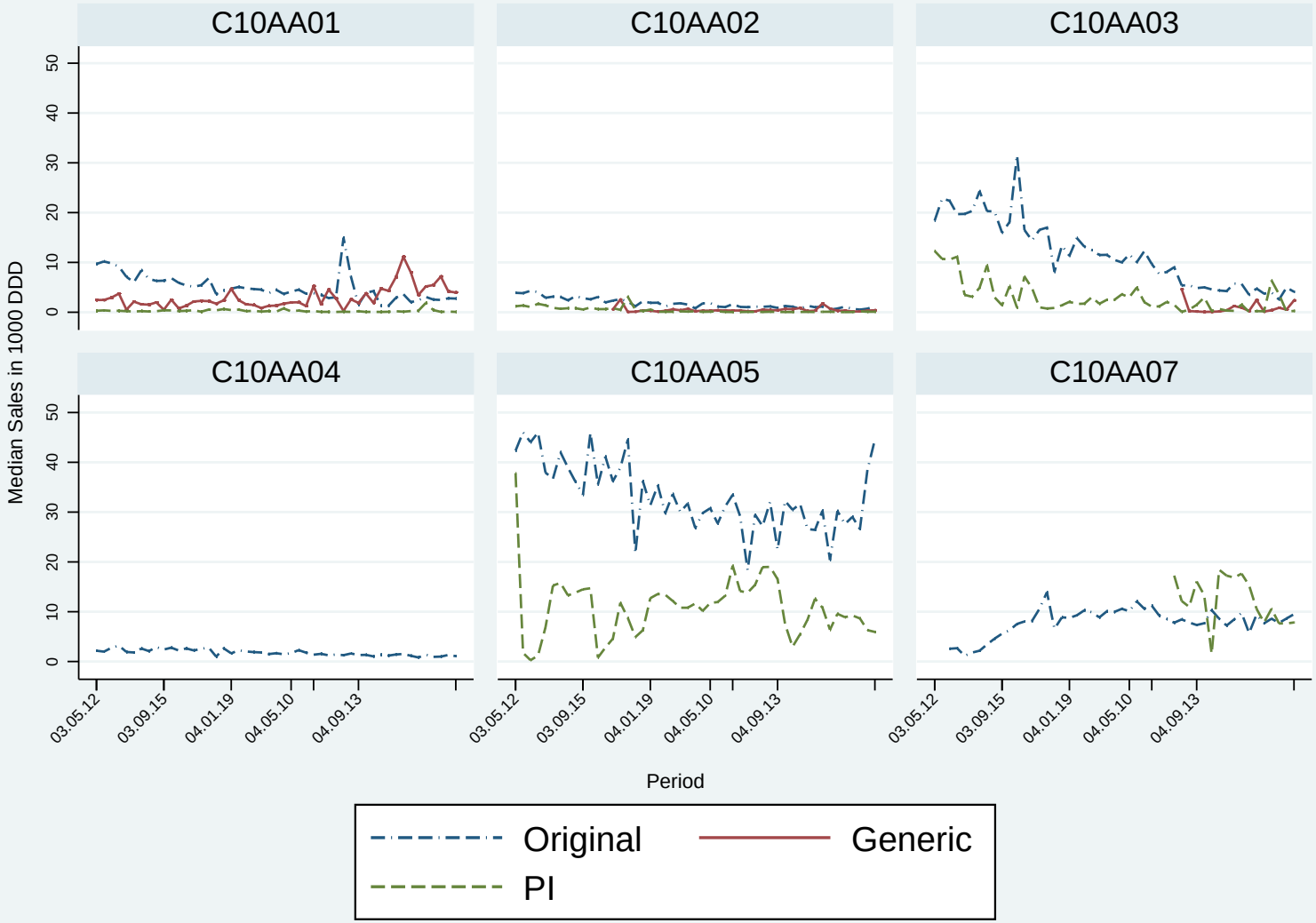
Conclusions

Removing parallel imports

- results in higher prices for original firms and generics firms (2.5% and 3.5%)
- leads to substitution from parallel imported products towards original products (Model considers consumer tastes - relevant to calculate welfare effects)
- results in an increase in consumer expenditures as well as government expenditures and firm profits (123%, 81%, 167%)
- decreases consumer surplus, leading to an overall decrease in welfare (around $9 million per year).

Thank you!

Median Sales



The market for Statins

ATC Code	Active Substance	Brand name and Firm	Obs.	Mean Number of		
				Firms	Presentations	Products
C10AA01	Simvastatin	Zocor	3'323	11.85	10.98	69.51
		Merck		(1.02)	(0.13)	(10.85)
C10AA02	Lovastatin	Mevacor	829	5.39	4.66	17.44
		Merck		(0.81)	(0.78)	(2.72)
C10AA03	Pravastatin	Pravachol	766	5.94	4	19.28
		Bristol-Myers Squibb		(2.06)	(0)	(8.13)
C10AA04	Fluvastatin	Lescol	490	2	6	10
		Novartis		(0)	(0)	(0)
C10AA05	Atorvastatin	Lipitor	611	3.03	8	12.57
		Pfizer		(0.44)	(0)	(1.11)
C10AA07	Rosuvastatin	Crestor	369	1.59	6	8.10
		AstraZeneca		(0.75)	(0)	(1.37)
Total			6'388	19.71	39.53	130.76
				(1.96)	(1.22)	(7.05)

Average prices

ATC Code	Total				Original Firms		Generic Firms		Parallel Importers	
	p^f	p^c	p^r	p^{cop}	p^f	p^{cop}	p^f	p^{cop}	p^f	p^{cop}
C10AA01	4.63	7.30	4.42	3.76	10.03	10.48	2.32	1.33	7.22	6.04
	4.94	7.18	3.91	5.77	4.96	7.24	2.01	1.97	6.12	7.11
C10AA02	7.08	10.79	9.16	3.47	11.44	7.74	3.98	1.46	8.03	3.03
	3.78	5.40	4.30	3.04	1.27	2.56	2.01	1.06	3.10	1.79
C10AA03	7.71	11.45	11.10	2.57	10.99	4.30	3.01	0.95	9.36	2.78
	4.41	6.31	6.20	1.81	1.74	1.99	1.15	0.33	4.14	1.18
C10AA04	8.27	12.69	12.66	2.56	7.77	2.41	9.02	2.78		
	2.14	3.56	3.56	0.72	2.27	0.76	1.67	0.57		
C10AA05	7.91	11.53	11.53	2.31	8.49	2.48			6.88	2.00
	2.99	4.50	4.50	0.90	3.51	1.06			1.10	0.33
C10AA07	4.92	7.19	7.19	1.44	5.03	1.47			3.24	0.92
	1.35	2.05	2.05	0.41	1.32	0.40			0.03	0.01
Total	5.93	9.06	7.31	3.21	8.65	4.83	3.07	1.42	7.64	4.40
	4.53	6.56	5.34	4.42	3.77	5.05	2.58	1.76	5.04	5.50

AMERICAN UNIVERSITY

Discussion of "Parallel Trade of Pharmaceuticals: The Danish arket for Statins", by Susan J. Méndez

Robin L. Lumsdaine

Joint AHEW and 7th Australasian Econometrics and Health Economics Workshop

Honolulu

December 12, 2015

KOGOD SCHOOL of BUSINESS

This paper

- Considers the impact of a parallel trade ban on the market for statins
- Structural model enables policy evaluation
- Concludes that banning parallel imports:
  - Increases profits for original producers and decreases for generic firms,
  - Increases governmental HC expenditures
  - Decreases consumer welfare

KOGOD SCHOOL of BUSINESS

Danish pharmaceutical market

- Generic substitution

  Required by law to dispense cheapest product among available substitutes (unless specifically requested)

- Retail price regulation: prices are identical nationwide

- Consumers pay more for off-patent products than on

KOGOD SCHOOL of BUSINESS

Modeling demand (1)

- Utility modeled as a function of observed/unobserved product characteristics

$$U_{ij} = X_j\beta - \alpha p_j^{cop} + \xi_j + \sum_g [d_{jg}\zeta_{jg}] + (1-\sigma)\varepsilon_{ij}$$

- ε_{ij} assumed iid extreme value, so that the consumer-specific part of the utility function is also.
- σ must lie between 0 and 1 for nested logit to be consistent with random utility maximization
- One outside option (non-statin), mean utility normalized to zero.

Modeling demand (2)

- IV to control for potential endogeneity between unobserved product characteristics and copayment and share
- Estimated by OLS and IV-nested logit

$$ln(s_j) - \ln(s_0) = X_j\beta - \alpha p_j^{cop} + \sigma ln\left(s_{j|g}\right) + \xi_j$$

- Own and cross-price elasticities

KOGOD SCHOOL of BUSINESS

Modeling Supply

- Firms choose pharmacy purchase price considering price of all other products
- Presence of generics or parallel trade enters through influence on market share?
- Linear system of prices are function of marginal cost and predicted markups
- Counterfactual calculation – assumes eliminating parallel imports does not alter consumers' tastes
- Consumer surplus

Thoughts

- Basics
  - What is sample size just 22 months?
  - Be clearer on details (e.g, # observations, what are J and G, what are the product characteristics?)
  - Label drugs rather than using ATC code name
  - Are results dominated by Zocor? What if take them out of sample?

- Beyond
  - Consumers can freely choose w/in same substitution ring – the role of marketing?
  - Possible extension: model price of drug while on patent as an option (impact of parallel trade varies over time as patent nearing expiry)?
  - Why don't original firms get in the parallel import business?

New approaches to estimating the child health-parental income relationship

Brenda Gannon, Mark Harris, Leandro Magnusson, David Harris, Brett Inder, Puskar Maitra, Luke Munford, Bruce Hollingsworth

AHEW & AWEHE Workshop, December 11-14 2015, Hawaii
Funding Acknowledged: ARC Discovery Grant

Background

- Much literature on child health-parental income gradient (e.g. Case, Lubotksy & Paxson, (2002) Currie, Shields & Wheatley Price (2007))

- Mechanism is theoretically represented by Grossman (1972) model $\Rightarrow$
  - Endowment of health (Case, Lubotksy & Paxson, (2002))
  - Access to health care (Apouey & Geoffard (2013))
  - Health shocks can have lifetime effects: health stock, human capital & labor market outcomes (Currie & Hyson, (1999), Case, Lee & Paxson, (2002), Currie (2004))

- Does the gradient vary by age?
  - UK: Currie et al. (2007) - small decrease after age 8 but Case, Lee & Paxson, (2008): - gradient increases up to age 12

How does the literature estimate the relationship?

- Ubiquitous approach - *age bands are exogenously fixed*
  - 0–3, 4–8, 9–12 and 13+
- Why these age bands?
  - No medical, theoretical, social, or empirical reason for these

What estimation methods are employed in the literature?

- Usual approach includes (log) parental income

$$H_i^* = \mathbf{x}_i'\beta + \gamma \ln y_i + \varepsilon_i \tag{1}$$

- *But* importantly splits the $\ln y$ variable into four groups such that

$$H_i^* = \mathbf{x}_i'\beta + \sum_{m=1}^{4} \gamma_m \left(\ln y_i \times D_m\right) + \varepsilon_i \tag{2}$$

- Any differences across γ_m are taken as *differential effects of parental income according to the age of the child*

Suggested approach 1: threshold effects

- Why not endogenously determine both the number and position of any discontinuities?
- Gannon, Harris[2] (2014) demonstrate how to consistently estimate relationships of this sort in nonlinear models (*e.g. OP*)
- Essentially a combination of grid-search techniques along with the use of information criteria

Suggested approach 1: threshold effects

- Consider estimation models of the form

$$H_i^* = \mathbf{x}_i'\beta + \sum_{m=1}^{M} \gamma_m \left(\ln y_i \times D_m\right) + \varepsilon_i \tag{3}$$

  - Estimate all possible $m^* = 0, 1, \ldots, M^*$ threshold models
  - Choose the one that minimises the BIC (the Bayesian Information Criteria)

- In practice, start with a small M^* : Approach *simultaneously consistently finds the optimal number and position of threshold effects* (nesting "none")

Suggested approach 2: a random parameters approach

- Re-write our basic health equation as

$$H_i^* = \mathbf{x}_i'\beta + \gamma_t \ln y_i + \varepsilon_i. \tag{4}$$

- The age-specific component will be composed of an average effect, plus an age-specific component $\rightarrow$

$$\gamma_t = \overline{\gamma} + \alpha_t \tag{5}$$

- α_t are random draws from $N\left(0, \sigma_\alpha^2\right)$; and $\overline{\gamma}$ is the average effect of income

- Interpreted as a generalisation of threshold approach: but is more parsimonious as estimation only involves one additional parameter $\left(\sigma_\alpha^2\right)$

Suggested approach 2: a random parameters approach

- Model can be estimated by simulation, where we draw $r = 1, \ldots, R$ normal variates of α_t from $N\left(0, \sigma_\alpha^2\right)$; using Halton draws
- Approach mirrors a panel data set-up
  - but t indexes the panel and not i
- The likelihood for a group of t observations will be the product of the sequence of the OP probabilities corresponding to the observed health outcome
- And these probabilities will depend on the r^{th} draw of α_t, α_t^r:

$$p_i\left(\beta, \sigma_\alpha^2, \alpha_t^r\right) = \prod_{j=0}^{J-1} \left[\Pr\left(H_i = j \,|\ln y_i, \mathbf{x}_i, \alpha_t^r\right)\right]^{d_{ij}} \qquad (6)$$

Suggested approach 2: a random parameters approach

- The simulated log likelihood function is therefore

$$\ell^* \left(\beta, \overline{\gamma}, \sigma_\alpha^2\right) = \sum_{t=0}^{T} \log \left\{ \frac{1}{R} \sum_{t=1}^{R} \left[\prod_{i=1}^{N_t} p_i \left(\beta, \sigma_\alpha^2, \alpha_t^r\right) \right] \right\} \qquad (7)$$

- *Ex post*, conditional on the data, age group-specific estimates of γ_t are available (Train, 2003, Greene, 2007) as

$$\widehat{\gamma}_t = \widehat{\overline{\gamma}} + \frac{1}{R} \sum_{r=1}^{R} \alpha_t^r \omega_t^r, \qquad \text{where } \omega_t^r = \frac{\prod_{i=1}^{N_t} p_i \left(\hat{\beta}, \hat{\sigma}_\alpha^2, \alpha_t^r\right)}{\frac{1}{R} \sum_{r=1}^{R} \left[\prod_{i=1}^{N_t} p_i \left(\hat{\beta}, \hat{\sigma}_\alpha^2, \alpha_t^r\right) \right]} \qquad (8)$$

The Data

- Health Surveys for England (HSE), Pooled 2008 – 2012; $N \approx 9,500$
- Repeated cross-section annual surveys designed to measure health and health related behaviours in adults and children
- For each household, we match child-parent data
- SAH with 4 ordered categories - from very good to bad health
- $Income = \ln(\text{real family income})$
- Plus standard set of controls:
    - age, sex, ethnicity, household size, age of parents, absence of father from household, parental education, employment status and year dummies.

Descriptive Statistics, 2008 - 2012

Variable	Mean	Std. Dev.	Minimum	Maximum
Self Reported Health	1.439	0.612	1	4
Very Good	62%			
Good	33%			
Fair	4%			
Bad/Very Bad	1%			
Child's Age	7.982	5.141	0	17
Log of Family Income	10.127	0.847	5.382	11.837
Male	0.505	0.500		
Mother is employed	0.652	0.476		
Father is employed	0.569	0.497		
is absent	0.371	0.483		
Log of Household size	1.330	0.266	0.693	2.303
Number of Observations	9,613			

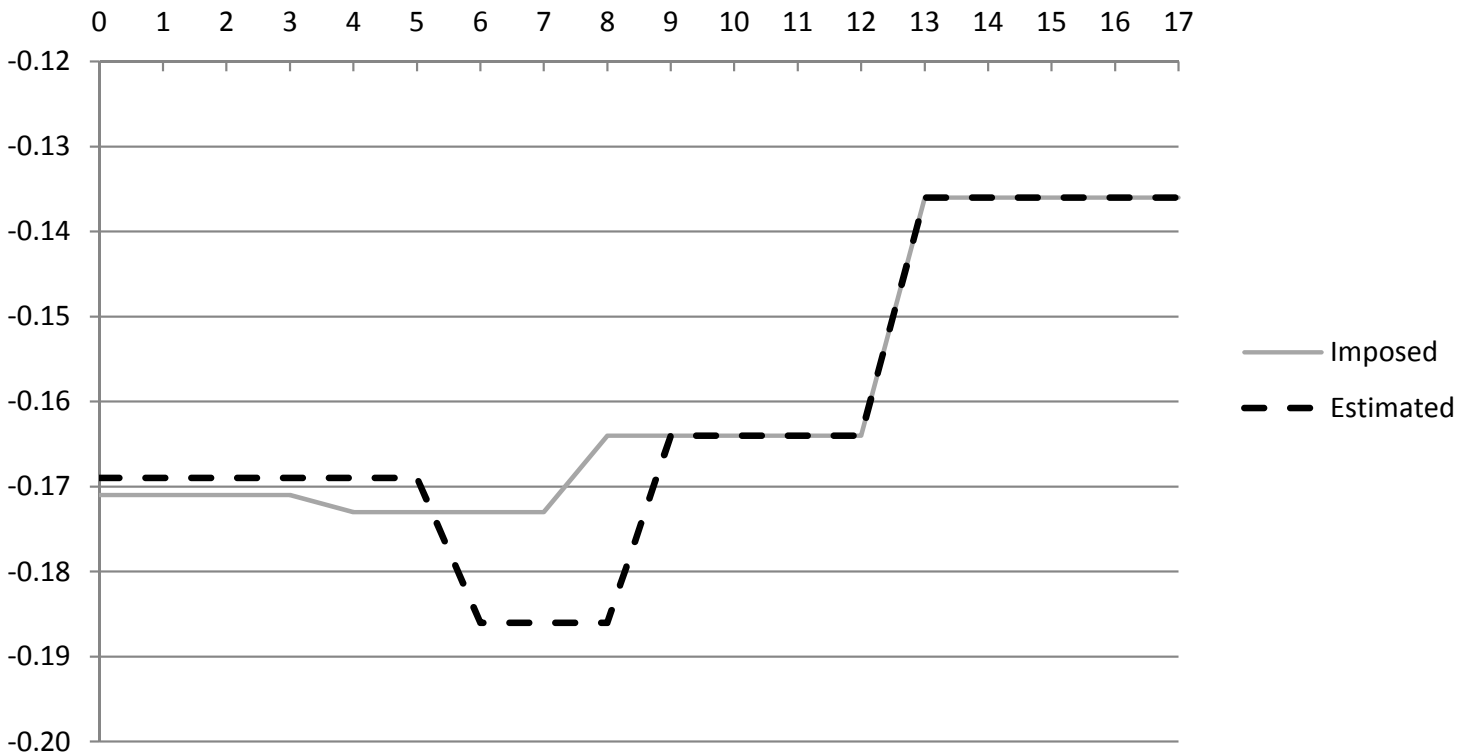
Threshold versus RP results

1. Four regimes; thresholds at ages 6, 8 and 12
2. Gradient is negative; $\gamma < 0 \Rightarrow \uparrow$ *income* $\rightarrow \uparrow$health (coded from good to bad)
3. γ Gets less negative with age $\Rightarrow$ income effects diminish as the child ages
4. *RP* and *AR* (1) *RP* very similar
   - $\widehat{\sigma}_{\alpha} = 0.018^{***}$;
   - $\widehat{\sigma}_{\alpha} = 0.016^{***}$ and $\widehat{\rho} = -0.246^{*}$

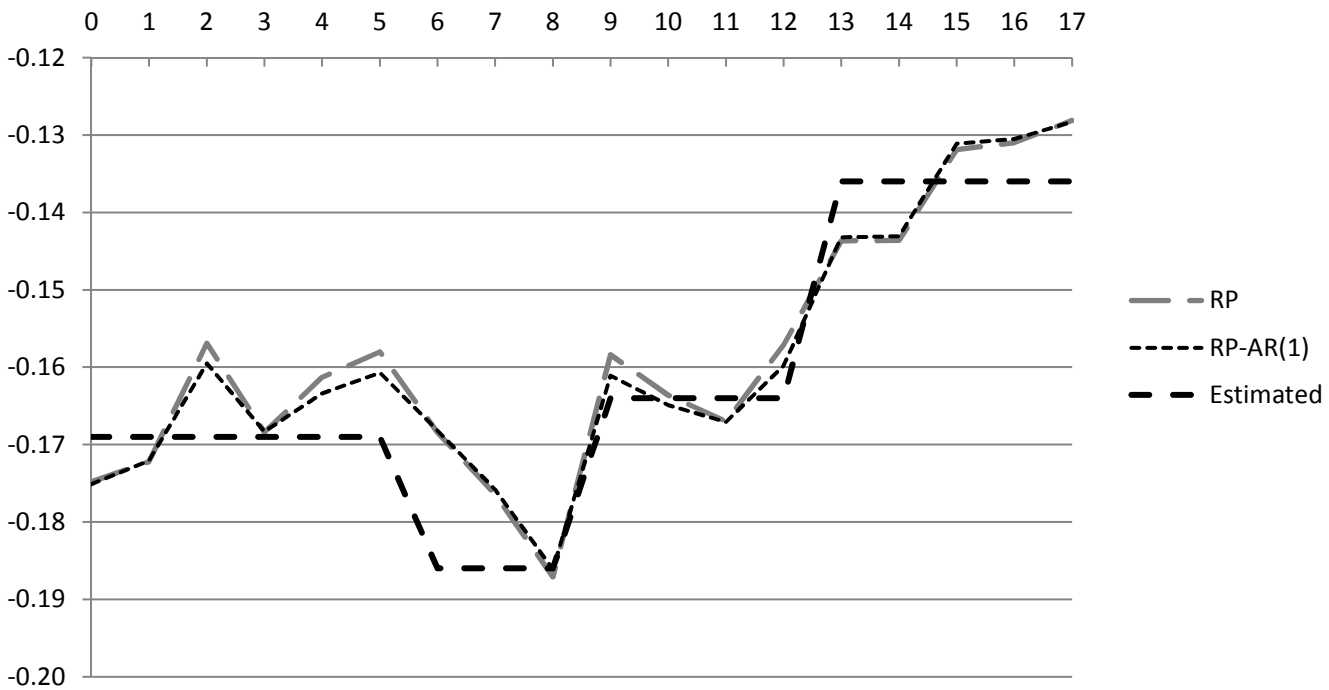
2008-2012 Optimal Values (Estimated Regimes)

	2	3	4	5
BIC(M)	15794.77	15794.84	**15789.79**	15789.86
$\widehat{\gamma}_1$	-0.1693	-0.1695	**-0.1685**	-0.1687
$\widehat{\gamma}_2$	-0.1355	-0.1442	**-0.1862**	-0.1864
$\widehat{\gamma}_3$		-0.1297	**-0.1641**	-0.1642
$\widehat{\gamma}_4$			**-0.1359**	-0.1446
$\widehat{\gamma}_5$				-0.1301
Number of Observations	9,613			

Threshold Effects



RP Effects


0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
-0.12
-0.13
-0.14
-0.15
-0.16
-0.17
-0.18
-0.19
-0.20
RP
RP-AR(1)
Estimated

Cohort effects

- Now we extend the analysis to allow for both endogenously determined age and cohort effects
- Consider *all* possible age thresholds, combined with *all* possible cohort effects!
- Thus we extend the generic model to additionally include the $c = 1, 2, \ldots, C$ cohort effects

$$H_i^* = \mathbf{x}_i'\beta + \sum_{m=1}^{M}\sum_{c=1}^{C}\gamma_{mc}\left(\ln y_i \times D_m \times D_c\right) + \varepsilon_i \qquad (9)$$

- Finding the optimal number of age-group and cohort effects $\left\{\widehat{M}, \widehat{C}\right\}$ proceeds as before, now with
- $D_M = 1 - D_1 - \ldots - D_{M-1}$ and $D_C = 1 - D_1 - \ldots - D_{C-1}$

Cohort effects

- Extend the RP approach for simultaneous cohort and child age thresholds:

$$\gamma_{tc} = \overline{\gamma} + \alpha_{tc} \tag{10}$$

- So γ_{tc} is specific to *every* age-cohort pairing
- Obtain individual (cohort-age) specific RP's *ex post*, as before

Cohorts

Born	1991 C0	1992 C1	1993 C2	1994 C3	1995 C4	1996 C5	1997 C6	1998 C7	1999 C8	2000 C9	2001 C10	2002 C11	2003 C12	2004 C13	2005 C14	200 C1
Age																
0																
1																
2																200
3															2008	200
4														2008	2009	201
5													2008	2009	2010	201
6												2008	2009	2010	2011	201
7											2008	2009	2010	2011	2012	
8										2008	2009	2010	2011	2012		
9									2008	2009	2010	2011	2012			
10								2008	2009	2010	2011	2012				
11							2008	2009	2010	2011	2012					
12						2008	2009	2010	2011	2012						
13					2008	2009	2010	2011	2012							
14				2008	2009	2010	2011	2012								
15			2008	2009	2010	2011	2012									
16		2008	2009	2010	2011	2012										
17	2008	2009	2010	2011	2012											

