Single agent dynamic models — applications in health economics

Paul Schrimpf

UBC
Economics 565

March 16, 2021

Section 1

Fang and Wang (2015)

Fang and Wang (2015)

"Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions"

- This paper: identification of discount factor (and more)
  - Key restriction: variables that shift expectations, but do not enter current payoff
- Find substantial present bias and naivety in mammography screening

Model

- Dynamic discrete choice, $i \in \{0, \ldots, I\}$

$$u_i(x, \epsilon) = u_i(x) + \epsilon_i$$

- Infinite time horizon, hyperbolic discounting

$$U_t(u_t, u_{t+1}, \ldots) = u_t + \beta \sum_{k=t+1}^{\infty} \delta^{k-1} u_k$$

  - Present bias $\equiv \beta$
  - Discount factor $\equiv \delta$
- Time t self believes that future selves will make choices with present bias $\tilde{\beta} \in [\beta, 1]$ and discounting δ
  - Completely naive if $\tilde{\beta} = 1$
  - Sophisticated if $\tilde{\beta} = \beta$

Model

- Continuation strategy profile $\sigma_t^+ = \{\sigma_k\}_{k=t^\infty}$,
- Continuation utility:

$$V_t(x_t, \epsilon_t; \sigma_t^+) = u_{\sigma_t(x_t,\epsilon_t)}(x_t, \epsilon_{\sigma_t(x_t,\epsilon_t)t}) + \delta \mathrm{E}\left[V_{t+1}(x_{t+1}, \epsilon_{t+1}; \sigma_{t+1}^+)\right|$$

- Perceived continuation strategy

$$\tilde{\sigma}_t(x_t, \epsilon_t) = \arg\max_i u_i(x_t, \epsilon_{it}) + \tilde{\beta}\delta \mathrm{E}\left[V_{t+1}(x_{t+1}, \epsilon_{t+1}; \tilde{\sigma}_{t+1}^+) | x_t, i\right]$$

- Perception-perfect strategy with partial naivety

$$\sigma_t^*(x_t, \epsilon_t) = \arg\max_i u_i(x_t, \epsilon_{it}) + \beta\delta \mathrm{E}\left[V_{t+1}(x_{t+1}, \epsilon_{t+1}; \tilde{\sigma}_{t+1}^+) | x_t, i\right]$$

Identification 1

- Assumptions:
  - Stationarity
  - Conditional independance: $\pi(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t, d_t) = q(\epsilon_t)\pi(x_{t+1}|x_t, d_t)$
  - ϵ_t iid extreme value
- Perceived long-run choice-specific value function:

$$V_i(x) = u_i(x) + \delta \sum_{x'} V(x')\pi(x'|x, i)$$

where $V(x) =$ expected value from following perceived continuation strategy

$$V(x) = \mathrm{E}[V_{\tilde{\sigma}(x,\epsilon)}(x) + \epsilon_{\tilde{\sigma}(x,\epsilon)}]$$

Identification 2

- Current choice-specific value function

$$W_i(x) = u_i(x) + \beta\delta \sum_{x'} V(x')\pi(x'|x, i)$$

- Naively perceived next-period choice-specific value function

$$Z_i(x) = u_i(x) + \tilde{\beta}\delta \sum_{x'} V(x')\pi(x'|x, i)$$

- Definition of W, Z

$$Z_i(x) - u_i(x) = \frac{\tilde{\beta}}{\beta}[W_i(x) - u_i(x)]$$

Identification 3

- Distribution of ϵ implies

$$P_i(x) = \frac{e^{W_i(x)}}{\sum_j e^{W_j(x)}}$$

and

$$V(x) = \log\left(\sum_i e^{Z_i(x)}\right) + (1 - \tilde{\beta})\delta \sum_j \frac{e^{Z_j(x)}}{\sum_k e^{Z_k(x)}} \sum_{x'} V(x')\pi(x'|x, j$$

- Given $\delta, \beta, \tilde{\beta}$, variant of usual inversion of choice probabilities identifies $u_i(x) - u_0(x)$
- Assumption: $\exists x_1 \neq x_2$ such that (i) $u_i(x_1) = u_i(x_2) \forall i$ and $\pi(x'|x_1, i) \neq \pi(x'|x_2, i)$ for some i and $(I + 1) \times |\mathcal{X}| \geq 4$

Identification of $\delta, \beta, \tilde{\beta}$

- Collect equations relating $u, \beta, \tilde{\beta}, \delta$ to observed π, P

$$G(\underbrace{u, \beta, \tilde{\beta}, \delta}_{x,\ \dim n=(I+1)|\mathcal{X}|+3} ; \underbrace{\pi, P}_{s=|\mathcal{X}|(|\mathcal{X}|-1)(I+1)+(I+1)|\mathcal{X}|}) = \underbrace{0}_{b,\ \dim m=(I+1)|\mathcal{X}|+|\mathcal{X}_e||\mathcal{X}_r|(I+1)}$$

- Transversality theorem: let $G : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}^m$ by $\max\{n-m, 0\}$ times continuously differentiable. Suppose 0 is a regular value of G, i.e. $G(x, b) = 0$ implies rank $DG_{x,b} = m$, then generically in b, $G(\cdot, b) : \mathbb{R}^n \to \mathbb{R}^m$ has 0 as a regular value, i.e. rank $D_x G_{x,b} = m$
- G satisfies these conditions, but $n < m$, so 0 is never a regular value, so generically in b, $G(x, b) \neq 0$
- Paper says, so generically $G(x, b) = 0$ has no solution, except at true x^*, but ...
- Theorem says except at true b^*, given true b^*, theorem says nothing about how many x satisfy equation ...
- Introduce obviously non-identified reparameterization, e.g. $\beta = \beta_1 + \beta_2$, where does argument breakdown?
- Theorem does imply that model is falsifiable

Identification of $\delta, \beta, \tilde{\beta}$

- I believe discount factors are identified here, but proof appears incomplete
- See Abbring and Daljord (2019) for details of the problem
- Abbring and Daljord (2017) and Abbring, Daljord, and Iskhakov (2018) identification results for similar models
- Related identification results:
  - Magnac and Thesmar (2002)
  - Bajari et al. (2013)
  - An, Hu, and Ni (2014)
  - Dubé, Hitsch, and Jindal (2014)
  - Yao et al. (2012)
  - Komarova et al. (2018)

Estimation

- Maximum pseudo-likelihood
  1. Estimate $\hat{u}_i(x; \beta, \tilde{\beta}, \delta)$ using choice probabilities and Bellman equations
  2. Maximize pseudo likelihood $\hat{P}_i(x; \hat{u}_i(x; \beta, \tilde{\beta}, \delta), \beta, \tilde{\beta}, \delta)$

Empirical application: mammography

- Data from HRS, women 51-64
- Instantaneous utility from getting mammogram:

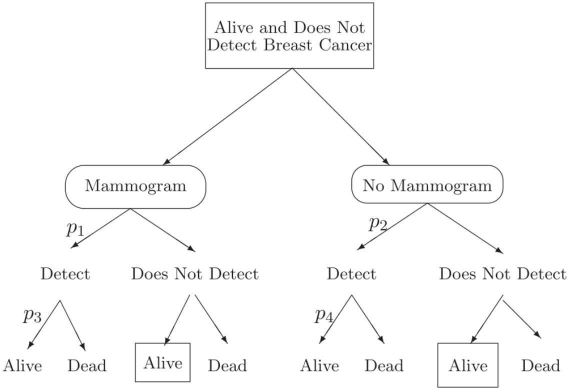
$$u_1(x) - u_0(x) = \alpha_0 + \alpha_1 BadHealth + \alpha LogIncome$$

- Demographics excluded from payoffs, but enter transition probabilities

TABLE 2
SUMMARY STATISTICS OF KEY VARIABLES IN THE ESTIMATION SAMPLE

Variable	Mean	Std. Dev.	Minimum	Maximum	Obs.
Mammogram	0.763	0.426	0	1	11,447
Bad health	0.217	0.412	0	1	11,447
Married	0.714	0.452	0	1	11,447
White (non-Hispanic)	0.798	0.401	0	1	11,447
High school or higher	0.796	0.402	0	1	11,447
Age	57.82	3.95	51	64	11,447
Death	0.014	0.117	0	1	11,447
Insurance	0.721	0.449	0	1	11,447
Household income ($1,000)	50.95	67.29	0.101	2136	11,447
Log of household income	10.354	1.053	4.615	14.575	11,447
Mother still alive or died after age 70	0.768	0.422	0	1	11,447
Mother education (high school or higher)	0.431	0.495	0	1	11,447
Father still alive or died after age 70	0.625	0.484	0	1	11,447
Father education (high school or higher)	0.404	0.491	0	1	11,447
Bad health ($t+1$)	0.237	0.425	0	1	11,289
Household income ($t+1$) ($1,000)	50.297	184.589	0.103	17,600	11,289
Log of household income ($t+1$)	10.307	1.036	4.638	16.684	11,289

NOTES: The last three variables in the table are observed only for those who survive to the second period.



