Optimal taxation and insurance using machine learning

Maximilian Kasy

Department of Economics, Harvard University

May 29, 2018

Introduction

- How to use (quasi-)experimental evidence when choosing policies, such as
  - tax rates,
  - health insurance copay,
  - unemployment benefit levels,
  - class sizes in schools, etc.?
- Answer in this paper: Maximize posterior expected welfare.
- Answer combines
  1. optimal policy theory (public finance),
  2. machine learning using Gaussian process priors.
- Application: coinsurance rates, RAND health insurance experiment.

Contrast with "sufficient statistic approach"

- Standard approach in public finance:
  1. Solve for optimal policy in terms of key behavioral elasticities at the optimum ("sufficient statistics").
  2. Plug in estimates of these elasticities,
  3. Estimates based on log − log regressions.
- Problems with this approach:
  1. Uncertainty: Optimal policy is nonlinear function of elasticities. Sampling variation therefore induces systematic bias.
  2. Relevant dependent variable is expected tax base, not expected log tax base.
  3. Elasticities are not constant over range of policies.
- Posterior expected welfare based on nonparametric priors addresses these problems.
- Tractable closed form expressions available.

Optimal insurance and taxation

- (Baily, 1978; Saez, 2001; Chetty, 2006)
- Example: Health insurance copay.
- Individuals i, with
  - Y_i health care expenditures,
  - T_i share of health care expenditures covered by the insurance,
  - $1 - T_i$ coinsurance rate,
  - $Y_i \cdot (1 - T_i)$ out-of-pocket expenditures.
- Behavioral response:
  - Individual: $Y_i = g(T_i, \varepsilon_i)$.
  - Average expenditures given coinsurance rate: $m(t) = E[g(t, \varepsilon_i)]$.
- Policy objective:
  - Weighted average utility, subject to government budget constraint.
  - Relative value of \$ for the sick: λ.
  - Marginal change of $t \rightarrow$ mechanical and behavioral effects.

Social welfare

- Effect of marginal change of t:
  - Mechanical effect on insurance budget: $-m(t)$
  - Behavioral effect on insurance budget: $-t \cdot m'(t)$
  - Mechanical effect on utility of the insured: $\lambda \cdot m(t)$
  - Behavioral effect on utility of the insured: 0
    By envelope theorem (key assumption: utility maximization)
- Summing components:

$$u'(t) = (\lambda - 1) \cdot m(t) - t \cdot m'(t).$$

- Integrate, normalize $u(0) = 0$ to get social welfare:

$$u(t) = \lambda \int_0^t m(x)dx - t \cdot m(t).$$

Experimental variation, GP prior

- n i.i.d. draws of (Y_i, T_i), T_i independent of ε_i
- Thus

$$E[Y_i|T_i = t] = E[g(t, \varepsilon_i)|T_i = t] = E[g(t, \varepsilon_i)] = m(t).$$

- Auxiliary assumption: normality, $Y_i|T_i = t \sim N(m(t), \sigma^2)$.
- Gaussian process prior:

$$m(\cdot) \sim GP(\mu(\cdot), C(\cdot, \cdot)).$$

- Read: $E[m(t)] = \mu(t)$ and $\mathrm{Cov}(m(t), m(t')) = C(t, t')$.

Posterior

- Denote $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, $\boldsymbol{T} = (T_1, \ldots, T_n)$,

$$\mu_i = \mu(T_i), \qquad C_{i,j} = C(T_i, T_j), \qquad C_i(t) = C(t, T_i).$$

- μ, $\boldsymbol{C}(t)$, and $\boldsymbol{C}$: vectors and matrix collecting these terms.
- Posterior expectation of $m(t)$:

$$\begin{aligned}\widehat{m}(t) &= E[m(t)|\boldsymbol{Y}, \boldsymbol{T}] \\ &= E[m(t)|\boldsymbol{T}] + \mathrm{Cov}(m(t), \boldsymbol{Y}|\boldsymbol{T}) \cdot \mathrm{Var}(\boldsymbol{Y}|\boldsymbol{T})^{-1} \cdot (\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{T}]) \\ &= \mu(t) + \boldsymbol{C}(t) \cdot \left[\boldsymbol{C} + \sigma^2 \boldsymbol{I}\right]^{-1} \cdot (\boldsymbol{Y} - \mu).\end{aligned}$$