Cohort-age threshold results

- Optimal overall model is $\left\{\widehat{M}, \widehat{C}\right\} = 4, 2$
  - and positioned exactly the same as before: $<= 6$, $7 - 8$, $9 - 12$ and $>= 13$

- The only cohort effect is at the 20th cohort, born 2010

1. For children for born pre 20^{th} cohort, the optimal number of age-groups are 4

2. For those above the 20^{th} cohort?
   - 2.1 by definition, can only be aged 0-2
   - 2.2 optimal number of age groups is 5 (split 0-6 into 0-2 and 2-6)

Policy implications

- Clearly identified health-income inequalities for young children; in particular $0 - 2$
- Recent UK policy on giving children "the healthiest start possible" (The Marmot Review, 2010)
- Recent developments include:
  - double number of places on Family Nurse Partnership (support for young first time parents until their child is 2)

Concluding summary

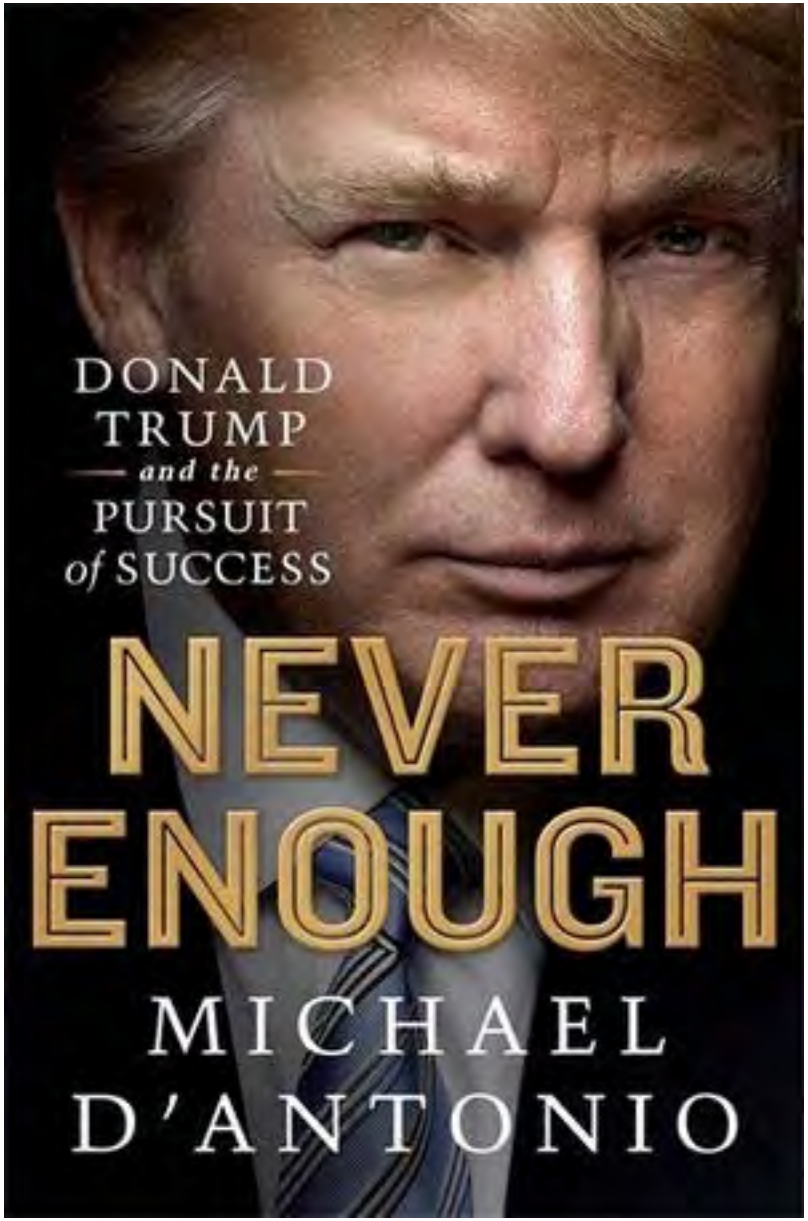
- In summary we considered two new alternative approaches for estimating the differential income-child health gradient
- Importantly these were both extended to allow for cohort effects
- The standard age groups imposed previously are approximately the same but do disguise significant heterogeneity before age 8
- The divergent effects we find however, at very young ages are very important in policy terms.

Discussion of:

"New Approaches to Estimating the Child Health-Parental Income Relationship"

by Gannon et al. (2015)

Discussion by Michael Shields, Monash University


DONALD
TRUMP
and the
PURSUIT
of SUCCESS
NEVER
ENOUGH
MICHAEL
D'ANTONIO

Contributions

1. Outlines two new modelling approaches for estimating the relationship between (ordinal) child health status and household income using cross-sectional data, paying attention to *thresholds* in the relationship across age

- Treats income effect as a random parameter, where the randomness is related to the specific age of the child
- Allows more flexibility and parsimony, arguing that using pre-defined age groupings used in previous studies might give an incomplete picture
- Implicitly suggests that these are better than simply estimating separate models for each age, or including interactions between age and income

2. Incorporates the identification of cohort effects in the modelling

3. Uses pooled cross-sections from the Health Survey for England (2008-12)

4. Discusses policy implications of the income gradient

Issues

- Extensive inter-disciplinary literature on the socioeconomic gradient in health – many measures, most correlational, few causal, in the UK

- Understanding the dynamics of health is important, including identifying 'sensitive' periods, and the extent and causes of inequalities in childhood (childhood predicts adulthood, intergenerational aspects of health)

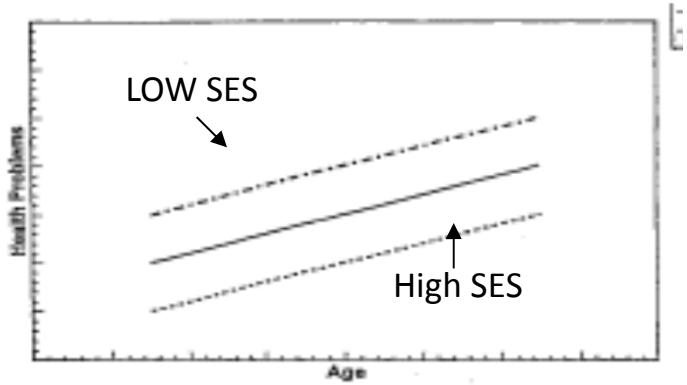
- Differences found across countries in age profile of the income (related) gradient – seminal study in economics by Case et al. (2002) for the US documented increasing income gradient with age

- Chen et al. (2002) provide a review of the large literature going back to the late 1960s that has examined how socioeconomic differences in children's health might change with age

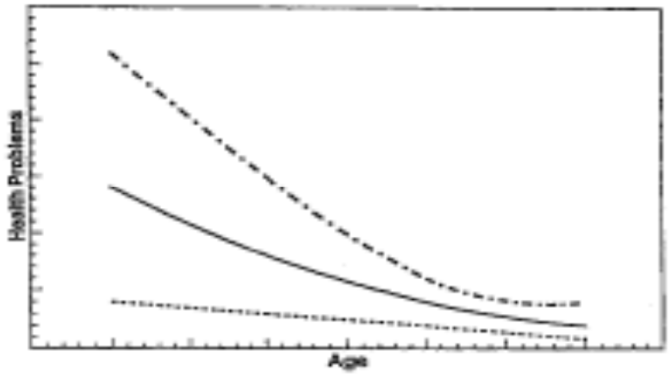
  - There is no "one-fit" answer

Possible Developmental Models
(Chen et al., 2002)

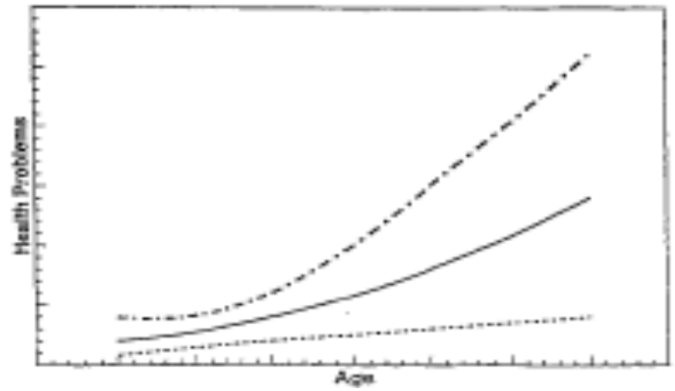
1. Childhood-adolescent persistent model
   - income gradient remains constant



2. Childhood limited model
   - income gradient decreases with age



3. Adolescent-emergent model
   - income gradient increases with age



<u>For example, think about most common conditions</u>:
Physical (e.g. asthma, accidents/injuries, eye problems)
Mental health (e.g. ADHD)

Motivation for this Paper

"Typically the extant literature has used the following age groupings 0–3, 4–8, 9–12 and 13 and higher. There is, however, no particular medical or theoretical reason for choosing these particular age bands."

0-3:	Pre-School / Nursery
4-7:	Infant School
8-11:	Junior School
11-16:	Secondary School
16-18:	6th Form

Roughly Correspond to schooling age in UK

Question: What can explain why the gap between children living in low income and high income households can increase or decrease significantly within one year? Currently the paper does not have such a discussion.

Comment: Include more detailed discussion of economic, medical or developmental theories etc.., to suggest hypotheses that the statistical analyses test
e.g. probability of diagnosis between poor and rich changes with age?

Estimated Income Gradient in Child Health by Age Group for England

Age Grouping	**0-3**	**4-8**	**9-12**	**13-15(17)**
Propper et al. (2004) General Household Survey (2000-2002)	-0.209 (0.042)	-0.185 (0.051)	-0.083 (0.047)	-0.061 (0.044)
Currie et al. (2007) Health Survey for England (1997-2002)	-0.146 (0.040)	-0.212 (0.028)	-0.196 (0.031)	-0.174 (0.034)
Case et al. (2008) Health Survey for England (1997-2005)	-0.141 (0.029)	-0.207 (0.022)	-0.229 (0.025)	-0.180 (0.029)
	0-3	**4-8**	**9-12**	**13-17**
Gannon et al. (2015) Health Survey for England (2008-2012)				
Age Group (Imposed Age Groupings)	-0.171***	-0.173***	-0.164***	-0.136***
	0-5	**6-8**	**9-12**	**13-17**
Age Group (Endogenous Age Groupings)	-0.169***	-0.186***	-0.164***	-0.136***

Consistent evidence that the UK looks different to the results found for the US

Other British Studies

West (1997, 1991 Census) and West and Sweeting (2004)
Find evidence of equalisation of health in adolescence in Scotland

Emerson et al. (2005, Mental Health Surveys)
Low household income associated with poorer health for 13 out of 24 indicators
"Little evidence of any systematic differences in the extent of health inequalities across age groups (5-10, 11-15)".

Propper et al. (2007, Avon Cohort - LSPAC)
Find no evidence that the association steepens with age between 0 and 7 years

Kruk (2013, Millennium Cohort)
"Higher household income increases the probability that children fully recover from some diseases within a given period".

However: A number of studies have found increasing socioeconomic gradients in health and health-related behaviours in adulthood (at least until retirement) in UK
For example:

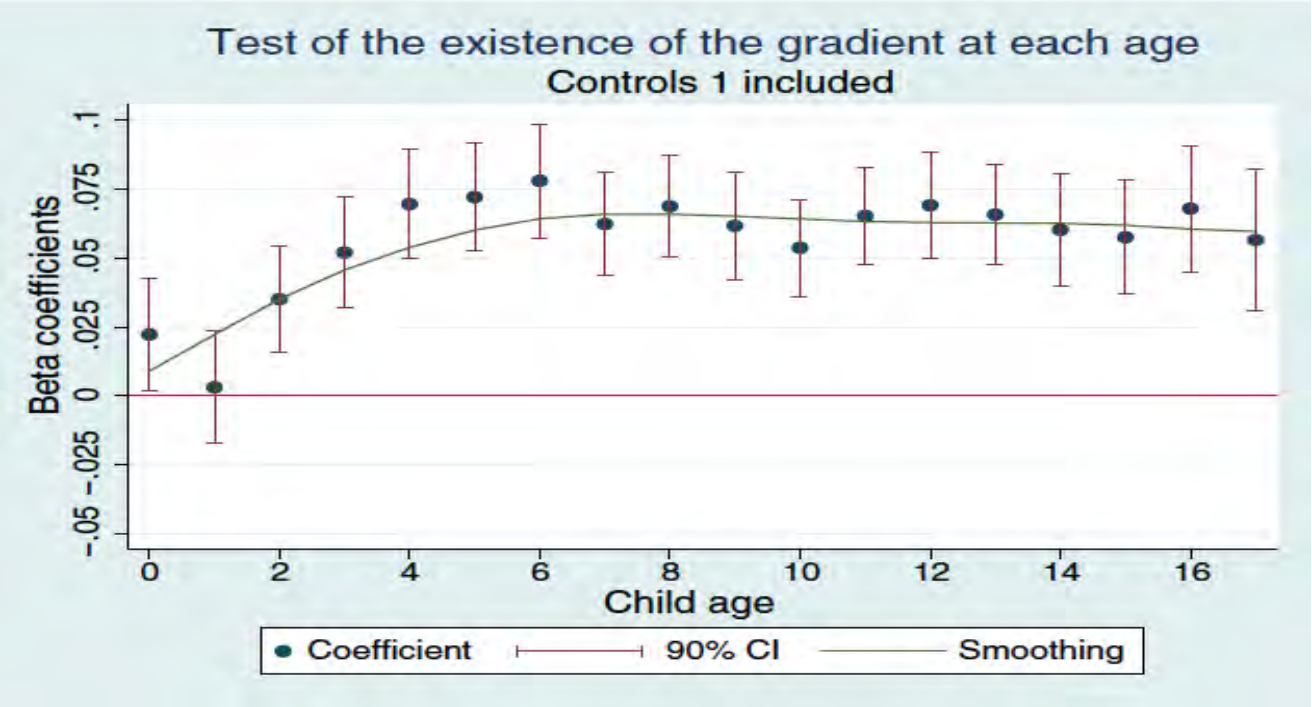
- Farrell et al. (2014) – increasing gap in physical inactivity
- Schurer et al. (2014) – increasing gap in self-reported bodily pain

Apouey and Geoffard (2013, JHE)

Arguably use better data than the HSE used in the current paper

- UK Family and Children Survey (FACS, 2001-08)
- Larger sample (79,000 observations)
- Better income measure (averaged over some years)
- More consistent child health measure (always answered by parent)

Linear probability of models of poor child health with (log) income/age interactions



Shows increasing income gradient until around age 4, then constant from 5-17

Causal UK Studies

Lindeboom et al.(2008, JHE, National Child Development Study)
Use changes in the compulsory school leaving age and finds that parental education (at least one additional year) does not explain child health. They also "conclude from this that the effects of parental income on child health are at most modest."

Kuehnle (2014, JHE, Millennium Cohort Study)
Uses (exogenous) variation in local area labour markets to instrument for household income

"We find that income has a very small but significant causal effect of subjective children health and no significant effect on chronic health conditions"

Potential Measurement Error in Household Income in the HSE?

Ideally want a measure of poverty persistence, or a measure of 'averaged' income

Household income measured in bands in HSE
No wealth or consumption data
About of 18% cases income missing

<u>Intergenerational income (correlation) literature</u>

- 0.2 increasing to up to 0.6 with better income measure (between fathers and sons)
- Best measured in mid-40's (13-17 children)

<u>Consumption is greater than income suggests</u>

- Brewer et al. (EJ, forthcoming)
- Bottom 1% income found to have consumption close to the median household
- Difficult for those receiving welfare to record income

<u>Reverse Causality</u>

- Literature showing labour supply response
- Mothers dropping out of work to look after child
- Issue most important for young pre-school children (the 0-3 age group estimates)

HSE (2010)

(D) Total Household Income	Freq.	Percent	Cum.
Not applicable	103	1.74	1.74
<£520	5	0.08	1.83
£520<£1,600	5	0.08	1.91
£1,600<£2,600	19	0.32	2.23
£2,600<£3,600	12	0.20	2.43
£3,600<£5,200	68	1.15	3.58
£5,200<£7,800	171	2.89	6.47
£7,800<£10,400	260	4.39	10.87
£10,400<£13,000	289	4.89	15.75
£13,000<£15,600	268	4.53	20.28
£15,600<£18,200	273	4.61	24.90
£18,200<£20,800	262	4.43	29.33
£20,800<£23,400	267	4.51	33.84
£23,400<£26,000	230	3.89	37.73
£26,000<£28,600	175	2.96	40.69
£28,600<£31,200	205	3.47	44.15
£31,200<£33,800	127	2.15	46.30
£33,800<£36,400	179	3.03	49.32
£36,400<£41,600	291	4.92	54.24
£41,600<£46,800	242	4.09	58.33
£46,800<£52,000	281	4.75	63.08
£52,000<£60,000	327	5.53	68.61
£60,000<£70,000	210	3.55	72.16
£70,000<£80,000	159	2.69	74.85
£80,000<£90,000	133	2.25	77.10
£90,000<£100,000	94	1.59	78.68
£100,000<£110,000	60	1.01	79.70
£110,000<£120,000	27	0.46	80.16
£120,000<£130,000	30	0.51	80.66
£130,000<£140,000	36	0.61	81.27
£140,000<£150,000	21	0.35	81.63
>=£150,000	126	2.13	83.76
Do not know	465	7.86	91.62
Refused	496	8.38	100.00
Total	5,916	100.00	

Measurement Inconsistency in Child Health in the HSE

For children aged 0-12, a parent (mostly mothers) assess their child's general health status on an ordinal scale
(Very good = 62%; Good = 33%; Fair = 4%, Bad/Very Bad = 1%)

For children aged 13-17, the child assesses their own health on the same scale. This might explain the drop in the income gradient for older children in the HSE, compared to the FACS data used by Apouey and Geoffard (2013)

Johnston et al. (2014), for example, compare estimated income gradients in childhood mental health using (SDQ) evaluations for parents, teachers and children. Find that income gradient is smallest using child's own responses

Table 6. Estimated coefficients of log-income from ordered probit models†

	Parent, (1)	*Teacher, (2)*	*Child, (3)*
Emotional SDQ	−0.159‡	−0.261‡	−0.115‡
	(0.033)	(0.035)	(0.032)
Conduct SDQ	−0.205‡	−0.136‡	−0.054§
	(0.034)	(0.038)	(0.032)
Hyperactivity SDQ	−0.143‡	−0.163‡	−0.010
	(0.032)	(0.033)	(0.032)

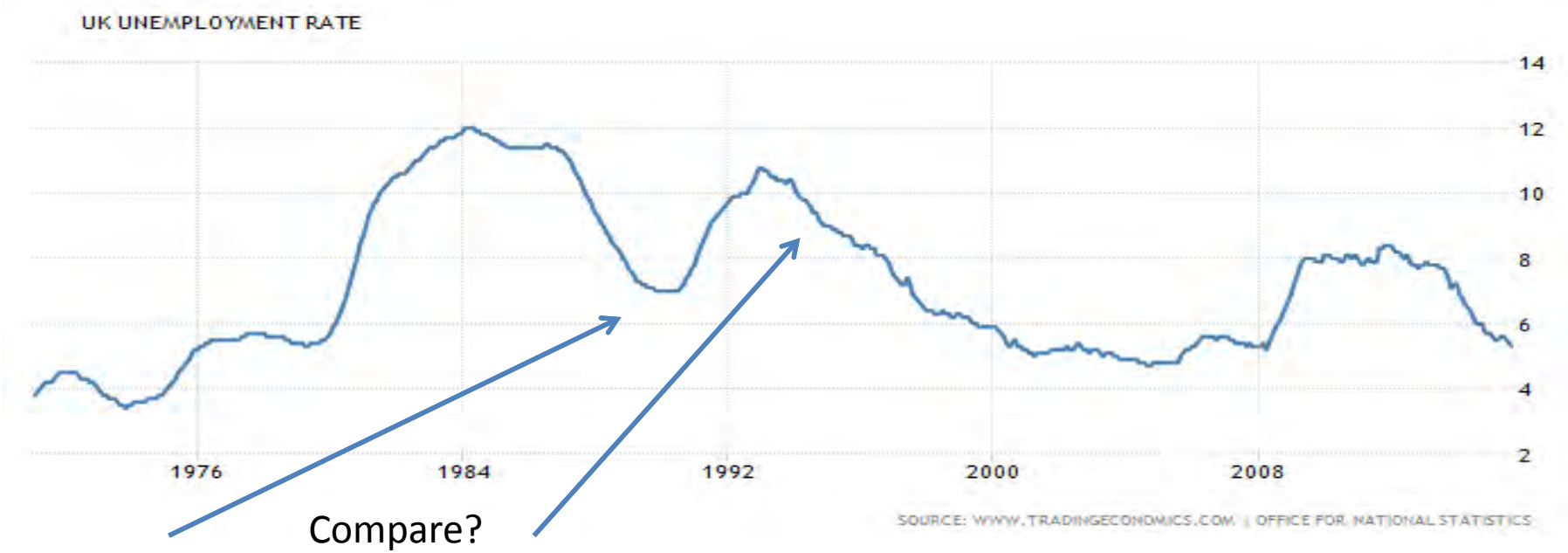
What does the Income Gradient tell us?

Income can be correlated with many plausible factors that can impact of child health, but the policy implications differ greatly. For example, what about a policy (increase in the minimum wage, or unemployment benefits) of 20%:

Might Improve	Fixed
Reduce financial stress	Cognitive ability of parents
Improve parental health	Parental time and risk preferences
Improve housing quality	Parental tastes (smoking, alcohol)
Buy better diet	Family and friends (and peers)
Pay for more physical activity	Culture
Increase access to health care?	
Higher quality schooling?	

Studies have also found that who receives the income makes a difference to what it is spent on

Cohort Effects in the Income/Child Health Relationship



Born	1991 C0	1992 C1	1993 C2	1994 C3	1995 C4	1996 C5	1997 C6	1998 C7	1999 C8	2000 C9	2001 C10	2002 C11	2003 C12	2004 C13	2005 C14	2006 C15	2007 C16	2008 C17	2009 C18	2010 C19	2011 C20	2012 C21
Age																						
0																		2008	2009	2010	2011	2012
1																	2008	2009	2010	2011	2012	
2																2008	2009	2010	2011	2012		
3															2008	2009	2010	2011	2012			
4														2008	2009	2010	2011	2012				
5													2008	2009	2010	2011	2012					
6												2008	2009	2010	2011	2012						
7											2008	2009	2010	2011	2012							
8										2008	2009	2010	2011	2012								
9									2008	2009	2010	2011	2012									
10								2008	2009	2010	2011	2012										
11							2008	2009	2010	2011	2012											
12						2008	2009	2010	2011	2012												
13					2008	2009	2010	2011	2012													
14				2008	2009	2010	2011	2012														
15			2008	2009	2010	2011	2012															
16		2008	2009	2010	2011	2012																
17	2008	2009	2010	2011	2012																	

Sample sizes

Literature – van den Berg / Lindeboom etc…

Threshold found for previous recession?

Who were most affected by the GFC?

Use local area economic indicators

Placebo test?

Some Potential Suggestions (1)

1. Place more emphasis on new models, and their potential use in a wide-range of applications in health economics, and perhaps less on new contribution to the substantial child health literature given data issues and non-causal design

"Two New Approaches to Modelling Important Thresholds in Health Economics"

2. Discuss economic/medical/child development theories at start of paper to better motivate the study, and statistically estimated age thresholds

3. Expand literature review if cohort effects are the main new result

4. Discuss measurement error issues in household income
What about functional form – for example, using income deciles (thresholds)
What about household-equivalence income if household size differs across SES

5. Maybe drop ages (13-17) because of change in respondent – or argue should be a threshold found at age 12/13 years. Or possibly use the FACS data instead, which also has a better income measure. Directly compare results found with a simpler interactions specification (or separate models) to highlight the practical value of the new models

Some Potential Suggestions (2)

5. Provide more detailed analysis of the cohort effects of GFC; placebo tests; compare to recession in 1990's, perhaps use macroeconomic indicators at birth across English local areas

6. Think of providing policy-relevant marginal effects and good-of-fit measures

- Not causal, but ME of a 20% increase in income for the lowest decile of income
- How important is household income in explaining the overall variation in child health?

7. Discuss the potential for the models to be extended to the case of panel data, and allowing for endogeneity of household income (IV), given more recent literature and the availability of new cohort data for the UK that tracks the same child

2SLS vs 2SRI: Appropriate methods for rare outcomes and/or rare exposures

Anirban Basu, Norma Coe

University of Washington, Seattle

Background

- The economics literature is teeming with applications where linear probability models (LPM) are used on binary outcomes.
- In case of instrumental variables methods,
- both the binary treatment (in 1st stage) and the binary outcome (in 2nd stage) modeled using LPM as in two-stage least squares (2SLS) estimators.
- Concerns
  - propensity to misspecify the underlying average structural function of the binary data generating process.
  - can make out of range predictions

Background

- Alternatively, non-linear models, such as those using probit or logit links functions can be used
  - do not allow for out of range predictions
  - in most cases provides good fit
- Issues
  - 2SLS linear models proposed for binary outcomes (Angrist and Pischke, 2009)
  - Predictor substitution approach is problematic when non-linear models such as probit and logistic are used in the second stage (Terza et al. 2008)
  - Instead, a control function approach, with residual inclusion is proposed (Blundell and Powell 2003, 2004)
- Other estimators, not studied here:
  - GMM approaches (McCarthy and Tchernis 2011)
  - semi-parametric estimators (Abadie 2003; Abrevaya et al. 2009, Chiburis 2010; Shaikh and Vytlacil 2011).

