Expanding Health Insurance with Mandate and Subsidy: Theory and Evidence from Massachusetts

Hongming Wang*

May 2019

Abstract

What is the proper scope of social insurance, and what motivates government mandate and subsidization of health insurance? This paper explores two rationales: adverse selection in insurance premium, and the social cost of uncompensated care. I assess both rationales as potential justification of the 2006-2007 insurance expansion in Massachusetts. I derive and calculate the motivating benefits relative to the cost of expanding insurance with policy incentives. I find adverse selection alone can justify the mandate penalty in this context, and the social cost of uncompensated care justifies the subsidy generosity with small-to-zero premium benefit. Incremental expansion is desirable from a pure efficiency standpoint, and becomes more desirable with equity.

Keywords: penalty, subsidy, public finance, adverse selection, uncompensated care

JEL Codes: I11, I13, I18

*Department of Economics, University of Southern California. Email: hongminw@usc.edu

1 Introduction

Governments worldwide invest large amounts of resources in the health, education, and well-being of its citizens. In the United States, health care spending exceeds 18% of GDP in 2017. The 2010 Affordable Care Act expanded health insurance coverage with premium subsidy and mandate penalty, which is expected to increase health care spending to 20% of GDP by 2020 (Keehan *et al.*, 2011). In contrast, the benefits and beneficiaries of insurance expansion, and the motivation they present for government mandate and subsidization of health insurance relative to cost, remain less understood.

The canonical motivation for government intervention in the insurance market is adverse selection (Akerlof 1970; Rothschild and Stiglitz 1976). Selection drives out low-cost individuals and increases insurance premium. In private insurance markets, welfare loss from the pricing inefficiency alone tends to be small, however, and does not fully justify government subsidy of premium (Einav *et al.*, 2010). Moreover, while information asymmetry can potentially explain the lack of private insurance for certain risks (Hendren, 2017), the sweeping nature of a mandate also limits the empirical variation useful for understanding the role of adverse selection, if any, in these contexts.

A second motivation recognizes the fact that the uninsured do not bear the full financial risk of health care expenditure; much of the cost is borne by third-party payers (Finkelstein *et al.*, 2015). When the government cannot pre-commit to not providing care to the uninsured—effectively implementing an implicit social insurance mandate, demand for formal insurance falls below own expected cost imposed on society (Finkelstein *et al.*, 2017). Expanding formal insurance lowers the social cost of uncompensated care, and depending on the distribution of the cost across economy, may outperform the implicit safety-net coverage in the public finance of the mandate.

This paper evaluates both inefficiencies in determining the desirable scope of social insurance. I first examine the motivations they present for a universal insurance mandate, under the stylized setting of perfect competition and zero behavioral responses. In this case, the pricing externalities are only subject to the resource cost of insurance. Adverse selection alone presents weak justification for a mandate, unless full redistribution across risk in addition to income is desirable (Blomqvist and Horn, 1984). Formalizing an implicit safety-net mandate with tax-financed subsidy always improves welfare, if the tax base achieves better distribution of healthcare costs than the risk. In lieu of uncompensated care, a subsidized universal mandate maximizes welfare with equity.

Although both adverse selection and uncompensated care potentially motivate an insurance mandate, I assess their empirical relevance in rationalizing the 2006-2007 insurance expansion in Massachusetts. I focus on two policy instruments that expanded formal insurance in the state: subsidy of insurance premium, and tax penalty on uninsured individuals ineligible for subsidy. I develop a welfare framework where the incentive effect on take-up, the selection effect on cost, and the resulting pricing effect on premium versus the surcharge burden of uncompensated care, provide sufficient statistics to characterize and compare the motivating benefits. The net benefit relative to cost indicates the desirability of incremental expansion of insurance using policy generosities.

The cost-benefit framework traces out the social externality of policy incentives across economy. The take-up response changes the cost composition in formal insurance and uncompensated care. Adverse selection implies expanding formal insurance lowers the average cost in both. The cost change is passed onto insurance premium and the charity surcharge, according to premium regulations and the public finance of charity care. In practice, I consider a range of pass-through from costs to prices. In particular, I shut down the effect on either premium or charity surcharge to examine the relevant rationale across expansion groups.

The flexible range of pass-through nests behavioral responses from insurers and enrollees to policy incentives. Insurer capture of policy generosity weakens the link between cost and premium. I calibrate the mark-up response to subsidy over costs using estimates in Jaffe and Shepard (2018). Spending increase in formal insurance lowers the net saving in uncompensated care. The spending increase is approximately 25% based on moral hazard estimates in Massachusetts (Chandra *et al.*, 2011), although I consider larger spending responses across contexts. The pricing benefits are subject to the cost of expanding insurance using policy incentives, or the fiscal externality on the government budget. With price-linked subsidy and penalty, the fiscal cost decreases with the premium benefit.

Although behavioral responses to subsidies and taxes are well examined in the literature, hence allowing for a pure calibration exercise on welfare, I estimate incentive effects specific to the Massachusetts expansion for non-elderly adults in the American Community Survey (ACS). The estimates serve two purposes. First, they show that incentive effects in Massachusetts fall within the range of existing evidence under similar contexts. The cost-benefit calculation provides informative bounds on the potential net benefit that applies in Massachusetts. Second, they reveal heterogeneous effects across sub-groups.

I exploit differences in subsidy generosity across income groups, within rating communities where premium does not vary by individuals. Given premium, lower-income individuals are eligible for subsidy at a higher percent of premium. Following standard practice, I correct for the selection response to subsidy with simulated generosity from a pre-reform reference sample unaffected by the subsidy (Currie and Gruber 1996a; Currie and Gruber 1996b). The simulation isolates generosity differences across demographic groups within rating communities. I use the simulated measure to instrument for the endogenous subsidy exposure of Massachusetts enrollees. The cross-group variation in subsidy (and hence premium) has been widely applied to estimate demand in the individual market (Tebaldi, 2017), and the incentive effect on take-up (Frean *et al.*, 2017).

Based on the simulated measure, increasing subsidy generosity by 10 percentage points (above the 70% baseline) is estimated to increase take-up by 1 percentage point over 2008-2011, and by 1.7 percentage point in 2011. The magnitude aligns well with the take-up response around subsidy discontinuities in Massachusetts (Finkelstein *et al.*, 2017). Employment response to subsidy is indistinguishable from zero, but increases significantly with subsidy in the near-elderly (55-64). Take-up increases the most with subsidy at younger ages (below 30). In both groups, coverage by employer-sponsored insurance (ESI) decreases with subsidy. By contrast, estimates based on the endogenous exposure measure are wrong-signed for the effect on take-up, and indicate substantially larger non-employment effect than the existing literature.

Combining behavioral responses with the composition effect on costs, I calculate the pricing benefits of a dollar increase in policy incentives. The trade-off with the fiscal externality evaluates the desirability of further expansion using stronger incentives. I approach the trade-off first from a pure efficiency standpoint: expanding formal insurance alleviates the adverse selection in premium and the social cost of uncompensated care, but does not serve redistribution purposes through the tax-subsidy incentives or the resulting effects on prices. In this case, net benefit from incremental expansion is close to zero for the range of estimates considered. If only the pricing inefficiencies are considered, then at 95% insurance rate in Massachusetts, benefits from additional take-up do not recover the cost of expansion with more generous incentives.

Between the two motivations, uncompensated cost saving amounts to 7 cents per dollar increase in subsidy. Allowing for substantial mark-up capture, the premium benefit that is around 2 cents on a dollar of subsidy (or 30% of the full benefit without mark-up capture) would still recover the fiscal externality on pure efficiency grounds. Accounting for redistribution preferences for the low-income, the direct benefit on subsidy recipients increases, lowering the fiscal externality term. Expanding subsidy generosity and take-up in the low-income group becomes desirable even with a small dosage of equity consideration.

By contrast, the mandate penalty is primarily motivated by the selection effect on premium. Enrollees responsive to the penalty are 70% less costly than existing enrollees in unsubsidized individual plans. The potential premium benefit, absent any mark-up capture, is around 2.6 cents on a dollar of penalty. Avoided uncompensated cost is 1 cent on a dollar, driven by high insurance rate and low uninsured cost in the unsubsidized population. Government collects little revenue from the penalty due to the take-up response. Shutting down the access to uncompensated care in this population, premium benefit alone recovers the cost of penalty on the uninsured, if the mark-up capture is close to zero. Factoring in uncompensated care benefits and equity, desirability of a higher tax penalty depends on the relative welfare weights between the uninsured and the insured patients bearing excess burden, and is less clear-cut than the subsidy transfer to the low-income. In general, the efficiency argument for the penalty decreases with the premium benefit, and the equity argument weakens with smaller social cost of uncompensated care.

Earlier studies have separately identified the benefit to uncompensated payers in the subsidized market (Finkelstein *et al.*, 2017), and the selection effect on premium in the individual market (Hackmann *et al.*, 2015). This paper develops a framework to understand the desirable scope of social insurance, motivating both benefits as potential rationales. I then empirically assess the welfare implications using incremental expansion in Massachusetts. Quantified welfare effects reveal the desirability of expanding current scope of coverage, the relevant rationale, and the relative importance of efficiency versus equity arguments. Differences across expansion groups have further implications for the global design of social insurance.

This paper is more broadly related to understanding the *universal* nature of government-mandated social insurance. The trade-off with private options has long received theoretical attention (Vickers and Yarrow 1991; Hart *et al.* 1997), with rising empirical interest following the privatization of social insurance around the world (Gruber 2017; Reynolds 2013). Landais *et al.* (2017) shows risk-based selection alone does not rationalize mandated unemployment insurance in Sweden. This paper suggests a universal health insurance mandate may ultimately involve redistribution arguments, but does not directly test for its desirability. Instead, incremental expansion featuring flexible policy incentives potentially improves welfare on pure efficiency grounds.

2 Massachusetts health insurance reform

Massachusetts enacted its comprehensive health reform law, Chapter 58 of the Acts of 2006, in April, 2006. The law brings together individuals, private insurers and employers to improve health care access in the state, and is the blueprint of the 2010 Patient Protection and Affordable Care Act (ACA) enacted nationwide in 2010. Key provisions of the law are an individual and employer mandate, subsidized insurance to the low-income through Medicaid and private insurance markets, and rating regulation. Together, the structure of the reform resembles a "three-legged stool". I introduce each component of the law below.

2.1 Mandate

Compared to subsidized insurance and rating regulation, the individual mandate is an innovative and controversial part of the law without a historic precedent. The mandate requires individuals over eighteen years old to purchase health insurance, or face a tax penalty. The insurance purchased should meet the "minimum creditable coverage" standards, satisfied by most public and commercial health insurance but not plans that only cover catastrophic events. The penalty was first implemented in 2007. Individuals without proof of eligible insurance by December, 2007 lose their personal income tax exemption. In later years, the tax penalty equals 50% of the monthly premium in the lowest-cost plan available for the individual, adjusted by the number of uninsured months during the year.

The penalty is waived for low-income groups below 150% FPL, a population eligible for free insurance either from the Medicaid or from the Exchange marketplace. For middle- to high-income groups, the penalty increases the outside cost of uninsurance, increasing the willingness to pay for insurance. Focusing exclusively on the market of unsubsidized private insurance, (Hackmann *et al.*, 2015) finds the mandate increased the take-up rate to near-universal, and resulted in lower enrollee cost and insurer price.

A second mandate imposes fines on employers who fail to provide adequate funds for employee health insurance. The motivation is that uninsured workers generate excessive uncompensated care costs financed by society. Specifically, firms with more than 10 full-time workers are charged a "fair share contribution" fee of $295 per uninsured worker, where the dollar amount reflects the expected charity care cost incurred by these uninsured workers. In addition, for employers whose workers generate particularly high costs of charity care (more than $50,000 annually), and firms who do not provide health insurance on a pre-tax basis, an additional "free rider surcharge" applies. The exact amount of penalty depends on a formula that weighs the employer cost of sponsorship against the charity cost to the state. Due to administrative costs to firms, however, the employer mandate was repealed in July, 2013.

2.2 Subsidized insurance

The second component of the law addresses the financial difficulty of purchasing health insurance that impedes take-up in the low-income group. Previously, individuals with income below 133% FPL are covered in the Medicaid program (MassHealth). Enrollment in

MassHealth also needs to satisfy certain demographic criteria that typically leave childless adults ineligible for coverage, regardless of income. To make insurance more affordable for more individuals, including those in the low- to middle-income groups, the state instituted the Exchange program, called the Commonwealth Care (CommCare), that subsidizes the insurance premium to eligible individuals below 300% FPL. Recipients are not eligible for Medicaid coverage or employer sponsorship.

To determine subsidy, enrollees are first assigned an "affordable" amount of out-of-pocket cost for purchasing insurance. Individuals are not required to pay for insurance premium above their affordability limit. Affordability is set as an increasing step function over ranges of income in percent of FPL. For instance, for individuals with income less than 150% FPL, affordability is zero, and premium is fully subsidized. Affordability then increases with income. In 2011, for instance, it increases to $39 per month in the 150-200% bracket, $77 per month for the 200-250% bracket, and $116 for the 250-300% bracket. For individuals with income over 300% FPL, subsidy no longer applies. A separate program, called the Commonwealth Choice, provides private insurance plans for individuals who wish to purchase from the Exchange but are ineligible for subsidy.

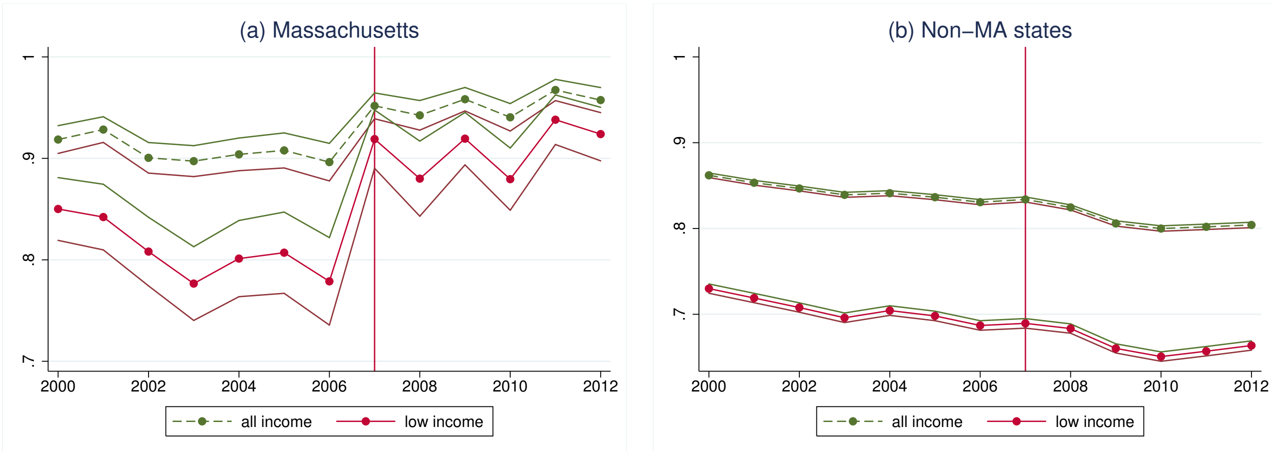
The difference between enrollee cost (affordability) and the premium price set by insurers is the subsidy. In percent of premium, subsidy is roughly 90% of premium in the 150-200% income bracket, 80% for the 200-250% bracket, and 70% for the 250-300% bracket. The increased generosity of subsidy explains the substantial coverage expansion in the low-income eligible population. As Figure 1 shows, the below 300% FPL group has significantly lower coverage rate before the reform, but increased coverage rate sharply after. Specifically, coverage in the low-income group surged to catch up with the average rate in Massachusetts, whereas the trending remain parallel in the national sample across income.[1]

2.3 Rating regulation

The final piece of the law is regulation on premium. Although formally a component of the law, the major regulation on premium, such as guarantee issue and community rating, is already effective in Massachusetts before the passage of Chapter 58. Guarantee issue ensures that individuals with pre-existing conditions are not denied coverage. Community rating in Massachusetts allows premium to differ by enrollee age and location of residence, but not by other demographics. Building on existing regulation, Chapter 58 further requires that the maximum premium variation across age does not exceed a ratio of 2, further limiting the the extend of premium discrimination. In addition, Chapter 58 merged the small-group and non-group risk pools in July 2007. Premium of individual plans and small-group plans are subject to the same set of regulation based on the joint risk pool. Since previously the individual market has low market share and high costs, the merger significantly reduced the premium of individual plans without meaningfully increasing the rate for group plans (Graves and Gruber, 2012). I take the rating regulation

[1]Similarly, the Massachusetts Health Reform Survey estimates that uninsurance below 300% FPL is 23.8% in fall, 2006, and decreased to 12.9% in fall, 2007. Above 300% FPL, uninsurance decreased from 5.2% in 2006 to 2.9% in 2007.