Posterior expected welfare

- Recall: $u(t)$ is a linear functional of $m(\cdot)$,

$$u(t) = \lambda \int_0^t m(x)dx - t \cdot m(t).$$

- Thus:

$$v(t) = E[u(t)] = \lambda \int_0^t \mu(x)dx - t \cdot \mu(t),$$ and

$$D(t,t') = \mathrm{Cov}(u(t), m(t'))) = \lambda \cdot \int_0^t C(x,t')dx - t \cdot C(t,t').$$

- Notation: $\boldsymbol{D}(t) = \mathrm{Cov}(u(t), \boldsymbol{Y}|\boldsymbol{T}) = (D(t,T_1), \ldots, D(t,T_n))$

- Posterior expected welfare:

$$\widehat{u}(t) = E[u(t)|\boldsymbol{Y}, \boldsymbol{T}] = \nu(t) + \boldsymbol{D}(t) \cdot \left[\boldsymbol{C} + \sigma^2 \boldsymbol{I}\right]^{-1} \cdot (\boldsymbol{Y} - \mu).$$

- Derivative:

$$\frac{\partial}{\partial t}\widehat{u}(t) = \nu'(t) + \boldsymbol{B}(t) \cdot \left[\boldsymbol{C} + \sigma^2 \boldsymbol{I}\right]^{-1} \cdot (\boldsymbol{Y} - \mu)$$

where

$$B(t,t') = \frac{\partial}{\partial t} D(t,t') = (\lambda - 1) \cdot C(t,t') - t \cdot \frac{\partial}{\partial t} C(t,t').$$

- Bayesian policymaker maximizes posterior expected welfare:

$$\widehat{t^*} = \widehat{t^*}(\boldsymbol{Y}, \boldsymbol{T}) \in \underset{t}{\operatorname{argmax}}\ \widehat{u}(t).$$

- First order condition:

$$\frac{\partial}{\partial t}\widehat{u}(\widehat{t^*}) = E[u'(\widehat{t^*})|\boldsymbol{Y}, \boldsymbol{T}] = \nu'(\widehat{t^*}) + \boldsymbol{B}(\widehat{t^*}) \cdot \left[\boldsymbol{C} + \sigma^2 \boldsymbol{I}\right]^{-1} = 0.$$

Prior specification, covariates

- Choice of covariance kernel:
  Squared-exponential, plus diffuse linear trend (popular in ML).

$$C(t_1, t_2) = v_0 + v_1 \cdot t_1 t_2 + \exp\left(-|t_1 - t_2|^2/(2l)\right).$$

- Covariates and conditional independence:
  - If exogeneity holds only conditional on covariates or control functions, then $T_i \perp \varepsilon_i | W_i$
  - Extend above analysis for $k(t, w) = E[Y|T = t, W = w]$.
  - Gaussian process prior for $k(t, w)$.
  - Dirichlet prior for P_W.

Application: The RAND health insurance experiment

- Cf. Aron-Dine et al. (2013).
- Between 1974 and 1981,
  representative sample of 2000 households,
  in six locations across the US.
- Families randomly assigned to
  plans with one of six consumer coinsurance rates.
- 95, 50, 25, or 0 percent,
  2 more complicated plans (I drop those).
- Additionally: randomized Maximum Dollar Expenditure limits,
  5, 10, or 15 percent of family income,
  up to a maximum of \$750 or \$1,000.
  (I pool across those.)

Table: Expected spending for different coinsurance rates

	(1) Share with any	(2) Spending in $	(3) Share with any	**(4) Spending in $**
Free Care	0.931	2166.1	0.932	2173.9
	(0.006)	(78.76)	(0.006)	(72.06)
25% Coinsurance	0.853	1535.9	0.852	1580.1
	(0.013)	(130.5)	(0.012)	(115.2)
50% Coinsurance	0.832	1590.7	0.826	1634.1
	(0.018)	(273.7)	(0.016)	(279.6)
95% Coinsurance	0.808	1691.6	0.810	1639.2
	(0.011)	(95.40)	(0.009)	(88.48)
family x month x site fixed effects	X	X	X	X
covariates			X	X
N	14777	14777	14777	14777