Background

- The comparative performance of 2SLS vs 2SRI for binary outcome and binary treatment has not been studied extensively, especially when either the outcome or the treatment is rare.

- GOALS OF THIS WORK:
- Extensive simulations are carried out to answer these questions, including the degree of rarity in outcomes/treatment required for 2SLS to break down
- Application to the effects of long-term care on health.

Instrumental Variable Analysis

Two-stage Methods:

Stage 1: Run propensity model same as before, but after adjusting for IVs

$\log\left(\frac{p}{1-p}\right) = \alpha_0 + \alpha_1 X + \alpha_2 Z$, where p = E(D|X,Z)

Predict the estimated propensity score for each subject in the sample = $\hat{p}(x,z)$ (exogenous part)

Compute residual for each individual: $\hat{r} = D - \hat{p}(x,z)$ (endogenous part)

$\hat{\alpha}_2$

Instrumental Variable Analysis

Two-stage Methods:

Stage 2a: (Predictor Substitution) Run outcomes regression with

$\hat{p}(x,z)$ REPLACING *D*:

$$Y = \beta_0 + \beta_1 \hat{p} + \beta_2 X + \varepsilon$$

OR

Stage 2b: (Residual Inclusion) Run outcomes regression with

$\hat{r} = D - \hat{p}(x,z)$ as one of the covariates

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 \hat{r} + \varepsilon$$

Instrumental Variable Analysis

Two-stage Methods:

Stage 2a: (Predictor Substitution) Run outcomes regression with

$\hat{p}(x,z)$ REPLACING *D*:

OR ~~$Y = g(\beta_0 + \beta_1 \hat{p} + \beta_2 X) + \varepsilon$~~ FOR NON-LINEAR MODELS

Stage 2b: (Residual Inclusion) Run outcomes regression with

$\hat{r} = D - \hat{p}(x,z)$ as one of the covariates

$$Y = g(\beta_0 + \beta_1 D + \beta_2 X + \beta_3 \hat{r}) + \varepsilon$$

Terza, Basu & Rathouz JHE 2008

Theoretical foundations

- Structural response model

  $y_i = 1\{y_i^* > 0\}$, where

  $y_i^* = x_i\beta + u_i$

If u_i was independent of x_i, a single index regression model

$E(y_i \mid x_i) = G(x_i\beta)$ $\quad$ $G(a) = Pr\{-u_i \leq a)$

Could be used to obtain consistent estimates of β

- Problem: u_i likely not independent of x_i, because some component of x_i , say d_i, are determined jointly with y_i^* .
- i.e. $x_i = (d_i, w_i)$ and

  $y_i = 1\{d_i\beta_1 + w_i\beta_2 + u_i > 0\}$

Theoretical foundations

- Let the reduced form of d_i be given as

$$d_i = E(d_i | w_i, z_i) + v_i$$

$$= \lambda(w_i, z_i) + v_i$$

z_i = vector of instruments, and $E(v_i \mid w_i, z_i) = 0$.

Theoretical foundations

Approach 1 (Fully parametric):

Joint distribution of the structural error term u_i and the reduced form error term v_i were parametrically specified (e.g. Normal), and

$\lambda(w_i, z_i)$ is parametrically specified,

$$E(y_i \mid d_i, w_i, v_i) = Pr(u_i > -d_i\beta_1 - w_i\beta_2 \mid v_i)$$
$$= \Phi(d_i\beta_1 + w_i\beta_2 + \rho v_i)$$

Then β, $\lambda(.)$ and ρ can be estimated using maximum likelihood estimation. (e.g. Bivariate probit regression (Heckman 1978)

Theoretical foundations

Approach 2 (Semi-parametric):

Insert the reduced form for d_i in the structural model for y_i.

$$y_i = 1\{d_i\beta_1 + w_i\beta_2 + u_i > 0\}$$
$$= 1(\lambda(w_i, z_i)\,\beta_1 + w_i\beta_2 + u_i + v_i\,\beta_1 > 0\}$$
$$= 1(\lambda(w_i, z_i)\,\beta_1 + w_i\beta_2 + (e_i) > 0\}$$

Semiparametric estimation of this regression function will yield consistent estimates for β, only if e_i is independent of z_i . HOWEVER,

- This is difficult to maintain as z_i may be dependent on v_i.
- Moreover, the "fitted value" approach does not yield a model for G(.) of the error term $-u_i$.

Theoretical foundations

Approach 3 (Semi-parametric):

Use estimates of reduced form error tem v_i as “control variables”.

Identification is through a distributional exclusion restriction. Therefore

$$E(y_i \mid d_i, w_i, v_i) = Pr[-u_i \leq d_i\beta_1 + w_i\beta_2 \mid d_i, w_i, v_i\}$$

$$= F(d_i\beta_1 + w_i\beta_2 \mid v_i)$$

Where F(.) is the conditional c.d.f. of $-u_i$ given v_i

The Marginal distribution function G(.) with respect to $-u_i$ could be identified

$$G(d_i\beta_1 + w_i\beta_2) = \int F(d_i\beta_1 + w_i\beta_2 , v_i)F_v$$

Blundell & Powell 2003, 2004

Theoretical foundations

Approach 3 (Semi-parametric):

Identification is through a distributional exclusion restriction:

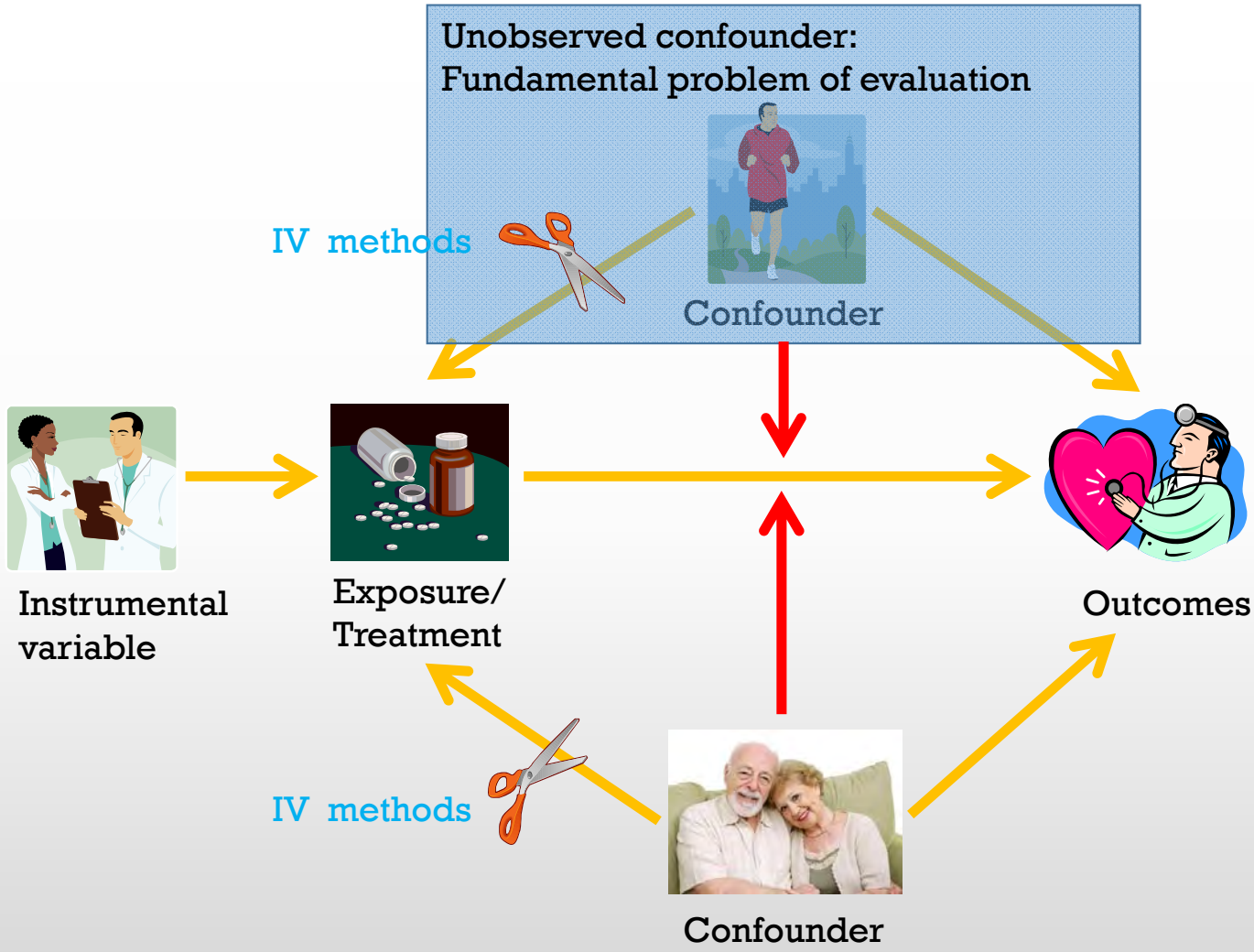
$u_i \mid d_i, w_i, z_i \sim u_i \mid d_i, w_i, v_i \sim u_i \mid v_i$

Under what form of v_i does this exclusion restriction hold?

- Raw-scale? (as defined here)
- Standardized form?
- Other forms?

Literature

- Underdeveloped literature on understanding what is the right "control function"
- Garrido et al. (2012) compared results from 2SRI models with different versions of residuals when applied to health expenditure data.
  - They found that results varied widely
  - They did not do any simulations to show which one is better
  - They raised the concern that raw residuals may not be the right control function variable.


Unobserved confounder:
Fundamental problem of evaluation
IV methods
Confounder
Instrumental variable
Exposure/ Treatment
Outcomes
IV methods
Confounder

Problems of Local Average Treatment Effect (LATE)

- LATE - the average treatment effect for individuals who would change their treatment choice when *instrument level* moves
  - But who are these "marginal" patients?
  - At what margins will my policy induce changes in treatment choices?
- IV effects (combinations of LATEs) are misleading as they can be arbitrarily weighted.
- However, in some cases where a binary IV is related to a specific policy, LATE may be interpretable

Imbens and Angrist 1994; Heckman, 1997; McClellan, Newhouse, 1998; Harris, Remler, 1998; Heckman, Vytlacil, 1999

FOCUS OF THIS WORK

- CONSIDER THE MOST SIMPLEST CASE
  - Binary outcome (y_i)
  - Binary treatment (d_i)
  - Binary controls (w_i)
  - Binary instrument (z_i)
- Questions asked
  - Can linear approximation (2SLS) provide consistent estimates for LATE
  - What form of residuals are most suited in a 2SRI Probit-Probit approach
  - How do the results change if outcomes (y_i) and/ or treatment (d_i) becomes rare

SIMULATIONS

Data generating processes

Exposure (treatment) DGP

$D^* = \alpha_0 + \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \alpha_3 \cdot X_3 + \alpha_U \cdot X_U + \alpha_Z \cdot Z - \omega$

- $(\alpha_1, \alpha_2, \alpha_3) = (0.5, 1, 2)$, $\alpha_U = 1$, $\alpha_Z = 1$.
- Observed variables X_1, X_2, X_3 and Z are all binary variables with mean equal to 0.5. X_U ~Normal (0,1).
- ω ~ Normal(0,1)
- $D = (D^* > 0)$
- $Pr(D) = \Phi((\alpha_0 + 2.25)/\sqrt{3.5625})$.
- Vary α_0 to take on values of -2, -1.5, -0.3, 0.5, 1.5 so that they correspond to Pr(D) = 0.55, 0.70, 0.85, 0.93, and 0.98.

Data generating processes

Outcome generating processes

$Y^* = \beta_0 + \beta_D \cdot D + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_U \cdot X_U - \varepsilon$

- $(\beta_1, \beta_2, \beta_3) = (1,1,1)$, $\beta_U = 2$, $\beta_D = 1$. $\varepsilon \sim$ Normal(0,1) .
- $Pr(Y|D) = \Phi((\beta_0 + \beta_D \cdot D + 1.5)/\sqrt{5.75}))$

- Vary β_0 to take on values of -2, 0.5, 1.5, 2.5 so that they correspond to Pr(Y) = 0.5, 0.82, 0.93, and 0.96.

- LATE: $E_X\{[E(Y|Z=1, X) - E(Y|Z=0,X)] /[E(D|Z=1, X) - E(D|Z=0,X)]\}$

  $= E_X\{ [(E(D|Z=1,X)*E(Y|D=1,X) + E(1-D|Z=1,X)*E(Y|D=0,X)) -$

  $(E(D|Z=0,X)*E(Y|D=1,X) + E(1-D|Z=0,X)*E(Y|D=0,X))] /[E(D|Z=1, X) - E(D|Z=0,X)] \}$

LATE

	E(D|Z=1,X)	E(D|Z=0,X)	E(Y|D=1,X)	E(Y|D=0,X)
X1=0, X2=0, X3=0	$\Phi((\alpha_0 + 1)/\sqrt{2})$	$\Phi((\alpha_0 + 0)/\sqrt{2})$	$\Phi((\beta_0 + 1)/\sqrt{3})$	$\Phi((\beta_0 + 0)/\sqrt{3})$
X1=0, X2=0, X3=1	$\Phi((\alpha_0 + 3)/\sqrt{2})$	$\Phi((\alpha_0 + 2)/\sqrt{2})$	$\Phi((\beta_0 + 2)/\sqrt{3})$	$\Phi((\beta_0 + 1)/\sqrt{3})$
X1=0, X2=1, X3=0	$\Phi((\alpha_0 + 2)/\sqrt{2})$	$\Phi((\alpha_0 + 1)/\sqrt{2})$	$\Phi((\beta_0 + 2)/\sqrt{3})$	$\Phi((\beta_0 + 1)/\sqrt{3})$
X1=1, X2=0, X3=0	$\Phi((\alpha_0 + 1.5)/\sqrt{2})$	$\Phi((\alpha_0 + 0.5)/\sqrt{2})$	$\Phi((\beta_0 + 2)/\sqrt{3})$	$\Phi((\beta_0 + 1)/\sqrt{3})$
X1=0, X2=1, X3=1	$\Phi((\alpha_0 + 4)/\sqrt{2})$	$\Phi((\alpha_0 + 3)/\sqrt{2})$	$\Phi((\beta_0 + 3)/\sqrt{3})$	$\Phi((\beta_0 + 2)/\sqrt{3})$
X1=1, X2=0, X3=1	$\Phi((\alpha_0 + 3.5)/\sqrt{2})$	$\Phi((\alpha_0 + 2.5)/\sqrt{2})$	$\Phi((\beta_0 + 3)/\sqrt{3})$	$\Phi((\beta_0 + 2)/\sqrt{3})$
X1=1, X2=1, X3=0	$\Phi((\alpha_0 + 2.5)/\sqrt{2})$	$\Phi((\alpha_0 + 1.5)/\sqrt{2})$	$\Phi((\beta_0 + 3)/\sqrt{3})$	$\Phi((\beta_0 + 2)/\sqrt{3})$
X1=1, X2=1, X3=1	$\Phi((\alpha_0 + 4.5)/\sqrt{2})$	$\Phi((\alpha_0 + 3.5)/\sqrt{2})$	$\Phi((\beta_0 + 4)/\sqrt{3})$	$\Phi((\beta_0 + 3)/\sqrt{3})$

$$\text{LATE} = E_X\left\{\frac{E(Y|Z=1,X)-E(Y|Z=0,X)}{E(D|Z=1,X)-E(D|Z=0,X)}\right\}$$

	Exposure DGP (α_0)				
Outcomes DGP (β_0)	-2	-1.25	-0.3	0.5	1.5
-2	Pr(D) = 0.55 E(Y) = 0.51 LATE = 0.20	Pr(D) = 0.70 E(Y) = 0.54 LATE = 0.20	Pr(D) = 0.85 E(Y) = 0.57 LATE = 0.20	Pr(D) = 0.93 E(Y) = 0.57 LATE = 0.20	Pr(D) = 0.995 E(Y) = 0.58 LATE = 0.20
0.5	Pr(D) = 0.55 E(Y) = 0.82 LATE = 0.09	Pr(D) = 0.70 E(Y) = 0.84 LATE = 0.09	Pr(D) = 0.85 E(Y) = 0.86 LATE = 0.09	Pr(D) = 0.93 E(Y) = 0.87 LATE = 0.09	Pr(D) = 0.995 E(Y) = 0.89 LATE = 0.09
1.5	Pr(D) = 0.55 E(Y) = 0.93 LATE = 0.042	Pr(D) = 0.70 E(Y) = 0.93 LATE = 0.042	Pr(D) = 0.85 E(Y) = 0.93 LATE = 0.042	Pr(D) = 0.93 E(Y) = 0.95 LATE = 0.042	Pr(D) = 0.995 E(Y) = 0.95 LATE = 0.042
2.5	Pr(D) = 0.55 E(Y) = 0.96 LATE = 0.015	Pr(D) = 0.70 E(Y) = 0.96 LATE = 0.015	Pr(D) = 0.85 E(Y) = 0.96 LATE = 0.015	Pr(D) = 0.93 E(Y) = 0.98 LATE = 0.015	Pr(D) = 0.995 E(Y) = 0.98 LATE = 0.015

Simulations and Estimators

Simulations

- Monte-Carlo simulations performed over 1000 replicates of datasets , each of sample size 50000.
- For each replicate, 500 bootstrap samples were used to calculated std. error and coverage.

Estimators

- IV regression with LPM
- Probit-Probit 2SRI with
  - Raw residuals
  - Standardized residuals
  - Deviance residuals
  - Anscombe residuals
- Bi-variate Probit

Residuals

Standardized (Pearson) residuals:

$$(y_i - \widehat{y_i})/\sqrt{\{(1-\widehat{y_i})\,\widehat{y_i}\}}$$

Deviance Residuals:

$$\sqrt{2\left\{y_i log\left(\frac{y_i}{\widehat{y_i}}\right) + (1-y_i) log\left(\frac{1-y_i}{1-\widehat{y_i}}\right)\right\}}$$

Anscombe residuals

$$(A(y_i) - A(\widehat{y_i}))/[A'(\widehat{y_i})\sqrt{\{(y_i - \widehat{y_i})\,\widehat{y_i}\}}\,]$$

Bernoulli: $(B(y_i, \frac{2}{3}, \frac{2}{3}) - B(\widehat{y_i}, \frac{2}{3}, \frac{2}{3}))/[\sqrt{\{(1-\widehat{y_i})\,\widehat{y_i}\}}\,]^{-1/6}$, B() = Beta Function

Results: %Bias (CV) {Coverage Pr}

E(Y)	Estimators	Pr(D) = 0.55	Pr(D) = 0.70	Pr(D) = 0.85	Pr(D) = 0.93	Pr(D) = 0.995
0.50~0.60	Naïve Probit	187 (.01) {.00}	178 (.01) {.00}	156 (.01) {.00}	137 (.01) [.00}	126 (.01) {.00}
0.50~0.60	IV-LPM	**6 (.09) {.89}**	**-2 (.12) {.92}**	-26 (.23) {.63}	-53 (.59) {.53}	-69 (1.43) {.65}
0.50~0.60	2SRI	-45 (.20) {.01}	-33 (.17) {.15}	**7 (.12) {.89}**	50 (.11) {.18}	77 (.12) {.10}
0.50~0.60	2SRI - sres	19 (.09) {.54}	32 (.10) {.25}	56 (.11) {.11}	81 (.15) {.10}	95 (.16) {.12}
0.50~0.60	2SRI - dres	-106 (-1.69) {.00}	-102 (-7.89) {.00}	-61 (.42) {.04}	11 (.18) {.9}	66 (.14) {.20}
0.50~0.60	2SRI - ares	-90 (1.08) {.00}	-84 (.75) {.00}	-49 (.31) {.13}	**8 (.19) {.91}**	**56 (.15) {.37}**
0.50~0.60	Bi.Probit	-18 (.10) {.44}	-18 (.11) {.44}	-17 (.15) {.71}	-20 (.3) {.87}	-21 (.55) {.93}
0.80 ~0.90	Naïve Probit	269 (.01){.00}	346 (.01) {.00}	446 (.01) {.00}	533 (.01) {.00}	587 (.02) {.00}
0.80 ~0.90	IV-LPM	10 (.16) {.89}	57 (.11) {.12}	116 (.10) {.00}	141 (.16) {.05}	140 (.26) {.39}
0.80 ~0.90	2SRI	**9 (.14) {.90}**	**-2 (.17) {.92}**	**-42 (.33) {.38}**	-66 (.77) {.30}	-63 (1.09) {.54}
0.80 ~0.90	2SRI - sres	93 (.06) {.00}	168 (.05) {.00}	247 (.06) {.00}	315 (.08) {.00}	366 (.08) {.00}
0.80 ~0.90	2SRI - dres	-69 (.64) {.06}	-94 (2.95) {.00}	-113 (-1.23) {.00}	-104 (-5.33) {.02}	-73 (1.23) {.36}
0.80 ~0.90	2SRI - ares	-44 (.32) {.31}	-50 (.34) {.18}	-53 (.37) {.20}	**-39 (.40) {.62}**	**-9 (.42) {.91}**
0.80 ~0.90	Bi.Probit	7 (.12) {.91}	8 (.12) {.93}	9 (.13) {.89}	8 (.18) {.95}	9 (.27) {.94}

Results: %Bias (CV) {Coverage Pr}

E(Y)	Estimators	Pr(D) = 0.55	Pr(D) = 0.70	Pr(D) = 0.85	Pr(D) = 0.93	Pr(D) = 0.995
0.9 ~ 0.95	Naïve Probit	349 (.02) {.00}	490 (.02) {.00}	707 (.02) {.00}	932 (.02) {.00}	1094 (.02) {.00}
0.9 ~ 0.95	IV-LPM	**5 (.28) {.94}**	77 (.16) {.27}	202 (.12) {.00}	321 (.14) {.00}	381 (.19) {.02}
0.9 ~ 0.95	2SRI	71 (.11) {.06}	57 (.15) {.33}	-26 (.36) {.83}	-108 (-3.69) {.13}	-132 (-1.1) {.14}
0.9 ~ 0.95	2SRI - sres	149 (.06) {.00}	252 (.06) {.00}	407 (.06) {.00}	540 (.08) {.00}	636 (.09) {.00}
0.9 ~ 0.95	2SRI - dres	-13 (.33) {.95}	-59 (.69) {.51}	-108 (-3.03) {.02}	-121 (-1.22) {.02}	-111 (-3.06) {.18}
0.9 ~ 0.95	2SRI - ares	13 (.22) {.91}	**-3 (.26) {.93}**	**-12 (.29) {.92}**	**-6 (.37) {.93}**	**16 (.42) {.96}**
0.9 ~ 0.95	Bi.Probit	37 (.15) {.60}	38 (.14) {.54}	39 (.14) {.61}	39 (.19) {.74}	40 (.25) {.85}
0.95~0.98	Naïve Probit	490 (.02) {.00}	734 (.02) {.00}	1165 (.02) {.00}	1691 (.03) {.00}	2116 (.03) {.00}
0.95~0.98	IV-LPM	**-2 (.59) {.94}**	97 (.33) {.66}	314 (.20) {.03}	614 (.18) {.00}	837 (.20) {.01}
0.95~0.98	2SRI	172 (.10) {.00}	159 (.15) {.02}	70 (.28) {.71}	-150 (-1.11) {.26}	-215 (-.43) {.08}
0.95~0.98	2SRI - sres	242 (.06) {.00}	375 (.06) {.00}	675 (.07) {.00}	963 (.09) {.00}	1166 (.11) {.00}
0.95~0.98	2SRI - dres	86 (.23) {.45}	81 (.25) {.5}	-79 (2.45) {.67}	-126 (-1.85) {.26}	-133 (-1.58) {.29}
0.95~0.98	2SRI - ares	109 (.17) {.2}	**16 (.49) {.9}**	**45 (.37){.87}**	**72 (.34) {.84}**	**96 (.42) {.88}**
0.95~0.98	Bi.Probit	89 (.24) {.45}	92 (.19) {.35}	94 (.20) {.32}	98 (.23) {.43}	100 (.29) {.66}

Preliminary conclusions

In the presence of binary risk factors only:

- IV-LPM appears to the least biased estimator as long as the treatment mean is about 50%, irrespective of the rarity of the outcome.
- IV-LPM methods can produce substantially biased results as the rarity of treatment increases and are doubly affected with the rarity of both outcomes and treatment increases.
- 2SRI approach with raw scale residuals can provide less bias in this scenario up to where the treatment mean is around 15% and outcome mean is at least 20%.
- If either the treatment mean or the outcome mean fall below these levels, 2SRI with anscombe residuals appear to produce the least bias in treatment effects estimates.
- With rare outcomes (<10%) and rate treatment (<10%), coverage probabilities appear to deteriorate slightly, even for 2SRI with anscombe residuals.