Figure 1: Insurance coverage trends



Notes. Figure compares coverage trends in Massachusetts (panel a) with the rest of the US states (panel b), for the full sample and the low income group where family income is less than or equal to 300% FPL. Coverage rates are aggregated from micro data for the 27-64 age group in the CPS March supplement, adjusted by insurance weights. 95% confidence intervals are plotted.

as given for my 2008-2013 sample period, but exploit policy variation in subsidy and penalty to analyze the effects on enrollees and society in general.

Table 1 gives a sense of the relative contribution of different programs to the increase in coverage rate in Massachusetts. Consistent with Figure 1, Commonwealth Care is the leading contributor to new enrollments in the first two years of the reform: of the 442,000 new enrollees by July 2008, around 40% received premium assistance from the program. Including subsidized enrollees in the free MassHealth aprogram, more than half (68%) of new enrollees received some premium assistance before take-up.

Table 1: New enrollment by source of coverage

	6/30/2006	12/31/2006	6/30/2007	12/31/2007	6/30/2008	diff. from 6/30/06
Private Group	4,274,000	4,338,000	4,378,000	4,406,000	4,421,000	147,000
Individual Purchase	40,000	39,000	36,000	65,000	80,000	40,000
MassHealth	705,000	741,000	732,000	765,000	785,000	80,000
Commonwealth Care	0	18,000	80,000	158,000	176,000	176,000
Total	5,020,000	5,136,000	5,226,000	5,394,000	5,462,000	442,000

Notes: Table shows administrative enrollment counts published in Health Care in Massachusetts: Key Indicators, November 2008. These numbers exclude Medicare enrollees; the MassHealth category only includes enrollees who list MassHealth as the primary insurer. For more details on the administrative records used in compiling the numbers, see the original report at http://archives.lib.state.ma.us/bitstream/handle/2452/36763/ocn232606916-2008-11.pdf?sequence=1&isAllowed=y.

2.4 Uncompensated care programs

Outside the main frame of Chapter 58, the state's Uncompensated Care Pool (UCP) is a safety net insurance program that disburses the cost of care incurred by the uninsured population. The program is relevant for the motivation of Chapter 58, because a key premise of the reform is that formal insurance subsidy lowers the social cost of charity care, solving a "free-rider problem". The problem originates from the 1986 Emergency Medical Treatment and Active Labor Act (EMTALA), which mandated hospitals and ambulatory services to provide emergency care to the uninsured regardless of ability to pay. When the care receiver does not pay for the medical costs she incurs, the cost is borne by third-party members. Before Chapter 58, the UCP is the program that finances the charity care to the insured in the state.

The UCP charges assessments on providers and private sector employers who contribute to the financing of charity care. The assessment can be considered as a tax on the revenues of firms. In 2005, UCP was billed \$739 million for uncompensated care, and paid out \$530 million through assessment and general fund appropriation.[2]

Chapter 58 renamed the UCP program to the Health Safety Net (HSN), and reformed the financing of charity care. First, Chapter 58 received permission from the federal government in 2005 to redirect the funding for uncompensated care programs to premium subsidy on the Commonwealth, under the premise that the cost of expanding subsidy programs is partially offset by savings in uncompensated care. The additional costs are to be shared evenly between the state and the federal government. The reduction in the charity care cost is significant. Within five years of the reform, charity cost decreased by \$243 million from \$739 million in 2005 to \$496 million in 2011.[3]

Under Chapter 58, HSN reimburses providers of uncompensated care with a mix of assessments, surcharges and revenues from the general fund. A service surcharge is applied to payments made by insured patients for medical services. Assessments on hospitals are uncompensated profit loss for treating the uninsured. In 2011, for example, service surcharge on payments to hospitals and assessment on hospital profits each contribute \$160 million to the fund, with an additional \$100 million from other revenue sources.

Although displaced funding from the charity care program lowers the financing burden of subsidy, the Commonwealth Care program still requires generous contribution from the state and federal government. The additional costs when replacing safety net coverage with subsidized insurance represents the efficiency cost of the reform, or the fiscal externality due to behavioral responses. In 2011, \$913 million is budgeted for CommCare subsidies to 176,500 enrollees.[4] Over time, as enrollment increased, CommCare spending increased from \$628 million (with a \$472 million budget) in 2008[5] to \$872 million (with a \$913 million budget) in 2011.

[2]Number based on Health Safety Net (Uncompensated Care Pool) annual reports, available at http://archives.lib.state.ma.us/handle/2452/392113

[3]UCP paid \$530 million in 2005, and \$412 million in 2011. In addition to reduced UCP spending, hospitals bear smaller profit loss for the residual uncompensated care cost not paid for by the government.

[4]The budget combines the Commonwealth Care and the Commonwealth Care Bridge, a similarly structured subsidy program enrolling eligible legal immigrants. Source: http://budget.digital.mass.gov/bb/h1/fy11h1/exec_11/hbuddevhc.htm

[5]Source: http://budget.digital.mass.gov/bb/h1/fy10h1/exec10/hbuddevhc.htm

3 Motivating insurance expansion

Although the motivations for expanding health insurance programs are a subject of constant debate, many social insurance programs are universal in scope. Individuals contribute to the financing of the program and receive transfer benefits once eligible. Participation is often mandatory or assumed by default. One typical example is the unemployment insurance program. Given universal enrollment, researchers are interested in the optimal benefit design that balances equity and efficiency for the economy. The case of health insurance expansion offers a unique opportunity to examine the desirability of universal coverage, and to shed light on the optimal scope of social insurance when take-up is voluntary and incomplete.

This section develops a stylized framework to understand the desirability of a health insurance mandate. I focus on the corner solution where individuals with zero demand for insurance are required to enroll. When premium reflects the average cost of enrollees, a universal mandate mitigates adverse selection and lowers premium for high-cost individuals. The desirability of a mandate then depends on the society's redistribution preferences across risk. Moreover, when the society cannot avoid providing uncompensated care to the uninsured, a formal insurance mandate improves upon the status quo if it achieves better distribution of healthcare costs in the economy. I delineate conditions under which formal insurance broadens the base of redistribution to the uninsured, and lowers the excess financing burden on third-party payers including patients.

3.1 Environment

Consider a unit mass of individuals heterogeneous in health type μ and labor productivity ν. $\mu \in [0,1]$ is the probability of staying healthy, and $1-\mu$ is the risk of illness, in which case medical care such as hospitalization is required to restore health. The cost of the medical care is M. Type μ has expected cost $(1-\mu)M$.

Workers produce output of value w. The opportunity cost of working, such as the value of lost leisure and home production, is captured in $g(\frac{1}{\nu})$ and heterogeneous in type $\nu \in [0,1]$. I assume $g'(\cdot) > 0$ and $g''(\cdot) > 0$, or the cost of working decreases (and decreases faster) with productivity. I further assume that $g(1) = 0$ and $g(+\infty) = +\infty$, or the highest productivity type always works, and the lowest type never works.

Distribution of types follows the density function $f(\nu,\mu)$. I assume $f(\nu,\mu)$ is strictly positive everywhere, and continuously differentiable in both arguments. The correlation between health and productivity type is arbitrary.

Risk-averse individuals have von Neumann-Morgenstern utility over consumption. Consumption varies with medical cost in the probabilistic health event. Individuals can insure against the uncertainty with health insurance. Insurance requires premium payment in both states of the world, but no additional payment for medical cost M when sick. Uninsured patients access similar medical services, and the cost is paid through a combination of individual earning (if any) and government transfer payments.

Government determines the boundary of the insurance market. $hi(\nu,\mu) \in \{0,1\}$ gives the insurance state of type (ν,μ). A universal mandate implies $hi(\nu,\mu) = 1$. Given the market size, perfectly competitive insurers offer premium $p(\nu,\mu)$. Depending on the information feasible in the pricing, premium can reflect individual or average risk in the market. Government determines income transfer $t(\nu,\mu)$ and uncompensated care transfer $\Delta t(\nu,\mu)$ specific to uninsured patients.

Type (ν,μ) has expected utility

$$U(\nu,\mu) = \mu\, u(c_H(\nu,\mu)) + (1-\mu)\, u(c_S(\nu,\mu)) - e(\nu,\mu)\, g(\frac{1}{\nu}),$$

where consumption in the healthy state $c_H(\nu,\mu) = e(\nu,\mu)\cdot w + t(\nu,\mu) - hi(\nu,\mu)\cdot p(\nu,\mu)$. $e(\nu,\mu) \in \{0,1\}$ indicates employment choice. Consumption in the unhealthy state $c_S(\nu,\mu) = c_H(\nu,\mu) + \Delta t(\nu,\mu) - M + hi(\nu,\mu)\cdot M$, where $\Delta t(\nu,\mu)$ is the additional transfer needed to smooth consumption when the patient is uninsured, or the cost of uncompensated care. $\Delta t(\nu,\mu) = 0$ if $hi(\nu,\mu) = 1$.

State utility $u(\cdot)$ satisfies $u'(\cdot) > 0 > u''(\cdot)$. In addition, utility satisfies the Inada condition $u'(0) = +\infty$: marginal utility rises to infinity when agent consumes very little, and so does the social value of income redistribution to the very poor. The concavity implies that income support $t(\nu,\mu)$ and uncompensated care Δt are more generous to the low-consumption groups.

3.2 Adverse selection

I first examine adverse selection as a potential motivation for universal insurance. Adverse selection occurs either because risk is private information, or because price discrimination based on medical records is prohibited.[6] For simplicity, I assume perfect competition among insurers, so that premium equals the average cost of enrollees. I consider mark-up response to cost changes in the empirical analysis below.

It is easy to see that take-up is incomplete for low-risk individuals. The lowest risk type, in particular, has zero WTP for insurance. They do not enroll unless subsidy reduces the cost of premium to zero. With asymmetric information, however, risk-based subsidy is not feasible. Premium drives a wedge between the marginal and the average cost, resulting in inefficiently low take-up in the low-cost population.

In lieu of risk-based transfer, government arranges for means-tested transfers based on observed employment choice. The (low-income) non-employed individuals receive premium subsidy that lowers premium by a fraction λ_p, in addition to a cash transfer A. A is the benefit level of the unemployment insurance (UI) program, assumed mandatory in nature. Subsidized premium cost is $(1-\lambda_p)p$. Subsidy and cash transfers are financed with a lump-sum pay-roll tax τ on the (high-income) employed.

I analyze the desirability of extending health insurance to the lowest demand types, holding fixed the benefit design of the UI program. In this setting, the lowest demand types are the perfect health types who have zero medical expenditure risk. A mandate that enrolls the zero-demand types lowers the average cost of insurance in the economy, benefiting infra-marginal enrollees with higher demand for insurance. The social benefit

[6] The premium regulation reflects some redistribution preferences for the sicker population potentially facing unaffordable level of premium due to pre-existing conditions. When premium is perfectly discriminatory in costs, Appendix A.1 shows that mandatory formal insurance priced at expected cost dominates uncompensated care.

is weighed against the private cost of insurance on the perfect health margin, governed by the society's redistribution preferences across risk. Specifically, enrollment lowers the average risk in the insurance program r by

$$\frac{dr}{dn}\bigg|_{n=1} = -\frac{r}{i} \cdot F_\mu(1,1)$$

where $r = E[1-\mu | hi(\nu,\mu) = 1]$ is the average risk, $i = E[hi(\nu,\mu)] = F(1,n)$ the insurance rate, and $F_\mu(1,1)$ the mass the ultra-health types.[7]

The price change lowers the total cost of insurance by $i \cdot M \cdot \frac{dr}{dn}\Big|_{n=1} = p \cdot F_\mu(1,1)$. Numerically it is equal to the cost of premium borne by marginal enrollees in perfect health. A mandate reaching the ultra-margin can be considered as a transfer to infra-marginal enrollers with higher cost of coverage. Whether the transfer is desirable depends on the society's preference for redistribution across risk. The marginal versus infra-marginal trade-off can be expressed as

$$\frac{dW}{dn}\bigg|_{n=1} \propto \underbrace{p \cdot u'(c_e)}_{\text{social benefit}} + \underbrace{u(c_e) - u(c_e + p)}_{\text{marginal utility loss}} > 0$$

where the cost of premium in utility terms, $u(c_e) - u(c_e + p)$, is compared with the premium saving $p \cdot u'(c_e)$ to infra-marginal enrollees.[8] For a given price p, the trade-off boils down to the relative value of p to different members of society. In the most simplistic setting, standard result applies that egalitarian outcome is desirable with concave utility. Insurance is universal, and premium is equal to the average cost of the economy $\bar{p} = E[1-\mu]$. In particular, for a given set of redistribution preferences, the allocation can be implemented either through rating regulation that prohibits price discrimination, or through risk-based subsidy on premium that varies freely with risk.[9]

Formally, enrolling the highest health type $\mu = 1$ (when lower health types have already enrolled) is at least locally optimal, and is matched with full premium subsidy regardless of the type distribution $f(\nu,\mu)$, employment $e(\nu,\mu)$, or cash transfer A:

Proposition 1. *For an arbitrary income transfer A and employment choice $e(\nu,\mu)$,*

(a) with full insurance subsidy $\lambda_p = 1$, enrolling $\mu = 1$ increases welfare:

$$\frac{dW}{dn}\bigg|_{n=1} > 0$$

[7] F_μ is the marginal density function of μ.

[8] Subscript e denotes the employed population. The trade-off applies specifically to the employed when unemployed individuals receive fully subsidized insurance financed by an earning tax on workers.

[9] Of course, different pricing and subsidy regulations may have different efficiency implications depending on behavioral responses. In this section, I conceptualize the pricing benefits and costs of expanding insurance to the ultra-margin (i.e., a mandate), but abstract from behavioral responses to price changes on the infra-margin. For instance, employment rate is assumed invariant to changes in the tax burden of subsidy. The empirical framework in Section 4 fully accounts for these behavioral responses and efficiency costs.

(b) universal insurance implies full insurance subsidy:

$$\frac{dW}{d\lambda_p}\Big|_{n=1} \geq 0$$

I provide the proof in Appendix A.2. With utilitarian social welfare and risk averse individuals, complete redistribution across risk implemented by a mandate is socially desirable. Absent the mandate, adverse selection leads to incomplete take-up and premium priced above $\bar{p}$, an inefficient outcome when enrolling lower-risk individuals would have generated positive social surplus ($\frac{dW}{dn}\Big|_{n=1} > 0$).[10]

With more complicated preference structures, complete redistribution across risk may not be desirable. For example, if risk tolerance increases with risk, then the low-risk population can efficiently bear more risk without purchasing insurance (Appendix A.3). When redistribution is evaluated by the marginal utility of consumption, heterogeneity in earnings and consumption complicates the redistribution across risk through the correlation with risk (Appendix A.4). Moreover, tax incentives, such as a mandate penalty, have additional redistribution consequences that interact with an insurance mandate (Appendix A.5). I introduce these additional elements step by step in the Appendix.

3.3 Uncompensated care

Suppose the government provides uncompensated care to uninsured patients, and enroll the uninsured in formal insurance at the site of care. The non-employed can either purchase formal insurance with full premium subsidy, or remain uninsured but utilize uncompensated care free of charge. The government finances the premium subsidy with a linear tax on payroll, and a tax penalty imposed on the working uninsured. Specifically, let the tax penalty be $k \cdot p$, where k gives the level of penalty relative to premium.

The government finances the uncompensated care by levying surcharge fees on paying customers in healthcare. In the conceptual framework, I focus on a premium surcharge on enrollees in formal insurance, and a potential service surcharge on patients. When there is negative correlation between risk and productivity, patients on average have lower income than healthy individuals, implying greater burden of premium surcharge on patients. With service surcharge, the excess burden on patients further increases.

Relative to the implicit insurance of uncompensated care, expanding formal insurance with subsidy is desirable, if it improves the distribution of healthcare costs in the economy.[11] This is the case if subsidy is financed progressively over the tax base, and lowers the excess burden of uncompensated care on low-income groups. The following proposition formulates the distribution implications.

[10] With risk-based pricing, positive social surplus only requires WTP higher than marginal cost, which is satisfied given risk aversion. With risk pooling, it requires social benefit offset marginal utility loss.

[11] The conceptual analysis assumes away behavioral responses that might increase the cost and utilization of individuals who take-up formal insurance. I consider both behavioral responses and distribution (equity) in the empirical analysis.

Proposition 2. *If productivity is negatively correlated with risk ($Cov[\nu, \mu] > 0$), then tax-financed subsidy reduces the excess burden on patients, and replacing uncompensated care with subsidized insurance always improves welfare.*

Appendix A.6 provides the proof. The negative correlation implies patients on average have lower income than taxpayers. Expanding formal insurance shifts the cost from patients to workers, reducing the excess burden on low-income individuals. Financed by a linear tax over payroll, subsidy further distributes healthcare costs progressively over income. Therefore social insurance financed by tax-based subsidy achieves more equitable distribution of healthcare costs, improving the de facto coverage of uncompensated care. More generally, Appendix A.6 shows that tax-based subsidies on premium achieve better public finance of the safety-net mandate, if subsidies broaden the base of public finance and if redistribution is more valuable over productivity than health.

