Journal of Public Economics 167 (2018) 205–219

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Optimal taxation and insurance using machine learning — Sufficient statistics and beyond[☆]

Maximilian Kasy

Department of Economics, Harvard University, Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138, United States of America

ARTICLE INFO

Article history:
Received 2 April 2018
Received in revised form 19 July 2018
Accepted 4 September 2018
Available online xxxx

JEL classification:
H21
C11
C14

Keywords:
Optimal policy
Gaussian process priors
Posterior expected welfare

ABSTRACT

How should one use (quasi-)experimental evidence when choosing policies such as tax rates, health insurance copay, unemployment benefit levels, and class sizes in schools? This paper suggests an approach based on maximizing posterior expected social welfare, combining insights from (i) optimal policy theory as developed in the field of public finance, and (ii) machine learning using Gaussian process priors. We provide explicit formulas for posterior expected social welfare and optimal policies in a wide class of policy problems.

The proposed methods are applied to the choice of coinsurance rates in health insurance, using data from the RAND health insurance experiment. The key trade-off in this setting is between transfers toward the sick and insurance costs. The key empirical relationship the policy maker needs to learn about is the response of health care expenditures to coinsurance rates. Holding the economic model and distributive preferences constant, we obtain much smaller point estimates of the optimal coinsurance rate (18% vs. 50%) when applying our estimation method instead of the conventional "sufficient statistic" approach.

1. Introduction

How should empirical evidence be used to determine the optimal level of policy parameters such as tax rates, unemployment benefits, health insurance copay, and class sizes in school? A standard approach, labeled the "sufficient statistics approach" by Chetty (2009), uses the data to estimate a key behavioral elasticity, and then plugs this elasticity into formulas for optimal policy levels that are based on elasticities at the optimum. In this paper, an alternative approach is proposed and implemented in the context of choosing coinsurance rates for health insurance.[1]

1.1. Setup

This paper takes the perspective of a policy maker who wants to maximize some notion of social welfare. We assume that the policy maker observes (quasi-)experimental data that allow her to learn about some behavioral relationship that is relevant for her decision. We assume further that the policy maker acts as a Bayesian decision maker. This assumption implies that she uses the available data to form a posterior expectation of social welfare given each possible policy choice, and that she chooses the policy that maximizes this posterior expectation.

The imposition of some additional structure allows us to derive explicit analytic solutions to the policy maker's problem. In Section 2, we assume that social welfare takes a form common to many problems in public finance, where the key trade-off is between a weighted sum of private utilities and public revenues. The empirical relationship that the policy maker needs to learn in these settings is the response of the tax base to tax rates, or of insurance claims to coinsurance rates. In Section 3, we consider Gaussian process priors for this behavioral relationship. The combination of the structure of the objective function and the structure of these priors implies that we can explicitly derive and characterize posterior expected social welfare. In contrast to the sufficient statistics method as discussed in Chetty (2009), our approach does not rely on extrapolation

☆ I thank Hunt Allcott, Gary Chamberlain, Raj Chetty, Ellora Derenoncourt, Bryan Graham, Danielle Li, Nathan Hendren, Michael Kremer, José Luis Montiel Olea, John Rust, Jann Spiess, Matt Taddy, as well as seminar participants at Stanford, UC Berkeley, Georgetown, and Brown for helpful discussions and comments. This work was supported by NSF grant SES-1354144 "Statistical decisions and policy choice."
E-mail address: maximiliankasy@fas.harvard.edu.

[1] The coinsurance rate is the share of health care expenditures that the insured have to pay out of pocket.

https://doi.org/10.1016/j.jpubeco.2018.09.002
0047-2727/

using constant elasticity functional form assumptions,[2] and it takes uncertainty into account. The difference matters in practice, as we will see.

1.2. Contributions of this paper

This paper contributes to the literature in several ways. First, for empirical researchers working on issues of public policy, this paper leverages the statistical insights of a well developed literature on machine learning using Gaussian process priors, spline regression, and reproducing kernel Hilbert spaces. This paper provides a simple framework to derive optimal policy choices given available data. The practical relevance of such a framework is demonstrated by our empirical application, where we find very different levels of optimal policy relative to those suggested by a conventional estimation approach (leaving the economic model and distributive preferences the same). Second, for statistical decision theorists, this paper suggests a class of objective functions ("loss functions") for statistical decision problems that have a substantive justification in economic theory, and which contrast with conventional loss functions such as quadratic error loss and mis-classification loss. Third, for practitioners of machine learning, this paper suggests a class of applications of machine learning methods where new predictive procedures might fruitfully be leveraged for problems other than prediction.