Empirical Example

- Effect of long-term health insurance (LTHI) on outcomes
- Exposure to LTHI instrumented with Subsidy for LTHI

Empirical Example

Binary Variables	Mean (sd)
Outcomes	
Informal Care from Any Source	0.60 (0.49)
Informal Care from Child	0.43 (0.50)
Informal Care from other Relative	0.165 (0.37)
Home Health Care	0.068 (0.25)
Any Nursing Home Care	0.023 (0.15)
Treatment	
LTCI coverage	0.157 (0.364)
IV	
Subsidies	0.335 (0.472)

Binary Variables	Mean (sd)
Other covariates	
Marital status==2	0.11 (0.32)
Marital status ==3	0.17 (0.37)
Marital status==4	0.06 (0.24)
Female	0.56 (0.5)
No. of children==1	0.1 (0.3)
No. of children==2	0.31 (0.46)
No. of children==3	0.22 (0.42)
No. of children==4	0.13 (0.34)
No. of children==5	0.15 (0.36)
No. of children==6	0.01 (0.11)
Retired	0.47 (0.5)
Education category ==2	0.35 (0.48)
Education category ==3	0.26 (0.44)
Education category ==4	0.3 (0.46)
Income category==2	0.36 (0.48)
Income category==3	0.64 (0.48)
Race category ==2	0.06 (0.25)
Race category ==3	0.03 (0.18)
Fair/Poor health	0.17 (0.37)
[illegible]	[illegible]

Effects of long-term care insurance on different outcomes.

Outcomes→	Informal Care from Any Source	Informal Care from Child	Informal Care from other Relative	Home Health Care	Any Nursing Home Care
Estimators	Pr(Y) = 0.60	Pr(Y) = 0.43	Pr(Y) = 0.165	Pr(Y) = 0.07	Pr(Y) = 0.023
Naïve Probit	-0.037 (0.006)++	-0.032 (0.006)++	-0.015 (0.004)++	-0.005 (0.003)	0.001 (0.002)
IV-LPM	-0.302 (0.165)+	-0.329 (0.165)++	0.161 (0.114)	-0.252 (0.089)++	0.087 (0.055)
2SRI	**-0.319 (0.103)++**	**-0.238 (0.099)++**	-0.091 (0.062)	-0.142 (0.031)++	0.063 (0.097)
2SRI - sres	-0.118 (0.029)++	-0.074 (0.029)++	-0.06 (0.017)++	-0.028 (0.013)++	0.008 (0.012)
2SRI - dres	-0.392 (0.085)++	-0.28 (0.082)++	-0.126 (0.052)++	-0.127 (0.032)++	0.072 (0.102)
2SRI - ares	-0.297 (0.07)++	-0.198 (0.068)++	**-0.114 (0.038)++**	**-0.085 (0.026)++**	**0.038 (0.055)**
Bi.Probit	-0.283 (0.055)++	-0.179 (0.059)++	-0.147 (0.044)++	-0.117 (0.033)++	0.023 (0.028)

Pr(long-term care insurance) in these data = 0.157. 2SRI – sres: 2SRI with standardized residuals; 2SRI – dres: 2SRI with deviance residuals; 2SRI – ares: 2SRI with anscombe residuals;

$^{+}$ p-val$\leq$ 0.10; $^{++}$ p-val$\leq$0.05

Conclusions and future work

- Large discrepancies in empirical results depending on methods used.
- 2SRI with anscombe residuals more robust, especially with rare exposure & rare outcomes.
- Future simulations needed to explain potential discrepancies:
  - Polynomials of residuals
  - Interaction of residuals with d_i
- Important question to answer:
  - Can goodness of fit test guide choice of control functions?
- Future extensions to count and non-negative continuous outcomes

Colorado School of Public Health

Discussion of Basu and Coe, "2SLS vs. 2SRI"

Richard C. Lindrooth
Colorado School of Public Health
University of Colorado, Anschutz Medical Campus

The Question

- **With binary outcomes and binary endogenous variables AND categorical instruments/exogenous variables**

- **Is the 2SLS linear approximation better than 2SRI?**
- **Under what conditions?**

  - 2SRI First-stage residuals may violate the distributional exclusion restriction:
    - *Estimated residuals end up categorical and/or heteroskedastic*

  - 2SLS relies on less restrictive mean independence but also need "constant-effects assumption".

Should we "punt[1]" on the constant effects assumption and use 2SLS?

- **If mean independence is satisfied, then:**
  - If 0.75>prob(D)>0.25 2SLS looks pretty good
  - (though when there are no continuous regressors and the model is close to being saturated)
- **Suggestion 1:**
  - Include an specification with a continuous variable for contrast
  - Applied researchers (like me) need to be reminded when we should and should not punt[1]

Footnote 1: Angrist (2001) JBES

What happens when the treatment is rare but the outcome is not?

- **As treatment becomes rare (Pr(D)<0.25):**
  - the linear first-stage approximation becomes untenable;
  - 2SLS % bias surpasses 2SRI; and if Pr(D<0.15) then
  - 2SRI w/Anscombe residuals perform better than Raw residuals.
- **Suggestion #2:**
  - Examine nonlinear estimate of the conditional mean as a 2SLS instrument (Heckman, 1978; Kelijian, 1971; Angrist, 2001)
  - Retain the less restrictive mean independence
- **But:**
  - What if the first-stage specification of the CEF is not correct?

What if both the outcome (E(Y=0.8~.99)) & treatment (Pr(D<0.15) are rare?

- **As outcome becomes rare and treatment is also rare:**
  - 2SRI w/Anscombe residuals perform best
- **Suggestion #3:**
  - Dig into this more
  - Intuitive but seems rather *ad hoc*

Why is the answer important?

- **Practical implications:**
  - Survey data includes categorical variables
  - Study of LTC insurance and utilization of LTC a great example

- **Suggestion 4:**
  - Tie to underlying theory of LTC
  - Are assets the continuous confounder?
  - What is the interpretation of the LATE?
  - Does the policy-relevance/interpretation suffer in 2SLS vs. 2SRI?
  - What if you observe relevant continuous variables?

Positively Aware?
Expert Drug Reviews and the Downside of Information

Jorge Balat
Johns Hopkins University

Nicholas W. Papageorge
Johns Hopkins University

Shaiza Qayyum
Johns Hopkins University

Annual Health Econometrics Workshop 2015

What is the role of outside information when individuals make choices under uncertainty?

Recent literature has examined the impact of product reviews on product demand.

Product reviews can come from experts or other consumers.

Does outside information affect demand?

- Movie box-office sales (Moretti (2011); Liu (2006); Reinstein & Snyder (2005))
- Wine (Hilger et al. (2011); Dubois & Nauges (2010)).
- Restaurants, books, & hotels (Anderson & Magruder (2012); Chevalier & Mayzlin (2006); Mayzlin et al. (2014))

- In all the above examples, making the wrong choice is not particularly costly.
  - Do product reviews matter when the cost of wrong choice is substantial?
- How do reviews affect not only choices, but also outcomes.
  - Do product reviews affect economic outcomes?
- Product reviews can be subject to manipulation, & hence, misinform consumers (Mayzlin et al. (2014)).
  - What is the impact of information that might not be correct?

This paper.

$\rightharpoonup$ Assess the impact of HIV drug reviews on HIV drug demand.

Contribution.

- Study reviews in a context where the "wrong" choice is costly.
- Use rich consumer data to explore choices & outcomes.
- Exploit variation in reviews over product life-cycle.

Outline.

1. Merge datasets containing information about drug quality, consumption patterns & HIV doctor & activist reviews.
2. Show that expert reviews matter, even after controlling for drug quality.
3. Argue causality using a DiD strategy similar to Mayzlin et al.(2014).
4. Estimate a demand system to capture substitution patterns & bundling.
5. Use the estimated demand model to study the impact of reviews on treatment choices

What is HIV?

- HIV is a virus that attacks the immune system.
- Left untreated, HIV infection leads to immune system deterioration (known as AIDS).
- Life expectancy for a newly-infected individual is about 11 years.
- HIV/AIDS directly affects approximately 1.2 million Americans.
- Treatments developed in 1996 transformed HIV from a virtual death sentence into a manageable, but chronic disease.
- Drug companies spent 4.5 billion on consumer marketing in 2014.

Two data sources

1. Data from the Multi-Center AIDS Cohort Study (MACS).
2. Data on drug reviews from an HIV lifestyle magazine *Positively Aware*.

Multi-Center AIDS Cohort Study (MACS)

- Ongoing study of HIV-negative & HIV-positive (henceforth: *HIV−* & *HIV+*) homosexual & bisexual men.
- For this project, we focus on HIV+ men.
- Semi-annual visits generate information on:
  - Medical treatment choices.
  - Objective health (CD4 count) & mortality.
  - "Subjective" health (physical ailments, e.g., nausea).
  - Socio-demographic factors, such as education, race & employment.

Positively Aware Annual HIV Drug Guides

- Published annually since 1997 by Test Positive Aware Network (TPAN).
- Contributing writers include HIV physicians & activists.
- Features a page-by-page guide to all FDA approved antiretroviral drugs on the market.
- Circulation >100,000
- 75,000 copies distributed to more than 1,900 community-based organizations & 700 Walgreens pharmacies across the US.

Data Sources & Data Construction
Positively Aware Drug Guides

"Am I ready to take the medications, as prescribed, every day for the rest of my life?" When people ask me how I've lived so long, what I have to say is, "Good doctors who listen, taking my meds when and like I am supposed to, and bucking up when I don't like the side effects, *and*, yes, LUCK and a good attitude doesn't hurt either.

—*Greg*
Via the Internet

THANKS FOR THE DRUG GUIDE

I just wanted to write to thank you for putting together your annual Drug Guide.

As someone who was diagnosed in 2009, it has been an anchor in the storm for me.

The first issue of Positively Aware that I read was the January/February issue which, I have to say, blew my mind. I live in a small town and your article about access to care in rural areas really hit home. I travel almost 200 miles to see a doctor, but thanks to that article, I researched my area until I found the one HIV specialist within driving distance.

And then I got the Drug Guide. My doctor had put me on Atripla right away and though I've had no problem with it and my viral load has come down, reading about the other drugs, the side effects, and the drugs being developed really opened my eyes. Before, I guess I was just in shock and willing to do anything to stay healthy. After reading the Drug Guide, I'll do my "homework" before just agreeing to any particular treatment. It has also helped me in talking to my doctor, who told me he always has a copy on hand and the chart picturing the drugs hangs in his office.

You guys do a great job and I will look forward to reading every issue of Positively Aware.

—*Marlen P.*
Lacey, WA

A CAPTIVE AUDIENCE

Greetings from a California state prison. I am truly blessed to receive a subscription to your invaluable magazine.

Not only have I been HIV-positive since 2001, a prisoner, but I am also a male-to-female transgender person. Your magazine is surely a God-send.

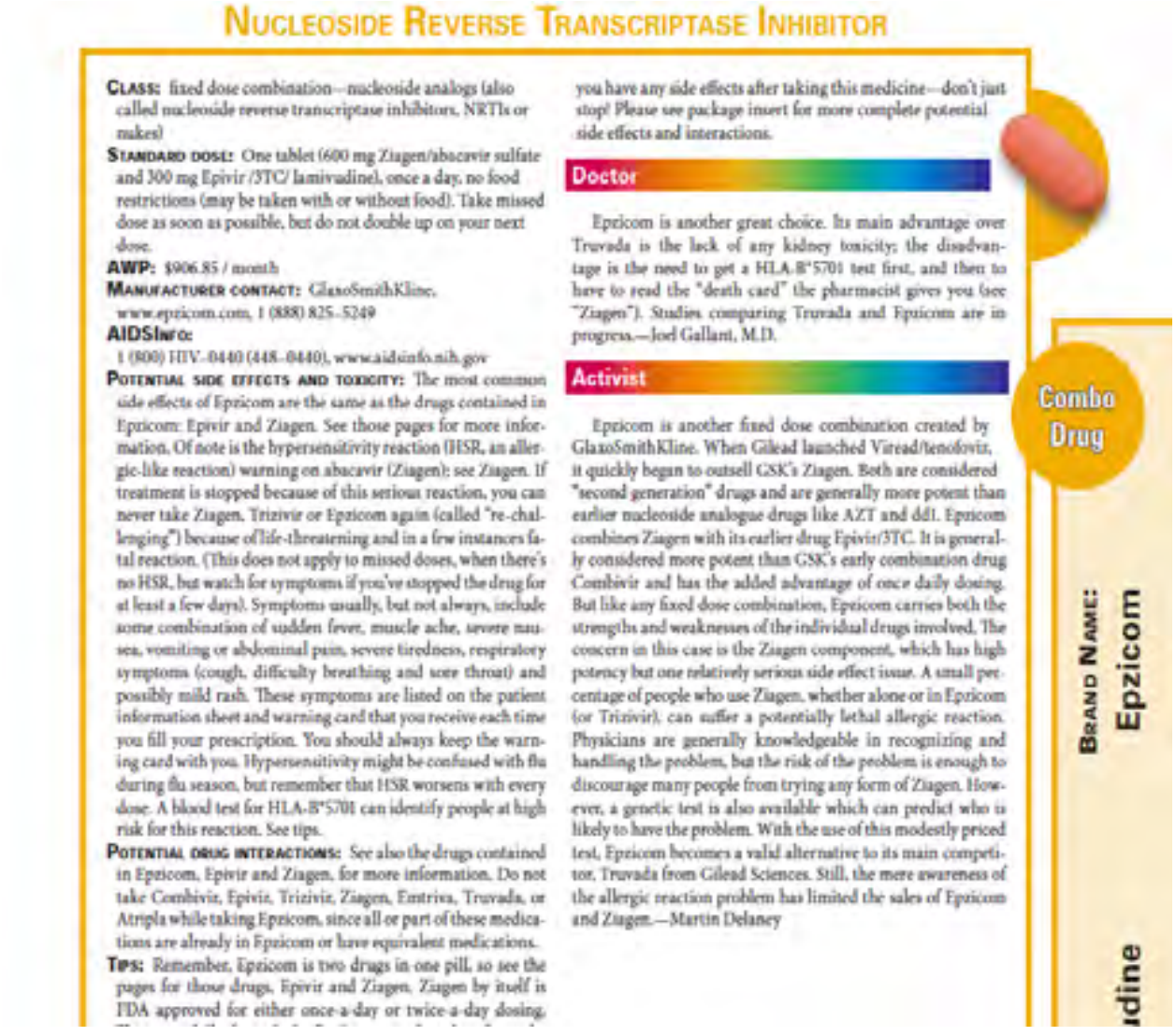
I was wondering if you could do an article on hormone treatment and HIV and also HIV meds. I know there are many who could use this information.

Thank you for all your insights, hints, and articles—I truly appreciate this much-needed resource.

We don't have a death sentence; we just share our bodies with a virus. I always say, "I don't live with HIV—it has to live with me!"

—*Nathan*

A screenshot from the 2008 Positively Aware HIV Drug Guide



Nucleoside Reverse Transcriptase Inhibitor

Class: fixed dose combination—nucleoside analogs (also called nucleoside reverse transcriptase inhibitors, NRTIs or nukes)

Standard dose: One tablet (600 mg Ziagen/abacavir sulfate and 300 mg Epivir /3TC/ lamivudine), once a day, no food restrictions (may be taken with or without food). Take missed dose as soon as possible, but do not double up on your next dose.

AWP: $906.85 / month

Manufacturer contact: GlaxoSmithKline, www.epzicom.com, 1 (888) 825–5249

AIDSInfo:
1 (800) HIV–0440 (448–0440), www.aidsinfo.nih.gov

Potential side effects and toxicity: The most common side effects of Epzicom are the same as the drugs contained in Epzicom: Epivir and Ziagen. See those pages for more information. Of note is the hypersensitivity reaction (HSR, an allergic-like reaction) warning on abacavir (Ziagen); see Ziagen. If treatment is stopped because of this serious reaction, you can never take Ziagen, Trizivir or Epzicom again (called "re-challenging") because of life-threatening and in a few instances fatal reaction. (This does not apply to missed doses, when there's no HSR, but watch for symptoms if you've stopped the drug for at least a few days). Symptoms usually, but not always, include some combination of sudden fever, muscle ache, severe nausea, vomiting or abdominal pain, severe tiredness, respiratory symptoms (cough, difficulty breathing and sore throat) and possibly mild rash. These symptoms are listed on the patient information sheet and warning card that you receive each time you fill your prescription. You should always keep the warning card with you. Hypersensitivity might be confused with flu during flu season, but remember that HSR worsens with every dose. A blood test for HLA-B*5701 can identify people at high risk for this reaction. See tips.

Potential drug interactions: See also the drugs contained in Epzicom, Epivir and Ziagen, for more information. Do not take Combivir, Epiviz, Trizivir, Ziagen, Emtriva, Truvada, or Atripla while taking Epzicom, since all or part of these medications are already in Epzicom or have equivalent medications.

Tips: Remember, Epzicom is two drugs in one pill, so see the pages for those drugs, Epivir and Ziagen. Ziagen by itself is FDA approved for either once-a-day or twice-a-day dosing.

you have any side effects after taking this medicine—don't just stop! Please see package insert for more complete potential side effects and interactions.

Doctor

Epzicom is another great choice. Its main advantage over Truvada is the lack of any kidney toxicity; the disadvantage is the need to get a HLA-B*5701 test first, and then to have to read the "death card" the pharmacist gives you (see "Ziagen"). Studies comparing Truvada and Epzicom are in progress.—Joel Gallant, M.D.

Activist

Epzicom is another fixed dose combination created by GlaxoSmithKline. When Gilead launched Viread/tenofovir, it quickly began to outsell GSK's Ziagen. Both are considered "second generation" drugs and are generally more potent than earlier nucleoside analogue drugs like AZT and ddI. Epzicom combines Ziagen with its earlier drug Epivir/3TC. It is generally considered more potent than GSK's early combination drug Combivir and has the added advantage of *once daily* dosing. But like any fixed dose combination, Epzicom carries both the strengths and weaknesses of the individual drugs involved. The concern in this case is the Ziagen component, which has high potency but one relatively serious side effect issue. A small percentage of people who use Ziagen, whether alone or in Epzicom (or Trizivir), can suffer a potentially lethal allergic reaction. Physicians are generally knowledgeable in recognizing and handling the problem, but the risk of the problem is enough to discourage many people from trying any form of Ziagen. However, a genetic test is also available which can predict who is likely to have the problem. With the use of this modestly priced test, Epzicom becomes a valid alternative to its main competitor, Truvada from Gilead Sciences. Still, the mere awareness of the allergic reaction problem has limited the sales of Epzicom and Ziagen.—Martin Delaney

Combo Drug

Brand Name: Epzicom

udine

Drug Level Information

- Manufacturer Name.
- Dosage Frequency.
- Food restrictions.
- Average Wholesale Price.
- Number of Side Effects.
- Number of Drug Interactions.
- Doctor's reviews about drug.
- HIV Activist's reviews about drug.

Construction of Doctor & Activist Index

- Rating $= 1$ if mostly negative words or phrases have been used to describe the drug

  "*There is* **not much to say** *about ddC anymore.*"

  "**hard to get excited** *about it, . . . often not prescribed.*"

  "*Invirase was* **extraordinarily weak**."

Construction of Doctor & Activist Index

- Rating = 2 if the doctor or activist points out the positve as well as the negative aspects of the drug, but does not give an absolute recommendation of whether the drug is good or bad

  "The new soft-gel formulation **achieves much better drug levels** *. . . but if you are going to use Fortovase as a sole PI,* **you will have to take a lot of pills***."*

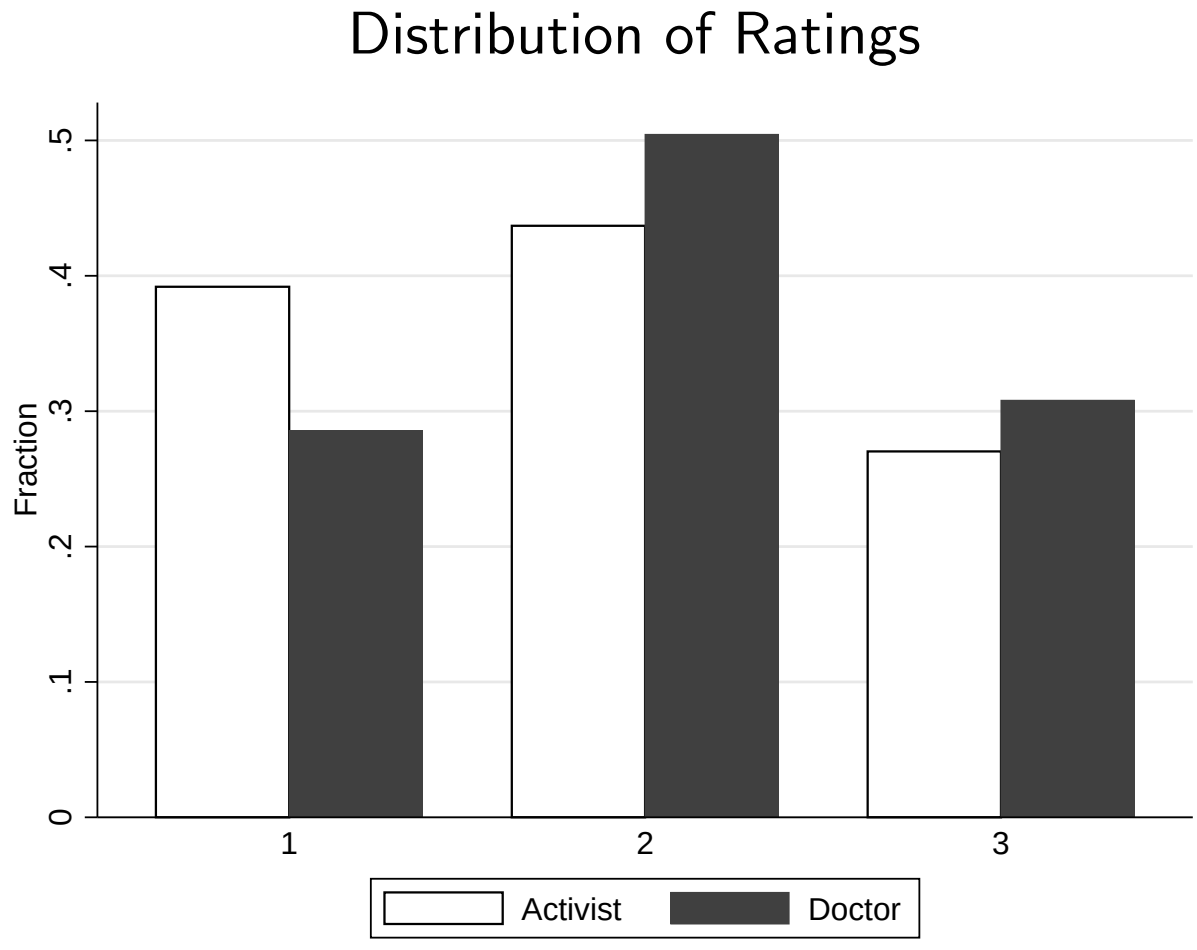
  "It **may not be the best bet** *to include in first-line treatment . . . but it remains a* **solid antiviral***."*

Construction of Doctor & Activist Index

- Rating = 3 to drugs with reviews that mostly use positive words to describe the drug e.g.

  "3TC is a **potent, convenient & well-tolerated** *drug."*

  "with its **minimal side effects, easy dosing schedule & high potency**, *3TC may be the* **most useful of the nucleosides**"


Distribution of Ratings
Fraction
0
.1
.2
.3
.4
.5
1
2
3
Activist
Doctor

Drug-level analysis.

- → Directly relates drug reviews & drug market shares.
- → Exploits drug age to develop & D-i-D identification strategy.
- → Downside: HIV drugs are taken in combinations.
- → Ignores the structure of demand for HIV drugs.

Consumer/combination level analysis.

- Examines consumers choosing combinations.
- In other words: the impact of reviews on bundle choices.
- Allows for spillover effects.
- Also allows us to examine consumer outcomes.
- (Drug-level analysis wastes rich consumer data).

- There are 3 classes of HIV drugs:
  1. Nucleoside Reverse Transcriptase Inhibitors (NRTIs)
  2. Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs)
  3. Protease Inhibitors (PIs)
- Analysis is at the drug/year level (197 drug-year dyads).
  - 27 drugs.
  - 12 years (1997-2008).
- We construct class-specific market shares from the MACS dataset.
- For each drug, we also use the MACS data to construct
  - Probability of Non-decreasing CD4 count.
  - Probability of Not Suffering Physical Ailments.