Relative to adverse selection, uncompensated care provides stronger motivation for universal health insurance. Although both motivations justify enrolling the lowest-demand individuals, enrollment in formal insurance is desirable at *any* level of coverage with uncompensated care. Subsidized universal insurance is generally the global optimum. The stronger motivation reflects the fact that in addition to the premium benefit, the distribution of healthcare costs improves when the subsidy is financed over the broader tax base that lowers the burden on vulnerable individuals.

4 Incremental expansion in Massachusetts

The idea of subsidized universal health insurance has its closest analogue in Massachusetts following the 2006-2007 reform. Instead of a direct mandate, the state achieved near-universal insurance coverage using policy instruments such as premium subsidy and tax penalty. Exploiting behavioral responses to policy as "sufficient statistics" (Chetty 2006; Chetty and Finkelstein 2013), I formulate and calculate the pricing benefits and costs of expanding current scope of insurance in an empirical framework of the Massachusetts reform. I compare motivations for policy incentives targeting different expansion groups. I present the key elements of the framework below.

4.1 Setting

The framework adapts the optimal UI benefit design in Chetty (2006) to the case of health insurance subsidies and taxes. In a continuous time economy, individuals enter period $t \in [0, 1]$ with state vector ω_t. They choose employment $e_t(\omega_t) \in \{0, 1\}$ and health insurance $hi_t(\omega_t) \in \{0, 1, 2\}$. $e_t = 1$ for employees. Uninsured individuals have $hi_t = 0$. $hi_t = 1$ for ESI enrollees, and equals 2 for subsidized enrollees in the Medicaid and the Exchange program.

Standard in discrete choice models, I assume individuals are subject to an employment shock ϕ_t each period. The binary employment outcome e_t is determined from a single-index equation $m_t^e(s_t, \phi_t; \omega_t)$, where s_t is the continuous search effort chosen optimally by individuals in state ω_t.[12] Similarly, individuals are subject to a taste shock η_t in the insurance choice hi_t. Specifically, the taste shock enters the WTP for insurance. Depending on the sign of η_t, individuals can either over- or under-insure, for a given state and cost of coverage. Empirically, the taste shock can represent behavioral features (Kling *et al.* 2012; Barseghyan *et al.* 2013) and optimization frictions (Handel and Kolstad 2015; Bhargava *et al.* 2017) associated with low take-up rates in targeted groups.

In each state, forward-looking individuals make optimal employment and insurance choices subject to friction ϕ_t and η_t.[13] The state vector ω_t contains choices and outcomes in previous periods, and price variables such as insurance premium, subsidy and a lump-sum payroll tax which individuals take as given. I assume a uniform subsidy generosity that covers λ_p of premium, so that enrollees cost is $(1-\lambda_p)p$, where p is the premium price before subsidy.[14] The subsidy is financed by a small tax burden on payroll, averaging τ_{pb} per worker, and the penalty $k \cdot p$ per uninsured individual ineligible for subsidy.

Let $\lambda_i = Pr\{hi_t(\omega_t) = i\}$, $i = 0,1,2$ denote insurance rates in the population. Let $e = E[e_t(\omega_t)]$ denote the employment rate. The public finance of premium subsidy implies

$$e \cdot \tau_{pb} + \lambda_0 \cdot k \cdot p \;=\; \lambda_2 \cdot \lambda_p \cdot p + \tau_1 \cdot \lambda_1 \cdot p \qquad (BC.1)$$

where λ_1 is the tax deduction of ESI. Budget constraint states that a lump-sum tax on payroll, net of ESI deduction and the tax penalty, finances the subsidy expenditures in public insurance programs. The condition relates an increase in subsidy generosity λ_p to the implied change in tax burden τ_{pb}, holding constant other insurance policies (k and λ_1). Specifically, the relationship $\frac{d\tau_{pb}}{d\lambda_p}$ is derived from total differentiation of Equation *BC.1*.

Let λ_1^e measure employees enrolled in ESI. They purchase insurance for own coverage and for non-employed dependent enrollees. The premium payment is deductible from the tax. The following condition captures the transfer payment in ESI

$$\lambda_1^e \cdot \tau_{pr} \;=\; (1-\tau_1) \cdot \lambda_1 \cdot p \qquad (BC.2)$$

where τ_{pr} is the private transfer made by ESI payers. I assume ESI payers are limited to working enrollees in the main analysis. I allow for increasing contribution by dependent enrollees in the robustness analysis.

Uncompensated care in Massachusetts is financed by a surcharge fee on paying patients and an assessment on providers. I model the share of total uncompensated cost financed by patient surcharge to be α. Per capita excess burden uc_p is given by

$$\lambda_{>0}^0 \cdot uc_p \;=\; \alpha \cdot \lambda_0^0 \cdot g \cdot OOC \qquad (BC.3)$$

[12]Employment $e_t = 1$ if the underlying equation crosses a threshold, for example, $m_t^e(s_t, \phi_t; \omega_t) \geq 0$. The continuous search effort s_t allows Envelope Theorem to be applied when employment outcome is binary and not continuously differentiable.

[13]That is, choices of job search effort and WTP for insurance are optimal in expectation of the friction shocks ϕ_t and η_t.

[14]The assumption implies the income composition of recipients stays approximately the same when subsidy becomes more generous. Consistent with this view, I do not find significant employment responses to generosity variation in Section 5.

where λ_0^0 ($\lambda_{>0}^0$) measures uninsured (insured) patients in hospital settings. OOC is the enrollee cost sharing in formal insurance. Informal coverage is less generous, reflected in parameter $g \leq 1$.[15] Individuals can choose to finance the cost sharing from medical debt $db_t(\omega_t)$. The uninsured can face a higher interest rate and cost of borrowing, if they accumulate greater amount of medical debt, an externality noted in (Miller, 2016).[16] I consider the potential gains to credit access as an additional benefit of formal insurance.

Premium deviates from the average cost in markets with few insurers. I model the market power of insurers using the mark-up parameter $\beta(\lambda_p, k)$. Depending on the market power, premium is set above the average cost at

$$p = \left[1 + \beta\left(\lambda_p, k\right)\right] \cdot r(\lambda_0) \qquad (BC.4)$$

where average cost $r(\lambda_0)$ varies depending on the selection into formal insurance. Adverse selection implies lower average cost in formal insurance as take-up increases. Since subsidy and penalty are linked to insurer-set prices, β is allowed to vary with policy parameters directly as a result of pricing responses to policy (Jaffe and Shepard, 2018).[17]

4.2 Welfare

Period utility

$$u_t(\omega_t) = \sum_{i=0}^{1}\sum_{j=0}^{2} Pr_t^{ij}(\omega_t)\left[\mathbf{E}_g u(c_t^{ijg}(\omega_t)) - 1_{\{i=1\}}\, g\left(\frac{1}{\nu}\right)\right] \qquad (1)$$

reflects expected utility from consumption over health state g, weighted by probabilistic employment i and insurance j outcome in $Pr_t^{ij}(\omega_t)$ from friction ϕ_t and η_t. Workers pay a fixed cost of employment decreasing in ν.

Across periods, individuals maximize life-cycle utility $U = \int_t \int_{\omega_t} u_t(\omega_t)\mathrm{d}\mathbf{F}_t(\omega_t)\mathrm{d}t$ with optimal choices of saving, medical debt, employment and insurance, subject to the intertemporal budget constraint and friction shocks.

Let the maximized individual utility be V. The government then chooses policy parameters $\mathbf{K} = (\lambda_p, k)$ to maximize a weighted sum of V and the burden of uncompensated care on providers $(1-\alpha)\cdot\lambda_0^0\cdot g\cdot OOC$. I assume the government is only concerned with minimizing the third-party cost of an unfunded mandate, but is otherwise uninterested in preserving industry profit. Assuming the relative weight on consumer welfare is ζ, social welfare

$$W = \zeta\cdot V - (1-\alpha)\cdot\lambda_0^0\cdot g\cdot OOC \qquad (2)$$

The decision structure implies a small policy change affects welfare only through prices and the friction, with no first-order impact through choice variables (Envelope Theorem).

[15] Higher spending in formal insurance is also consistent with the moral hazard effect of take-up, which I use to parametrize g for the main analysis.

[16] A growing body of evidence shows public insurance programs improve financial well-being across contexts. See Allen *et al.* (2017), Brevoort *et al.* (2017), Argys *et al.* (2017), Dranove *et al.* (2016), Gross and Notowidigdo (2011), Gallagher *et al.* (2018), Hu *et al.* (2016), among others.

[17] I focus on the average premium charged by insurers in Massachusetts. The vast majority of Exchange enrollees (90%) are served by three insurers with near-identical premium.

Behavioral responses to policy affect welfare through 1) the effect on prices, and 2), the effect on the distribution of employment and insurance outcomes through friction. In general, the following proposition is true.

Proposition 3. *The welfare effect of a small increase in policy* $\mathbf{K}$ *equals*

$$\frac{\mathrm{d}W}{\mathrm{d}\mathbf{K}} = \zeta \frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}}\bigg|_{\rho^{ij}} + \zeta \frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}}\bigg|_{u'} - \frac{\mathrm{d}(1-\alpha)\lambda_0^0}{\mathrm{d}\mathbf{K}} \cdot g \cdot OOC \tag{3}$$

where $\rho^{ij} = \mathbf{E}_t \mathbf{E}_{\omega_t} 1_{\{e_t=i, hi_t=j\}}$ *is the distribution of employment (i) and insurance (j) outcomes. In first-order approximation,*

$$\frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}}\bigg|_{u'} \approx \theta^{ij}(\mathbf{K}) \frac{\mathrm{d}\rho^{ij}}{\mathrm{d}\mathbf{K}} \cdot \Delta u \bigg|_{\theta^{ij}(\mathbf{K})} \tag{4}$$

where $\frac{\mathrm{d}\rho^{ij}}{\mathrm{d}\mathbf{K}}$ *is the policy effect on outcome, of which fraction* $\theta^{ij}(\mathbf{K})$ *moved due to the policy effect on friction, experiencing utility difference* Δu.

I discuss details in the proof in Appendix A.7. Intuitively, $\frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}}\Big|_{\rho^{ij}}$ captures the pricing externality on consumption through the budget constraints. In addition to changes in the state utility, the state distribution varies both because of individual choice and the policy effect on friction. For a policy effect $\frac{\mathrm{d}\rho^{ij}}{\mathrm{d}\mathbf{K}}$ estimated from data, friction accounts for $\theta^{ij}(\mathbf{K})$ of the total effect. Equation 4 captures the first-order utility change for frictional movers in $\theta^{ij}(\mathbf{K})\frac{\mathrm{d}\rho^{ij}}{\mathrm{d}\mathbf{K}}$.

From Equation 3, I derive closed-form formulas for the pricing benefits on premium and the burden of uncompensated care. I similarly derive the fiscal cost on the government for implementing the policy, and the fiscal externality net of the recipient valuation. Comparing the pricing benefits with the fiscal externality determines the desirability of expanding current scope of insurance,[18] and the motivations for different expansion policies. I formulate the welfare benefits and costs for premium subsidy and the mandate penalty below.

4.3 Premium subsidy

Consider a small increase in subsidy generosity $\mathrm{d}\lambda_p$. Enrollee cost decreases by $p \cdot \mathrm{d}\lambda_p$. New enrollees $-\mathrm{d}\lambda_0$ change the average cost $r(\lambda_0)$ in formal insurance. Adverse selection implies lower enrollee cost with higher insurance rate, or $\varepsilon_{r,\lambda_0} > 0$. In addition to the cost composition effect, mark-up set by insurers responds to subsidy generosity. Total effect on premium (*BC*.4) is given by

$$\frac{\mathrm{d}\log p}{\mathrm{d}\lambda_p} = \frac{\varepsilon_{r,\lambda_0}}{\lambda_0} \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p} + \frac{1}{1+\beta} \cdot \frac{\partial \beta}{\partial \lambda_p} \tag{5}$$

[18] I subject the evaluation to potential over-insurance from behavioral friction as an additional source of negative externality.

The pricing effect applies to all payers of insurance premium, adjusted by the out-of-pocket share and the payer's marginal utility. Standardizing the welfare metric in terms of a dollar increase in worker earning, or $\zeta = \frac{1}{u'(c_{1..})}$, the premium benefit from a dollar increase in subsidy equals

$$\begin{aligned} \frac{\mathrm{d}W_p}{\mathrm{d}\lambda_p\, p} = & -\frac{\mathrm{d}\log p}{\mathrm{d}\lambda_p}\left[\lambda_2\left(\frac{u'(c_{.2.})}{u'(c_{1..})}(1-\lambda_p)+\lambda_p\right)\right. \\ & \left. +\lambda_1\left(\frac{u'(c_{11.})}{u'(c_{1..})}(1-\tau_1)+\tau_1\right)+\lambda_0\, k\left(\frac{u'(c_{.0.})}{u'(c_{1..})}-1\right)\right] \end{aligned} \tag{6}$$

Enrollment $-\mathrm{d}\lambda_0$ reduces the social cost of uncompensated care by $g\cdot ri(\lambda_0)\cdot\left(1+\varepsilon_{ri,\lambda_0}\right)\cdot \mathrm{d}\lambda_0$,[19] of which fraction $1-\alpha$ accrues to providers as avoided profit loss. The remainder lowers patient service surcharge by $\mathrm{d}uc_p$. In total the saving is valued at

$$\begin{aligned} \mathrm{d}W_{UC} = & -g\cdot ri\cdot\left(1+\varepsilon_{ri,\lambda_0}\right)\cdot \mathrm{d}\lambda_0 \\ & -\underbrace{\alpha\cdot\left[\frac{u'(c_{.>0,0})}{u'(c_{1..})}-1\right]\cdot g\cdot ri\cdot\left(1+\varepsilon_{ri,\lambda_0}\right)\cdot \mathrm{d}\lambda_0}_{\text{excess burden}} \\ & -\underbrace{\alpha\cdot\frac{u'(c_{.>0,0})}{u'(c_{1..})}\cdot g\cdot ri\cdot\left(\frac{\lambda_0}{1-\lambda_0}-\varepsilon_{r,\lambda_0}\right)\cdot \mathrm{d}\lambda_0}_{\text{selection}} \end{aligned} \tag{7}$$

where $\frac{u'(c_{.>0,0})}{u'(c_{1..})}-1$ captures excess burden on patients. The third term adjusts for the selection effect on per capita burden. Normalized in dollar units of subsidy, the welfare benefit on the social cost of uncompensated care equals $\frac{\mathrm{d}W_{UC}}{\mathrm{d}\lambda_p\, p}$.

Recipients receive a mechanic reduction in premium cost from an increase in λ_p. The benefit per dollar equals

$$\frac{\mathrm{d}W_B}{\mathrm{d}\lambda_p\, p} = \frac{u'(c_{.2.})}{u'(c_{1..})}\cdot\lambda_2 \tag{8}$$

the standard valuation of a dollar transfer to beneficiaries from workers. Raising the subsidy transfer through taxes imposes efficiency costs. In addition to the mechanic cost of current beneficiaries $\lambda_2\cdot p\cdot \mathrm{d}\lambda_p$, new enrollees raise the cost by $(\lambda_p+k)\cdot p\cdot \mathrm{d}\lambda_0$. Switchers from ESI raise the government cost by $(\lambda_p-\tau_1)\cdot p\cdot \mathrm{d}\lambda_1$, and adjust the burden of private transfer. Employment response e adjusts the tax burden. In total, the behavioral responses

[19] $ri(\lambda_0)$ is the average cost of uninsured if enrolled in formal insurance. Cost of charity care to third-party payers is lower at a fraction g.

imply the fiscal cost of raising a subsidy dollar equals

$$\frac{\mathrm{d}W_C}{\mathrm{d}\lambda_p p} = -\lambda_2 + \underbrace{(\lambda_p + k)\frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p}}_{\text{new enrollees}} + \underbrace{\left[\lambda_p - \tau_1 - \overbrace{\frac{u'(c_{11\cdot})}{u'(c_{1\cdot\cdot})}(1-\tau_1)\right]\frac{\mathrm{d}\lambda_1}{\mathrm{d}\lambda_p} + \frac{u'(c_{11\cdot})}{u'(c_{1\cdot\cdot})}(1-\tau_1)\frac{\lambda_1}{\lambda_1^e}\frac{\mathrm{d}\lambda_1^e}{\mathrm{d}\lambda_p}}^{\text{private transfer } \frac{\mathrm{d}\tau_{pr}}{\mathrm{d}\lambda_p}}}_{\text{selection from ESI}}$$
$$+ \frac{\tau_{pb}}{p}\frac{\mathrm{d}e}{\mathrm{d}\lambda_p} \tag{9}$$

The difference between the fiscal cost of subsidy (Equation 9) and the recipient valuation (Equation 8) gives the fiscal externality.