NOTES: (1) $p_1 > p_2$: mammogram can detect breast cancer at its early stage; $p_3 > p_4$: survival rate is higher when breast cancer is detected at earlier stage. (2) The states with rectangular frame box are those in which she will keep making decisions on whether to undertake mammography.

FIGURE 1

THE TIMELINE FOR MAMMOGRAPHY DECISIONS

TABLE 3

MAMMOGRAM CHOICES BY EACH SHORT PANEL DEFINED BY CONSECUTIVE WAVES OF HRS

Panels	Total	No	Yes
Wave 3→4	3,899	1,059 (27.16%)	2,830 (72.84%)
Wave 4→5	636	135 (21.23%)	501 (78.77%)
Wave 5→6	3,476	771 (22.18%)	2,705 (77.82%)
Wave 6→7	45	7 (15.56%)	38 (84.44%)
Wave 7→8	3,372	741 (21.98%)	2,631 (78.02%)
Wave 8→9	29	5 (17.24%)	24 (82.76%)
Total	11,447	2,718	8,729

TABLE 4

DETERMINANTS OF MAMMOGRAPHY DECISIONS: THE CHOICE PROBABILITIES FROM LOGIT REGRESSION

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Bad health	0.146**	0.113*	0.161***	0.099*	0.095	0.079
	(0.054)	(0.063)	(0.057)	(0.056)	(0.058)	(0.059)
Log income	0.253***	0.264***	0.252***	0.290***	0.287***	0.324***
	(0.027)	(0.028)	(0.026)	(0.025)	(0.026)	(0.026)
Married	0.116***	0.106*	0.086	0.063	0.080	0.127**
	(0.056)	(0.058)	(0.053)	(0.053)	(0.055)	(0.055)
White	−0.265***	−0.206***	−0.269***	−0.212***	−0.270***	−0.244***
	(0.060)	(0.063)	(0.056)	(0.056)	(0.059)	(0.059)
Insurance	0.537***	0.513***	0.519***	0.558***	0.563***	
	(0.055)	(0.056)	(0.052)	(0.052)	(0.053)	
High school	0.344***	0.324***	0.370***			0.414***
	(0.061)	(0.063)	(0.056)			(0.060)
Mother70	−0.090*		−0.058			−0.089*
	(0.054)		(0.051)			(0.054)
MotherHighSchool	0.112**				0.179***	0.115**
	(0.050)				(0.048)	(0.049)
Father70		0.089*		0.082*		
		(0.048)		(0.044)		
FatherHighSchool		0.081				
		(0.051)				
Age	0.012**	0.011*	0.010*	0.010*	0.011**	0.009
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Constant	−2.607***	−2.800***	−2.460***	−2.718***	−2.733***	−2.918***
	(0.448)	(0.467)	(0.426)	(0.430)	(0.442)	(0.446)
Pseudo-R^2	0.043	0.042	0.040	0.037	0.040	0.034

NOTES: (1) Robust standard errors are in parenthesis; (2) the included variables in each specification correspond to those in Table 6; (3) *, **, and *** represent statistical significance at 10%, 5%, and 1%, respectively.

TABLE 5
DETERMINANTS OF PROBABILITY OF DYING IN TWO YEARS FROM LOGIT REGRESSIONS

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Mammogram	−0.459***	−0.446***	−0.435***	−0.442***	−0.436***	−0.446***
	(0.172)	(0.181)	(0.159)	(0.161)	(0.170)	(0.170)
Bad health	1.750***	1.725***	1.696***	1.631***	1.698***	1.780***
	(0.187)	(0.189)	(0.177)	(0.173)	(0.184)	(0.186)
Log income	−0.208**	−0.225***	−0.173**	−0.161**	−0.166**	−0.230***
	(0.085)	(0.082)	(0.081)	(0.079)	(0.085)	(0.079)
Married	0.092	0.136	−0.067	−0.072	0.093	0.052
	(0.189)	(0.195)	(0.173)	(0.171)	(0.186)	(0.188)
White	−0.137	−0.088	−0.148	−0.162	−0.202	−0.110
	(0.187)	(0.199)	(0.168)	(0.170)	(0.181)	(0.186)
Insurance	−0.151	−0.151	−0.122	−0.066	−0.139	
	(0.190)	(0.196)	(0.177)	(0.178)	(0.184)	
High school	0.273	0.446**	0.116			0.251
	(0.198)	(0.207)	(0.173)			(0.192)
Mother70	−0.273		−0.239			−0.286
	(0.179)		(0.165)			(0.178)
MotherHighSchool	0.122				0.154	0.099
	(0.190)				(0.185)	(0.188)
Father70		−0.257		−0.165		
		(0.172)		(0.154)		
FatherHighSchool		−0.254				
		(0.204)				
Age	0.075***	0.050**	0.073***	0.072***	0.079***	0.079***
	(0.021)	(0.022)	(0.020)	(0.020)	(0.021)	(0.021)
Constant	−6.890***	−5.412***	−6.891***	−6.905***	−7.528***	−6.962***
	(1.534)	(1.586)	(1.391)	(1.454)	(1.552)	(1.516)
Pseudo-R^2	0.106	0.101	0.106	0.099	0.102	0.107

NOTES: (1) Robust standard errors are in parenthesis; (2) the included variables in each specification correspond to those in Table 6; (3) ** and *** represent statistical significance at 5% and 1%, respectively.

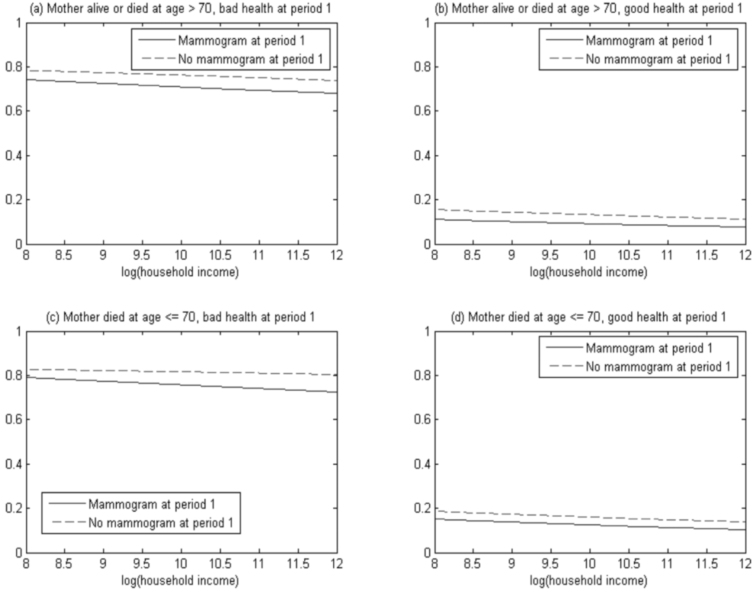


Figure 2

nonparametric estimate of the probability of bad health as a function of logincome, conditional on mother70, badhealth, and mammogram in previous period