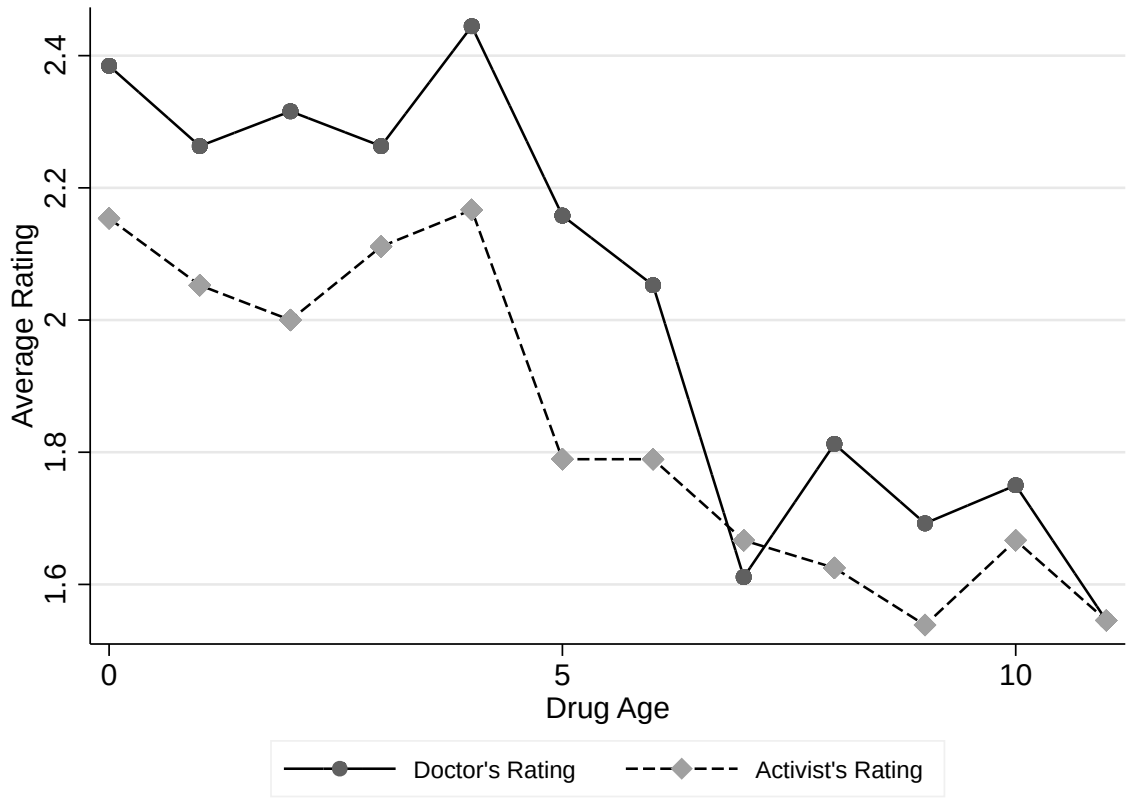
Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Figure: Reviews over Drug Life-Cycle



Why do reviews decline over a drug's life-cycle?

- Quality Change
- Learning
- Deflation due to technological advancements

Figure: Probability of Non-decreasing CD4 count

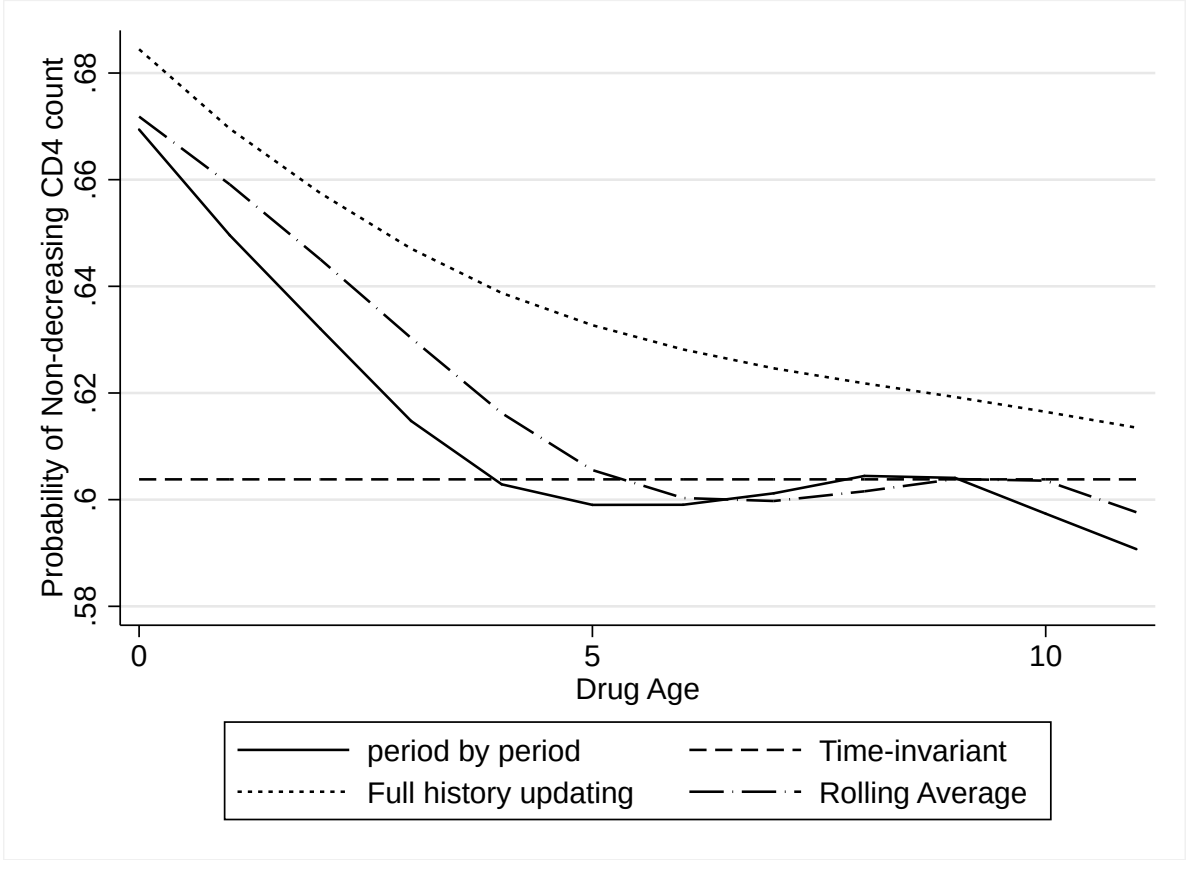
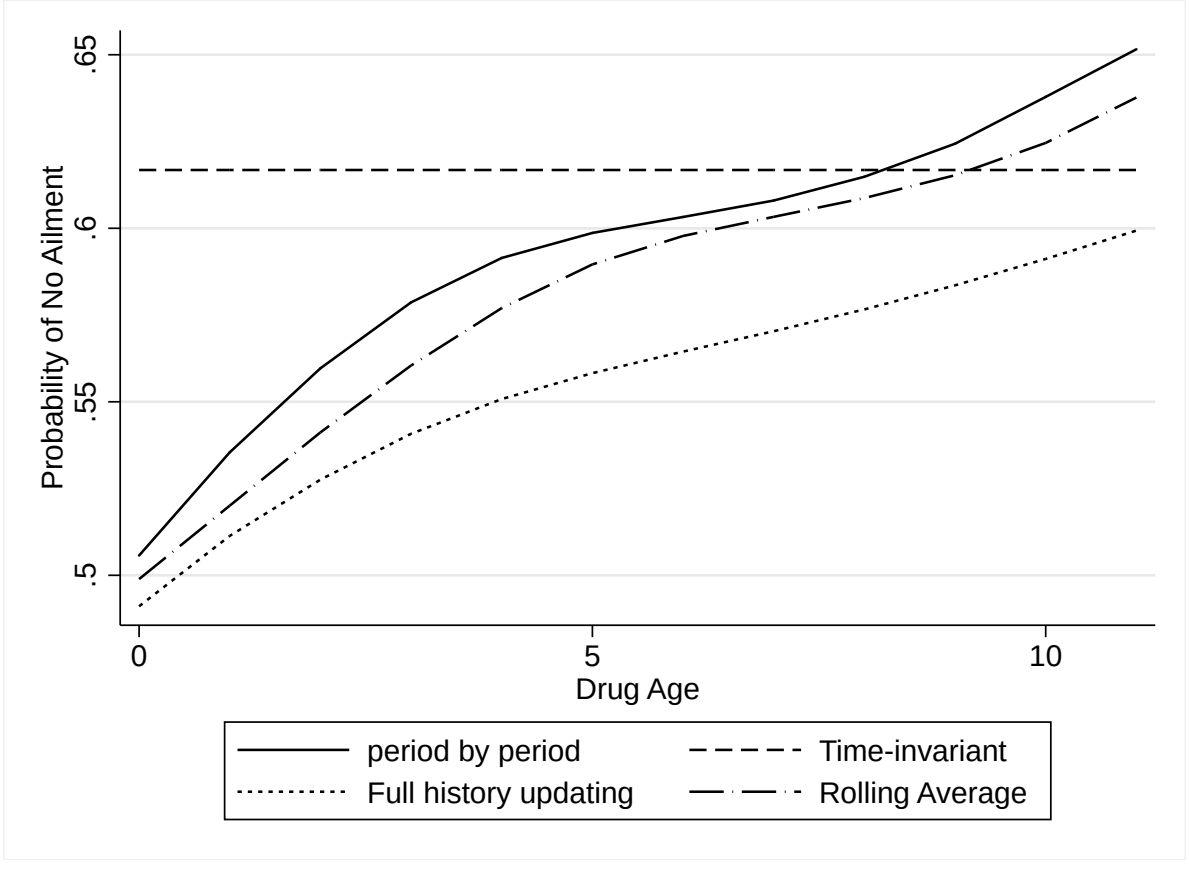


Figure: Probability of No Ailment



Relating Reviews with Quality Over Time

	a_3	d_3	And	a_3	d_3	And
Rolling Average No Ailment	2.03	0.44	-0.85			
	(2.02)	(1.83)	(2.12)			
Rolling Average CD4	5.66**	2.32	4.49*			
	(2.50)	(2.03)	(2.42)			
Old (Drug Age $>$ 4 years)	-5.61	-7.03*	-13.38**	-4.60	-5.97*	-12.45**
	(4.68)	(3.91)	(6.39)	(4.08)	(3.58)	(5.96)
Rolling Avg CD4 $\times$ Old	-3.90	-1.00	-4.89			
	(6.95)	(5.92)	(9.74)			
Rolling Avg No Ailment $\times$ Old	11.34**	10.87***	24.38***			
	(4.89)	(4.13)	(8.00)			
Relative Rolling Avg No Ailment				1.77	-1.23	-0.45
				(1.13)	(1.06)	(1.21)
Relative Rolling Avg CD4				2.72*	2.55**	2.54*
				(1.47)	(1.18)	(1.40)
Relative Rolling Avg CD4 $\times$ Old				0.03	-1.92	-1.08
				(3.09)	(3.10)	(4.66)
Relative Rolling Avg No Ailment $\times$ Old				3.69	6.97***	12.23***
				(2.48)	(2.22)	(4.23)
N	195	195	195	195	195	195

* p $<$ 0.10,** p $<$ 0.05, ***p $<$ 0.01; standard errors in parentheses

Correlations between Activist & Doctor Rating over Drug's

Life-cycle

Age 0-2	Age 3-5	Age 6-8	Age 9-11
Full Sample of Drugs			
0.416*	0.418*	0.6274*	0.864*
Subsample: Drugs with Age $<$ 9 years			
0.490*	0.478*	0.731*	-

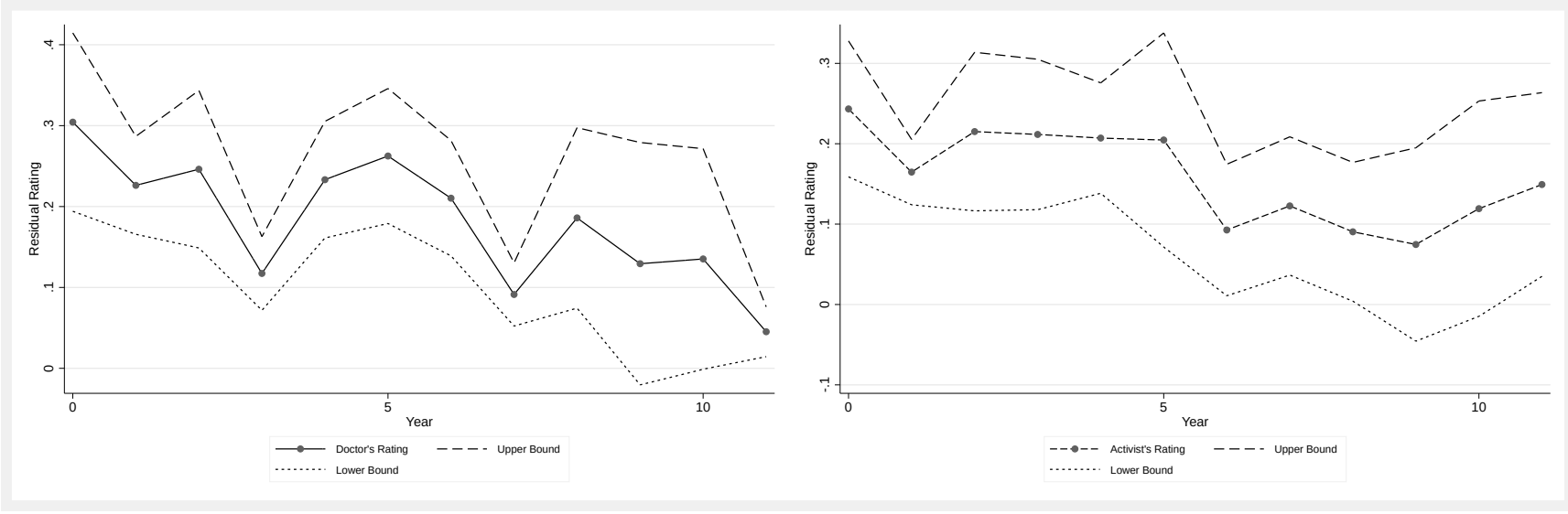
The table reports pairwise correlations; * $p < 0.05$

Relating Reviews & Rival Drugs' Quality

	[a3]	[d3]	[And]
Rolling Average CD4	6.37***	5.63***	5.80**
	(2.28)	(1.99)	(2.37)
Rolling Average No Ailment	2.00	0.06	0.95
	(1.82)	(1.64)	(1.92)
Mean Rolling Average of CD4 of Rival Drugs	-19.28	-3.66	-4.88
	(14.45)	(13.71)	(15.28)
Mean Rolling Average of No Ailment of Rival Drugs	-12.51*	4.87	-3.08
	(7.49)	(6.99)	(8.15)
N	195	195	195

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses

Do these Explanations Explain Declining Reviews



Note: The figure plots the residual after the probit regression of reviews on own and rival quality measures, and doctor and activist fixed effects.

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Market Share Regressions with reviews

	[1]	[2]	[3]	[4]	[5]	[NW]
Doctor's Rating = 3	0.133***	0.129***	0.085***	0.088***	0.080***	0.080***
	(0.032)	(0.033)	(0.024)	(0.025)	(0.026)	(0.019)
Activist's Rating=3	0.132***	0.119***	0.088***	0.074***	0.060**	0.060**
	(0.032)	(0.033)	(0.024)	(0.025)	(0.025)	(0.025)
No Ailment		0.115		0.253**	-0.038	-0.038
		(0.147)		(0.115)	(0.147)	(0.099)
CD4 count		0.244		0.073	0.093	0.093
		(0.199)		(0.157)	(0.162)	(0.147)
# of 3-star Rivals (Activist)			-0.030	-0.029	-0.080	-0.080*
			(0.052)	(0.052)	(0.053)	(0.042)
# of 2-star Rivals (Activist)			-0.008	-0.007	-0.072	-0.072*
			(0.054)	(0.054)	(0.056)	(0.043)
# of 1-star Rivals (Activist)			0.004	0.002	-0.056	-0.056
			(0.049)	(0.049)	(0.050)	(0.038)
# of 3-star Rivals (Doctor)			-0.035	-0.034	0.022	0.022
			(0.052)	(0.052)	(0.053)	(0.039)
# of 2-star Rivals (Doctor)			-0.030	-0.030	0.036	0.036
			(0.051)	(0.051)	(0.054)	(0.041)
# of 1-star Rivals (Doctor)			-0.019	-0.021	0.034	0.034
			(0.051)	(0.051)	(0.052)	(0.038)
Drug Quality Controls	N	N	N	N	Y	Y
N	196	196	196	196	196	196
R^2	0.310	0.319	0.637	0.648	0.696	
Adjusted R^2	0.295	0.297	0.618	0.625	0.659	
Prob > F	-	0.2608	-	0.0414	0.812	0.7883

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses.

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

Do drug reviews have a causal effect?

- The idea: information for old drugs is less costly to obtain.
- We assess the impact of reviews for new drugs.
- Control: Reviews for the same drug when it is older.
- Similar to a difference-in-differences.

Do drug reviews have a causal effect?

- → Reviews & quality evolve over time.
- → We control for rival quality & objective drug quality.
- → Recall: reviews seem more aligned with objective quality as drugs age.
- → If reviews have greater impact when a drug is new, it means that reviews matter precisely when:
  1. They are less aligned with objective quality.
  2. They deviate from later-life-cycle reviews (after we have controlled for rival qualities).

Differences-in-differences empirical specification:

$$
\begin{aligned}
s_{jt} &= \beta_0^d + \beta_1^d X_{jt}^d + \sum_{k=1}^{5} \beta_{2k}^d t_k \\
&+ \beta_3^d R_{jt}^D + \sum_{k=1}^{5} \beta_{4k}^d (t_k \times R_{jt}^D) \\
&+ \beta_3^d R_{jt}^A + \sum_{k=1}^{5} \beta_{4k}^d (t_k \times R_{jt}^A) + u_{jt}^d
\end{aligned}
$$

where

- s_{jt} are class-specific market shares for drug j at time t.
- R_{jt}^D indicates high doctor reviews for drug j at time t.
- R_{jt}^A indicates high activist reviews for drug j at time t.
- t_k are dummies for years 1, 2, 3, 4 & above 4.
- X^d includes controls for drug quality from both data sources.

Difference-in-Differences

	[Doctor]	[Activist]	[Doc and Act]
Age 1 × d_3	0.170**		0.007
	(0.078)		(0.085)
Age 2 × d_3	0.040		-0.001
	(0.065)		(0.066)
Age 3 × d_3	0.048		0.068
	(0.063)		(0.062)
Age 4 × d_3	0.006		0.017
	(0.066)		(0.065)
Above Age 4 × d_3	-0.001		-0.024
	(0.052)		(0.052)
Age 1 × a_3		0.312***	0.302***
		(0.074)	(0.084)
Age 2 × a_3		0.143**	0.135**
		(0.063)	(0.067)
Age 3 × a_3		-0.042	-0.059
		(0.065)	(0.067)
Age 4 × a_3		0.013	0.041
		(0.061)	(0.063)
Above Age 4 × a_3		0.055	0.059
		(0.049)	(0.052)
Drug Quality Controls	Y	Y	Y
State of the Market Controls	Y	Y	Y
N	196	196	196
R^2	0.732	0.756	0.774
Adjusted R^2	0.685	0.714	0.725

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses.

Using drug-level analysis, we address the following questions:

1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

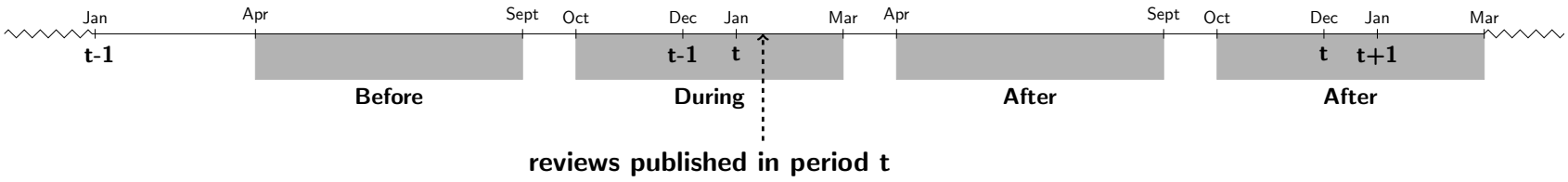
Using drug-level analysis, we address the following questions:

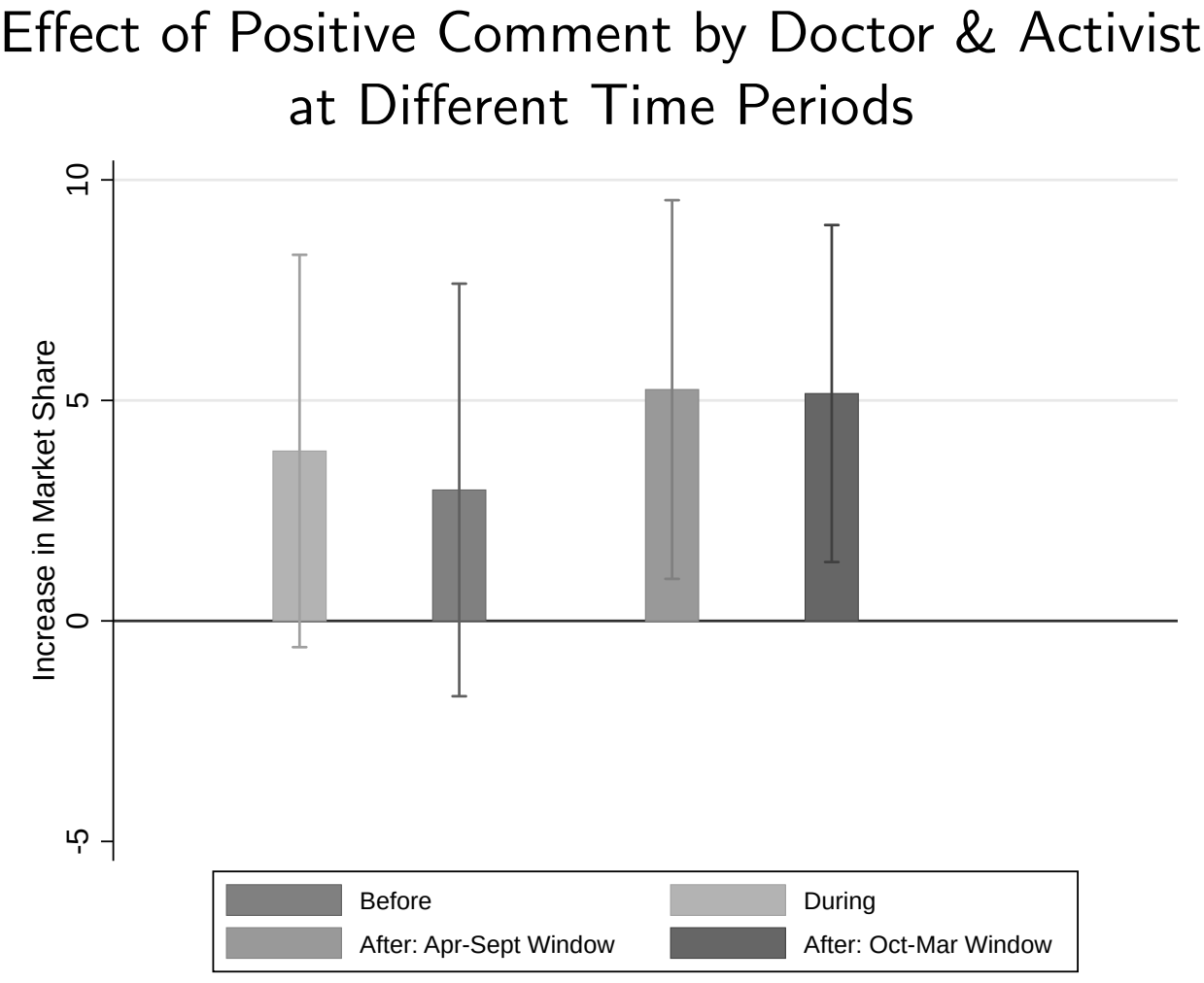
1. How do reviews evolve over the life-cycle?
2. Do reviews predict market share?
3. Do reviews affect market share?
4. Can we rule out alternative stories, e.g., social learning?

D-i-D estimates are consistent with

- → Social Learning (Banerjee (1992); Bikhchandani et al. (1992); Moretti (2011)).
- → Direct impact of magazine reviews.