Weighing the fiscal externality with the pricing benefits, the welfare effect of premium subsidy is

$$\frac{\mathrm{d}W}{\mathrm{d}\lambda_p p} = \frac{\mathrm{d}W_P}{\mathrm{d}\lambda_p p} + \frac{\mathrm{d}W_{UC}}{\mathrm{d}\lambda_p p} + \frac{\mathrm{d}W_B}{\mathrm{d}\lambda_p p} + \frac{\mathrm{d}W_C}{\mathrm{d}\lambda_p p} + \frac{\zeta}{p}\Pi\left(\frac{\mathrm{d}\pi}{\mathrm{d}\lambda_p}\right) + \frac{\zeta}{p}\frac{\mathrm{d}V}{\mathrm{d}\lambda_p}\bigg|_{u'} \tag{10}$$

where $\Pi\left(\frac{\mathrm{d}\pi}{\mathrm{d}\lambda_p}\right)$ evaluates the credit benefit of insurance, and $\frac{\mathrm{d}V}{\mathrm{d}\lambda_p}\Big|_{u'}$ the utility loss from friction.

4.4 Mandate penalty

Consider a small increase in penalty $\mathrm{d}k$. Suppose new enrollees $-\mathrm{d}\lambda_0$ affect the average cost in formal insurance according to elasticity $\varepsilon^k_{r,\lambda_0}$. Allowing for mark-up adjustments, the effect on premium is

$$\frac{\mathrm{d}\log p}{\mathrm{d}k} = \frac{\varepsilon^k_{r,\lambda_0}}{k}\cdot\frac{\mathrm{d}\lambda_0}{\mathrm{d}k} + \frac{1}{1+\beta}\cdot\frac{\partial\beta}{\partial k} \tag{11}$$

Across payers, the premium benefit is valued at

$$\frac{\mathrm{d}W_p}{\mathrm{d}k\,p} = -\frac{\mathrm{d}\log p}{\mathrm{d}k}\left[\lambda_2\left(\frac{u'(c_{\cdot 2\cdot})}{u'(c_{1\cdot\cdot})}(1-\lambda_p)+\lambda_p\right)\right.$$
$$\left. + \lambda_1\left(\frac{u'(c_{11\cdot})}{u'(c_{1\cdot\cdot})}(1-\tau_1)+\tau_1\right) + \lambda_0 k\left(\frac{u'(c_{\cdot 0\cdot})}{u'(c_{1\cdot\cdot})}-1\right)\right] \tag{12}$$

New enrollees lower the social cost of uncompensated care according to Equation 7. The benefit per dollar of penalty is $\frac{\mathrm{d}W_{UC}}{\mathrm{d}k\,p}$.

The cost of penalty on the uninsured is valued at

$$\frac{\mathrm{d}W_B}{\mathrm{d}k\,p} = -\frac{u'(c_{\cdot 0\cdot})}{u'(c_{1\cdot\cdot})}\lambda_0 \tag{13}$$

where $\frac{u'(c_{.0.})}{u'(c_{1..})}$ adjusts for the utility loss from taxing the uninsured (relative to workers) to increase government revenue. In addition, the government loses revenue $(\lambda_p + k) \cdot p \cdot \mathrm{d}\lambda_2$ when the uninsured take-up subsidized insurance, and gives out tax rebate $(\tau_1 + k) \cdot p \cdot \mathrm{d}\lambda_1$ when the uninsured take-up ESI. Total fiscal cost of increasing the mandate penalty is,

$$\frac{\mathrm{d}W_C}{\mathrm{d}k\,p} = \lambda_0 \underbrace{-(\lambda_p + k)\frac{\mathrm{d}\lambda_2}{\mathrm{d}k}}_{\text{new subsidy enrollees}} \underbrace{-\left[\tau_1 + k + \frac{u'(c_{11.})}{u'(c_{1..})}(1-\tau_1)\right]\frac{\mathrm{d}\lambda_1}{\mathrm{d}k} + \frac{u'(c_{11.})}{u'(c_{1..})}(1-\tau_1)\frac{\lambda_1}{\lambda_1^e}\frac{\mathrm{d}\lambda_1^e}{\mathrm{d}k}}_{\text{new ESI enrollees}}$$
$$+ \frac{\tau_{pb}}{p}\frac{\mathrm{d}e}{\mathrm{d}k} \tag{14}$$

The mechanic burden (Equation 13) net of the behavioral effects in Equation 14 gives the fiscal externality of tax penalty. The total welfare effect of raising a dollar penalty is

$$\frac{\mathrm{d}W}{\mathrm{d}k\,p} = \frac{\mathrm{d}W_P}{\mathrm{d}k\,p} + \frac{\mathrm{d}W_{UC}}{\mathrm{d}k\,p} + \frac{\mathrm{d}W_B}{\mathrm{d}k\,p} + \frac{\mathrm{d}W_C}{\mathrm{d}k\,p} + \frac{\zeta}{p}\Pi\left(\frac{\mathrm{d}\pi}{\mathrm{d}k}\right) + \frac{\zeta}{p}\frac{\mathrm{d}V}{\mathrm{d}k}\Big|_{u'} \tag{15}$$

4.5 Model discussion

The framework exploits policy incentives to understand the social implications of insurance expansion on price, cost and welfare. The evaluation is internally valid for incremental expansion from the near-universal baseline in Massachusetts, but does not directly assess the desirability of a universal mandate. The evaluation is also specific to the analysis of externality, specifically the mechanism and distribution of effects across economy. I discuss some of these modeling assumptions below.

Uncompensated care – Uncompensated Care Pool in Massachusetts is funded by an assessment on hospitals and a service surcharge on paying patients. When total cost exceeds the program budget, hospitals bear extra cost as profit loss. The share of assessment in the finance of uncompensated care is available in the program reports.

The balance sheet may not accurately reflect the economic incidence of charity costs. In general, there is strong evidence that hospitals experience substantial uncompensated cost saving immediately after Medicaid expansions (Dranove *et al.* 2016; Blavin 2016). The effect on profit, however, depends on the incidence on insurers and enrollees who also bear the burden of uncompensated care. If service fees are irresponsive to the total cost ($\frac{\mathrm{d}uc_p}{\mathrm{d}\mathbf{K}} = 0$), then benefits accrue solely to hospitals as higher profit, and $\alpha = 0$. In Massachusetts, the surcharge burden is set to match the financing need of the charity care program, and went down from 2.90% in 2005 to 1.87% in 2013, suggesting at least some benefits to patients. Nonetheless, I vary α between $[0, 1]$ in the valuation of Equation 7.

In addition to the service surcharge, hospitals can negotiate higher rebates from insurers to lower the assessment burden on profit. The loss of rebate increases the enrollee cost of premium. With complete pass-through of assessment on enrollees, hospital profit is not affected by uncompensated care. The benefit to enrollees and patients is

$$\mathrm{d}W_{UC} = -g \cdot ri\left[(1-\alpha)\frac{u'(c_{.>0.})}{u'(c_{1..})}\left(\frac{1}{1-\lambda_0} + \varepsilon_{ri,\lambda_0}\right) + \alpha\frac{u'(c_{.>00})}{u'(c_{1..})}\left(\frac{1}{1-\lambda_0} + \varepsilon_{ri,\lambda_0} - \varepsilon_{r,\lambda_0}\right)\right]\mathrm{d}\lambda_0 \tag{16}$$

I discuss the robustness to different distribution of uncompensated care costs in the welfare calculation.

Private transfer – In the model, private transfer refers to ESI coverage of non-employees who receive premium transfer from current employees on the job, rather than subsidy from the government. The distinction is relevant for understanding the fiscal externality of subsidy, which affects welfare through interaction with private insurance transfers.

Measuring private transfer is empirically challenging. Enrollees in extended ESI coverage receive subsidy from the Medical Security Program at similar rate as Commonwealth Care.[20] Since these enrollees would have paid full premium, reported ESI need not represent private transfer to non-employees. This results in an over (under) estimate of private transfer (subsidy) recipients. In particular, change in private transfer recipients

$$\frac{\mathrm{d}\lambda_1^{1-e}}{\mathrm{d}\mathbf{K}} = (1-\vartheta)\cdot\frac{\widehat{\mathrm{d}\lambda_1^{1-e}}}{\mathrm{d}\mathbf{K}} - \widehat{\lambda_1^{1-e}}\cdot\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{K}}$$

where ϑ is the share of subsidy recipients among self-reported ESI enrollees $\widehat{\lambda_1^{1-e}}$. The bias in $\frac{\widehat{\mathrm{d}\lambda_1^{1-e}}}{\mathrm{d}\mathbf{K}}$ increases with ϑ. I discuss alternative valuation of $\frac{\mathrm{d}\lambda_1^{1-e}}{\mathrm{d}\mathbf{K}}$ in the empirical analysis.

Friction – The welfare framework admits two sources of friction shock. The first shock ϕ_t affects employment outcome through demand-side behavioral response to policy. The second shock η_t represents behavioral features in the WTP for insurance that deviates from the expected utility (EU) decision model.

Overall evidence of strong employment effect of subsidy, both in Massachusetts and more recently in the ACA, is weak. In the welfare calculation, I focus on the behavioral friction on insurance choice. More specifically, to form a stress test on the net benefits of insurance expansion, I focus on the case where η_t increases with policy generosity and generates over-demand for insurance. The sub-optimal insurance take-up by low-risk individuals results in utility loss in $\left.\frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}p}\right|_{u'}$, which I quantify using taste-shock estimates in Saltzman (2018).

5 Estimation

I apply the framework to conduct a cost-benefit analysis of premium subsidy λ_p and mandate penalty k. To estimate the incentive effect of subsidy generosity, I exploit differential subsidy eligibility across demographics within pricing communities.[21] I quantify the incentive effect of mandate penalty based on enrollment and cost variation in the unsubsidized individual market (Hackmann *et al.*, 2015).

This section estimates the incentive effect of subsidy generosity. I instrument subsidy exposure in Massachusetts with simulated generosity measures from a reference national sample not subject to the policy. Estimated effects are comparable to those found under

[20] Unemployed individuals not eligible for ESI extension can also purchase subsidized insurance directly from the Medical Security Program.

[21] Similar identifying strategy has been adopted to study insurer pricing response in Commonwealth Care (Jaffe and Shepard, 2018), and subsidy design in the ACA Exchange (Tebaldi, 2017).

similar contexts in the literature, and a pure calibration exercise would leave welfare analysis largely unchanged. Still, the micro evidence is directly relevant for the policy impact in Massachusetts, as I discuss below.

5.1 Sample summary

Estimation sample consists of adults of age 27-64 living in Massachusetts in 2008-2011 waves of the American Community Survey (ACS). Health insurance variables are added to the survey in 2008, including any insurance, Medicare, Medicaid, other types of public insurance, employer sponsored insurance (ESI), and privately purchased insurance plans. I assume ESI is the primary insurance whenever it is reported.

Individuals report labor force participation–either employed or searching for employment, and current employment status. I focus on any insurance, insurance type (ESI or other), employment, and ESI interacted with employment as main outcome variables relevant for welfare.

Table 2 summarizes the estimation sample.[22] A quarter of the sample does not have ESI, and may qualify for subsidized insurance from Medicaid or the individual market. These individuals are relatively young, single, more likely to be ethnic minorities, and have lower education and income. Average subsidy rate in this group is 68% (69% excluding the uninsured): enrollees pay about 30% of the lowest premium applicable.

5.2 Subsidy rate

I calculate subsidy exposure for all individuals in the estimation sample using the Schedule HC Worksheets and Tables. The document is the official guideline for determining mandate penalty, affordability, and subsidy on Commonwealth Care. Figure 2 shows a snapshot of relevant tables in 2011. The left side determines affordability based on family income. The right side gives the lowest unsubsidized premium across age band and region.

Premium contributed by subsidy enrollees equals affordability.[23] I therefore construct subsidy rate

$$subs = 1 - \frac{affordability}{market\_rate}$$

or the discount provided by the subsidy relative to the full price. I discuss the measurement of the numerator and the denominator below.

Numerator: affordability – Affordability is zero for low-income individuals below 150% FPL. In 2011, single individuals with income less than \$16,344 and married couples with family income less than \$22,068 fall in this range.[24]. At higher income, affordability increases at 200%, 250% and 300% FPL, beyond which no subsidy applies.

[22] I exclude young adults below age 27 eligible for dependent coverage from the ACA dependent mandate (Akosa Antwi *et al.*, 2013). I also exclude 2,863 group-quarter inmates (2.1% of the 27-64 sample). Family income relevant for tax-filing and subsidy application is difficult to construct for these individuals.

[23] Premium contribution is zero for the below 150% FPL group. Between 150-300% FPL, Commonwealth Care provides at least one plan (CeltiCare) that charges subsidized premium at affordability level.

[24] Married couples living in the same household are instructed to combine individual income when calculating affordability and subsidy. The instruction on the Schedule HC reads: "If married filing separately and living in the same household, each spouse must combine their income figures from their separate U.S.

Table 2: Summary Statistics

	Full Sample N=132,360		No ESI N=30,389	
	mean	s.d. error	mean	s.d. error
Demographics				
age	45.39	0.034	44.81	0.074
male	0.48	0.0016	0.48	0.0035
race				
=White	0.83	0.0013	0.73	0.0032
=Black	0.061	0.00086	0.095	0.0021
=other	0.11	0.0011	0.18	0.0028
Hispanic	0.080	0.0010	0.16	0.0027
education				
=less than high school	0.072	0.00093	0.18	0.0028
=high school	0.30	0.0015	0.41	0.0034
=some college	0.62	0.0016	0.41	0.0034
married	0.60	0.0016	0.39	0.0033
have child below 18	0.38	0.0016	0.32	0.0032
Insurance outcome				
have any insurance	0.95	0.00084	0.80	0.0029
have ESI	0.74	0.0015	0	–
Labor/insurance outcome				
in labor force	0.83	0.0012	0.64	0.0033
employed	0.77	0.0014	0.51	0.0035
worked last year	0.83	0.0012	0.62	0.0033
in labor force + ESI	0.66	0.0016	0	–
in labor force + no ESI	0.17	0.0013	0.64	0.0033
not in labor force + ESI	0.077	0.00081	0	–
not in labor force + no ESI	0.094	0.0010	0.36	0.0033
income in % FPL	567.21	1.73	267.06	2.33
subsidy rate	0.29	0.0014	0.68	0.0028
simulated subsidy rate	0.33	0.00074	0.48	0.0016

Notes: Table summarizes estimation sample of non-institutionalized Massachusetts residents aged 27-64 in 2008-2011, adjusted by ACS sampling weights. Income as percentage of FPL is calculated by summing individual income in a tax filing unit, or nuclear families of parents/care-takers and dependent children below age 18. I then apply subsidy schedules to these units to calculate subsidy rate. Simulated subsidy rate is calculated to reflect a schedule's generosity over a fixed national sample. Details are explained in the main text.

Figure 2: Affordability and Premium in 2010 Schedule HC Worksheets and Tables

Table 3: Affordability

Individual or Married Filing Separately (no dependents)		
a. Federal adjusted gross income		b. Monthly premium
From	To	
$ 0	$16,344	$ 0
$16,345	$21,780	$ 39
$21,781	$27,228	$ 77
$27,229	$32,676	$116
$32,677	$39,215	$175
$39,216	$44,443	$235
$44,444	$54,900	$354
$54,901	Any individual with an annual income over $54,900 is deemed to be able to afford health insurance.	

Married Filing Jointly with no dependents or Head of Household/ Married Filing Separately with one dependent		
a. Federal adjusted gross income		b. Monthly premium
From	To	
$ 0	$22,068	$ 0
$22,069	$29,424	$ 78
$29,425	$36,780	$154
$36,781	$44,136	$232
$44,137	$55,113	$315
$55,114	$65,611	$422
$65,612	$86,607	$589
$86,608	Any couple with an annual income over $86,607 is deemed to be able to afford health insurance.	

Married Filing Jointly with one or more dependents or Head of Household/Married Filing Separately with two or more dependents		
a. Federal adjusted gross income		b. Monthly premium
From	To	
$ 0	$ 27,804	$ 0
$27,805	$ 37,068	$ 78
$37,069	$ 46,332	$154
$46,333	$ 55,596	$232
$55,597	$ 73,688	$373
$73,689	$ 94,742	$586
$94,743	$115,796	$849
$115,797	Any family with an annual income over $115,796 is deemed to be able to afford health insurance.	