TABLE 6

PARAMETER ESTIMATES FOR THE INSTANTANEOUS UTILITY FUNCTION AND TIME PREFERENCE PARAMETERS UNDER SIX DIFFERENT SETS OF EXCLUSIVE RESTRICTION VARIABLES

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (A) Instantaneous Utility Function Parameters						
Bad health	−0.434***	−0.724***	−0.138***	−0.913***	−0.335***	−0.472***
	(0.114)	(0.260)	(0.061)	(0.120)	(0.053)	(0.103)
Log income	1.177***	1.167***	1.346***	1.153***	1.265***	1.280***
	(0.031)	(0.104)	(0.062)	(0.106)	(0.032)	(0.072)
Constant	−0.811***	−0.928***	−2.732***	−0.926***	−1.722***	−2.014***
	(0.256)	(0.199)	(0.064)	(0.163)	(0.054)	(0.091)
Panel (B) Time Preference Parameters						
δ	0.681***	0.792***	0.741***	0.947***	0.759***	0.764***
	(0.123)	(0.098)	(0.058)	(0.100)	(0.020)	(0.058)
β	0.679***	0.791***	0.679***	0.508***	0.578***	0.762***
	(0.187)	(0.193)	(0.298)	(0.109)	(0.074)	(0.185)
$\tilde{\beta}$	1.000***	1.000***	0.984***	1.000***	1.000***	1.000***
	(0.282)	(0.247)	(0.496)	(0.105)	(0.027)	(0.281)
Panel (C) Hypothesis Tests						
$H_0 : \beta = 1$	Reject	Reject	Reject	Reject	Reject	Reject
$H_0 : \tilde{\beta} = \beta$	Reject	Reject	Reject	Reject	Reject	Reject
Exclusion Variables:						
White	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Married	Yes	Yes	Yes	Yes	Yes	Yes
HighSchool	Yes	Yes	Yes	No	No	Yes
Insurance	Yes	Yes	Yes	Yes	Yes	No
Mother70	Yes	No	No	No	No	Yes
MotherHighSchool	Yes	No	Yes	No	Yes	Yes
Father70	No	Yes	No	Yes	No	No
FatherHighSchool	No	Yes	No	No	No	No

NOTES: (1) The last panel indicates the exclusive restriction variables used in the specification in that column, with "Yes" meaning the variable is used and "No" otherwise; (2) standard errors for parameter estimates are in parenthesis, and *** represents statistical significance at 1%; (3) for hypothesis tests reported in panel C, all are rejected with p-value less than 0.01.

TABLE 7

MAMMOGRAPHY COMPLIANCE RATES PREDICTED BY THE MODEL AND IMPLIED BY DIFFERENT COUNTERFACTUAL EXPERIMENTS

	(1)	(2)	(3)	(4)	(5)	(6)
Data	0.76236	0.76307	0.75892	0.75841	0.76152	0.76179
Model	0.76440	0.76603	0.75781	0.76015	0.76100	0.76259
Counterfactual experiments:						
[1] No naivety: $\tilde{\beta} = \beta[= \hat{\beta}]$	0.76442	0.76607	0.75786	0.76061	0.76104	0.76262
[2] No naivety and no present bias: $\tilde{\beta} = \beta = 1$	0.78673	0.79121	0.78796	0.85860	0.79025	0.78498

NOTES: (1) The sample sizes slightly vary as we change the set of the exclusive restriction variables, which explains the changes in the mammography compliance rates in the data; (2) the exclusive variables used in each column correspond to those of the same column in Table 6.

Section 2

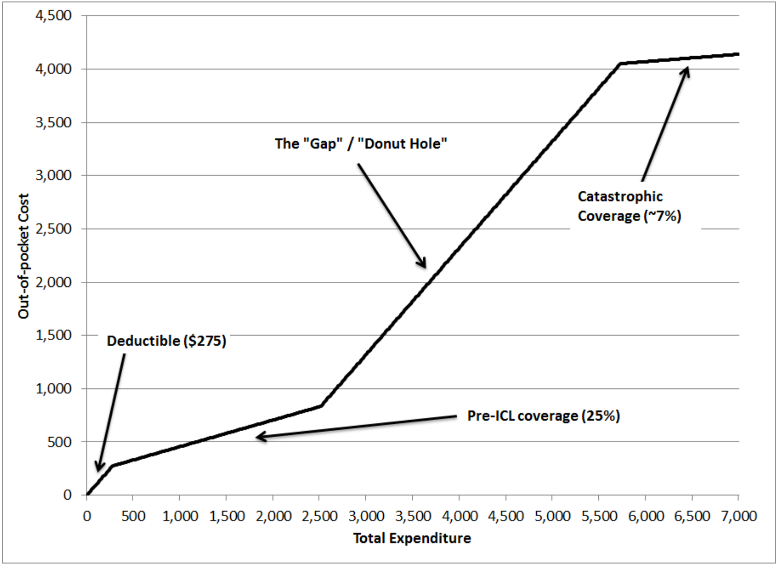
Einav, Finkelstein, and Schrimpf (2015)

The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D

Medicare part D

- Medicare Part D
  - The largest expansion of Medicare since inception
  - 32 million beneficiaries, 11% of Medicare spending
  - Typical coverage highly non-linear
- Many planned and potential changes
  - Under ACA, "donut hole" will be "filled" by 2020
- Main objectives:
  - Assess the contract design impact on drug spending ("moral hazard")
  - Estimate the spending effects of the proposed changes in Part D contracts
  - Conceptually: analyze healthcare utilization under non-linear contracts

Standard Coverage in Medicare part D (in 2008)



Approach

- We use “moral hazard”to mean the effect of out-of-pocket price on health care spending.

1. Descriptive analysis of moral hazard
   - Prescription demand is responsive the out-of-pocket price
   - Individuals are forward looking when choosing prescriptions
2. Structural dynamic model of prescription demand to quantify descriptive results and for counterfactual analysis of alternative contract designs

Related literatures

- Large recent literature on Medicare Part D, mostly focusing on the quality of plan choice (Heiss et al. 2010, 2012; Abaluck and Gruber 2011; Ketcham et al. 2012)
- Large venerable literature on response of healthcare spending to insurance contracts (“moral hazard”)
  - Only recently has attention focused on non-linear nature of contract (Bajari et al. 2011; Kowalski 2011; Marsh 2011; Aron-Dine et al. 2012)
- “Bunching” response to progressive income tax (Saez 2010; Chetty et al. 2011; Chetty 2012)
  - In our context, need to account for dynamics

Setting

- Part D introduced in 2006, covering approximately 30M eligible individuals
- Government sets standard plan, but actual plans often provide different coverage
- Individuals eligible the month they turn 65, and then make plan choices prior to every calendar year

Data and sample

- 20% random sample of all Part D-covered individuals (2007 - 2009)
- Baseline sample is about one-quarter of full sample
  - Restrict attention to those 65+, not dual eligibles, not entitled to low income subsidies, in stand-alone PDPs

Sample	Full Sample	Baseline Sample
Obs. (beneficiary years)	16,036,236	3,898,247
Unique beneficiaries	6,208,076	1,689,308
Age	70.9 (13.3)	75.6 (7.7)
Female	0.60	0.65
Risk score	n/a	0.88 (0.34)

Spending patterns

Sample	Full Sample	Baseline Sample
Annual Total Spending		
Mean	2,433	1,888
Std. Deviation	4,065	2,675
Pct with no spending	7.35	5.65
25th pctile	378	487
Median	1,360	1,373
75th pctile	2,942	2,566
90th pctile	5,571	3,901
Annual Out of Pocket Spending		
Mean	418	778
Std. Deviation	744	968