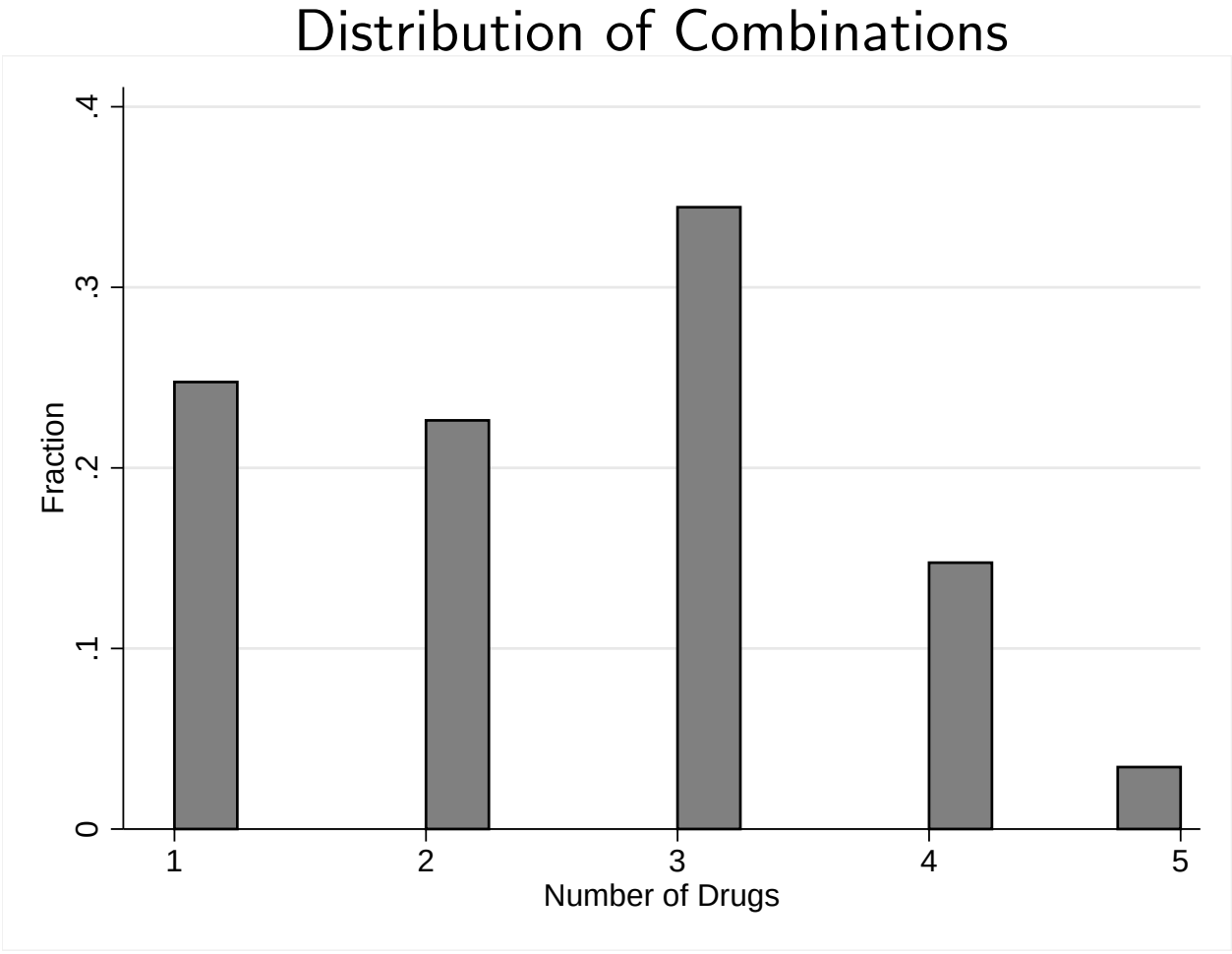
Timeline of Events




Effect of Positive Comment by Doctor & Activist at Different Time Periods
Increase in Market Share
10
5
0
-5
Before
During
After: Apr-Sept Window
After: Oct-Mar Window

Consumption in bundles:

- → Our estimates until now suggest a causal impact of reviews.
- → However, our estimates are difficult to interpret.
- → HIV drugs are consumed in bundles.
- → What is the impact of reviews on combination choice?
- → If we study combination choice, we can incorporate individual characteristics.


Distribution of Combinations
Fraction
0
.1
.2
.3
.4
1
2
3
4
5
Number of Drugs

Data:

- A total of 1,248 drug combinations observed in the data.
- Define a fringe category, in which we aggregate all combos taken by < 25 individuals.
- Drop observations with missing value for demographics, CD4 counts & ailment reports.
- Unbalanced panel of 13,472 observations (1,268 individuals followed biannually from 1997 till death).
- 19.6% of sample taking no HIV medication.
- Evolving choice set. Minimum alternatives $= 21$; Maximum alternatives $= 55$.

Descriptive Statistics: Individuals.

	Mean	S.Dev	Min	Max
Age	47.15	8.21	19.5	80
Lagged CD4 count	533.01	285.5	0	3819
AIDS	.20	.40	0	1
Work Full-time	.54	.50	0	1
White	.54	.50	0	1
High School	.19	.39	0	1
College	.50	.50	0	1
Real Income	3.65	2.09	.60	6.59
N	13,472			

Note: Real Income has been divided by 10,000.

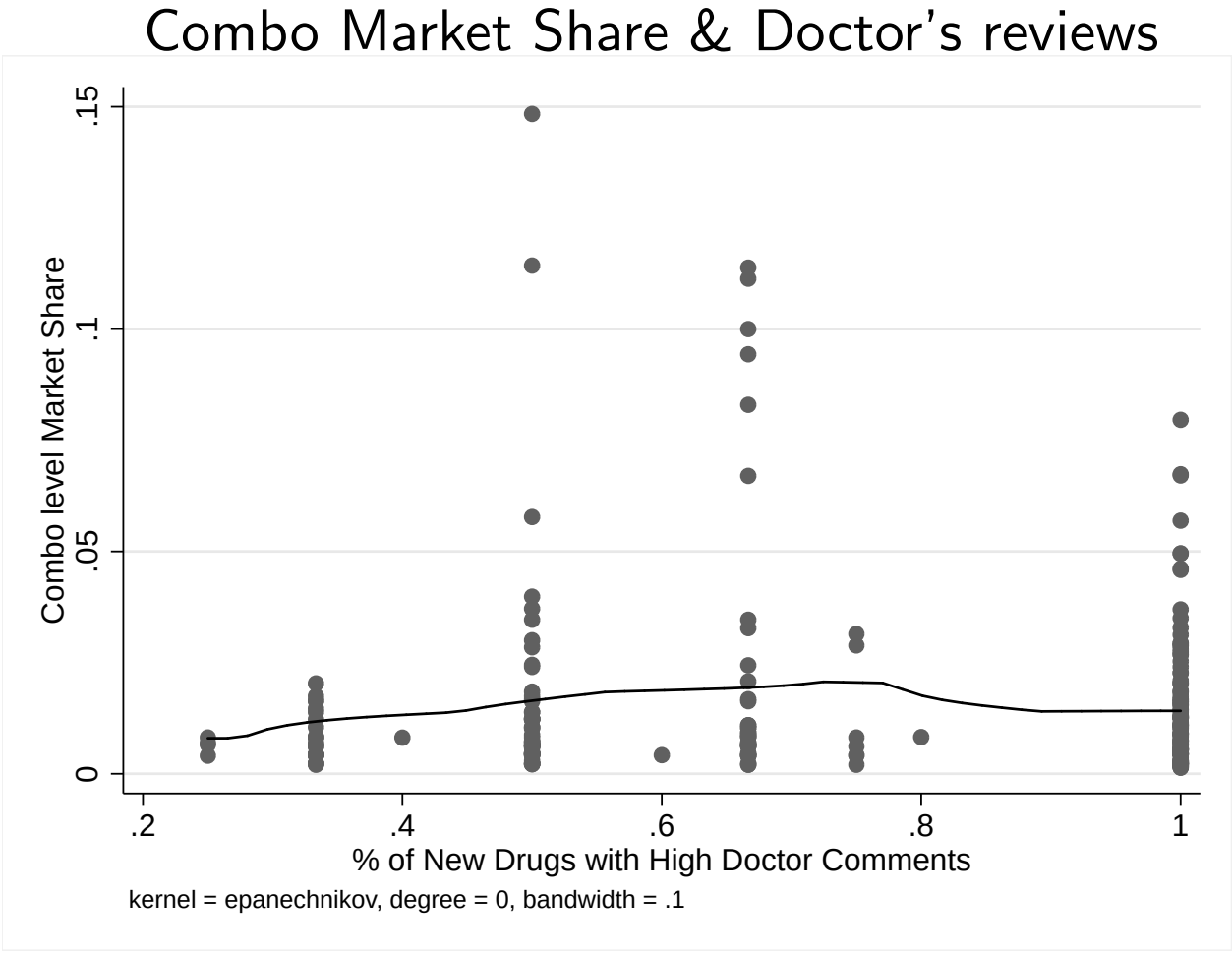
- Define a drug to be new if it has been on the market for less than three years. Then, for each combination, we define:
  - % of new drugs in the combination
  - % of new drugs that have a rating of 3 by doctor.
  - % of new drugs that have a rating of 3 by activist.
  - % of old drugs that have a rating of 3 by doctor.
  - % of old drugs that have a rating of 3 by activist.

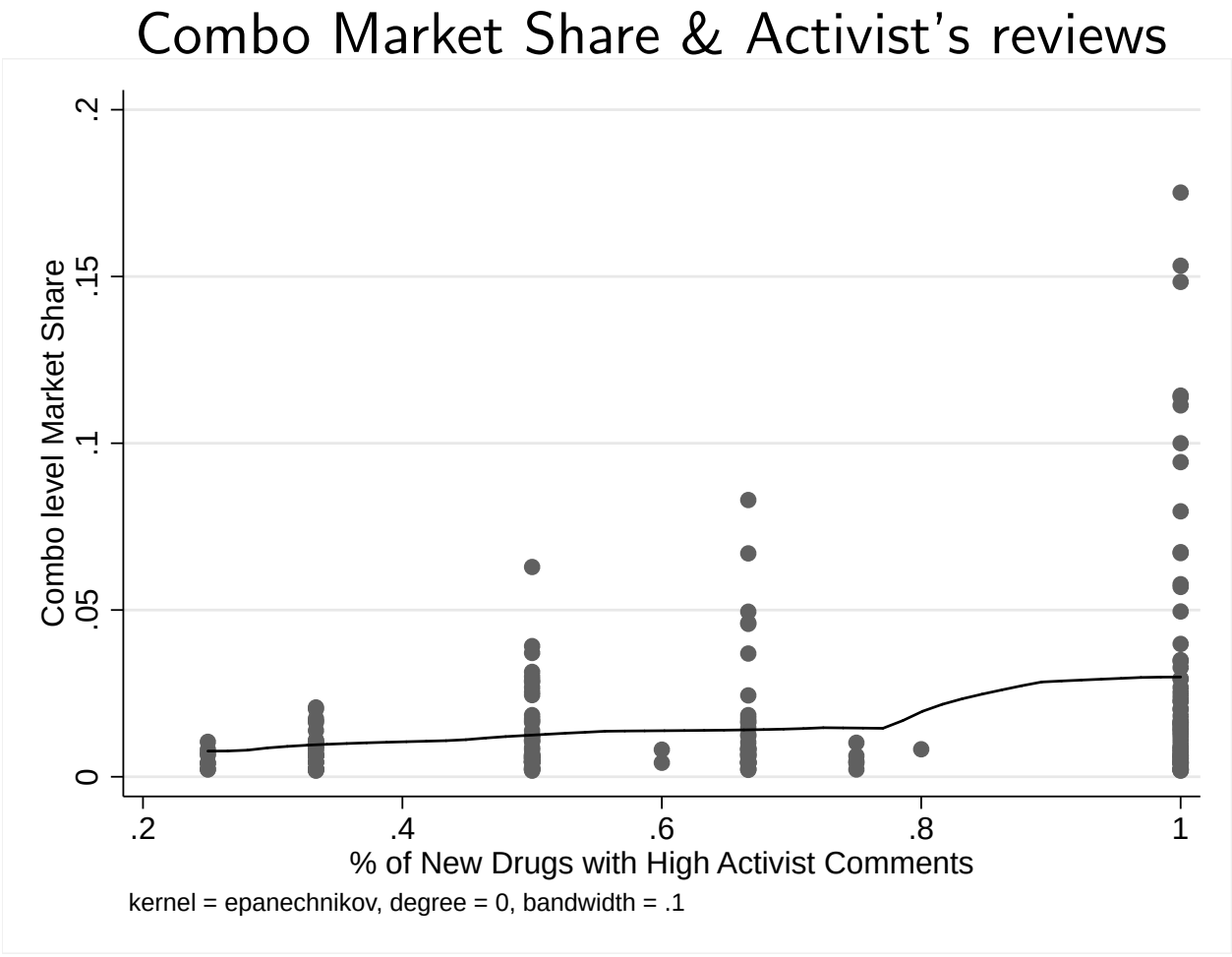
Combo Characteristics - Excluding Fringe & No Drug.

	Mean	S.Dev	Min	Max
Probability of No Ailment	.60	.20	0	1
Probability of Non-dec CD4 count	.55	.15	0	1
% of New Drugs in Combo	.22	.38	0	1
% Drugs with High Doctor reviews	.38	.34	0	1
% Drugs with High Activist reviews	.33	.30	0	1
% of New Drugs with High Doctor reviews	.18	.34	0	1
% of New Drugs with High Activist reviews	.12	.28	0	1
% of Old Drugs with High Doctor reviews	.28	.34	0	1
% of Old Drugs with High Activist reviews	.24	.30	0	1

Fringe Characteristics

	Mean	S.Dev	Min	Max
Probability of No Ailment	.57	.02	.52	.60
Probability of Non-dec CD4 count	.57	.02	.55	.61
% of New Drugs in Combo	.39	.38	0	1
% Drugs with High Doctor reviews	.39	.32	0	1
% Drugs with High Activist reviews	.36	.27	0	1
% of New Drugs with High Doctor reviews	.36	.43	0	1
% of New Drugs with High Activist reviews	.23	.37	0	1
% of Old Drugs with High Doctor reviews	.23	.31	0	1
% of Old Drugs with High Activist reviews	.31	.36	0	1


Combo Market Share & Doctor's reviews
Combo level Market Share
0
.05
.1
.15
.2
.4
.6
.8
1
% of New Drugs with High Doctor Comments
kernel = epanechnikov, degree = 0, bandwidth = .1


Combo Market Share & Activist's reviews
Combo level Market Share
0
.05
.1
.15
.2
.2
.4
.6
.8
1
% of New Drugs with High Activist Comments
kernel = epanechnikov, degree = 0, bandwidth = .1

The utility for individual i of choosing alternative j is defined as:

$$U_{ij} = V_{ij} + \epsilon_{ij}, \tag{1}$$

$$V_{ij} = X'_{ij}\alpha + W'_i\beta. \tag{2}$$

where

- ϵ_{ij} is a random variable unobserved by the econometrician,
- X_{ij} includes probability of no ailment, probability of non-decreasing CD4 count, % of new drugs in combinations, & measures for doctor & activist reviews,
- W_i includes demographics such as age, race, education, income & AIDS status.

The probability that individual i chooses the j^{th} alternative is given by:

$$p_{ij} = \frac{e^{X'_{ij}\alpha + W'_i\beta}}{\sum_{l=1}^{m} e^{X'_{il}\alpha + W'_i\beta}}, \quad j = 1, \ldots, m. \tag{3}$$

where m is the total number of alternatives.

	[1]	[2]	[3]	[4]	[5]
Probability of No Ailment	0.48***		0.52***	0.43***	0.42***
	(0.09)		(0.10)	(0.10)	(0.10)
Probability of Non-dec CD4	0.47***		0.48***	0.53***	0.60***
	(0.14)		(0.14)	(0.15)	(0.14)
Age	-0.06***	-0.06***	-0.06***	-0.06***	-0.06***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AIDS	0.42***	0.11*	0.35***	0.28**	0.28**
	(0.11)	(0.06)	(0.11)	(0.12)	(0.12)
Working Full-time	-0.61***	-0.63***	-0.68***	-0.72***	-0.72***
	(0.09)	(0.05)	(0.09)	(0.09)	(0.09)
High School	-0.71***	-0.78***	-0.75***	-0.76***	-0.75***
	(0.11)	(0.06)	(0.11)	(0.11)	(0.11)
College	0.44***	0.31***	0.42***	0.40***	0.40***
	(0.10)	(0.06)	(0.11)	(0.11)	(0.11)
White	1.00***	0.90***	0.92***	0.81***	0.82***
	(0.09)	(0.05)	(0.09)	(0.10)	(0.10)
% of drugs with High Doctor reviews		-0.03	0.30***		0.36***
		(0.04)	(0.07)		(0.08)
% of drugs with High Activist reviews		0.88***	0.68***		0.41***
		(0.04)	(0.07)		(0.08)
% of New Drugs				-1.04***	-1.31***
				(0.15)	(0.16)
% of Old Drugs with High Doctor reviews				0.27***	
				(0.08)	
% of New Drugs with High Doctor reviews				-0.19	-0.23
				(0.13)	(0.15)
% of Old Drugs with High Activist reviews				0.72***	
				(0.08)	
% of New Drugs with High Activist reviews				2.56***	2.33***
				(0.15)	(0.17)

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses.

- We start with a sort of case study.
- Combivir was introduced in 1998.
- It was likely under-rated during its first years.
- It was expected to be mediocre, but performed very strongly.
- In contrast to most drugs, reviews rose as it aged.

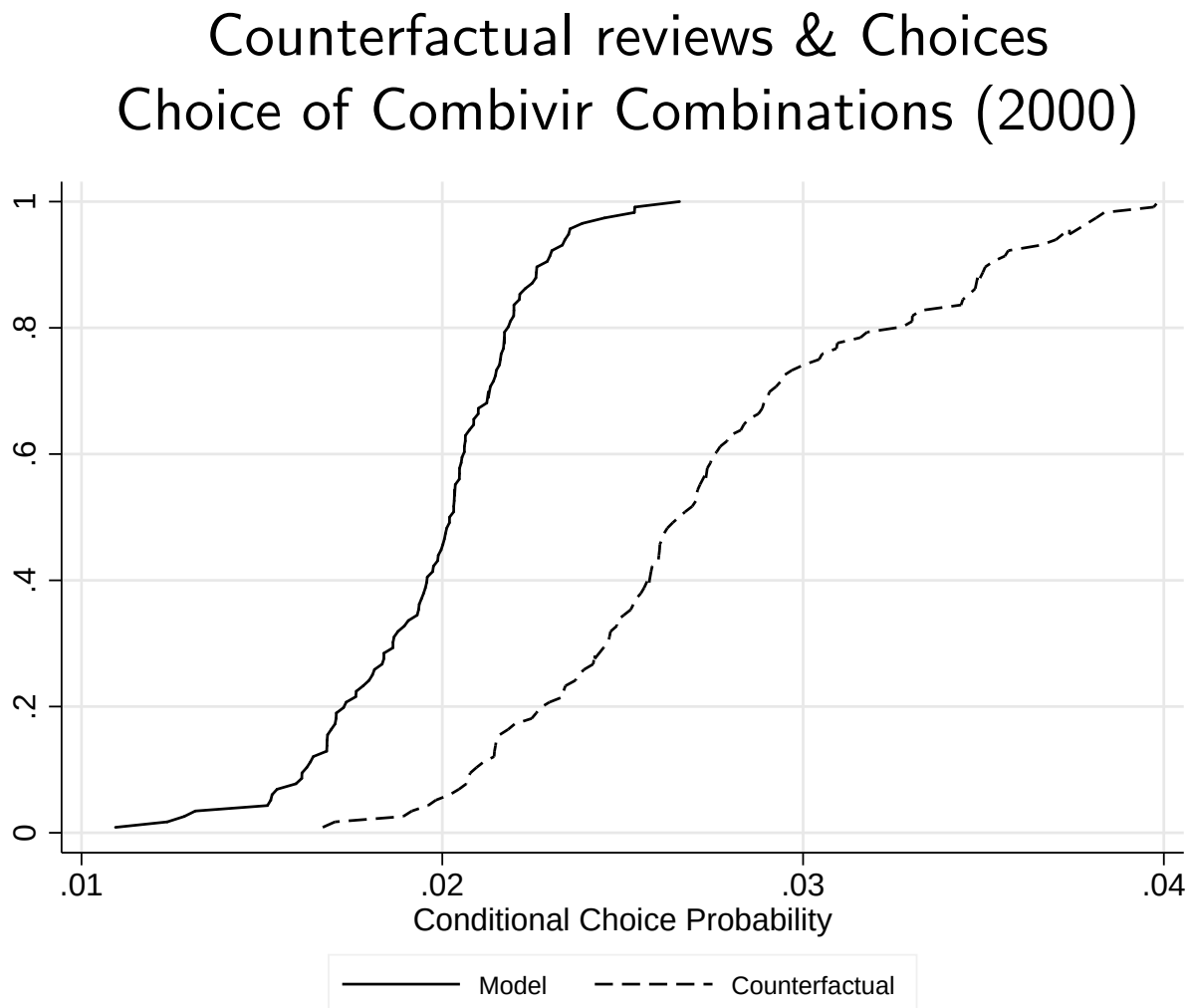
- We ask: what happens if ratings are 3 rather than 2?
- We focus on the year 2000.
- A caveat: this is (at best) "partial equilibrium" analysis.
- The reason: we have no model for how reviews are generated.
- Therefore, we do not know, for example, how other reviews would change in response to a counterfactual Combivir review.

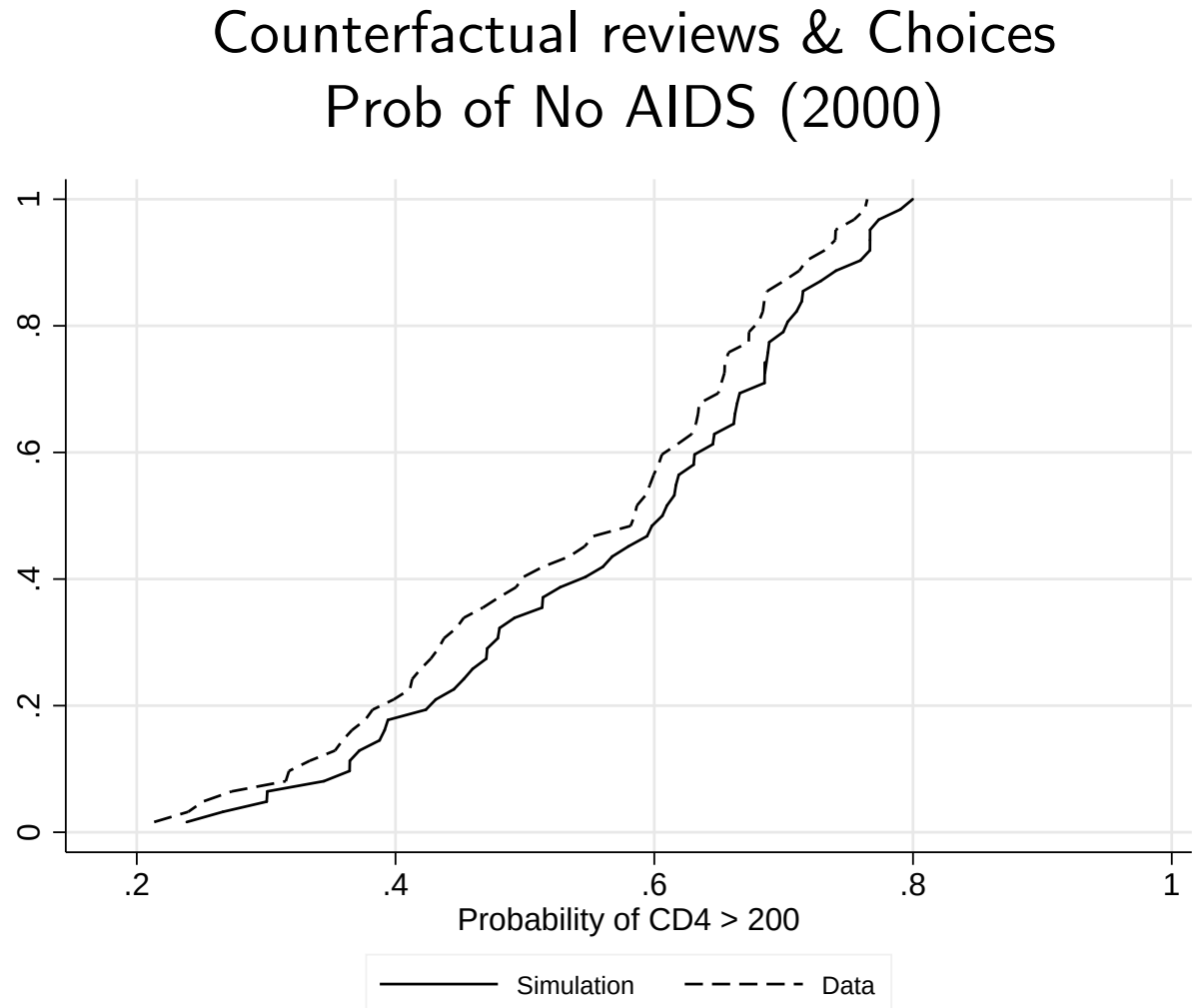
We consider:

1. The impact of counterfactual reviews on choices.
2. The impact of counterfactual reviews on outcomes.
   - → Focus: one-period-ahead health.
3. Distributional effects.
   - → Focus: college educated versus not college educated.

Procedure:

1. Estimate choice-specific health outcome probabilities from the sample.
2. Compute combination choice probabilities from the estimated choice model.
3. Simulate 10,000 consumers with year 2000 observables.
4. For each draw:
   1. Simulate choices using computed combination choice probabilities.
   2. Simulate health transitions conditional on simulated choice using estimated health outcome probabilities.
   3. Simulate predicted health outcomes conditional on simulated choice.
5. Repeat, starting at step 2, but changing Combivir review.