Table 4: Premiums

Region 1. Berkshire, Franklin and Hampshire Counties			
Age	Individual[1]	Married couple[2] (no dependents)	Family[3]
0–26	$164	$328	$ 846
27–29	$258	$516	$ 875
30–34	$270	$540	$ 887
35–39	$291	$582	$ 887
40–44	$316	$632	$ 922
45–49	$372	$744	$1,011
50–54	$455	$910	$1,137
55+	$455	$910	$1,173

Region 2. Bristol, Essex, Hampden, Middlesex, Norfolk, Suffolk and Worcester Counties			
Age	Individual[1]	Married couple[2] (no dependents)	Family[3]
0–26	$165	$330	$ 719
27–29	$238	$476	$ 719
30–34	$241	$482	$ 860
35–39	$266	$532	$ 899
40–44	$282	$564	$ 952
45–49	$319	$638	$1,061
50–54	$404	$808	$1,255
55+	$416	$832	$1,305

Region 3. Barnstable, Dukes, Nantucket and Plymouth Counties			
Age	Individual[1]	Married couple[2] (no dependents)	Family[3]
0–26	$164	$328	$ 709
27–29	$229	$458	$ 724
30–34	$229	$458	$ 910
35–39	$261	$522	$ 932
40–44	$297	$594	$ 959
45–49	$328	$656	$1,050
50–54	$384	$768	$1,238
55+	$396	$792	$1,269

1. Includes married filing separately (no dependents).
2. Rates for a married couple are based on the combined monthly premium cost of individual plans for each spouse, rather than the cost of a two-person (or self plus spouse) plan.
3. Head of household or married couple with dependent(s).

WS-3

Note: snapshot of page 3 in 2011 Schedule HC Worksheets and Tables, available at https://www.mass.gov/files/documents/2016/08/sr/sch-hc-wksht-tables.pdf

To determine income as percentage FPL, I identify family units relevant for tax and subsidy filing using relationship pointers developed by (Ruggles *et al.*, 2018). Because adult children living with parents cannot be claimed as dependents, I split multi-generation households into single-generation families. I combine total personal income reported by parents to construct family income,[25] which I transform to percentage FPL based on poverty guidelines.[26] Affordability is assigned according to the Worksheets and Tables.

Denominator: market premium rate – Premium is community-rated and varies only by region and age band in a year. In 2011, premium is highest in the Berkshire-Franklin-Hampshire region.[27] Within region, premium is twice as large for the near-elderly (55+) as for young adults (27-29). Married couples combine individual premium for family coverage. I assign premium to individuals based on year, location and age.

I match rating regions to public use micro-data area (PUMA) in ACS. PUMAs are contiguous geographic units build on census tracts.[28] 52 PUMAs in Massachusetts are mapped imperfectly to 14 counties, and then assigned to one of the three rating regions.

When PUMA straddles two rating regions, I calculate average premium weighted by population share cross region, and assign the average to all individuals in the PUMA.[29]. Similar strategy is adopted in Frean *et al.* (2017). In Appendix I show results are unaffected if I assign PUMA to the larger region, or simply drop these PUMAs from the sample.

Comparing with administrative records – Based on calculated subsidy rate in the estimation sample, Commonwealth Care enrollees receive an average subsidy rate of 91% in 2011.[30] Based on administrative records, in 2011, subsidy rate calculated using the lowest priced plan (subsidized premium equals affordability) is $1 - \frac{\$46}{\$405} = 89\%$, fairly similar to the rate calculated from ACS.[31] The small difference may be attributable to young adults 19-26 of age not included in the estimation sample, and to noisier income measures in ACS.

5.3 Empirical strategy

The denominator in the subsidy rate captures policy variation in community rating. The numerator captures policy variation in affordability. Within community, insurer premium does not vary; enrollee cost differs by income according to the affordability schedule.

returns when completing this worksheet." The instruction applies to Line 11: Eligibility for Government Subsidized Health Insurance.

[25] Reported personal income in ACS includes total pre-tax income and losses from all sources, and is used to approximate federal adjusted gross income (AGI). It may differ from AGI because capital losses are not fully deductible from AGI, some social security income is tax-exempt and not included in AGI, and because other AGI deductions are not measured in ACS.

[26] Poverty guidelines are published by the Department of Health and Human Services.

[27] The region has the lowest average premium in 2010, and the median premium across regions in 2009.

[28] PUMAs differ widely in geographic size, but generally have population between 100,000 to 200,000.

[29] Although residents' exact location within PUMA is unknown, the split of PUMA population by county is available from `https://usa.ipums.org/usa/volii/2000pumas.shtml`. The weighting generates 7 new rating regions affecting 14% of the state population.

[30] I identify CommCare enrollees in ACS has having income below 300% FPL, insured from a non-ESI source, and not eligible for Medicaid (below 133% FPL with children).

[31] See Table 1 in Finkelstein *et al.* (2017).

The joint variation suggests incentive effect of subsidy policy is identified off income and demographic differences in subsidy exposure within rating communities, as implemented in Jaffe and Shepard (2018) and Tebaldi (2017).

In the regression analysis, I absorb premium variation in the denominator with community fixed effects, and affordability variation in the numerator with main effects of income. Interactive variation between premium and affordability identifies the causal impact of subsidy exposure. However, with selection responses to affordability, subsidy rate calculated from realized income does not isolate exogenous policy incentives, resulting in biased estimates due to reverse causality.

Selection over affordability may occur, when reported income exhibits excess mass on one side of subsidy eligibility.[32] Among eligibles, selection occurs when behavioral responses lead to changes in income and realized subsidy. For example, if subsidy relieves the "job lock" of near-elderly eligibles, lower retirement income qualifies them for higher subsidy. Furthermore, equilibrium effects on wage, employment and ESI offer bias the incentive measure based on realized income distribution.

I address the selection bias exploiting the fact that remaining states did not undergo the reform – calculated subsidy rate from the national sample is exempt from the selection responses in the Massachusetts sample. The simulation quantifies exogenous, sometimes complex, policy variation in a single parameter (Currie and Gruber 1996a; Currie and Gruber 1996b; Cutler and Gruber 1996). I discuss the application to subsidy generosity below.

5.4 Simulated generosity

I simulate two instruments to measure generosity. The main instrument captures community rating variation *and* affordability across demographics. Specifically, I construct instrument $subiv$ as the follows:

$$subiv_{dapt} = 1 - \frac{1}{|\mathbb{N}_{da}|} \sum_{i \in \mathbb{N}_{da}} \frac{affordability_{it}}{market\_rate_{apt}}.$$

Applying a PUMA(p)-year(t) schedule to the entire simulation sample, I assign market premium based on age, and affordability based on income. I average individual rates within demographic(d)-age(a) to generate $subiv_{dapt}$. The idea is to capture affordability variation through baseline income differences across demographics, rather than selected income responses in Massachusetts.[33]

The simulation sample consists of continental US states in 2005-2006, the baseline economy unaffected by subsidy incentives.[34] Within community apt, generosity varies

[32]Using tax return data, Heim *et al.* (2017) finds bunching below 400% FPL, the phase-out threshold of ACA premium subsidy. At the subsidy phase-in level (138% FPL) in non-expansion states, Kucko *et al.* (2018) finds bunching above the threshold, but does not detect meaningful labor adjustments by wage and salary workers.

[33]For a similar strategy exploiting demographic differences in assets to quantify bankruptcy protection laws, see Mahoney (2015).

[34]I choose this time period to avoid the confounding effects of the economic downturn.

across 144 demographic groups in d.[35] Table 3 shows subsidy is more generous for minorities, the less educated, and singles. Generosity varies substantially more across groups. In Berkshire-Franklin-Hampshire, for example, college-educated White males 30-34 of age married without children receive 7.5% subsidy in 2011 (baseline income 743% FPL). In this age group, African American single mothers with less than high-school education receive 98% subsidy (income 63% FPL).

Despite the rich variation, causal interpretation needs to assume that demographic outcomes would have trended similarly absent the policy. If differences in outcomes are explained by unobserved factors correlated with policy, the instrument is invalid.[36] To assess the extent of omitted variable bias, I consider a second instrument

$$sublean_{apt} = 1 - \frac{1}{|\mathbb{N}_a|} \sum_{i \in \mathbb{N}_a} \frac{affordability_{it}}{market\_rate_{apt}}.$$

where affordability varies over age as with the rating regulation. Within age band, generosity does not differ by demographics. The instrument then identifies policy incentive under weaker assumption than the main instrument, and is plausibly more exogenous.

Between the two instruments, the demographic variation provides over-identification. In specification tests, if the policy incentive appears uncorrelated with unobserved differences by age *and* demographics, then both instruments are likely exogenous. For power concerns I focus more on the main instrument, although results are qualitatively similar using either or both instruments.

I use the 2005-2006 US sample to simulate both $subiv$ and $sublean$. In principle, instruments simulated from the pre-reform Massachusetts sample could similarly address the selection response to subsidy, and possibly provide a stronger correlation with the endogenous exposure in the first stage. In practice, however, because the pre-reform sample does not contain sufficient observations for all demographic groups, the main instrument $subiv$ cannot be computed solely from Massachusetts. I compute the second instrument $sublean\_ma$ using the 2005-2006 sample in Massachusetts. I show over-identified estimates using $sublean\_ma$ as the second instrument in Appendix Table B4.

5.5 Econometric model

I instrument endogenous exposure $subs_{ipt}$ with generosity $subiv_{dpt}$. In the reduced form,

$$\begin{aligned} y_{iapt} = {} & \beta \cdot subiv_{d(i)apt} + \chi_1 \cdot incb_{d(i)} + \rho_a + \phi_p + \tau_t + \rho_{b(a)} \cdot \phi_{r(p)} \cdot \tau_t \\ & + \rho_{b(a)} \cdot \phi_{r(p)} \cdot incb_{d(i)} + \phi_{r(p)} \cdot \tau_t \cdot incb_{d(i)} + \rho_{b(a)} \cdot \tau_t \cdot incb_{d(i)} \\ & + \phi_{r(p)} \cdot \tau_t \cdot X_{d(i)} + \gamma_1 \cdot UE_{b(a)t} + \phi_{r(p)} \cdot \tau_t \cdot X_{d(i)} \cdot UE_{b(a)t} + \epsilon_{iapt} \end{aligned} \tag{17}$$

I control for the mains effects of age ρ_a, PUMA ϕ_p, year τ_t, and baseline income by demographics $incb_{d(i)}$. I further control for all three-way interactions across the four,[37]

[35] I include gender, race (White, Black, other), Hispanic origin, education levels (< high school, high school, some college), marital status and presence of children in the simulation. Hispanic origin leads to small number of observations in certain cells. Dropping Hispanic origin barely changes results.

[36] In this case, fixed effects may control for baseline differences across demographics, but cannot address time-varying confounds correlated with policy.

[37] I use more aggregated level of age band $b(a)$ and region $r(p)$ in the interaction.

Table 3: Demographic variation in subsidy rate

	Observation	subsidy rate		simulated subsidy rate	
		mean	s.d. error	mean	s.d. error
age					
27-29	8,454	0.40	0.0057	0.45	0.0028
30-34	14,340	0.33	0.0043	0.39	0.0024
35-39	15,407	0.30	0.0041	0.35	0.0022
40-44	18,400	0.28	0.0038	0.33	0.0019
45-49	20,440	0.26	0.0035	0.30	0.0018
50-54	20,423	0.26	0.0035	0.29	0.0018
55-64	34,896	0.27	0.0027	0.32	0.0014
male	62,612	0.27	0.0020	0.32	0.00097
female	69,748	0.31	0.0020	0.35	0.0011
race					
=White	113,212	0.25	0.0015	0.30	0.00074
=Black	6,518	0.50	0.0066	0.52	0.0032
=other	12,630	0.46	0.0048	0.47	0.0025
Hispanic origin	8,163	0.59	0.0057	0.61	0.0027
non-Hispanic origin	124,197	0.26	0.0014	0.31	0.00071
education					
=less than high school	7,831	0.69	0.0055	0.74	0.0019
=high school	38,556	0.41	0.0027	0.46	0.0012
=some college	85,973	0.18	0.0015	0.23	0.00057
married	85,843	0.18	0.0015	0.22	0.00065
not married	46,517	0.46	0.0025	0.51	0.0011
have dependent children	51,822	0.28	0.0022	0.32	0.0013
no dependent children	80,538	0.30	0.0018	0.34	0.00091

Notes: Table shows subsidy rate by demographics in the estimation sample. Endogenous subsidy rate is calculated from reported family income. Simulated subsidy rate is calculated from a baseline national sample, and assigned to Massachusetts individuals based on rating community and demographics. Details of the simulation are in the main text.

generating community-level fixed effects and differential income trends over region-year, age-year, and age-region.

To alleviate omitted variable bias, I control for unemployment rate at the same level of the instrument.[38] Specifically, I include age-specific unemployment rate $UE_{b(a)t}$, interacted with level effects of demographic variables $X_{d(i)}$ and region-year indicators. Less aggressive controls of employment shocks yield similar estimates.

Table 4 shows the first stage. All columns control for main effects of PUMA, age, year, and income, demographic variables, and region-year fixed effects. Column 3-4 in addition include unemployment rate interacted with demographic variables. Identifying power in *sublean* weakens with added controls. Column 5 corresponds to the main specification with full set of interactions,[39] where the F-statistic is above 500. Because the first-stage estimate is nearly one, I focus on reduced-form results in what follows.

Table 4: First stage: endogenous subsidy predicted by simulated generosity

	(I)	(II)	(III)	(IV)	(V)
subiv		0.94***		0.94***	0.99***
		(0.044)		(0.045)	(0.042)
sublean	1.04**	0.24	0.86*	0.085	
	(0.42)	(0.43)	(0.47)	(0.48)	
region-year FE	Y	Y	Y	Y	Y
region-year-age FE					Y
UE			Y	Y	Y
F-stat	6.06	228.36	3.41	227.99	549.49
R^2	0.28	0.29	0.28	0.29	0.29

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table shows the first-stage regression of endogenous subsidy rate on simulated instruments. In all specifications I include main effects of PUMA, age, year, and income, region-year fixed effects, and demographic variables. Column 3-4 additionally include age-band unemployment rate interacted with demographic variables. Column 5 corresponds to the main specification with full interaction terms. Robust standard errors clustered at the level of PUMA in the parenthesis.

[38] If the macroeconomic impact across demographics is intermediated by the subsidy policy–for example, individuals with higher subsidy are less likely to lose coverage when they lose jobs, then the instrument is correlated with employment shocks in the error term.

[39] In this specification, community-level fixed effects sweep out the variation in the lean instrument. I hence base over-identified results on specifications in column 3-4.

5.6 Results

Table 5 shows estimated effect of subsidy incentives. In Panel A, endogenous exposure is significantly associated with lower insurance coverage and employment. In particular, employment would drop by 4 percentage points following a ten percentage point increase in subsidy. In panel B, reduced-form estimates suggest a modest effect on insurance take-up, and a null effect on employment.

Labor participation is similarly unresponsive to subsidy. Stratified by year, the take-up effect is small in the first two years but increases over time.[40] In 2011, ten percentage point increase in subsidy raises coverage by 1.7 percentage points. Employment outcomes vary slightly over years, but on average, evidence of significant labor supply and employment effect of subsidy is minimal.

ESI coverage of workers decreases by 0.26 percentage point per subsidy percent. Relative to the OLS estimate, the reduced form suggests larger selection on ESI benefit rather than employment. Appendix Table B1 further suggests ESI benefit decreased more for young workers than middle-age workers. At the near-elderly, employment selection becomes significant, and ESI decreases more. The overall pattern is consistent with sorting of young workers to small firms less affected by the mandate (Aizawa, 2017), and with the "retirement lock" of ESI alleviated by pubic subsidy.

ESI decreases more for the non-employed. To the extent that insurance is sponsored by previous employers for which unemployed enrollees pay full premium, or receive state subsidy, actual selection from private transfer is smaller. Measuring λ_1^{1-e} as ESI enrollees out of work for at least a year (and hence less likely enrolled in ESI continuation plans), estimated selection is -0.28 percentage point per subsidy percent. For enrollees out of work for five years, the estimate is further lower at -0.19 (Appendix Table B2). I calculate welfare based on different estimates of private transfer crowd-out.

5.7 Robustness

The instruments pass the joint exogeneity test in Panel C of Table 5. The test is based on the specification with main effects of location, year, age, and demographic variables interacted with unemployment rate. Appendix Table B3 presents similar results for basic specification without unemployment controls, and separate 2SLS estimates using either instrument. Although the weaker instrument produces larger and less precise estimates, over-identified estimates appear similar across specifications.

Appendix Table B4 examines the robustness of over-identified results when the second instrument is simulated from the Massachusetts sample. Due to sample size limitation, I only simulate generosity across age, location, and year, but not across demographics. When instrumented jointly with *sublean*, the effects are larger but none significant, similar to results in Appendix Table B3. When instrumented jointly with *subiv*, effects on insurance and employment outcomes are comparable to those in Panel C, Table 5. Effect on ESI selection is concentrated in the non-employed, but insignificant for workers, different from the main results in Table 5. For the calculation of welfare, I follow the main estimates that

[40] Studying the ACA Exchange, Frean *et al.* (2017) shows subsidy increased take-up by 0.051 percentage point per subsidy percent in 2014, and 0.089 in 2015.