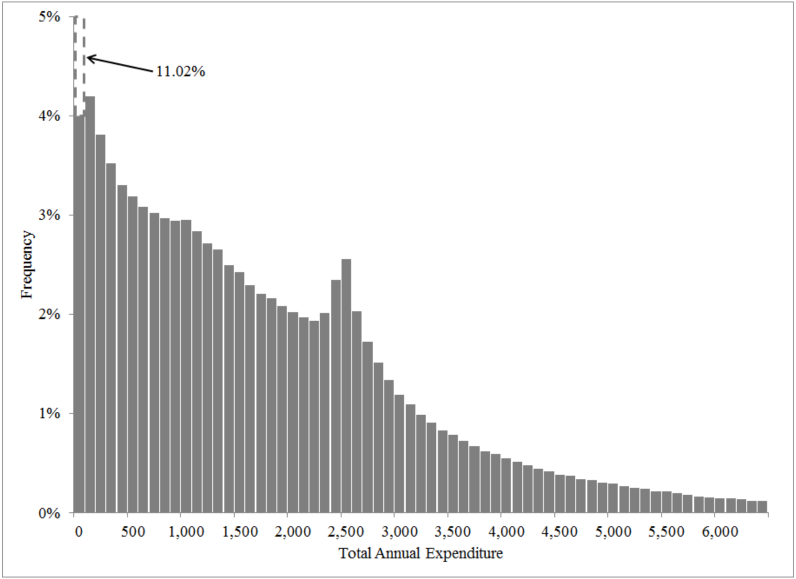
Actual contracts

Sample[a]	Bunching Sample		65 y.o. Sample	
	Deductible plans	No Ded. plans	Deductible plans	No Ded. plans
Obs. (beneficiary years)	1,038,228	2,922,570	28,958	111,516
Deductible Amount	265.9	0	257.1	0
Fraction w/ standard Ded.	XX	--	XX	--
Deductible Coins. Rate	0.88	--	0.85	--
Has standard ICL	1.00	0.98	1.00	0.97
ICL Amount	2,522.6	2,535.1	2,516.4	2,526.7
Pre-ICL Coins. Rate	0.26	0.37	0.27	0.37
Some Gap Coverage	0.01	0.17	0.00	0.12
Gap Coins. Rate	0.88	0.95	0.88	0.96
Gap Coins. Rate if some coverage	XX	XX	XX	XX
Catastrophic Amount	4,059.7	4,090.6	4,048.3	4,079.1
Catastrophic Conis. Rate	0.07	0.07	0.07	0.07

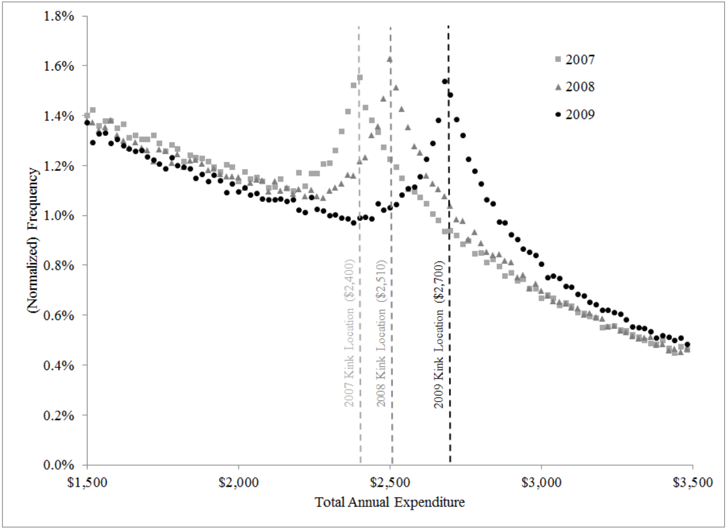
Static price response: bunching at the kink

- Sharp increase in price when go into donut hole
    - On average price rises from 34 to 93 cents for every dollar
- Standard economic theory: with convex preferences smoothly distributed in population, should see bunching at the convex kink

Bunching at kink I: 2008 spending distribution



Bunching II: year-to-year movement in kink

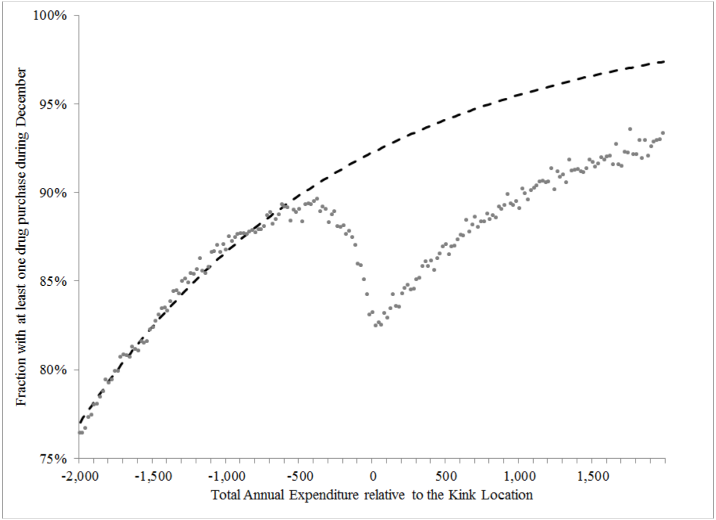


Heterogeneity across individuals

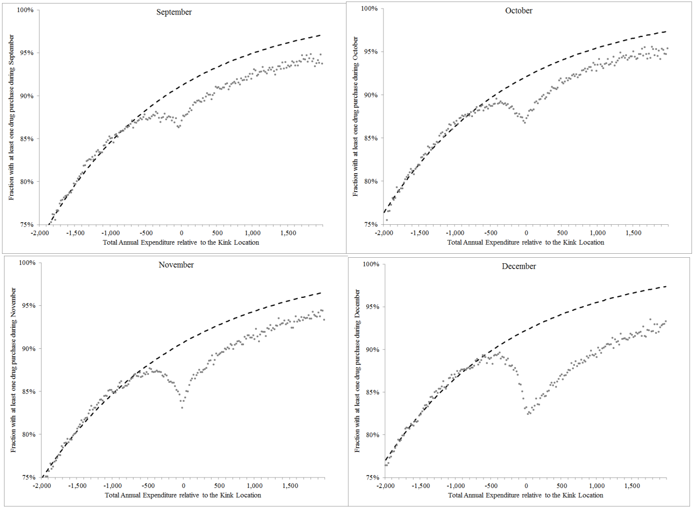
Population	Excess Mass
All	0.291
Year	
2006	0.088
2007	0.150
2008	0.213
2009	0.293
Gender	
Male	0.348
Female	0.262
Age group	
66	0.519
67	0.426
68-69	0.383
70-74	0.334
75-79	0.255
80-84	0.194
85+	0.136

Population	Excess Mass
Risk Score Quartile	
healthiest	0.448
less healthy	0.155
sicker	0.250
least healthy	0.346
Number of HCCs	
0	0.837
1	0.494
2	0.191
3	0.197
4	0.236
5+	0.316

Timing of purchases (December)



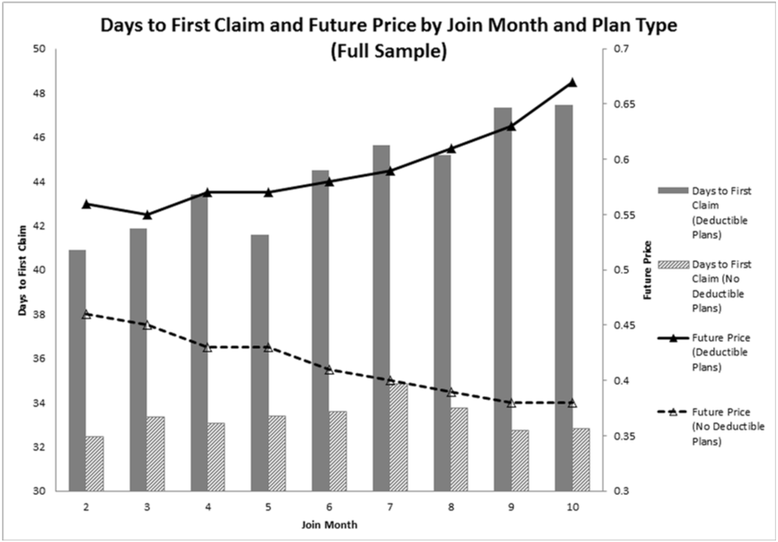
Timing of purchases (Sept - Dec)