Counterfactual reviews & Choices
Choice of Combivir Combinations (2000)
0
.2
.4
.6
.8
1
.01
.02
.03
.04
Conditional Choice Probability
Model
Counterfactual


Counterfactual reviews & Choices
Prob of No AIDS (2000)
0
.2
.4
.6
.8
1
.2
.4
.6
.8
1
Probability of CD4 > 200
Simulation
Data

Average change under the counterfactual

	Difference
Pr[No AIDS]	1.2%
Pr[No Ailments]	-0.5%

Note: We compare the model & counterfactual probabilities for individuals who switch to Combivir under the counterfactual ratings.

- Allow correlation patterns (ditch the *IIA* assumption) in our demand analysis.
- Interact reviews with observables (e.g., college) in the choice model.
- Use an IV for reviews in the demand model.
  - Current candidate: quality of new drugs that come onto the market.
- Counterfactuals:
  - What if reviews are no longer published?
  - What if only doctor reviews are published?
  - What if only activist reviews are published?
  - Are there distributional effects?

X

Combination Choice Model - without Fringe

	[1]	[2]	[3]	[4]	[5]
Probability of No Ailment	1.30***		1.37***	1.29***	1.28***
	(0.13)		(0.13)	(0.13)	(0.13)
Probability of Non-dec CD4	0.68***		0.69***	0.82***	0.92***
	(0.17)		(0.18)	(0.19)	(0.19)
Age	-0.07***	-0.07***	-0.07***	-0.07***	-0.07***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AIDS	0.20	-0.18***	0.08	0.12	0.11
	(0.12)	(0.07)	(0.13)	(0.13)	(0.13)
Working Full-time	-0.59***	-0.62***	-0.68***	-0.66***	-0.66***
	(0.10)	(0.05)	(0.10)	(0.10)	(0.10)
High School	-0.69***	-0.81***	-0.76***	-0.76***	-0.76***
	(0.12)	(0.07)	(0.12)	(0.12)	(0.12)
College	0.39***	0.23***	0.35***	0.36***	0.36***
	(0.11)	(0.06)	(0.11)	(0.11)	(0.11)
% of drugs with High Doctor reviews		0.07*	0.32***		0.28***
		(0.04)	(0.09)		(0.10)
% of drugs with High Activist reviews		1.20***	1.08***		1.07***
		(0.05)	(0.09)		(0.09)
% of New Drugs				-0.94***	-1.38***
				(0.17)	(0.19)
% of Old Drugs with High Doctor reviews				0.33***	
				(0.10)	
% of New Drugs with High Doctor reviews				0.20	0.45**
				(0.16)	(0.18)
% of Old Drugs with High Activist reviews				1.08***	
				(0.09)	
% of New Drugs with High Activist reviews				1.66***	0.84***
				(0.19)	(0.21)

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses.

Combination Choice Model - without Demographics

	[1]	[2]	[3]	[4]	[5]
Probability of No Ailment	0.79***		0.76***	0.65***	0.67***
	(0.09)		(0.09)	(0.09)	(0.09)
Probability of Non-dec CD4	-0.94***		-0.95***	-0.76***	-0.73***
	(0.12)		(0.12)	(0.13)	(0.13)
% of drugs with High Doctor reviews		-0.50***	-0.18**		-0.09
		(0.03)	(0.07)		(0.08)
% of drugs with High Activist reviews		0.33***	0.16**		-0.05
		(0.04)	(0.07)		(0.08)
% of New Drugs				-1.74***	-1.72***
				(0.16)	(0.16)
% of Old Drugs with High Doctor reviews				-0.19**	
				(0.08)	
% of New Drugs with High Doctor reviews				-0.19	-0.13
				(0.13)	(0.14)
% of Old Drugs with High Activist reviews				0.25***	
				(0.08)	
% of New Drugs with High Activist reviews				1.94***	1.95***
				(0.16)	(0.17)

* $p < 0.10$,** $p < 0.05$, ***$p < 0.01$; standard errors in parentheses.

- Simulate drug choices using conditional choice probabilities under the original & counterfactual probabilities.
- Predict the following conditional transition probabilities:
  - $\Pr_i$[No Aids | Aids, Prob of Non-dec CD4, W_i] $= p_1$
  - $\Pr_i$[No Ailments | Ailments, Prob of No Ailment, W_i] $= p_2$
  - $\Pr_i$[No Aids | No Aids, Prob of Non-dec CD4, W_i] $= p_3$
  - $\Pr_i$[No Ailments | No Ailments, Prob of No Ailment, W_i] $= p_4$
- Calculate unconditional transition probabilities Pr[No Aids | Aids], Pr[No Ailments | Ailments], Pr[No Ailments | No Ailments], Pr[No Aids | No Aids].

$$p_1 = \Phi(W_i'\beta_1^{H_1 1} + \mathbf{1}(CD4_{t-1} < 250)\beta_2^{H_1 1} + \Pr(\text{Non-dec CD4})\beta_3^{H_1 1}), \tag{4}$$

$$p_2 = \Phi(W_i'\beta_1^{H_2 1} + \mathbf{1}(\text{Ailments}_{t-1} = 1)\beta_2^{H_2 1} + \Pr(\text{No Ailments})\beta_3^{H_2 1}), \tag{5}$$

$$p_3 = \Phi(W_i'\beta_1^{H_1 2} + \mathbf{1}(CD4_{t-1} > 250)\beta_2^{H_1 2} + \Pr(\text{Non-dec CD4})\beta_3^{H_1 2}), \tag{6}$$

$$p_4 = \Phi(W_i'\beta_1^{H_2 1} + \mathbf{1}(\text{Ailments}_{t-1} = 0)\beta_2^{H_2 1} + \Pr(\text{Non-dec CD4})\beta_3^{H_2 1}), \tag{7}$$

where W_i includes demographic characteristics of individual i.

Investment in Health and Human Capital over the Life-Cycle

Elena Capatina, Michael Keane, and Shiko Maruyama

The Joint 2015 Annual Health Econometrics Workshop and the 7th Australasian Workshop on Econometrics and Health

December 12, 2015

Well-Documented Health-SES Gradient

- Strong association between Health and socioeconomic status (SES)
  - Also known as **health-wealth gradient** or health-wealth nexus
  - Reported in many fields: economics, sociology, public health
- Robust regardless of measures used:
  - For health: mortality, subjective health, morbidity, etc.
  - For SES: wealth, income, education, etc.
- Robust regardless of countries, cohorts, sub-populations
- But why?

Causal Mechanism Still Controversial

Disentangling the causal mechanism is challenging because of potential simultaneity

- Story 1: Health, $H \Rightarrow$ Wealth $\uparrow$
  - Good health allows people to become rich...?
  - $H \Rightarrow$ high productivity, low medical expenditure
- Story 2: Wealth $\Rightarrow$ $H \uparrow$
  - Wealth allows people to stay healthy...?
  - Access to expensive care
  - Better housing, better neighborhood, better nutrition
  - Better mental health
- Also potential omitted factors (such as IQ, risk aversion, or intertemporal discounting)

The Role of Health in One's Life-Cycle

- The relationship is even more complicated due to the time dimension:
  - both health and wealth are (1) time-persistent and (2) the consequences of past behaviors
  - people make decisions today taking into account future consequences
  - **Health investment** ... choosing healthy behavior and avoiding risky behavior

This Project

- We have very little quantitative knowledge about the causal mechanisms behind the health-wealth nexus
- This project aims to study interlink between health capital, human capital, and wealth,
- based on a dynamic life-cycle framework in a much more elaborate way than previous studies

In particular, 4 novel key features:

1. Detailed and carefully defined "health"
2. Two types of human capital: health capital and work experience
3. Endogenous *health investment* (in the spirit of Grossman, 1972)
4. Medical costs occur as exogenous income shock not as health investment

Today's Talk

- Provide an overview of the project by:
  1. Showing a simple life-cycle model with "health capital"
  2. Discuss how we incorporate "detailed health" and "endogenous health" and why they are novel and important
- Results still preliminary — No discussion today

Setup

A textbook dynamic life-cycle model with a small twist of health depreciation

- Males age 18-99
- 2 Education types: (1) College and (2) High School
- Uncertainty in health deterioration $\Rightarrow$ "life expectancy"
  - At age 100, everyone dies

Utility

- Utility function:

$$u(c_t) = b + \frac{c_t^{1-\sigma}}{1-\sigma}$$

  - $u'(c_t) > 0$ and $u''(c_t) < 0$ when $c_t > 0$
  - $b > 0$ is a constant utility one receives as long as alive
  - Utility does NOT depend on health
  - Assume: $\sigma = 0.75$

Health

Health simply governs mortality

- Health level, $H_t \in \{0, 1, ..., 20\}$, where 20 represents the best health and 0 death
- $H_{18} = 20$
- Health transition: $H_{t+1} = \begin{cases} H_t - 1 \text{ with probability } \delta \\ H_t \text{ with probability } 1\text{-}\delta \end{cases}$
  - $\Rightarrow$ when $H_t = 1$, mortality is δ
  - Assume: $\delta = 0.33$
- Its implication is only for mortality and longevity
  - No direct effect on individual's utility, quality of life, productivity, wage, or labour supply
- Education plays no role for mortality

Labor Supply and Budget Constraint

- Everyone works full-time until age 65 (2,000 hours p.a.).
  - Hourly wage equation:

$$w_t\,(\text{College}) = \begin{cases} 20 \text{ if age} < 65 \\ 0 \text{ if age } \geq 65 \end{cases}$$

$$w_t\,(\text{High School}) = \begin{cases} 15 \text{ if age} < 65 \\ 0 \text{ if age } \geq 65 \end{cases}$$

- Intertemporal budget constraint:

$$A_{t+1} = (1 + r)A_t + w_t \cdot 2000 - c_t$$

  - No borrowing

Dynamics

- Choice variables $=\{c_t\}$
- State variables $=\{A_t, H_t, age\}$
- Health deterioration is the only source of uncertainty
- Value function at age t:

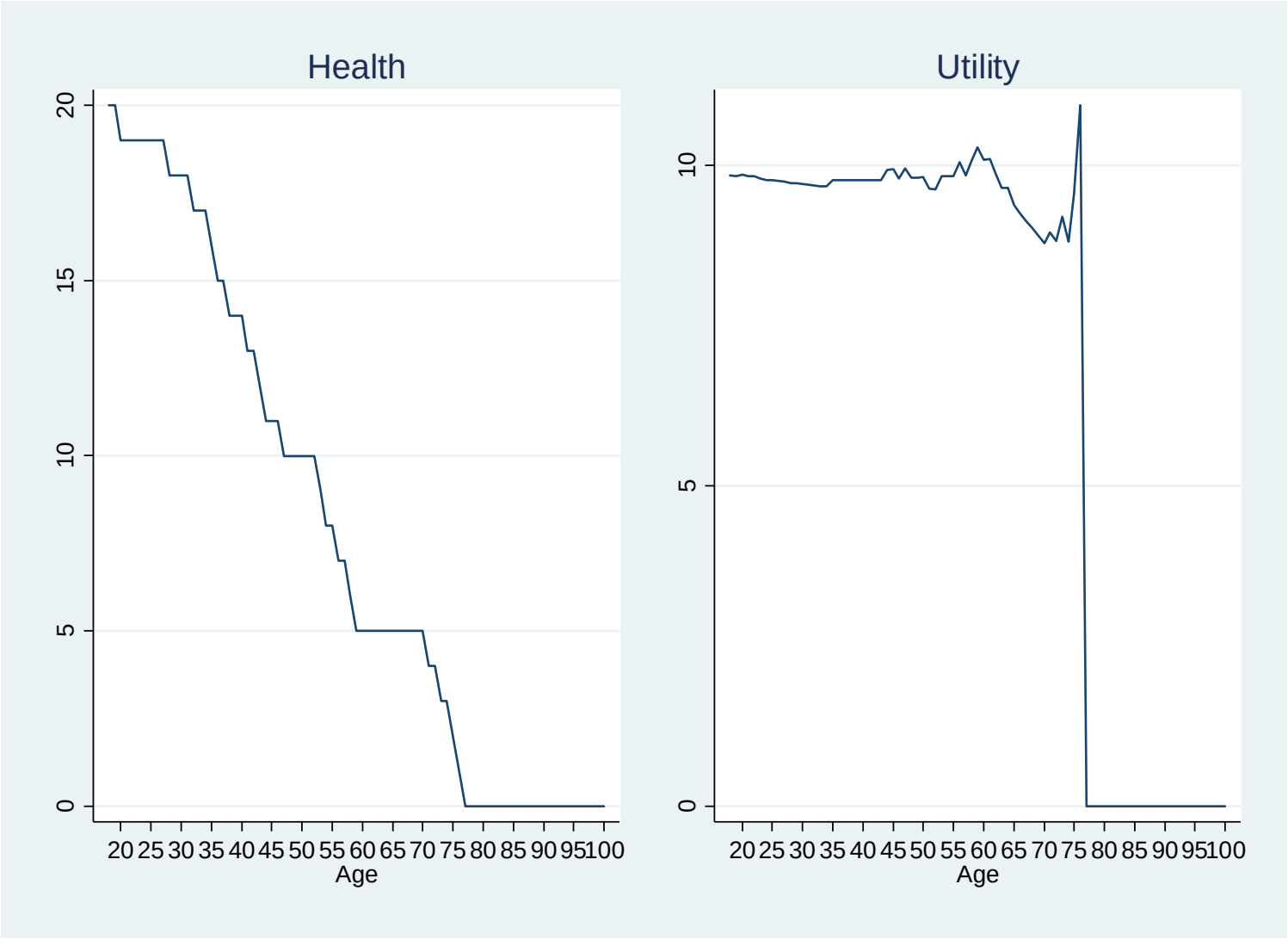
$$V(A_t, H_t, t) = \max_{c_t} \{u(c_t) + \beta EV(A_{t+1}, H_{t+1}, t+1)\}.$$

- Intertemporal discounting: $\beta = 0.975$
- Interest rate: $r = 0.025$
- These two offset each other $\Rightarrow$ Optimal to consume the same amount every period

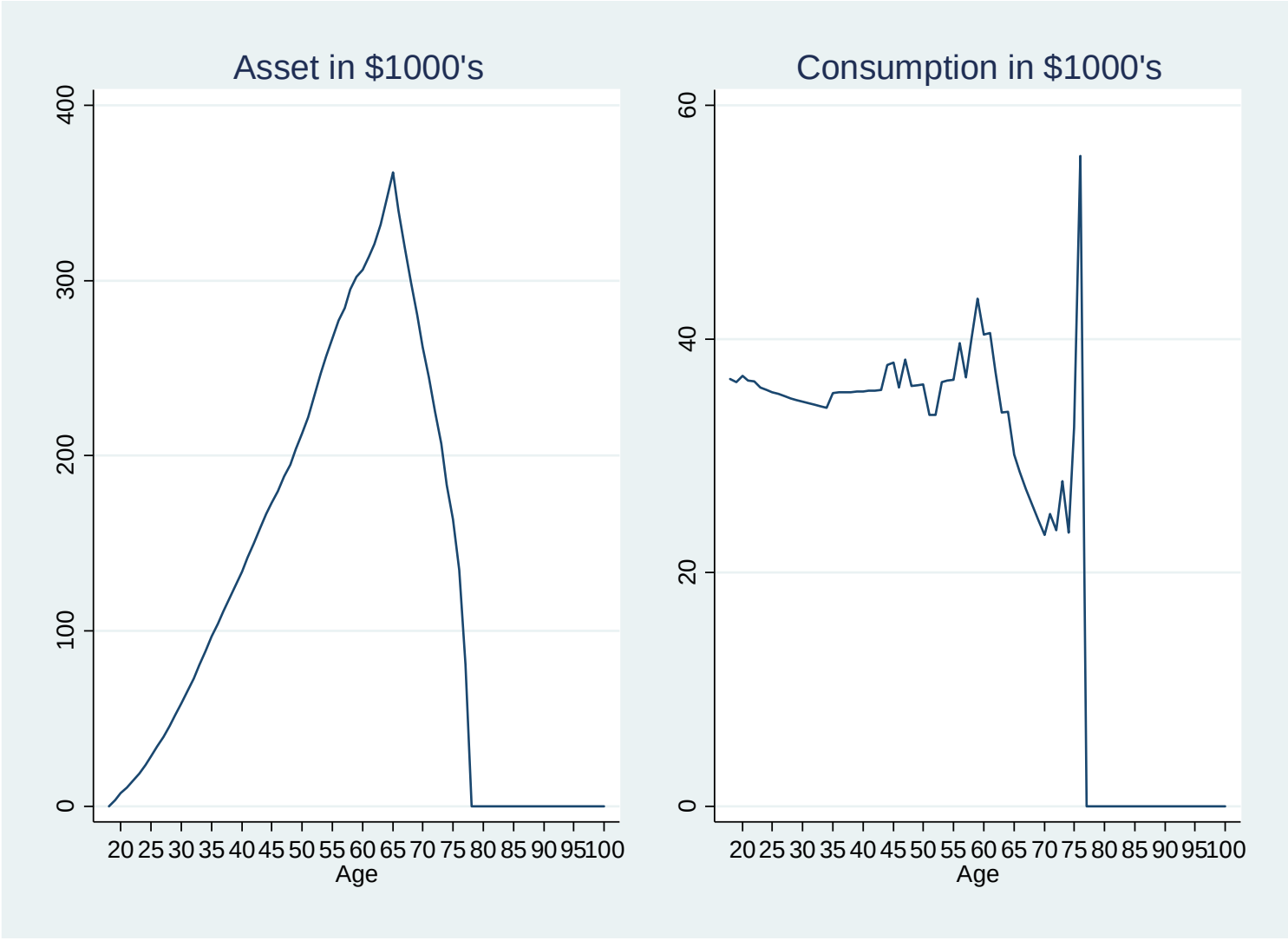
Simulation Setup

- Simulate the life-cycle of individuals from age 18 to death
- Asset: discretized into 500 equidistant grid points: $\left\{A^1, ..., A^{500}\right\} = \{0, ..., \text{ maximum amount}\}$.
- Discrete choice dynamic programing (DCDP)
  - First, solve the dynamic program from age 100 backward
  - Then simulate the life-cycle for each individual
  - Then for the entire economy
  - (In the paper, we estimate parameters by repeating this process)
- The results below are based on 50,000 simulated individuals

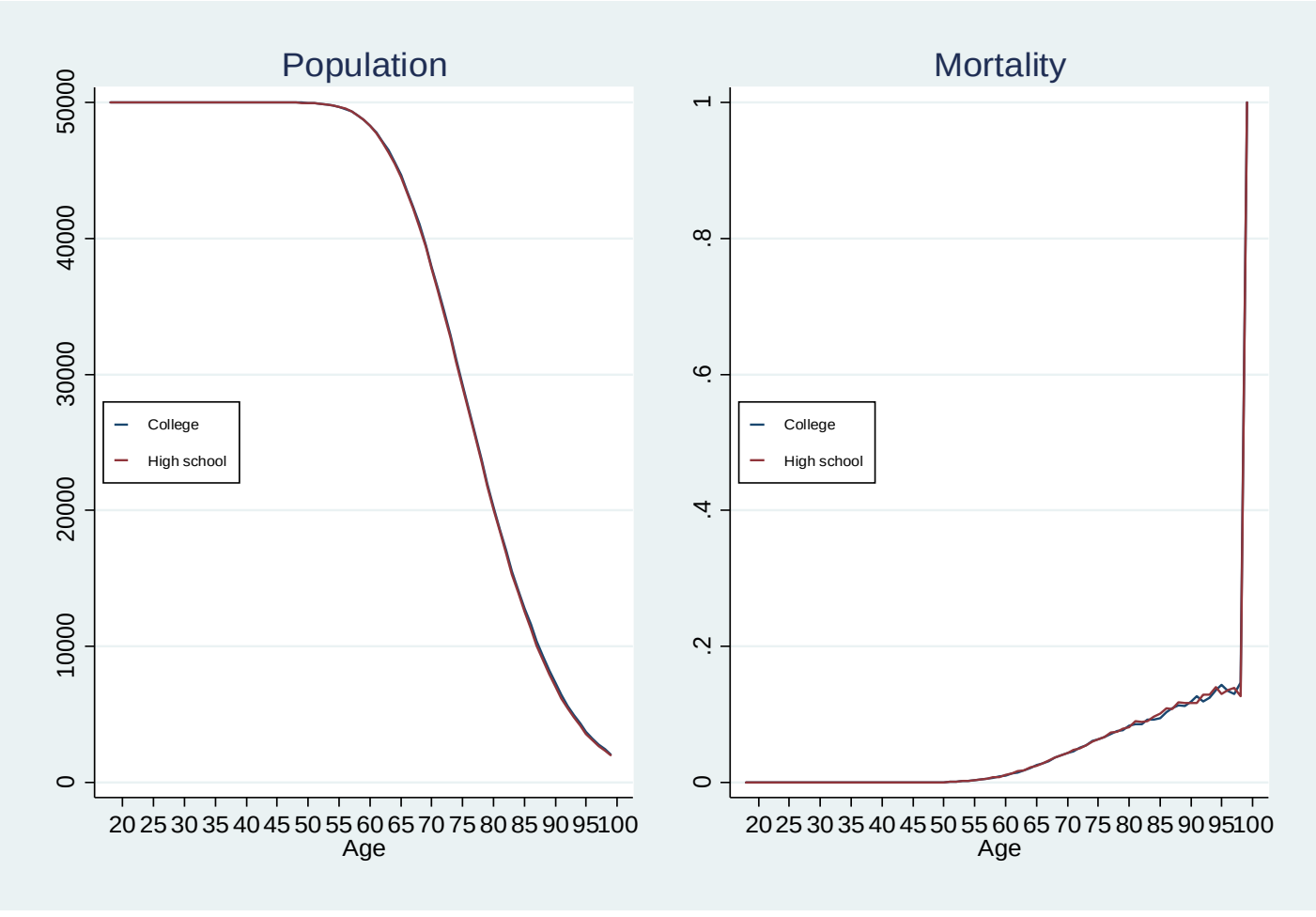
Example: Results of Individual 1



Example: Results of Individual 1

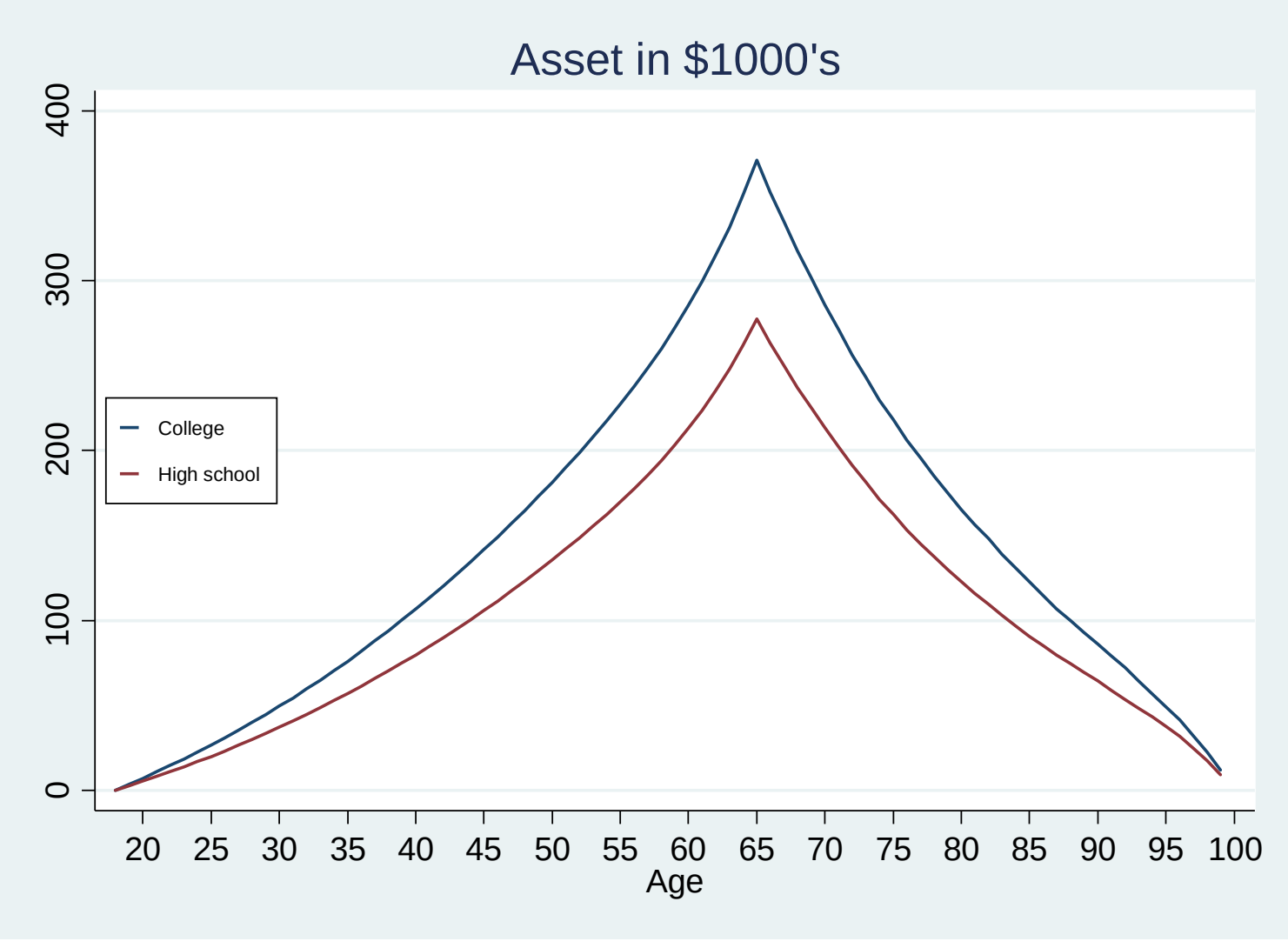


Results: Entire Economy (50,000 Simulated Individuals)