1.3. Application

In Section 4, the proposed approach is applied to the problem of setting coinsurance rates in health insurance. Lowering coinsurance leads to more redistribution from healthy contributors to those in need of health care. However, it also increases insurance costs, both mechanically and through the behavioral response of possibly increased health care spending. We use data from the RAND health insurance experiment in order to estimate this behavioral response. We then use the estimated relationship to determine the optimal coinsurance rate. We find an optimal coinsurance rate of 18%. This contrasts markedly with the optimal coinsurance rate of 50% suggested by the conventional sufficient statistics approach under otherwise identical assumptions. Both of these numbers are based on the (arbitrary) normative assumption that the marginal value of a US$ for the sick is 1.5 times the marginal value of a US$ for the insurance provider.[3] For a range of alternative assumptions about this relative marginal value, we find the same qualitative comparison. The expected welfare loss per capita of using the sufficient statistics plug-in approach, and thus a coinsurance rate of 50% rather than the optimal 18%, is equal to 98 US$. For a hypothetical population of one million insurees, using the plug-in approach would thus result in a welfare loss of almost 100,000,000 US$.

1.4. When the difference to sufficient statistics matters most

The approach proposed here yields the same answer as the sufficient statistics approach under three conditions: (i) The sample is very large so that estimation uncertainty is negligible, (ii) the functional form imposed to estimate sufficient statistics (e.g., linearity of average log expenditures in the log coinsurance rate) is correctly specified, and (iii) the residuals of the regression used to estimate sufficient statistics are homoskedastic. When these conditions are violated, the estimated optimal policies can differ substantially.

Condition (i) might not matter much for estimates based on IRS data, say, but is more salient for estimates based on experimental data. Condition (ii) might be less of an issue when the optimal policy lies inside the observed range of policy levels, because misspecifications are more easily diagnosed in this case. This condition is however very important when the optimal policy lies near the boundary or outside the observed range. Condition (iii) presumably matters in most settings. Violations of all three conditions explain the difference of estimated optimal policy levels in the health insurance application.

In Section 5, we discuss these conditions in detail, and make the case that our approach is preferred when the conclusions differ. There are strong normative arguments for the expected welfare (i.e., Bayesian) approach for (policy) decision making under uncertainty. This differs notably from other statistical problems where the main goal is interpersonal replicability, and where frequentist approaches might be preferred. Lastly, not relying on functional form assumptions is key since generically such assumptions will be violated, distorting policy decisions when imposed.

1.5. Literature

This paper draws on two distinct literatures, (i) optimal policy theory as discussed in the field of public finance, and (ii) statistical decision theory and machine learning using Gaussian process priors. Models of optimal policy in public finance have a long tradition going back at least to the discussion in Samuelson (1947) of social welfare functions, with classic contributions including Mirrlees (1971) and Baily (1978). The empirical implementation of such models using "sufficient statistics" is discussed in Chetty (2009) and Saez (2001). Gaussian process priors and nonparametric Bayesian function estimation are discussed extensively in Williams and Rasmussen (2006). Gaussian process priors are closely related to spline estimation and reproducing kernel Hilbert spaces, as discussed in Wahba (1990). When controlling for covariates, we also make use of Dirichlet process priors, which are reviewed in Ghosh and Ramamoorthi (2003). The related problem of assigning treatment optimally, maximizing the posterior expectation of average observed outcomes, has been considered in Dehejia (2005) and Chamberlain (2011).

1.6. Road map

The rest of this paper is structured as follows. Section 2 briefly reviews the theory of optimal insurance and optimal taxation, and reformulates the solution to these problems in a form amenable to our approach. Section 3 states our assumptions on the data generating process and the prior. We then derive simple closed form expressions for posterior expected social welfare and for the first order condition characterizing the optimal policy choice. Section 4 applies the proposed approach to data from the RAND health insurance experiment and provides estimates of the optimal coinsurance rate. Section 5 provides an extended discussion comparing our proposed approach to the sufficient statistics approach. Section 6 discusses a number of extensions of our framework, including conditional exogeneity, optimal experimental design for policy, and an alternative class of social welfare functions involving production. Section 7 concludes. The appendix discusses technical details, including the envelope theorem, a generalization of our setup involving affine operators, additional models of optimal taxation covered by our framework, explicit weight functions for our application, approximations using equivalent kernel weights, and numerical examples comparing our approach to the sufficient statistics approach. Code implementing the proposed methods and replicating the figures in this paper is available at https://github.com/maxkasy/optimaltaxationusingML.

[2] We allow for arbitrary (smooth) variation of elasticities across policy levels. Optimal tax theory does not restrict us to assume elasticities are constant. The difficulties involved in interpolation and extrapolation relying on an assumption of constant elasticities have been recognized in the literature, of course. While contributions such as Gruber (1997) do calculate globally optimal policies, more recent papers often prefer to only evaluate marginal deviations from the status quo, to avoid undue extrapolation.

[3] The choice of such welfare weights based on normative considerations is discussed in Saez and Stantcheva (2016).