Forward looking moral hazard

- Individuals become eligible when they turn 65; contract resets on January 1st
- Compare spending in first month of eligibility for people who turn 65 in February versus October – same spot price, but very different expected future prices

Forward looking moral hazard



Model 1

- Risk-neutral forward-looking individual faces uncertain health shocks
- Prescriptions are defined by (θ, ω),
  - $\theta > 0$ is the prescription's (total) cost
  - $\omega > 0$ is the monetized cost of not taking the drug
  - Arrive at weekly rate λ, drawn from $G(\theta, \omega) = G_2(\omega|\theta)G_1(\theta)$
  - λ follows a Markov process $H(\lambda|\lambda')$
- Insurance defines $c(\theta, x)$ – the out-of-pocket cost associated with a prescription that costs θ when total spending so far is x
- Individuals choose to fill each prescription or not
- Flow utility

$$u(\theta, \omega; x) = \begin{cases} -c(\theta, x) & \textit{if filled} \\ -\omega & \textit{if not filled} \end{cases}$$

Model 2

- Bellman equation given by

$$v(x, t, \lambda_{t+1}) = E_{\lambda|\lambda_{t+1}} \left[\begin{array}{c} (1-\lambda)\delta v(x, t-1, \lambda) + \\ \lambda \int \max \left\{ \begin{array}{l} -c(\theta, x) + \delta v(x+\theta, t-1, \lambda), \\ -\omega + \delta v(x, t-1, \lambda) \end{array} \right\} dG(\theta, \omega) \end{array} \right]$$

with terminal condition $v(x, 0) = 0$ for all x

Three key economic objects

- Statistical description of distribution of health shocks: λ and $G_1(\theta)$
- "Primitive" price elasticity capturing substitution between health and income: $G_2(\omega|\theta)$
  - If $\omega \geq \theta$, fill even if have to pay full cost
  - If $\omega < \theta$, fill only if some portion of cost (effectively) paid by insurer
  - Convenient to think about the ratio ω/θ
  - Loosely, identified off the bunching
- Extent to which individuals understand and respond to dynamic incentives in non-linear contract: $\delta \in [0, 1]$
  - "Full" myopia ($\delta = 0$): don't fill if $\omega < c(\theta, x)$
  - Dynamic response ($\delta > 0$): utilization depends on both spot and future price
  - δ is context specific! ... Captures salience, discounting, and perhaps liquidity constraints
  - Loosely, identified off the timing patterns

Parameterization 1

- $G_1(\theta)$ is lognormal: $\log\theta \sim N(\mu, \sigma^2)$.
- $\omega|\theta$ is stochastic, and is drawn from a mixture distribution:
  - $\omega \geq \theta$ with probability $1 - p$ (prescription is filled for sure)
  - $\omega \sim U[0, \theta]$ with probability p (decision responds to price)
- λ can obtain two values, and follows a 2-by-2 transition matrix
- Heterogeneity modeled using a finite (5 types) mixture:
  - Individual is of type m with probability $\pi_m = \exp\left(z_i'\beta_m\right) / \sum_{k=1}^{M} \exp\left(z_i'\beta_k\right)$
  - Almost all parameters vary with type: $\lambda_{m,low}$, μ_m, σ_m^2, p_m (exception: δ, $\lambda_{high}/\lambda_{low}$, $\lambda-$transition)
  - Baseline z_i: constant, risk score and 65-year-old indicator

Parameterization 2

- Allowing for heterogeneity in both individual health (λ, μ, σ) and in responsiveness of individual spending to cost-sharing (p)
- Do not use panel nature across years (although risk scores introduce some serial correlation within individuals across years)

Intuition for identification

Three key objects:

- Claim sizes θ distributions identified from observed claims (panel data allows identification despite unobserved types and selection, similar to Kasahara (2009), Hu and Shum (2012), and Sasaki (2012))
- Moral hazard (p's and λ's) identified from bunching described earlier.

Estimation 1

- Moments chosen based on identification intuition
  - Summary statistics (mean, standard deviation, $\mathbb{P}=0$) of total spending
  - Bunching: histogram of total spending near ICL
  - Timing patterns for individuals around and below the kink
- Calculate objective function using simulation
  1. Given parameters solve for value function using backward induction
     - Choose grid of values of x_g, set $v(x,0;m)=0 \forall x$
     - Given $v(x,t-1;m)$, calculate

$$v_{g,m} = (1-\lambda_m)\delta v(x_g, t-1) + \lambda \mathrm{E}\left[\max\left\{\begin{array}{r}-oop(\theta, x_j) + \delta v(x_g+\theta, t \\ -\omega + \delta v(x_j, t-1; m\end{array}\right.\right.$$

     - Set $v(x,t;m)=$ shape preserving cubic spline interpolation of $\{x_g, v_{g,m}\}$, repeat until maximum t

Estimation 2

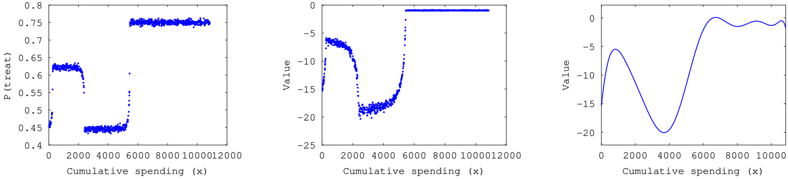
- Backward induction can amplify approximation errors so important to calculate $\mathrm{E}[\max]$ accurately (our distributional assumptions give an analytic expression conditional on θ and we use quadrature to integrate over θ) and a good interpolation method of v (polynomials, Fourier series, and non shape preserving splines fail spectacularly here; linear interpolation is okay)

2. Draw m, sequences of θ, ω and simulate the model
3. Compute moments from observed data and simulated data

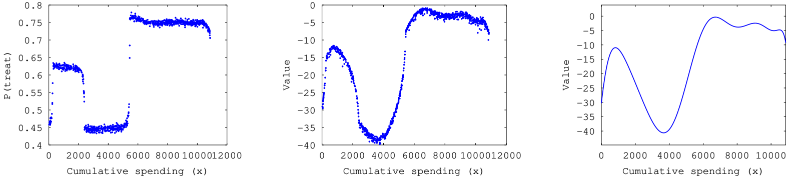
Poor approximation with polynomials

$\mathrm{P}(treat)$ | v_g | $v(x,t)$

$t = 1$



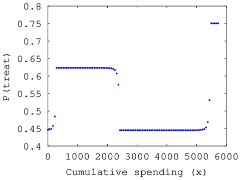
$t = 2$

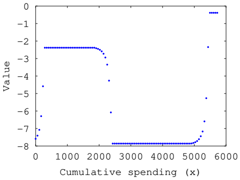


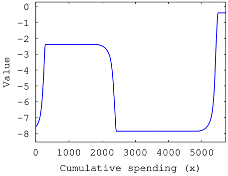
Value function with shape preserving cubic splines

$P(treat)$ | v_g | $v(x,t)$

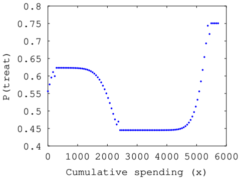
$t = 1$

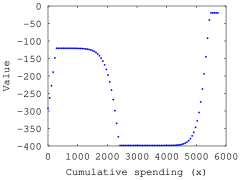


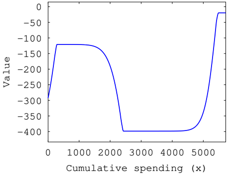




$t = 52$







Estimation 1

- Objective function minimized using CMA-ES
  - Uses quadratic approximation of objective to guide search, so converges more quickly than most random or global minimization algorithms (e.g. simulated annealing, genetic algorithms, pattern search)
  - Introduces randomness so able to escape local minima unlike deterministic algorithms (e.g. Nelder-Mead, (Quasi)-Newton, conjugate gradient)
  - Good performance, especially on non-convex problems compared to other algorithms

Parameter estimates

δ	0.961
	(0.00186)