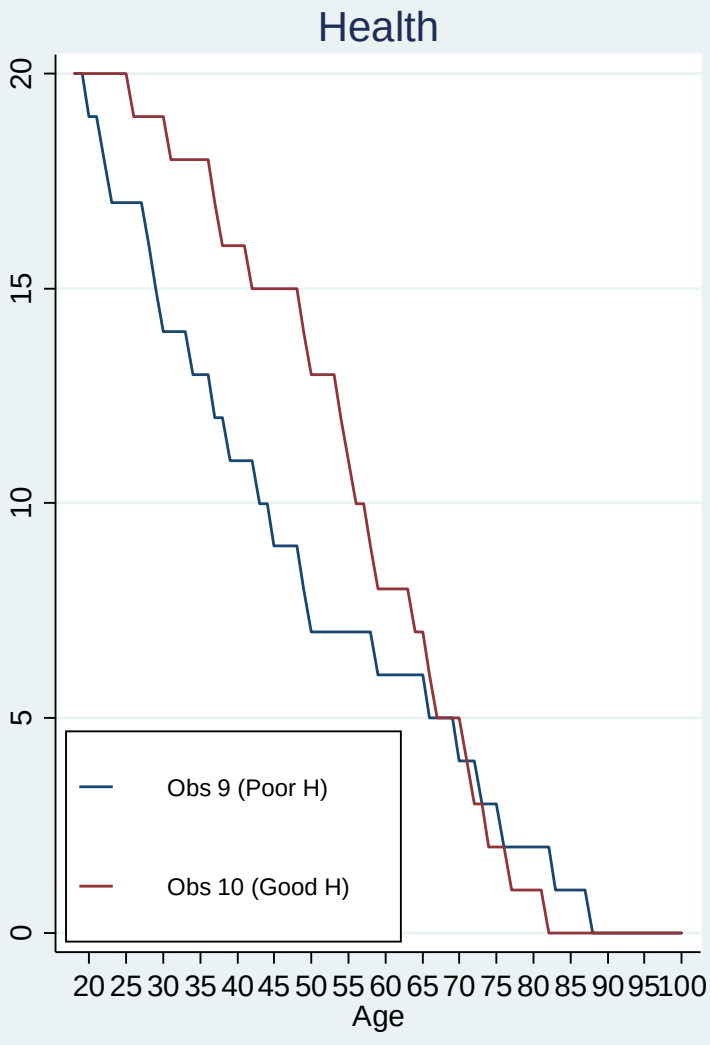


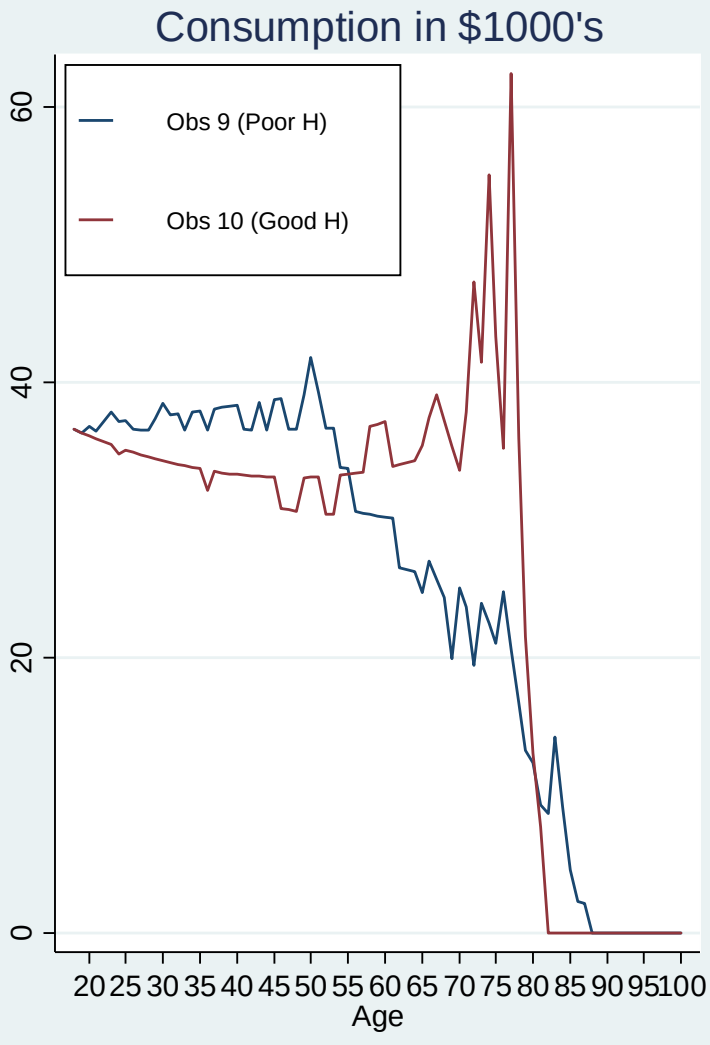
- Average life-expectancy: 77.34.

Results: Entire Economy (50,000 Simulated Individuals)

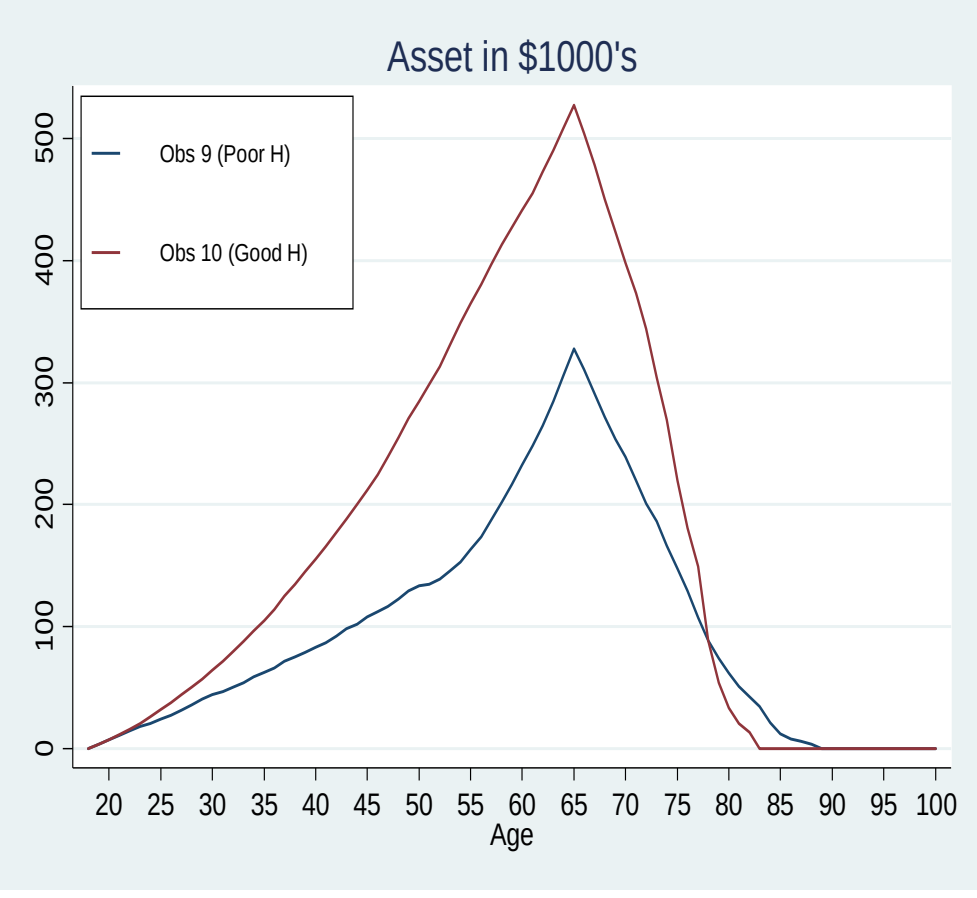


Results: Comparing Healthy and Sick Individuals





Why the Health-Wealth Gradient in this Model?



- NOT "The rich can buy health"
- NOT "Good health helps earning money"
- NOT "Good health minimizes health expenditures"
- NOT "Education or preferences"
- Simply, the healthy save because they expect to live long

Model Overview

- We build on Capatina (2015, JME), "Life-Cycle Effects of Health Risk", which quantified the roles of health in a life-cycle framework
- Population: US male household-heads age 25-99
- Partial equilibrium
- Choice: consumption/saving, health investment, labor supply (full-time / part-time / not working)
- State variables: assets, human capital, health capital (health, H, and health risk, R)
  - Human capital is simply years of work experience
- Uncertainty in job offer, health, medical expenditure, mortality
- Other model features: four education groups (exogenous), exogenous retirement age, social security (Medicaid, Medicare), health insurance, and unobserved heterogeneity

Utility

- In each period, individuals receive utility from being alive, consumption, leisure, and health

$$u(c, l, H) = b + \frac{c^{(1-\sigma)}}{1-\sigma} + \alpha \frac{l^{(1-\gamma)}}{1-\gamma} + \eta H$$

- l ... leisure:
  - $[1 - \text{hours specified in the job offer} - \text{time cost of health investment}]$
  - time endowment is normalized to 1

Data

- Medical Expenditure Panel Survey (MEPS)
  - 2-year "rotating" panel. It covers all ages
  - Rich information about health and health expenditure
  - Use waves 2000 to 2012
- For each person, 5 interview rounds during 2 years
  - $\Rightarrow$ 3 points for stock variables, e.g. H_t, H_{t+1}, H_{t+2}
  - $\Rightarrow$ 2 points for flow variables, c_t, c_{t+1}

Empirical Procedure

- (1) Construct variables from data
  - Health variables, labor supply, etc
  - DCDP, everything is discretized
- (2) Estimate exogenous transition process
  - Health transition regressions
  - Medical expenditure regressions
  - Survival regression
- (3) Calibration
  - Some parameter values taken from existing studies
  - Add standard errors by method of simulated moments (future)
- (4) Counterfactual simulations
  - quantify the contributions of various causal channels

Literature

There are many studies with health in the dynamic life-cycle framework, but health has been modeled in a highly stylized way.

- Studies on retirement:
  - Rust and Phelan (1997): ECMA: Health binary (good or bad)
  - French, Jones (2011) ECMA: Health binary
- Studies on saving and medical expenditure risk
  - Palumbo (1999) REStud: Health: good/fair/poor
  - De Nardi, French, Jones (2010) JPE: Health binary
- Khwaja (2010) JEcmt: Insurance choice. Health 5 categories
- Capatina (2015) JME: Evaluating 4 distinct health effects. Health good/fair/poor

Literature

- All these studies construct health measures relying on "self-assessed health"
  - Useful but not perfect
- There are also macro calibration studies that model health even without using health data

Studies with Endogenous Health

There are a few life-cycle studies with endogenous health

- Many of them are theoretical or calibration ... they do not use any data of health / health investment
- Halliday et al (2015): Health investment is medical expenditures. Focus is more on explaining health expenditures
- Khwaja (2010); constructs health investment from the data of exercise, smoking, etc. Focus is on health insurance choice
- No discussion about how to tackle simultaneity bias

Our Approach to Health Process

2 state variables, 3 temporary shocks, and 1 choice variable

- (1) Health capital, H
- (2) *Underlying risk factor*, R
  - Based on BMI, hypertension, high cholesterol, etc
  - A person can be "healthy" at the moment but with a large future health risk
- (3) 3 types of *Manifestation* of health conditions
  1. long-term predictable, d^p: e.g. life-style diseases
  2. long-term unpredictable, d^u: e.g. some cancers
  3. short-term, s: e.g. infections
- (4) Health investment, inv, which requires time and money costs
- (5) Medical expenditures ... exogenous income shock

Evolution of Health

- $H' \Leftarrow H, d^p, d^u, inv, t, educ$
- $R' \Leftarrow R, H, inv, t, educ$
- $d^p \Leftarrow R, H, t, educ$
- $d^u, s \Leftarrow t$
- Survival $\Leftarrow H, t, d^p, d^u$
- OOP Medical expenditures $\Leftarrow H, d^p, d^u, s, t,$ insurance
- Actual hours worked $\Leftarrow H, d^p, d^u, s$
- H also affects utility and job offer (wage)

Constructing Health Variables

- construct H based on various health measures
  - Factor analysis for perceived health, mental health, disability variables
  - Although still one-dimensional, much finer and more accurate
- construct R based on BMI and certain ICD9 conditions (e.g. hypertension, high cholesterol)
- to construct d^p, d^u, and s, we have classified over 300 conditions (at the 3-digit ICD code level) into 3 groups with a help of an MD
- construct inv based on (1) medical investment, (2) physical activities, (3) diet, (4) smoking, (5) health-related risk attitude
  - An index is constructed for each of these 5 groups and standardized by age group
  - Then 5 indexed are averaged and discretized into 3 levels

3 Digit ICD Code

International Classification of Diseases, Revision 9 (1975)

[Return to International Classification of Diseases]

(001-009) Intestinal infectious diseases
001 Cholera
002 Typhoid and paratyphoid fevers
003 Other Salmonella infections
004 Shigellosis
005 Other food poisoning (bacterial)
006 Amoebiasis
007 Other protozoal intestinal diseases
008 Intestinal infections due to other organisms
009 Ill-defined intestinal infections
(010-018) Tuberculosis
010 Primary tuberculous infection
011 Pulmonary tuberculosis
012 Other respiratory tuberculosis
013 Tuberculosis of meninges and central nervous system
014 Tuberculosis of intestines, peritoneum and mesenteric glands
015 Tuberculosis of bones and joints
016 Tuberculosis of genito-urinary system
017 Tuberculosis of other organs
018 Miliary tuberculosis
(020-027) Zoonotic bacterial diseases
020 Plague
021 Tularaemia
022 Anthrax
023 Brucellosis
024 Glanders
025 Melioidosis
026 Rat-bite fever
027 Other zoonotic bacterial diseases
(030-041) Other bacterial diseases
030 Leprosy
031 Diseases due to other mycobacteria
032 Diphtheria
033 Whooping cough
034 Streptococcal sore throat and scarlatina
035 Erysipelas
036 Meningococcal infection

Advantages of Detailed Health Process

1. Helps identification (especially by modelling d^p, d^u, s)
   - Health declines and new disease develops stochastically. Their realizations provide a good source of identification
   - Many life-cycle studies aim to "disentangle interlink between health and wealth" without even mentioning simultaneity bias
2. More realistic stories
   - A sudden realization of a disease
     - E.g. What happens one is diagnosed diabetes? Does future risk increase savings or decrease? Why?
   - Nowadays, many people live for very long with disability, which can be captured in our model (low H, low R)
     - This may be the worst risk you want to avoid
   - How social security and health insurance matter for these situations?

Summary

Four novel features of this project:

1. Detailed and carefully defined "health"
2. Two types of human capital: health capital and work experience
3. Endogenous *health investment* (Grossman, 1972)
4. Medical costs occur as consequences of health not as health investment

- We can talk about interaction between (1) asset accumulation, (2) human capital accumulation, and (3) health capital decumulation.
  - Longer life expectancy and better life expectation provide greater incentive to invest and save today
  - How much "self-enforcing" and "cross-fertilizing" effects?
- Preliminary results suggest a significant role of health investment

Other Implications

- We can quantitatively disentangle the health-SES gradient — shed light on why the rich are healthier and live longer.
  - Two alternative hypotheses:
    1. Poor people die early because they cannot afford it.
    2. Poor people die early because they do not want to live long.
  - Very different policy implications.
- Health insurance, pension/annuity
  - E.g. What if changing lump-sum pension to annuity? Do people try harder to live long?
- Social welfare
- Implications for the intertemporal elasticity literature.

Investment in Health and Human Capital over the Life-Cycle

Authors: Elena Capatina, Michael Keane, and Shiko Maruyama

Discussant: Audrey Laporte
University of Toronto
and
Canadian Centre for Health Economics

December, 2015

Objective of the paper

- Ultimately to produce a unified model that can be used to project an individual through his entire life course.
- In particular, want to be able to trace out the health investment and human capital investment patterns and the feedback between them over time.
- Will allow for an examination of the impact of health shocks on different dimensions –health stock, assets, employment.
- This is an ambitious paper, and the authors are clear throughout that this is ongoing work and that they are still in the process of refining their model.
- Has the potential to make an important contribution to the literature.

General Theoretical backdrop

- $Max \int_0^T U(C,H,L)e^{-\varrho t}dt$

Choice variables = C, I, L

Subject to

$$H = h(K_H),\ h_K > 0, h_{KK} < 0$$

$$\dot{K}_H = g(I) - \delta_H K_H$$

$$\dot{K}_F = rK_F + Y(H) - C - p_I I$$

$$Y(H) = w[D - L - S(H(K_H)]$$

In this paper:

K_H is Grossman-type health capital, K_F is financial capital,

Paper also includes human capital, measured as work experience.

K_H is not observable, H (and R, not included here) is.

In the paper, Health Capital actually evolves according to a stochastic process.

Two kinds of shock:

Transitory as in Cropper (1977) Journal of Political Economy "Health, Investment in Health, and Occupational Choice"

Permanent as in Poisson shocks in Laporte and Ferguson (2007), Journal of Population Economics "Investment in health when health is stochastic"

Hamiltonian:

$$\mathcal{H} = U(C, H(K_H), L) + \Psi_H[I - \delta K_H] + \Psi_F[rK_F + w(H(K_H))[D - L - S(H(K_H)] - C - p_I I]$$

Necessary conditions:

$$U_C(C, H, L) - \Psi_F = 0$$

$$U_L - \Psi_F\left[w\left(H(K_H)\right)\right] = 0$$

$$\Psi_H - p_I \Psi_F = 0$$

And:

$$\dot{\Psi}_H = \varrho\Psi_H - [U_H H_K - \delta\Psi_H - \Psi_F\, wS_H H_K] = [\varrho + \delta]\Psi_H - [U_H - \Psi_F\, wS_H]H_K]$$

$$\dot{\Psi}_F = [\varrho - r]\Psi_F$$

Then substituting from:

$$\Psi_H = p_I \Psi_F$$

$$\Psi_F = U_C(C, H, L)$$

$$U_L = \left[w\left(H(K_H)\right)\right]\Psi_F$$

Lets us replace the Ψ terms with expressions in marginal utilities.

Converting to discrete time terms, this gives us, in general notation :

$$C_t = c\left(X_t, C_{t-1} I_{t-1} L_{t-1}, K_{F,t-1}, K_{H,t-1}\right)$$
$$I_t = I\left(X_t, C_{t-1} I_{t-1} L_{t-1}, K_{F,t-1}, K_{H,t-1}\right)$$
$$L_t = L\left(X_t, C_{t-1} I_{t-1} L_{t-1}, K_{F,t-1}, K_{H,t-1}\right)$$
$$K_{F,t} = f\left(X_t, C_{t-1} I_{t-1} L_{t-1}, K_{F,t-1}, K_{H,t-1}\right)$$
$$K_{H,t} = h\left(X_t, C_{t-1} I_{t-1} L_{t-1}, K_{F,t-1}, K_{H,t-1}\right)$$

K_H is Grossman health capital, unobservable, so replace it with observable health proxies: $H(K_H)$

Wind up with highly nonlinear set of equations representing individuals' optimizing choices:

$$C_t = c\left(X_t, C_{t-1}I_{t-1}L_{t-1}, K_{F,t-1}, H_{t-1}\right)$$
$$I_t = I\left(X_t, C_{t-1}I_{t-1}L_{t-1}, K_{F,t-1}, H_{t-1}\right)$$
$$L_t = L\left(X_t, C_{t-1}I_{t-1}L_{t-1}, K_{F,t-1}, H_{t-1}\right)$$
$$K_{F,t} = f\left(X_t, C_{t-1}I_{t-1}L_{t-1}, K_{F,t-1}, H_{t-1}\right)$$
$$H_t = h\left(X_t, C_{t-1}I_{t-1}L_{t-1}, K_{F,t-1}, H_{t-1}\right)$$

Applied economist's problem – how to estimate this and derive values for underlying parameters?

For regression-based approach with linearization, see:

Jones, Laporte, Rice, Zucchelli (2015): A Synthesis of the Grossman and Becker-Murphy Models of Health and Addiction: Theoretical and Empirical Implications Canadian Centre for Health Economics Working Paper

What this paper does

- Have a ***very non-linear*** system of interconnected auto-regressions equations.
- One approach is linearization (Jones, et al. 2015).
- The approach in this paper is calibration
- The appeal of calibration is that it should let you retain the non-linearity.

- As it stands the calibration in the paper is still a work in progress.

- Most of the preference parameters are given values taken from the literature.

- The estimation results reported in the paper are for the most part stand alone ordered probits.

- It's not easy at this point to judge whether the calibration exercise will be fruitful.

- In one way calibration has the same issues as FIML.
- Mis-specifications spill through all of the estimation
- Here to simplify the authors assume that utility is additive and that the consumption component of the utility function has a constant relative risk aversion form.
- There are a couple of other less obvious assumptions which might significantly bias the results of the calibration exercise.

Discount rate -β

- SES gradient in health is recognized but that ties into the education health gradient
  - gets us back to the Fuchs/Grossman question—is positive relationship between education & health is it because education directly improves health outcomes or because if you invest in education you are likely to invest in health

- Applying same discount rate to every individual in the model?
  - Means can't have people making different investment decisions in the model because they discount the future at different rates
  - Thus can't test one of the fundamental alternative hypotheses—attributing any observed effects to education differences without accounting for the effect of β

Cohort effects

- Trying to get at wage and asset profiles over the life course and trying to match cross-sectionally but without taking account of cohort effects
  - i.e. not taking account of the circumstances in which people would have been making their savings and investment decisions

- Smoking behaviour-graphs-reason 70 year olds less likely to smoke as seen from MEPs data is possibly because a lot of the smokers died—what does this say about the fidelity of the behavioural model?

Issues related to estimated equations

Example: specification of wage equation (page 20)

- $w(educ;HC;H;h) = \beta_0 + \beta_1 HC + \beta_2 HC^2 + \beta_3 HC^3 + \beta_4 I_{H=(A;G)} + \beta_5 I_{H=G} + \beta_6 I_{h=hrsPT}$

- If this is the take-home wage—shoudn‘t it include health insurance? for any given hrs worked & MP there will be a trade-off btw take-home wage and benefits.
- Where is region, industry etc.?
- Rather thin estimating model-risk of omitted variable bias?

Nature of the calibration exercise

- Calibration tries to find values for utility function parameters for example based on matching moments of simulated data with moments of actual data (here the moment is just the mean) and iterating until the best match has been found.

- Commonly pick values for some of the parameters straight out of the literature and use simulation methodology to find values of the remaining ones.

- At the present stage of this paper, judging from Table 14, virtually every parameter has been pulled from the literature.

- Not clear how much we are looking at actual calibration results and whether the policy scenarios rely at all on calibration results.

Implementation of policy scenarios

- What we would really like to be able to do with this model is to trace out the impact of a change occuring at some point in the person's life through the remainder of their life.

- Not clear whether the policy scenarios are making any use of the calibrated model—they seem basically to be marginal effects from the ordered probits.

- What can the policy scenarios tell us then about the value of the calibrated model (as per the title of the paper)?

Calibration versus Estimation?

These are different ways of trying to find the values of the coefficients.

$$C_t = C(X_t, C_{t-1}, I_{t-1}, L_{t-1}, K_{F,t-1}, H_{t-1})$$
$$I_t = I(X_t, C_{t-1}, I_{t-1}, L_{t-1}, K_{F,t-1}, H_{t-1})$$
$$L_t = L(X_t, C_{t-1}, I_{t-1}, L_{t-1}, K_{F,t-1}, H_{t-1})$$
$$K_{F,t} = f(X_t, C_{t-1}, I_{t-1}, L_{t-1}, K_{F,t-1}, H_{t-1})$$
$$H_t = h(X_t, C_{t-1}, I_{t-1}, L_{t-1}, K_{F,t-1}, H_{t-1})$$

One argument in favour of calibration is that can find coefficient values in much more non-linear functions i.e. don't have to linearize the utility function for example.

Given that in the end have to make so many simplifying assumptions about functional forms e.g. U additive, would we be better off using flexible functional forms? —like translog –or second order Taylor series approximation in whatever relation.

Have large enough datasets so perhaps instead of trying to use fully detailed models—use regression but on flexible functional forms recognizing that they are approximations?