Table 5: Estimated incentive effect of subsidy

	(I) any insurance	(II) employed	(III) in labor force	(IV) ESI + employed	(V) ESI + not employed
			Panel A: OLS		
subs	-0.071***	-0.41***	-0.30***	-0.55***	0.046***
	(0.0032)	(0.0071)	(0.0066)	(0.0071)	(0.0045)
R^2	0.083	0.21	0.18	0.29	0.054
			Panel B: reduced form		
subiv	0.10***	-0.055	-0.056	-0.26***	-0.33***
	(0.025)	(0.053)	(0.044)	(0.058)	(0.023)
R^2	0.071	0.090	0.10	0.13	0.054
2008	0.034	-0.049	-0.017	-0.27***	-0.37***
	(0.051)	(0.088)	(0.084)	(0.10)	(0.044)
2009	0.072*	-0.0063	-0.075	-0.22***	-0.38***
	(0.041)	(0.081)	(0.062)	(0.081)	(0.042)
2010	0.11***	-0.095	-0.074	-0.30***	-0.29***
	(0.037)	(0.063)	(0.055)	(0.067)	(0.036)
2011	0.17***	-0.059	-0.053	-0.25***	-0.31***
	(0.047)	(0.083)	(0.070)	(0.087)	(0.036)
			Panel C: over-identified 2SLS		
$\widehat{subs}$	0.16***	-0.056	-0.046	-0.32***	-0.26***
	(0.037)	(0.057)	(0.047)	(0.046)	(0.029)
F-stat	227.99	227.99	227.99	227.99	227.99
p-value	0.92	0.46	0.92	0.19	0.44
y mean	0.95	0.77	0.83	0.64	0.10

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table shows estimated incentive effect of premium subsidy. Panel A shows OLS estimates using endogenous subsidy rate *subs*. Panel B shows reduced-form effect of simulated generosity *subiv*, and year-specific effects interacting *subiv* with year dummies from a separate regression. Panel A and B are based on the main specification with full interaction terms. Panel C shows 2-stage least square estimates instrumenting *subs* with *subiv* and *sublean*, based on a specification with main effects of PUMA, age, year, income, and demographic variables interacted with unemployment rate. P-value from Hansen over-identification test is reported. Robust standard errors clustered at the level of PUMA in the parenthesis.

likely over-state the selection effect on private insurance. I then vary the size of selection and the case where subsidy does not affect ESI coverage in the robustness analysis.

Appendix Table B5 shows robustness to alternative treatment of border PUMAs. The main analysis first averages premium across rating regions before calculating subsidy. Results are nearly identical if subsidy is calculated for each region and then averaged. Assigning the PUMA to region with larger population yields similar estimates. Moreover, results remain significant at conventional level with standard error clustered at the level of region and age band.

I conduct permutation tests over the remaining 50 states that did not undergo the reform. Specifically, I randomly assign Massachusetts rating communities across location, year and age band in control states, and estimate pseudo policy effect using simulated generosity $sublean$. Appendix Figure C1 plots the empirical distribution of reduced-form estimates across states. Appendix Figure C2 plots the distribution based on $subiv$, where I also permute Massachusetts affordability across income (demographics). In most cases, the true effect in Massachusetts is significant at 95% level.

6 Calculation

I calibrate the enrollment effect on formal insurance pricing $\varepsilon_{r,\lambda_0}$ and uncompensated care $\varepsilon_{ri,\lambda_0}$ using administrative data on marginal and average costs. I calibrate insurer pricing distortion due to subsidy using simulated evidence from Jaffe and Shepard (2018), the credit channel from Brevoort *et al.* (2017), and consumption from the Consumer Expenditure Survey (CEX).

Pricing effect of subsidy – I focus on the Commonwealth Care program for the effect of subsidy on pricing. Enrollees face discrete changes in subsidy when income crosses thresholds at 150% FPL, 200% FPL, and 250% FPL. In 2011, larger subsidy below 150% FPL increased take-up from 70% to 94%, and lowered average monthly cost from \$380 to \$333 (**?**). Cost of new enrollees is $\frac{\$333\cdot 94\% - \$380\cdot 70\%}{94\% - 70\%} = \$196$ at 150% FPL. Similarly, cost of new enrollees is \$268.2 at 200% FPL, and \$280.86 at 250% FPL.

Across thresholds, new enrollees lower average cost by

$$\mathrm{d}r = \frac{\$377.3 - \$232.5 * 0.05\%}{94\% - 0.05\%} - \$377.3 = \$0.07$$

where \$377.3 is the cost of existing enrollees, and \$232.5 the cost of new enrollees. At 94% coverage baseline in the 19-64 population, implied cost elasticity is small

$$\varepsilon_{r,\lambda_0}^{\lambda_p} = \frac{\mathrm{d}r}{\mathrm{d}\lambda_0} \cdot \frac{\lambda_0}{r} = \frac{\$0.07}{0.05\%} \cdot \frac{6\%}{\$377.3} = 0.023$$

Recall from Equation 5 that premium also responds to mark-up adjustment subsidy:

$$\frac{\mathrm{d}p}{\mathrm{d}\lambda_p} = \varepsilon_{r,\lambda_0}^{\lambda_p} \cdot \frac{r}{\lambda_0} \cdot \left[1 + \beta + r \cdot \frac{\partial \beta}{\partial r}\right] \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p} + r \cdot \frac{\partial \beta}{\partial \lambda_p}$$

Two pieces of evidence potentially inform the magnitude of $\frac{\partial\beta}{\partial r}$ and $\frac{\partial\beta}{\partial\lambda_p}$. Jaffe and Shepard (2018) shows price-linked subsidy increased average CommCare premium by 5.4%, relative to a fixed subsidy. The distortion is largest for the cheapest plan directly linked to subsidy, where mark-up increased from 17.6% to 25%, the maximum allowed by law.[41] Average mark-up increased from 12.40% to 13.55% with linkage, implying $\frac{\partial\beta}{\partial\lambda_p} = \frac{1.15\%}{71\%} = 0.016$.

Hackmann *et al.* (2015) finds both mark-up and cost decreased in the unsubsidized individual market, implying $\frac{\partial\beta}{\partial r} > 0$. However, risk pooling with the small-group market may also lower mark-up in the individual market.[42] Evaluated at the lowest premium in 2011, where mark-up binds at the maximum 25%, and $\frac{\partial\beta}{\partial r} = \frac{\partial\beta}{\partial\lambda_p} = 0$ by the rating regulation, price effect

$$\frac{\mathrm{d}p}{\mathrm{d}\lambda_p} = 0.023 \cdot \frac{\$323.04}{6\%} \cdot \left[1 + 25\%\right] \cdot (-0.17) = -\$26.31$$

Implied saving is $\frac{\mathrm{d}p}{\mathrm{d}\lambda_p \cdot p} = -0.065$ per subsidy dollar.[43]

For average enrollee premium (\$424) and mark-up (11%), premium saving is similarly small at $\frac{0.023}{6\%} \cdot (-0.17) + \frac{0.016}{1+11\%} = -0.051$ per subsidy dollar, assuming $\frac{\partial\beta}{\partial r} = 0$. I shut down the premium response and calibrate $\frac{\mathrm{d}\log p}{\mathrm{d}\lambda_p} = 0$ in the baseline analysis. I re-calibrate the premium benefit and compare with uncompensated care for a range of mark-up responses in the discussion of reform rationales.

Pricing effect of mandate – Enrollment in unsubsidized individual market increased from 1% in 2006 to 3% by 2010.[44] Both the mandate penalty and the rating regulation potentially contribute to the increase with lower (net) cost of coverage. The maximum effect attributable to penalty is $\frac{\mathrm{d}\lambda_0}{\mathrm{d}k} = \frac{-2\%}{50\%} = -0.04$, where penalty is half of the cheapest plan premium.

Average cost decreased by 8.7% when take-up increased by 26.5 percentage points in the individual market (Hackmann *et al.*, 2015).[45] New enrollees cost $-\frac{\mathrm{d}r}{\mathrm{d}\lambda_0} = -\frac{\mathrm{d}\log r}{\mathrm{d}\lambda_0} \cdot r = \frac{8.7\%}{26.5\%} \cdot \frac{\$5{,}720}{12} = \$144.18$, one-third the baseline average cost $r = \frac{\$5{,}720}{12} = \$439.17$, and 40% less costly than new enrollees in the subsidized program.

Absent the mandate, if high-income low-cost individuals did not enroll, average cost would be higher by $\frac{\$377.3 \cdot 94\% - \$144.18 \cdot 2\%}{94\% - 2\%} - \$377.3 = \$5.07$. Cost elasticity of mandate $\varepsilon^k_{r,\lambda_0} = \frac{\$5.07}{2\%} \cdot \frac{6\%}{\$377.3} = 0.040$. Assuming penalty has no direct effect on pricing ($\frac{\partial\beta}{\partial k} = 0$), and the smaller mark-up is completely due to risk pooling (hence $\frac{\partial\beta}{\partial r} = 0$), premium decreases by

41 The minimal medical loss ratio law requires at least 80% of premium spent on medical claims, or a maximum mark-up of 25%.

42 As part of the 2006-2007 reform, Massachusetts pooled the risk base of small-group and non-group market. The pooling significantly reduced the premium of individual coverage, with much smaller effect on group insurance premium.

43 Premium of cheapest plan (CeltiCare) is \$403.8 per month, with cost \$323.04 (Jaffe and Shepard, 2018).

44 I divide enrollment counts in Key Indicators, available at `http://archives.lib.state.ma.us/bitstream/handle/2452/112747/ocn232606916-2011-05.pdf`, by population estimate from ACS.

45 The take-up estimate is specific to eligible individual market enrollees not offered ESI.

$\frac{\mathrm{d}p}{\mathrm{d}k \cdot p} = \frac{\varepsilon^k_{r,\lambda_0}}{\lambda_0} \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}k} = \frac{0.040}{6\%} \cdot (-0.04) = -0.027$ per subsidy dollar. The baseline analysis calibrates $\frac{\mathrm{d}\log p}{\mathrm{d}k} = 0$. I introduce the premium benefit and compare with uncompensated care in the discussion of reform rationales.

Uncompensated care – In 2011, uncompensated care to the uninsured and the underinsured totals \$496 million, of which 90% is spend on the 19-64 population, and 85% of beneficiaries have no primary source of insurance.[46] Average uncompensated cost is \$133.47 per uninsured.[47]

Assuming a 25% moral hazard of insurance, or 80% actuarial value of uncompensated care ($g = 0.8$),[48] average cost of enrolling the uninsured is $ri = \frac{\$133.47}{0.8} = \$166.84$. Subsidy enrollment lowers the cost by $\frac{\$166.84*6\%+\$232.5*0.05\%}{6\%+0.05\%} - \$166.84 = \$0.54$. Cost elasticity $\varepsilon^{\lambda_p}_{ri,\lambda_0} = \frac{\$0.54}{0.05\%} \cdot \frac{6\%}{\$166.84} = 0.39$. Mandate penalty lowers the cost by $\frac{\$166.84*6\%+\$144.18*2\%}{6\%+2\%} - \$166.84 = -\$5.67$, with cost elasticity $\varepsilon^{k}_{ri,\lambda_0} = \frac{-\$5.67}{2\%} \cdot \frac{6\%}{\$166.84} = -0.10$.

Around one-third of uncompensated cost is financed by patient surcharge $uc_p = \alpha \frac{\lambda_0}{1-\lambda_0} \frac{ri}{r} g\, OOC$, where $\alpha = 0.32$,[49] and $\frac{\lambda_0}{1-\lambda_0} \frac{ri}{r} = 0.028$ is the cost share per patient. Formal insurance enrollment lowers patient surcharge by 40% every ten percent subsidy,[50] whereas the effect on premium is only $\frac{\mathrm{d}\log p}{\mathrm{d}\lambda_p} \cdot 0.10 = -0.5\%$. A ten percent penalty lowers surcharge by 6.2% and premium by 0.27%, where new enrollees are less costly than the average uninsured.

The remaining cost is financed by state and federal funds (up to \$100 million), and assessment on provider.[51] Avoided externality on government budget and hospital profit is $(1-\alpha) \cdot g \cdot ri \cdot (1 + \varepsilon_{ri,\lambda_0})$ per enrollee, or 39% (6%) per ten percent subsidy (penalty).

Credit channel – Health insurance reduces medical debt, and improves consumption opportunity cross periods through the household balance sheet.[52] Moreover, interest payment on non-medical debt decreases as credit rating improves. I calibrate both effects based on Brevoort *et al.* (2017).

With optimal choice of medical debt, repayment generates no social benefit. Nonetheless, liquidity benefit to households is non-trivial. In the ACA context, quarterly saving of new medical debt is \$13.50 in states that expanded Medicaid, an increase in liquidity of

[46] Detailed accounts of cost and funding sources are in the annual report of Health Safety Net (HSN), the primary payer of uncompensated care in Massachusetts. The 2011 report is available at `https://www.mass.gov/files/documents/2016/07/tp/hsn11-ar.pdf`. 15% of the HSN cost is supplemental benefit to Medicaid, CommCare and private insurance enrollees. Adjusting for the supplement would increase the cost of an average insured by \$1.48.

[47] Total cost $\$496M * 90\% * 85\%$ divided by uninsured months $12 * 236,902$.

[48] The estimate is consistent with spending response to cost-sharing in Commcare (Chandra *et al.*, 2014), and experimental evidence from Medicaid lottery (Finkelstein *et al.*, 2012).

[49] Statutory contribution of service charge is \$160 million, about $\frac{\$160}{\$496} = 32\%$ of total cost.

[50] Note that $\frac{\mathrm{d}\log uc_p}{\mathrm{d}\mathbf{K}} = \frac{\mathrm{d}\log uc_p}{\mathrm{d}\lambda_0} \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}\mathbf{K}}$, where $\frac{\mathrm{d}\log uc_p}{\mathrm{d}\lambda_0} = \frac{1}{\lambda_0}\left(\frac{1}{1-\lambda_0} + \varepsilon^{\mathbf{K}}_{ri,\lambda_0} - \varepsilon^{\mathbf{K}}_{r,\lambda_0}\right)$.

[51] I only consider the mechanic cost externality on third-party payers. If the externality distorts hospital mark-up on reimbursed services, or crowds-out government funds that generate greater social value–e.g., investment in early education and transfer program to the poor, then the social cost of uncompensated care is under-estimated.

[52] Consumption increases less than the liquidity of $bd_t(\widetilde{\omega}_t)$, depending on non-medical borrowing adjustments $\dot{A}_t(\widetilde{\omega}_t)$.

$\frac{\$13.50}{3}\frac{1}{4.4\%}\frac{0.17}{\$424} = \frac{\$17.39}{\$424} = 0.041$ per subsidy dollar.[53]

The externality on non-medical interest payment is welfare-relevant, when individuals do not internalize the cost on credit rating and price. Rating improvement from Medicaid expansion lowers annual interest payment by \$14.60, a monthly saving of $\frac{\$14.60}{12}\frac{1}{4.4\%}\frac{0.17}{\$424} = 0.011$ per subsidy dollar. The calibration does not incorporate dynamic effect on net worth,[54] and may understate the credit benefit.

Consumption – To monetize welfare, price externalities are weighted by marginal utility relative to the reference group (workers). Assuming constant relative risk aversion (CRRA), $u'(c) = c^{-\gamma}$, welfare weight depends on consumption ratio given calibration of γ.

I measure consumption of 27-64 Massachusetts residents in the 2011 CEX panel. I infer insurance (ESI, public, uninsured) from premium expenditure, and hospital utilization ($g = 0$) from medical expenses.[55] Appendix Table B6 shows average non-medical consumption for beneficiary groups. Based on the statistics, private transfer externality is weighted by $1.03^{-\gamma}$, subsidy by $0.64^{-\gamma}$, penalty by $0.57^{-\gamma}$, surcharge by $0.77^{-\gamma}$, and interest saving by $0.89^{-\gamma}$.

More generally, society may approach insurance externalities based on concerns other than differences in marginal utility of consumption. In this case, a different set of general social welfare weights applies (Saez and Stantcheva, 2016). I discuss implications of different motivations and measurements of welfare weights.[56]

Friction – I focus on micro-level friction that places a wedge between consumption smoothing benefit of insurance (the welfare metric) and private utility generating observed patterns of take-up. Individuals optimize based on private utility, but welfare may increase or decrease depending on the direction of friction response to policy. I focus on the case of over-insurance, or potential welfare loss for marginal enrollees, to assess the robustness of standard welfare calculation from a frictionless model.