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
β_0	0.00	3.59	3.98	-4.37	-4.35
	(0.000000)	(0.001427)	(0.001682)	(0.000696)	(0.000594)
β_{RS}	0.00	-2.46	-2.85	4.10	6.18
	(0.000000)	(0.001378)	(0.002090)	(0.000774)	(0.000296)
β_{65}	0.000	-0.101	1.336	0.926	-1.596
	(0.00e+00)	(1.17e-02)	(8.85e-04)	(1.53e-14)	(4.13e-15)
μ	-0.00292	3.99789	2.94797	4.31604	4.29602
	(1.44e-05)	(1.32e-02)	(8.75e-05)	(1.04e-02)	(1.04e-02)
σ	2.373	1.180	1.582	0.419	1.431
	(0.000114)	(0.010186)	(0.004952)	(0.003278)	(0.005838)
p_ω	0.859	0.902	0.495	0.505	0.374
	(8.02e-05)	(9.24e-03)	(3.26e-03)	(4.26e-03)	(1.45e-03)
λ	0.0114	0.1432	0.6300	0.8817	0.4490
	(8.65e-06)	(3.17e-04)	(2.35e-03)	(5.83e-03)	(1.25e-03)
λ	0.010	0.127	0.557	0.779	0.397
	(0.000051)	(0.000609)	(0.001905)	(0.004545)	(0.001467)

λ transition probabilities

0.5518941	0.4354298
(0.00180)	(0.00195)
0.4481059	0.5645702
(0.00180)	(0.00195)

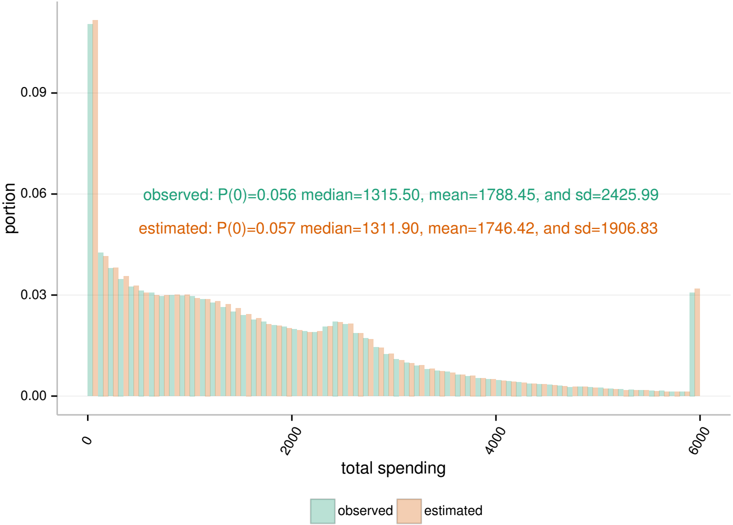
λ marginal probabilities

0.4928264	(0.00155)
0.5071736	(0.00155)

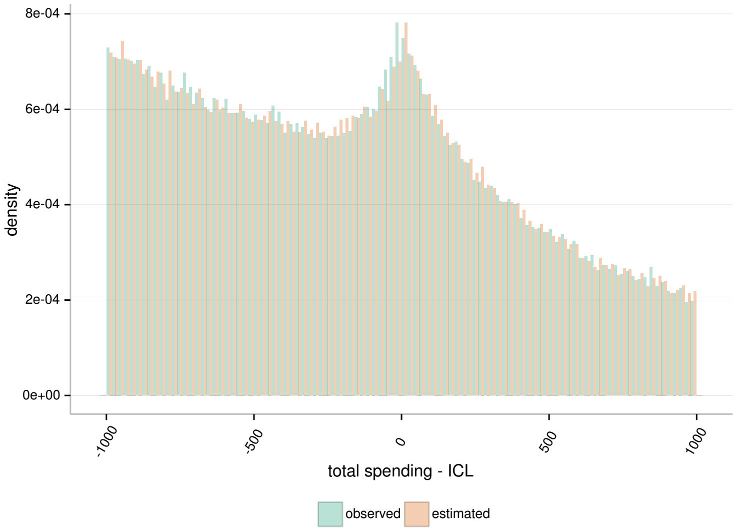
Parameter estimates: implied quantities

$\mathrm{P}(j)$	0.05	0.29	0.35	0.03	0.29
$\mathrm{P}(j \mid \text{age} = 65)$	0.00	0.14	0.86	0.00	0.00
$\mathrm{P}(j \mid \text{age} > 65)$	0.05	0.29	0.33	0.03	0.30
$\mathrm{E}\left[\frac{d\mathrm{P}(j \mid \text{rs})}{d\text{rs}}\right]$	0.01	-0.38	-0.51	0.06	0.83
$\mathrm{E}[\theta \mid j]$	16.65	109.36	66.65	81.76	204.42
$\text{s.d.}(\theta \mid j)$	278	190	223	36	531
$\mathrm{E}[\text{spend full ins.} \mid j]$	9.84	814.53	2183.38	3748.90	4772.61
$\mathrm{E}[\text{spend 0.25 coins.} \mid j]$	7.73	630.86	1913.16	3275.54	4326.28
$\mathrm{E}[\text{spend no ins.} \mid j]$	1.39	79.85	1102.50	1855.44	2987.30