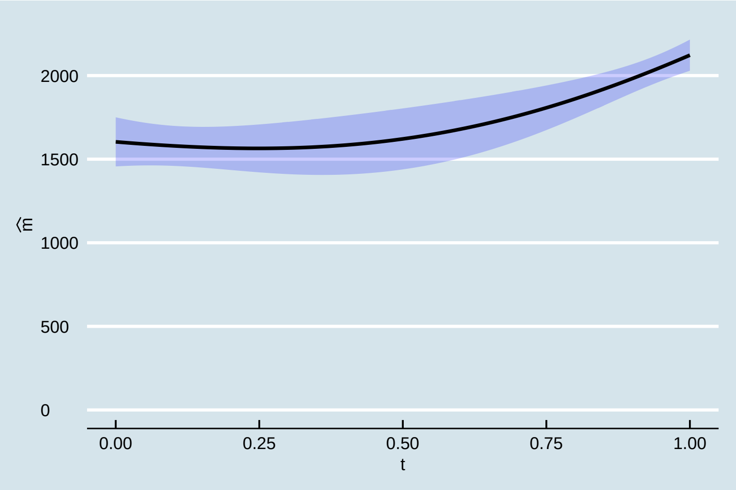
Assumptions

1. **Model**: The optimal insurance model as presented before
2. **Prior**: Gaussian process prior for *m*, squared exponential in distance, uninformative about level and slope
3. **Relative value** of funds for sick people vs contributors: $\lambda = 1.5$
4. Pooling data: across levels of maximum dollar expenditure

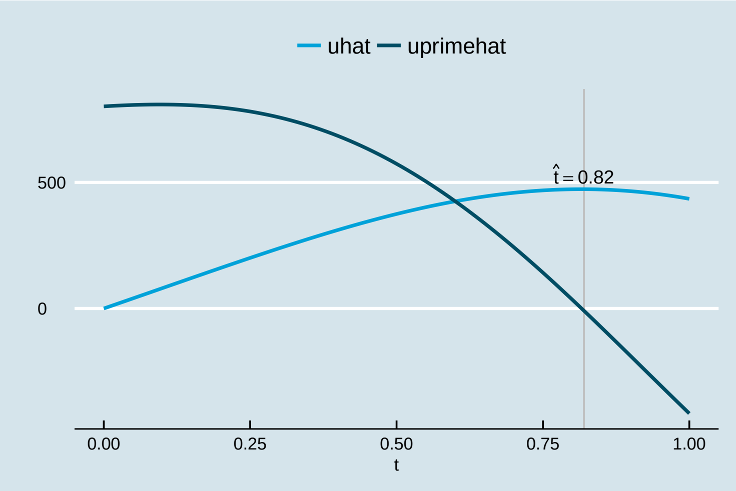
Under these assumptions we find:

Optimal copay equals 18%
(But free care is almost as good)

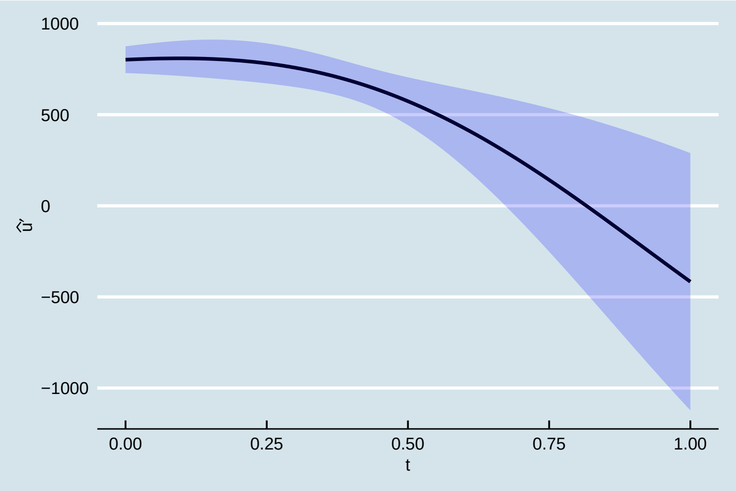
Posterior for *m* with confidence band



Posterior expected welfare and optimal policy choice



Confidence band for u' and t^*



Thank you!