I assume over-insurance is driven by "taste for complicance" that increases with policy generosity for a set of low-risk individuals. Saltzman (2018) estimates the taste shock in terms of WTP: the existence of mandate penalty increases WTP for insurance by \$13 (3% of premium) in the Washington Exchange, and \$64 (15% of premium) in California. The taste subsidy on demand is $\frac{3\%}{71\%} = 4.2\%$ the magnitude of premium subsidy in Washington, and 21% in California. Assuming $\theta_{\lambda_p}^{>0} \in [4.2\%, 21\%]$ of the total effect on take-up is driven by taste, welfare loss is

$$-\theta_{\lambda_p}^{>0} \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p} \cdot \left(\frac{u'(c_{.0.}) \cdot (1 - \lambda_p - k) \Big|_{\theta_{\lambda_p}^{>0}}}{u'(c_{1..})} + \frac{u'(c_{.01}) \cdot (1-\mu) \cdot \frac{bd_t + uc_p - (1-g)nM}{p} \Big|_{\theta_{\lambda_p}^{>0}}}{u'(c_{1..})} \right)$$

[53] The average saving is driven by a 4.4% take-up, adjusted by the subsidy effect on take-up.

[54] The calculation in Brevoort *et al.* (2017) holds fixed the contract term and the debt structure, simplifying from dynamic adjustments. When choices of medical and non-medical borrowing are assumed optimal, dynamic externality from interest saving is likely small.

[55] CEX estimate of insurance is 94%, and 12-month employment rate 82%, both very similar to ACS estimates (Table 2). Hospital utilization in CEX is 7.7%.

[56] For example, using food expenditure variation to reflect redistributive preference for subsistence consumption, social weight of externality is generally smaller (Appendix Table B6).

where I Taylor-expand the utility of taste-based enrollees at the optimal consumption when uninsured. Friction reduces consumption by the net cost of subsidized premium $(1-\lambda_p-k)\,p$. Utility change in the sick state is small if taste-based enrollees are characterized with low risk and low medical debt.[57] Focusing on the premium cost on consumption, the welfare loss is

$$-\,\theta_{\lambda_p}^{>0}\cdot\frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p}\cdot\frac{u'(c_{\cdot 0\cdot})\cdot\left(1-\lambda_p-k\right)\Big|_{\theta_{\lambda_p}^{>0}}}{u'(c_{1\cdot\cdot})} \;=\; 21\%\cdot 0.17\cdot 0.63^{-\gamma}\cdot\frac{1}{2} \;=\; \frac{0.018}{0.63^{\gamma}} \tag{18}$$

where (non-medical) consumption is evaluated at the uninsured average, subsidy assumed zero, and $\theta_{\lambda_p}^{>0}=21\%$ to over-state the welfare loss.

For mandate penalty, a 15% taste subsidy is $\theta_k^{>0}=\frac{15\%}{50\%}=30\%$ the magnitude of penalty on premium in the unsubsidized population. Friction is calculated as

$$-\,\theta_{k}^{>0}\cdot\frac{\mathrm{d}\lambda_0}{\mathrm{d}k}\cdot\frac{u'(c_{\cdot 0\cdot})\cdot\left(1-\lambda_p-k\right)\Big|_{\theta_{k}^{>0}}}{u'(c_{1\cdot\cdot})} \;=\; 30\%\cdot 0.04\cdot\frac{1}{2}\cdot 0.37^{-\gamma} \;=\; \frac{0.006}{0.37^{\gamma}} \tag{19}$$

using non-medical consumption.

7 Welfare

7.1 Subsidy

Absent friction to marginal enrollees, subsidy affects welfare through premium assistance to recipients, uncompensated care to third-party payers, interest charges on the uninsured, and the fiscal cost to taxpayers. Table 6 calculates welfare ignoring the efficiency gain on insurance premium. Welfare weights vary by the curvature in the utility function (γ), reflecting the value society derives from moving a dollar from high-income to low-income individuals. Given γ, society may value transfer to cover subsistence needs more than additional transfers–as an example, I consider food consumption.

With linear utility in consumption, social value of redistribution is zero. Compared to the fiscal cost of public funds, and the moral hazard of formal insurance, benefits to subsidy recipients and uncompensated care payers do not outweigh the cost. Further accounting for the external benefit on credit still yields a small welfare loss of 0.02 per dollar.

The baseline calculation does not consider the premium response to enrollment, and over-states the efficiency cost of subsidy. Nonetheless, a modest dose of equity consideration offsets the baseline cost. Fixing curvature at $\gamma=1$, when redistribution is based on total non-medical consumption, recipient valuation of a subsidy dollar, net of the welfare loss to taste-based enrollees, outweighs the fiscal cost.

[57] When $bd_t\big|_{\theta_{\lambda_p}^{>0}}\approx 0$, because formal insurance is more generous, $(1-g)\,nM>uc_p$, consumption increases for taste-based enrollees in bad health state.

Table 6: Welfare effect of premium subsidy

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$	(III) $\frac{dW_\Pi}{d\lambda_p \cdot p}$	(IV) $\frac{dW_C}{d\lambda_p \cdot p}$	(V) (1)+(2)+(3)+(4)	(VI) $\frac{dW_F}{d\lambda_p \cdot p}\Big\|_{u'}$	(VII) $\frac{dW}{d\lambda_p \cdot p}$
			Panel A: non-medical consumption				
$\gamma = 0$	0.232	0.074	0.011	-0.338	-0.021	-0.018	-0.039
$\gamma = 1$	0.368	0.082	0.013	-0.342	0.121	-0.028	0.093
$\gamma = 2$	0.585	0.092	0.014	-0.346	0.345	-0.045	0.3
$\gamma = 3$	0.928	0.105	0.016	-0.350	0.699	-0.071	0.628
			Panel B: food consumption				
$\gamma = 1$	0.249	0.079	0.011	-0.339	0	-0.018	-0.018
$\gamma = 2$	0.268	0.085	0.012	-0.339	0.026	-0.019	0.007
$\gamma = 3$	0.288	0.092	0.012	-0.340	0.052	-0.021	0.031

Notes: Table calculates the benefit of subsidy dollar to enrollees in column 1 (Equation 8), to uncompensated care payers in column 2 (Equation 7), interest saving in column 3 (Brevoort *et al.*, 2017), and the fiscal cost of financing in column 4 (Equation 9). Column 5 calculates the net benefit in column 1 to 4. Column 6 calculates the utility loss from over-insurance (Equation 18), where the preference shock for compliance increases with subsidy according to Saltzman (2018) . Total welfare effect in column 7 is intended as a lower bound. Welfare weights varies by the curvature parameter γ in the utility, and by the consumption base considered appropriate for redistribution.

When redistribution is based on food consumption only (Panel B), social benefit to recipients is perceived less than the cost to taxpayers, and redistribution is not a motivating rationale for subsidy. In this case, social cost of uncompensated care becomes a crucial element in the justification. In Panel B, saving in patient surcharge and provider assessment generates about one-third the direct benefit to enrollees, and roughly closes the gap with fiscal cost at low preference for redistribution ($\gamma = 2$).

Accounting for the interest saving to the uninsured rationalizes subsidy cost for $\gamma = 1$, although welfare loss from taste-based enrollment potentially offsets the benefit. Appendix Table B7 varies the incidence of uncompensated cost. Welfare benefit is largest with complete pass-through to patient, and smaller when uncompensated cost is financed by a premium tax on enrollees. Regardless of incidence, total welfare effect (including friction) balances out under conservative welfare weights for $\gamma = 2$.

7.2 Mandate penalty

I calculate welfare effect of mandate penalty for the unsubsidized population (income above 300% FPL), where subsidy is zero ($\lambda_p = 0$) and penalty is half the lowest premium ($k = \frac{1}{2}$). As with subsidy, the baseline calculation ignores the premium response to enrollment, and over-states the efficiency cost of policy. For the unsubsidized group, I further ignore any credit benefit to the uninsured.[58] Table 7 calculates the direct cost of penalty on the uninsured in column 1, saving to uncompensated care payers in column 2, and the fiscal impact on government budget in column 3.[59]

Absent redistribution concerns ($\gamma = 0$), a dollar increase in penalty raises government revenue by 0.003 dollar. Efficiency cost of new enrollees $k\frac{\mathrm{d}\lambda_0}{\mathrm{d}k} = -0.02$ nearly offsets the penalty increase on the uninsured.[60] Benefit to third-party payers is small relative to the subsidized population, reflecting the low cost of coverage in the high-income uninsured. Including this benefit still leaves out a net welfare loss of 0.01 per dollar. When uncompensated care is more restricted to the low-income uninsured, the efficiency loss increases.

With $\gamma > 1$, welfare further trades-off the dollar cost of public fund on the uninsured versus taxpayers. If payroll tax on workers is deemed more desirable than penalty, or if the uninsured receive higher welfare weights (based on non-medical consumption in Table 7), then welfare decreases with penalty. In contrast, food consumption implies public fund is slightly less costly if financed by penalty. The small redistribution value, however, does not offset efficiency cost: net effect on welfare is a 0.006 loss per dollar for $\gamma = 2$, similar to the case when $\gamma = 0$.

Therefore equity concerns and uncompensated care do not rationalize the baseline efficiency cost of penalty. The remaining gap (0.015 net of friction when $\gamma = 0$) is small, and is potentially offset by the efficiency gain in insurance premium excluded from the baseline analysis. I discuss the relative importance of uncompensated care and premium response to the efficiency argument of policies below. Appendix Table B9 shows the baseline calculation is robust to alternative incidences of uncompensated care.

7.3 Robustness

The baseline analysis calculates welfare for the 27-64 age group, combining ACS estimates with calibrated values from the literature. The robustness analysis conducts a full calibration exercise for the 18-64 age group based on (the range of) estimates in the literature. I continue to exclude any premium response to enrollment, but re-calculate the fiscal cost varying the incentive effects on insurance and employment, and the uncompensated care benefit varying the spending response to formal insurance.

7.3.1 Fiscal cost of subsidy

The fiscal cost of a dollar subsidy is calculated based on Equation 9. In the main analysis in Table 6, I calibrate $\frac{\mathrm{d}e}{\mathrm{d}\lambda_p} = 0$. Alternatively, I benchmark the employment response to the effect on take-up, and consider cases where zero (baseline), 50%, and 100% of new

[58] Credit cost borne by the uninsured is less than 20% the medical cost borne by third-party payers. In the high-income group, social cost of uncompensated care is calculated to be small, and the effect on credit is negligible.

[59] See Appendix Table B8 for consumption and associated welfare weights in this sample.

[60] In practice, contribution of mandate penalty to the Commonwealth Care trust fund is less than 1.5%.

Table 7: Welfare effect of mandate penalty

	(I) $\frac{dW_B}{dk \cdot p}$	(II) $\frac{dW_{UC}}{dk \cdot p}$	(III) $\frac{dW_C}{dk \cdot p}$	(IV) (1)+(2)+(3)	(V) $\frac{dW_F}{dk \cdot p}\Big\|_{u'}$	(VI) $\frac{dW}{dk \cdot p}$
			Panel A: non-medical consumption			
$\gamma = 0$	-0.023	0.011	0.003	-0.009	-0.006	-0.015
$\gamma = 1$	-0.062	0.013	0.003	-0.046	-0.016	-0.062
$\gamma = 2$	-0.168	0.014	0.003	-0.151	-0.044	-0.195
$\gamma = 3$	-0.454	0.016	0.003	-0.435	-0.118	-0.553
			Panel B: food consumption			
$\gamma = 1$	-0.022	0.012	0.003	-0.007	-0.006	-0.013
$\gamma = 2$	-0.021	0.012	0.003	-0.006	-0.006	-0.012
$\gamma = 3$	-0.020	0.013	0.003	-0.004	-0.005	-0.009

Notes: Table calculates the welfare effect of mandate, restricted to the unsubsidized population with income above 300% FPL. Column 1 calculates the penalty cost to the uninsured (Equation 13), column 2 calculates the benefit to uncompensated care payers (Equation 7), and column 3 the fiscal cost to the government (Equation 14). Column 4 gives the net benefit in column 1 to 3. Column 5 calculates the welfare loss from over-insurance (Equation 19), where the preference shock for compliance increases with penalty according to Saltzman (2018). Total welfare effect in column 6 is intended as a lower bound. Welfare weights varies by the curvature parameter γ in the utility, and by the consumption base considered appropriate for redistribution.

enrollees reduced employment. Calibrated $\frac{\mathrm{d}e}{\mathrm{d}\lambda_p} = 0$, -0.08, and -0.16, respectively.[61]

I calibrate $\frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p}$ based on the take-up response in the low-income eligible population not enrolled in ESI (?). Take-up is 19.3% lower across income thresholds where subsidy decreases by ten percentage points. Implied incentive effect is $\frac{19.3\%}{10\%} = 1.93$ per subsidy percent in the eligible population, and $-\frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p} = \frac{19.3\% * 8.3\%}{10\%} = 0.16$ in the 19-64 population, where 8.3% are eligible for subsidy.[62] The calibration is similar to the ACS estimate in 2011 (-0.17). Implied cost of new enrollees is $(\lambda_p + k) \cdot \frac{\mathrm{d}\lambda_0}{\mathrm{d}\lambda_p} = 0.139$.[63]

Moreover, selection from ESI increases taxpayer cost by $\lambda_p - \tau_1$, and decreases private payer cost by $1 - \tau_1$. With welfare weight $\frac{u'(c_{11.})}{u'(c_{1..})} \approx 1$, cost of insurance to society is lower by $1 - \lambda_p$, the cost sharing of subsidized enrollees. Switchers who are payers of private transfer (λ_1^e) increase the transfer cost τ_{pr} on remaining payers, mitigated by larger tax rebate τ_1.

τ_1 is calibrated at 62% in the main analysis, based on average income of Massachusetts ESI payers ($70,000 in 2011) and simulated ESI exemption in Gruber (2010).[64] Calibrated τ_1 is similar for average workers or ESI enrollees in Massachusetts. $\frac{\mathrm{d}\lambda_1^e}{\mathrm{d}\lambda_p} = -0.25$ in the ACS sample, $\frac{\mathrm{d}\lambda_1^{1-e}}{\mathrm{d}\lambda_p} = -0.27$ for ESI beneficiaries (enrollees not in labor force), and $\frac{\mathrm{d}\lambda_1}{\mathrm{d}\lambda_p} = -0.52$ in the main analysis.[65] Given null effect on employment, I calculate cost under alternative crowd-out in the robustness analysis.

Appendix Table B10 calculates welfare effect of subsidy, varying the cost calibration. With null effect on employment, result is comparable to the 27-64 group in Table 6. Employment response half the size of take-up increases taxpayer cost by 0.064 per subsidy

[61] Estimated effect of subsidy on employment tends to be small, except for sub-groups such as the near-elderly. One of the larger estimates suggests childless adults losing Medicaid eligibility in Tennessee (Garthwaite *et al.*, 2014) increased employment by 4.6 percentage points. In the Massachusetts context, losing subsidy on average decreases employment by $\frac{1}{2} * 74 * 0.16 = 5.92$ percentage points when employment effect is half the size of take-up, already larger than most estimates in the literature.

[62] Following ?, I construct eligible population in ACS as having income between 135% and 300% FPL and not enrolled in ESI. Eligibles are $\frac{333{,}964}{4{,}007{,}524} = 8.3\%$ of the 19-64 population. Average subsidy is 74% for those who enroll.

[63] Note that $k = \frac{1}{2}(1 - \lambda_p) = 0.13$ by regulation. Penalty is set at half of affordability, and subsidy lowers enrollee contribution to equal affordability.

[64] ESI exemption is also calculated in *Tax Expenditures for Health Care* submitted to the Senate Commission of Finance for the hearing "Health Benefits in the Tax Code: The Right Incentives" (https://www.jct.gov/publications.html?func=startdown&id=1193). ESI return to the $ 50,000-$74,999 income group is $3,106, or $\frac{\$3{,}106}{\$424*12} = 61\%$ of premium.

[65] ESI selection on the extensive margin is $-0.74 * 0.52 = -0.38$ per subsidy enrollment. The magnitude is in the middle-to-lower range of estimates surveyed in Gruber and Simon (2008), possibly due to the employer mandate. The remaining selection reflects (small) firm ESI offer and worker sorting across firm size and benefits (Aizawa, 2017). Sommers *et al.* (2018) provides empirical evidence from Massachusetts consistent with firm response to mandate and subsidy, and worker sorting (low ESI take-up) when eligible for subsidy (see also Lyons (2017) for increased subsidy take-up in Massachusetts small firms). Including ESI enrollees (59%) as eligibles (?), take-up inclusive of selection is approximately $(1 - 59\%) \cdot \frac{19.3\%}{10\%} = 0.79$ across income. Restricting selection to below 500% FPL (64% of 19-64 population), for example, implied take-up is $64\% \cdot (1 - 59\%) \cdot \frac{19.3\%}{10\%} = 0.52$.

dollar. Generous welfare weights are needed to balance the cost implied by an employment effect the full size of take-up, although actual employment effect is likely well below the calibrations considered.

I then shut down ESI selection from workers, non-workers, or both. Fiscal cost increases the most when selection is driven by ESI payers ($\frac{d\lambda_1^{1-e}}{d\lambda_p} = 0$), with the effect similar to an employment response half the size of take-up. By contrast, selection from private transfer beneficiaries lowers fiscal cost with increased enrollee share of premium. In this case, welfare effects from take-up and selection response roughly balance out, implying a near-zero efficiency cost (not including the external benefit on credit) when $\gamma = 0$.