Fit: Distribution of total costs

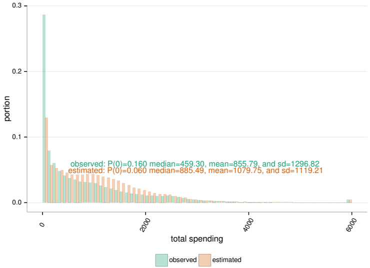


Fit: Total costs near ICL

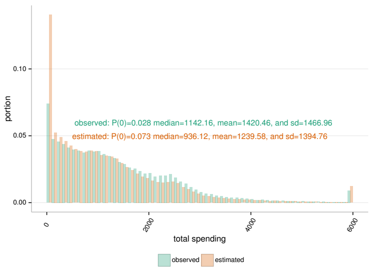


Fit: Risk score and total spending

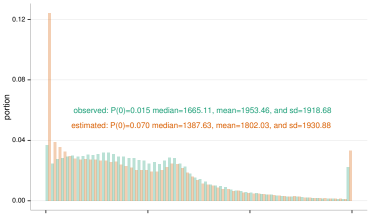
First Quartile



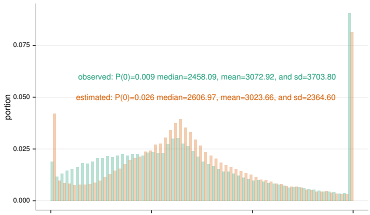
Second Quartile



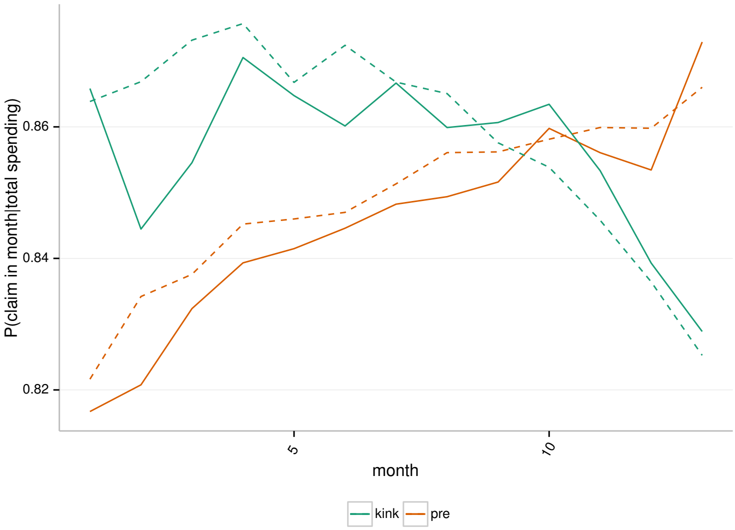
Third Quartile



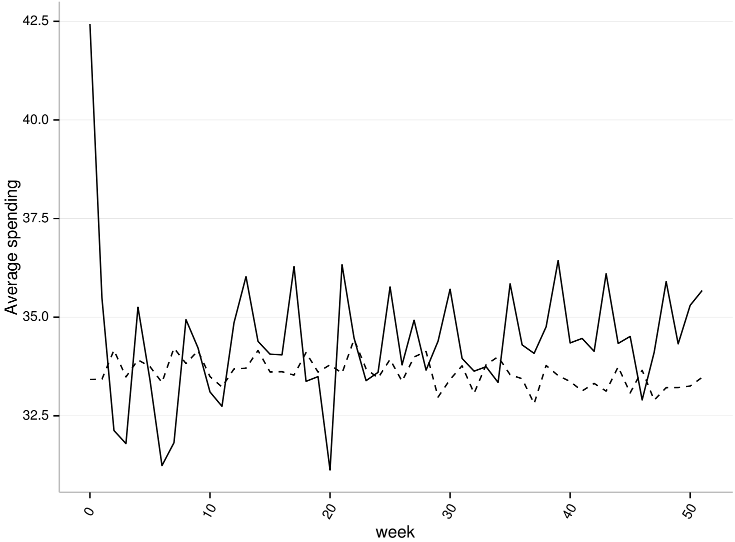
Fourth Quartile



Fit: Timing moments



Average weekly spending



Counterfactual: “filling the gap”

- Main counterfactual exercise considers “filling the gap” as specified by ACA by 2020:
  - Coinsurance rate in standard contract will remain at its pre-gap level (of 25%) until out of pocket spending puts individual at CCL
- First consider spending effect of “filling the gap” in the 2008 standard benefit design
  - On average, increases total spending by $204 (11.5%)

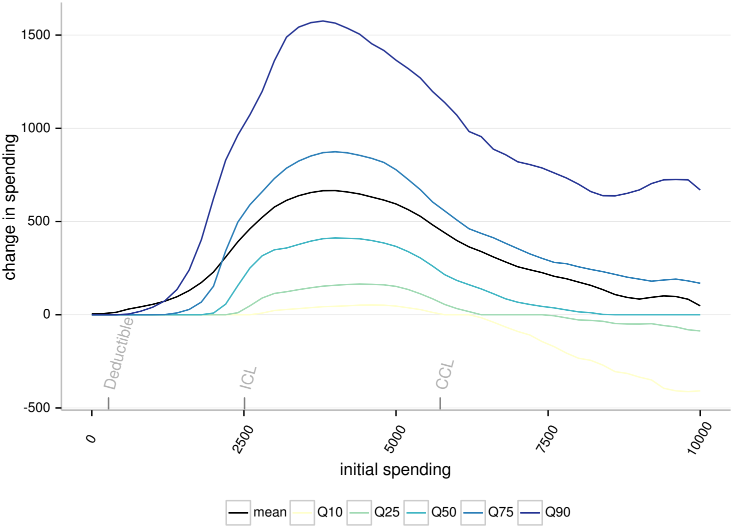
	mean	sd	Q25	Q50	Q75	Q90
baseline	1760	1924	402	1413	2513	3632
filled gap	1964	2127	407	1455	2862	4450

What has happened?

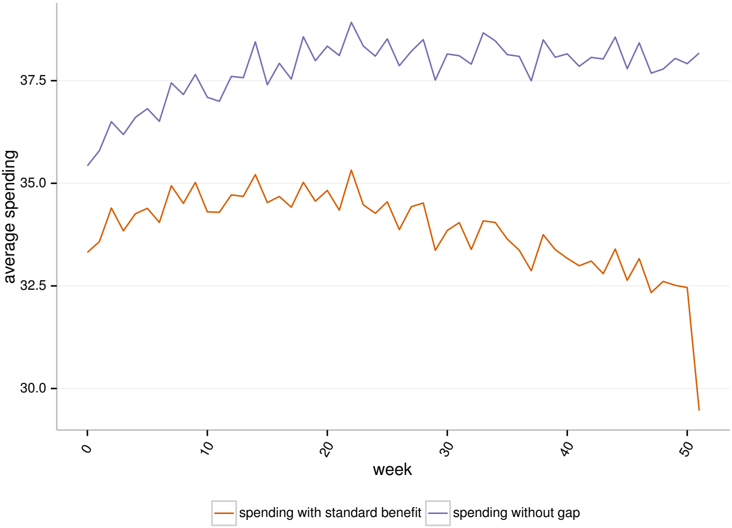
TODO: find data on spending per beneficiary by year (ideally leaving out subsidized beneficiaries, etc).

https://www.everycrsreport.com/reports/R40611.html

Counterfactuals: change in spending from filling gap



Counterfactuals: change in weekly spending from filling gap



Some subtle implications of non-linear contracts

- Change in spending by people far from gap / endogeneity of people at risk of bunching
  - Arises due to dynamic considerations
  - Estimate that about 25% of average $204/person increase in annual spending comes from ”anticipatory” response by people more than $200 below kink location
- “Filling” donut hole causes some people to *decrease* spending
  - CCL held constant with respect to out of pocket (vs. total) spending, so for some people marginal price actually rises
  - General point: with non-linear contracts, a given contract change can provide more coverage on margin to some individuals but less coverage to others

Policy-relevant counterfactuals

- Most people have more coverage than standard benefit
  - So effect will be lower (people have some gap coverage already)
- Analyze filling gap on existing contracts (ignore potential firm responses in contract design or beneficiary contract choice)
  - Filling gap increases spending by about $148 (8.5%)
  - But insurer (≈Medicare) spending rise by $253 (26%); absent behavioral response, insurer spending would increase by only $150

	Mean	Std. Dev.	25th pctile	Median	90th pctile	Mean OOP	Mean Insurer
Assign everyone to Standard 2008 contract:							
1 Baseline	1,760	1,924	402	1,413	3,632	809	951
2 "Filled" gap	1,964	2,127	407	1,455	4,450	655	1,309
Assign everyone observed (chosen) contract:							
3 Baseline	1,768	1,909	499	1,342	3,675	796	973
4 "Filled" gap	1,916	2,051	502	1,371	4,287	690	1,226

Robustness

- Results appear quite stable across a (limited) set of alternative specifications
- Explore sensitivity to:
  - Alternative number of discrete types (heterogeneity)
  - Set of covariates (e.g. gap coverage indicator as ”reduced form” way to capture potential plan selection)
  - Adding risk aversion via recursive utility
  - Letting ω vary more flexibly with θ

Robustness

TABLE VII

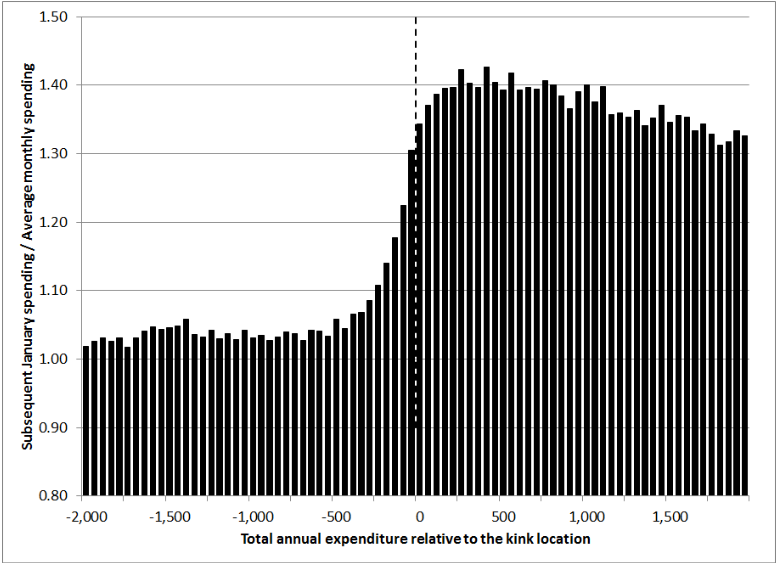
ROBUSTNESS OF THE KEY RESULTS TO VARIOUS MODELING ASSUMPTIONS

	Total spending			Insurer spending		
	Baseline	Filled gap	Change (%)	Baseline	Filled gap	Change (%)
(1) Baseline model	1,768	1,916	8.4	973	1,226	26.0
Number of types:						
(2) Three types	1,778	1,878	5.6	986	1,208	22.5
(3) Six types	1,760	1,937	10.1	967	1,240	28.2
Different sets of covariates:						
(4) Remove all covariates	1,783	1,892	6.1	976	1,209	23.9
(5) Add “no gap” covariate	1,758	1,946	10.7	968	1,246	28.7
Concave (risk averse) utility function:						
(6) CARA = exp(-6.5)	1,737	1,873	7.8	945	1,193	26.2
(7) CARA = exp(-9.5)	1,738	1,862	7.1	940	1,183	25.9
Distribution of ω:						
(8) ω correlated with θ	1,781	1,918	7.7	983	1,230	25.1

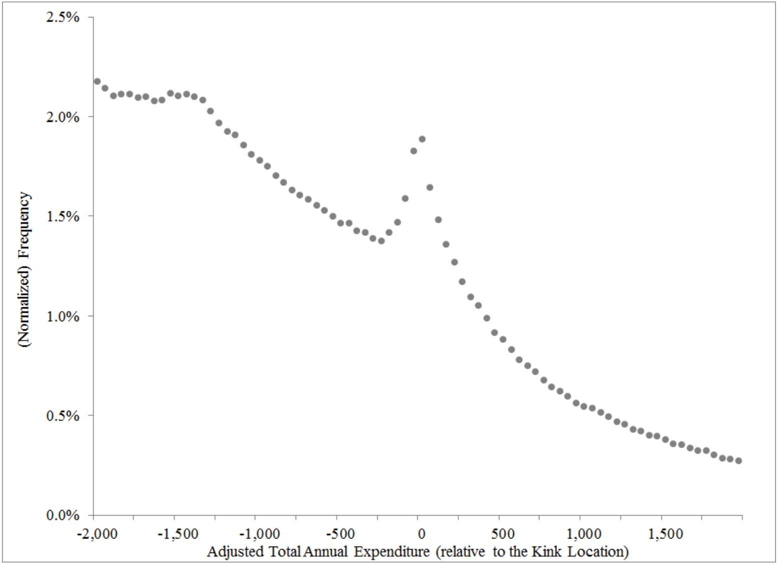
Cross-year substitution

- So far we treated each year of coverage in isolation
    - Typical in the literature
- But some of the effect could simply reflect substitution to next calendar year (as in Cabral, 2013)
- Consequences (for health and spending) may be less important if cross-year substitution is the main story

Evidence of cross-year substitution



But it does not explain all ...



Quantifying cross-year substitution

- Back of the envelope: January effect is $\approx$ \$40, so if applied to everyone would be about 25% of estimated average response
- Extend (or "tweak"!) the model by making terminal values depend on unfilled prescriptions, and allowing individual to fill in the beginning of next year
- Extension suggests overall effect may be reduced by 70%, from \$153 to \$44

		Mean	Std. Dev.	25th pctile	Median	90th pctile	Mean OOP	Mean Insurer
Assign everyone observed (chosen) contract:								
3	Baseline	1,765	1,874	502	1,341	3,684	796	969
4	"Filled" gap[a]	1,918	2,022	504	1,373	4,315	691	1,227
Assign everyone observed (chosen) contract, plus allow cros-year substitution:								
5	Baseline	1,770	1,912	501	1,337	3,749	807	963
6	"Filled" gap[a]	1,814	1,964	501	1,341	3,956	658	1,155

Section 3

Summary

Summary

- Spending response to non-linear health insurance contracts (vs ”an elasticity” with respect to ”the” price)
- Context: Medicare Part D (lots of current policy interest)
- Results:
  - “Filling the gap” (ACA) will increase Part D spending by about $150 per beneficiary (8.5%), and government spending by $250 (26%)
  - Importance of non-linearity of insurance contracts
    - Much of spending increase comes from ”anticipatory” behavior of individuals whose predicted spending is below the gap (would not be captured in static model)
  - A big part of the effect (but not all) could be explained by cross-year substitution

Normative analysis

- Focus of paper has been entirely positive
- Normative implications are more tricky
- Some of our findings regarding nature of response to kink may be useful for informally beginning to assess normative implications. e.g.,
  - Larger response by healthier individuals
  - Larger response of chronic (vs. acute) drugs
  - (Evidence on spillover effects to non-drug spending and health would also be useful)
- Conceptual question: optimality of drug consumption in absence of insurance?
  - Prices marked up above social MC (patents)
  - Failures of rationality may produce under-consumption w/o insurance?

Abbring, Jaap H. and Øystein Daljord. 2017. “Identifying the Discount Factor in Dynamic Discrete Choice Models.” URL https://faculty.chicagobooth.edu/oystein.daljord/research/papers/ddc.pdf.

———. 2019. “A Comment on ”Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting” by Hanming Fang and Yang Wang.”

Abbring, Jaap H., Øystein Daljord, and Fedor Iskhakov. 2018. “Identifying Present-Biased Discount Functions in Dynamic Discrete Choice Models .” URL https://www.semanticscholar.org/paper/Identifying-Present-Biased-Discount-Functions-in-%E2%88%97-Abbring-Daljord/c44616d28c151e4ddb0373fc331b818df02b1710.

An, Yonghong, Yingyao Hu, and Jian Ni. 2014. "Dynamic models with unobserved state variables and heterogeneity: time inconsistency in drug compliance." URL https://research.chicagobooth.edu/~/media/C321ABC7FAC5455886671CF64B84B82C.pdf.

Bajari, Patrick, Chenghuan Sean Chu, Denis Nekipelov, and Minjung Park. 2013. "A Dynamic Model of Subprime Mortgage Default: Estimation and Policy Implications." Working Paper 18850, National Bureau of Economic Research. URL http://www.nber.org/papers/w18850.

Dubé, Jean-Pierre, Günter J. Hitsch, and Pranav Jindal. 2014. "The Joint identification of utility and discount functions from stated choice data: An application to durable goods adoption." *Quantitative Marketing and Economics* 12 (4):331–377. URL https://doi.org/10.1007/s11129-014-9149-3.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D." *The Quarterly Journal of Economics* 130 (2):841–899. URL http://dx.doi.org/10.1093/qje/qjv005.

Fang, Hanming and Yang Wang. 2015. "Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions." *International Economic Review* 56 (2):565–596. URL http://dx.doi.org/10.1111/iere.12115.

Hu, Yingyao and Matthew Shum. 2012. "Nonparametric identification of dynamic models with unobserved state variables." *Journal of Econometrics* 171 (1):32 – 44. URL http://www.sciencedirect.com/science/article/pii/S0304407612001479.

Kasahara, Hiroyuki. 2009. "Temporary Increases in Tariffs and Investment: The Chilean Experience." *Journal of Business & Economic Statistics* 27 (1):113–127. URL http://amstat.tandfonline.com/doi/abs/10.1198/jbes.2009.0009.

Komarova, Tatiana, Fabio Sanches, Daniel Silva Junior, and Sorawoot Srisuma. 2018. "Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models." *Quantitative Economics* 9 (3):1153–1194. URL http://qeconomics.org/ojs/index.php/qe/article/view/618.

Magnac, Thierry and David Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes." *Econometrica* 70 (2):801–816. URL http://www.jstor.org.libproxy.mit.edu/stable/2692293.

Sasaki, Yuya. 2012. "Heterogeneity and selection in dynamic panel data." Working paper. URL http://www.sfu.ca/content/dam/sfu/economics/Documents/SeminarPapers/Sasaki_Nov_8_2012_seminar_paper.pdf.

Yao, Song, Carl F. Mela, Jeongwen Chiang, and Yuxin Chen. 2012. "Determining Consumers' Discount Rates with Field Studies." *Journal of Marketing Research* 49 (6):822–841. URL https://doi.org/10.1509/jmr.11.0009.