Shutting down all interaction with private transfer yields a cost of 0.381 per subsidy dollar,[66] which balances out benefits for γ between 2 and 3 under conservative weights. Additional selection improves welfare if it increases enrollee premium contribution without increasing the transfer burden on ESI payers. Welfare gain is smaller, however, if private beneficiaries already bear some cost of premium, and the fiscal saving from the selection is smaller.

7.4 Uncompensated cost saving

Since sizable benefits of (formal) insurance expansion accrue to third-party payers, I next assess the sensitivity of uncompensated cost saving to a range of moral hazard estimates, modeled as the relative generosity of informal insurance g. In the main analysis, $g = 0.8$ implies spending increases by 25% with formal insurance. In the robustness analysis, I re-calibrate uninsured cost elasticity $\varepsilon_{ri,\lambda_0}$ assuming $g = 1$ (no moral hazard) or $g = 0.7$ (43% spending increase).

Absent moral hazard, avoided uncompensated cost is 0.09 per subsidy dollar (Appendix Table B11). Including interest savings, benefits roughly equal the fiscal cost at $\gamma = 0$, implying small efficiency loss of subsidy. The efficiency cost increases with moral hazard. A 43% (25%) spending response lowers uncompensated cost saving by 0.027 (0.015) per subsidy dollar. In both cases, benefits outweigh the cost under conservative welfare weights and mild redistribution preference ($\gamma = 2$).

Moral hazard similarly has modest impact on the welfare of mandate penalty (Appendix Table B12). For the range of spending response considered, net welfare effect is near-zero for $\gamma = 3$ under conservative weights, but becomes more negative as the uninsured receive larger weights (consistent with lower consumption). Although moral hazard generally lowers the net benefit of formal insurance, welfare conclusion remains robust to large spending response in the calibration.

7.5 Reform rationale

The previous cost-benefit analysis assumes expansion has zero effect on premium. I find that the cost-effectiveness of premium subsidy does not particularly rely on the premium benefit. For mandate penalty, the premium benefit potentially plays a larger role. Here, to

[66] including only mechanic cost (0.242) and new enrollee cost (0.139)

formally understand the relative importance of premium versus uncompensated care for policies targeting different expansion groups, I include the premium benefit $\frac{dW_P}{d\mathbf{K}p}$.

Table 8 adjusts the fiscal cost of subsidy by the maximum premium benefit calculated from Equation 6 assuming zero mark-up response. I highlight the potential extent of cost saving attributable to the composition effect on premium. Absent equity considerations ($\gamma = 0$), saving in enrollee premium is $6.5\% \cdot 95\% = 0.062$ per subsidy dollar, or $\frac{0.062}{0.074} = 84\%$ of the saving to uncompensated care payers. The joint benefit on formal insurance pricing and uncompensated care is sufficient to dominate the fiscal cost, without resort to the credit benefit or redistribution preferences.

Table 8: Welfare effect of premium subsidy, allowing for premium benefit $\frac{dW_P}{d\lambda_p \cdot p}$

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$	(III) $\frac{dW_\Pi}{d\lambda_p \cdot p}$	(IV) $\frac{dW_C}{d\lambda_p \cdot p}$			(V) (1)+(2)+(3)+(4)		(VI) $\frac{dW}{d\lambda_p \cdot p}\Big\vert_{u'}$
				$\frac{d\log p}{d\lambda_p} = 0$	$= -6.5\%$	$\frac{dW_P}{d\lambda_p p}$	$\frac{d\log p}{d\lambda_p} = 0$	$= -6.5\%$	
				Panel A: non-medical consumption					
$\gamma = 0$	0.232	0.074	0.011	-0.338	-0.276	0.062	-0.021	0.041	-0.018
$\gamma = 1$	0.368	0.082	0.013	-0.342	-0.278	0.064	0.121	0.185	-0.028
$\gamma = 2$	0.585	0.092	0.014	-0.346	-0.278	0.068	0.345	0.413	-0.045
$\gamma = 3$	0.928	0.105	0.016	-0.350	-0.280	0.070	0.699	0.769	-0.071
				Panel B: food consumption					
$\gamma = 1$	0.249	0.079	0.011	-0.339	-0.277	0.062	0	0.062	-0.018
$\gamma = 2$	0.268	0.085	0.012	-0.339	-0.277	0.062	0.026	0.088	-0.019
$\gamma = 3$	0.288	0.092	0.012	-0.340	-0.278	0.062	0.052	0.114	-0.021

Notes: Table calculates the benefit of subsidy dollar to recipients in column 1, uncompensated care payers in column 2, and interest payment in column 3. Column 4 calculates fiscal cost assuming full mark-up response to cost saving ($\frac{d\log p}{d\lambda_p p} = 0$ as in the main analysis), or zero mark-up response implying $\frac{d\log p}{d\lambda_p p} = -6.5\%$. The difference in cost, $\frac{dW_P}{d\lambda_p p}$, gives the risk pooling benefit of premium. Column 5 calculates net benefit with and without the premium benefit. Column 6 shows the friction cost from the main analysis.

Adjusting for welfare weights increases the return to subsidy. The equity gain reflects rising benefit to low-income recipients relative to payers. The basic trade-off, however, centers around the *efficiency* gain of insurance pricing versus the enrollment cost. Although the premium benefit is weakened in less competitive markets, one-third of the potential benefit offsets the baseline efficiency loss without friction (-0.021), and two-thirds of the benefit offsets the loss including friction (-0.039). Corresponding mark-up adjustments range from $\frac{\partial \beta}{\partial \lambda_p} = 0.029$ to 0.049. For the small mark-up increase (0.016) in CommCare, including the premium benefit rationalizes subsidy on pure efficiency grounds.

Welfare argument for mandate penalty strengthens with the risk pooling benefit in premium (Table 9). Without the premium effect (Equation12), uncompensated care alone does not justify penalty: the 0.20 dollar net cost on the uninsured is twice the benefit to uncompensated care. Without uncompensated care, the premium benefit ($2.7\% \cdot 97.7\% = 0.026$ per penalty dollar) alone offsets the cost on the uninsured from an efficiency standpoint. Adding equity concerns, unless the society places substantially higher weights on the uninsured, return to penalty remains positive.

Table 9: Welfare effect of mandate penalty, allowing for premium benefit $\frac{dW_P}{dk\,p}$

	(I) $\frac{dW_B}{dk\cdot p}$	(II) $\frac{dW_{UC}}{dk\cdot p}$	(III) $\frac{dW_C}{dk\cdot p}$			(IV) (1)+(2)+(3)		(V) $\frac{dW_F}{dk\cdot p}\Big\vert_{u'}$
			$\frac{d\log p}{dk}=0$	$=-2.7\%$	$\frac{dW_P}{dk\,p}$	$\frac{d\log p}{dk}=0$	$=-2.7\%$	
			Panel A: non-medical consumption					
$\gamma=0$	-0.023	0.011	0.003	0.029	0.026	-0.009	0.017	-0.018
$\gamma=1$	-0.062	0.013	0.003	0.029	0.026	-0.046	-0.020	-0.028
$\gamma=2$	-0.168	0.014	0.003	0.030	0.027	-0.151	-0.124	-0.045
$\gamma=3$	-0.454	0.016	0.003	0.034	0.031	-0.435	-0.404	-0.071
			Panel B: food consumption					
$\gamma=1$	-0.022	0.012	0.003	0.029	0.026	-0.007	0.019	-0.018
$\gamma=2$	-0.021	0.012	0.003	0.028	0.025	-0.006	0.019	-0.019
$\gamma=3$	-0.020	0.013	0.003	0.028	0.025	-0.004	0.021	-0.021

Notes: Table calculates the penalty cost to beneficiaries (the uninsured) in column 1 and benefit to uncompensated care payers in column 2. Column 3 calculates the fiscal effect on government budget assuming full mark-up response to cost saving ($\frac{d\log p}{dk}=0$ as in the main analysis), or zero mark-up response implying $\frac{d\log p}{dk}=-2.7\%$. The difference in cost, $\frac{dW_P}{dk\,p}$, gives the risk pooling benefit of premium. Column 4 calculates the net benefit with or without the premium benefit. Column 5 shows the friction cost from the main analysis. Sample is restricted to the unsubsidized population above 300% FPL.

The baseline efficiency gain roughly offsets the friction loss of taste-based enrollees. Smaller benefit to third-party payers, consistent with more limited access to uncompensated care in the high-income, lowers the efficiency gain. Overall desirability of mandate penalty relies on the risk pooling benefit on premium, whereas uncompensated care is the leading motivation for subsidy. When both rationales are considered, including realistic friction in insurance pricing and take-up, subsidy and penalty in Massachusetts are justified on largely efficiency grounds for a range of plausible effect sizes.

7.6 Discussion

The efficiency argument follows from the incentive-based policy design. When subsidy is less than full, part of the social cost of insurance is internalized as enrollee cost sharing. With optimal take-up response, subject to only a small degree of behavioral preferences, the efficiency cost on the margin of choice is negligible. Unlike a universal mandate that trades-off marginal and infra-marginal benefits, the choice-based design is evaluated based on the net social benefit given behavioral responses.

The net benefit compares the efficiency gains of formal insurance with the implied fiscal cost of expansion using policy instruments. Perhaps not very surprisingly, four years into the reform, at a 95% insurance rate, the net benefit of further increasing the premium subsidy and the mandate penalty falls close to zero for a range of estimates. Put differently, efforts to achieve higher coverage rate with the current set of incentives are unlikely to be cost-effective.

While social externalities may motivate government expansion of health insurance, universal health insurance is hard to justify on pure efficiency grounds. In the cost-benefit framework, an increase in subsidy increases take-up, but also decreases the cost of social insurance internalized as enrollee cost-sharing. For policy generosities consistent with (near-)universal insurance,[67] the social benefit decreases below the current level whereas the social cost increases, implying universal insurance may not be desirable in this setting without some preference for redistribution.[68]

8 Conclusion

This paper analyzes two potential rationales for expanding health insurance in the US: adverse selection and uncompensated care. With adverse selection, social insurance is constrained by the desirable scope of redistribution across risk: a mandate maximizes welfare only when complete pooling is socially desirable. With uncompensated care, formal insurance improves welfare through the distribution of health care cost in the economy. When the scope and value of redistribution are larger across income than risk, tax-financed formal insurance mandate maximizes welfare.

I assess the welfare implications of two rationales exploiting policy incentives that incrementally expanded formal insurance in Massachusetts. Quantified welfare trade-off suggests the social cost of uncompensated care motivates premium subsidy without substantial redistribution preference for low-income recipients. Mandate penalty is motivated solely by the pooling benefit of premium, and the motivation is weakened by the friction loss to taste-based enrollees. Without strong redistribution preferences for the already insured, a universal mandate based on risk pooling alone may not be desirable.

The analysis showcases the trade-off in the global design of social insurance. Relative to flexible incentives across expansion groups, a universal mandate relies more on redistribution arguments and may be less viable in practice. Including both rationales, current level of penalty and subsidy in Massachusetts improves welfare from a pure efficiency standpoint. The analysis focused on the desirable scope of formal insurance. Additional trade-offs over health care quality and delivery, and contract designs when social insurance is delegated to private insurers, are important topics for future research.

References

Aizawa, N. (2017). Labor market sorting and health insurance system design. *mimeo*.

Akerlof, G. (1970). The market for lemons. *Quarterly journal of Economics*, **84** (3), 488–500.

[67] Based on WTP for insurance in the low-income uninsured, a 90% subsidy would still leave 1% of the population uninsured (Finkelstein *et al.*, 2017).

[68] Additional efficiency cost may arise with higher tax rates to finance the subsidy increase.

Akosa Antwi, Y., Moriya, A. S. and Simon, K. (2013). Effects of federal policy to insure young adults: evidence from the 2010 affordable care act's dependent-coverage mandate. *American Economic Journal: Economic Policy*, **5** (4), 1–28.

Allen, H., Swanson, A., Wang, J. and Gross, T. (2017). Early medicaid expansion associated with reduced payday borrowing in california. *Health Affairs*, **36** (10), 1769–1776.

Andrews, I. and Miller, C. (2013). Optimal social insurance with heterogeneity.

Argys, L. M., Friedson, A. I., Pitts, M. M. and Tello-Trillo, D. S. (2017). Losing public health insurance: Tenncare disenrollment and personal financial distress.

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, **103** (6), 2499–2529.

Barsky, R. B., Juster, F. T., Kimball, M. S. and Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, **112** (2), 537–579.

Bhargava, S., Loewenstein, G. and Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, **132** (3), 1319–1372.

Blavin, F. (2016). Association between the 2014 medicaid expansion and us hospital finances. *Jama*, **316** (14), 1475–1483.

Blomqvist, A. and Horn, H. (1984). Public health insurance and optimal income taxation. *Journal of Public Economics*, **24** (3), 353–371.

Brevoort, K., Grodzicki, D. and Hackmann, M. B. (2017). Medicaid and financial health.

Chandra, A., Gruber, J. and McKnight, R. (2011). The importance of the individual mandate—evidence from massachusetts. *New England Journal of Medicine*, **364** (4), 293–295.

—, — and — (2014). The impact of patient cost-sharing on low-income populations: evidence from massachusetts. *Journal of health economics*, **33**, 57–66.

Chetty, R. (2006). A general formula for the optimal level of social insurance. *Journal of Public Economics*, **90** (10-11), 1879–1901.

— and Finkelstein, A. (2013). Social insurance: Connecting theory to data. **5**, 111–193.

Cohen, A. and Einav, L. (2007). Estimating risk preferences from deductible choice. *American economic review*, **97** (3), 745–788.

Currie, J. and Gruber, J. (1996a). Health insurance eligibility and child health: lessons from recent expansions of the medicaid program. *Quarterly Journal of Economics*, **431**, 466.

— and — (1996b). Saving babies: the efficacy and cost of recent changes in the medicaid eligibility of pregnant women. *Journal of political Economy*, **104** (6), 1263–1296.

Cutler, D. M. and Gruber, J. (1996). Does public insurance crowd out private insurance? *The Quarterly Journal of Economics*, **111** (2), 391–430.

Diamond, P. A. (1967). The role of a stock market in a general equilibrium model with technological uncertainty. *The American Economic Review*, pp. 759–776.

Dranove, D., Garthwaite, C. and Ody, C. (2016). Uncompensated care decreased at hospitals in medicaid expansion states but not at hospitals in nonexpansion states. *Health Affairs*, **35** (8), 1471–1479.

Einav, L., Finkelstein, A. and Cullen, M. R. (2010). Estimating welfare in insurance markets using variation in prices. *The quarterly journal of economics*, **125** (3), 877–921.

Finkelstein, A., Hendren, N. and Luttmer, E. F. (2015). The value of medicaid: Interpreting results from the oregon health insurance experiment.

—, — and Shepard, M. (2017). Subsidizing health insurance for low-income adults: Evidence from massachusetts.

— and McGarry, K. (2006). Multiple dimensions of private information: evidence from the long-term care insurance market. *American Economic Review*, **96** (4), 938–958.

—, Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K. and Group, O. H. S. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, **127** (3), 1057–1106.

Frean, M., Gruber, J. and Sommers, B. D. (2017). Premium subsidies, the mandate, and medicaid expansion: Coverage effects of the affordable care act. *Journal of Health Economics*, **53**, 72–86.

Gallagher, E., Gopalan, R. and Grinstein-Weiss, M. (2018). The effect of health insurance on home payment delinquency: Evidence from aca marketplace subsidies.

Garthwaite, C., Gross, T. and Notowidigdo, M. J. (2014). Public health insurance, labor supply, and employment lock. *The Quarterly Journal of Economics*, **129** (2), 653–696.

Graves, J. A. and Gruber, J. (2012). How did health care reform in massachusetts impact insurance premiums? *American Economic Review*, **102** (3), 508–13.

Gross, T. and Notowidigdo, M. J. (2011). Health insurance and the consumer bankruptcy decision: Evidence from expansions of medicaid. *Journal of Public Economics*, **95** (7-8), 767–778.

Gruber, J. (2010). The tax exclusion for employer-sponsored health insurance.

— (2017). Delivering public health insurance through private plan choice in the united states. *Journal of Economic Perspectives*, **31** (4), 3–22.

— and Simon, K. (2008). Crowd-out 10 years later: Have recent public insurance expansions crowded out private health insurance? *Journal of health economics*, **27** (2), 201–217.

Hackmann, M. B., Kolstad, J. T. and Kowalski, A. E. (2015). Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, **105** (3), 1030–66.

Handel, B. R. and Kolstad, J. T. (2015). Health insurance for" humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, **105** (8), 2449–2500.

Hart, O., Shleifer, A. and Vishny, R. W. (1997). The proper scope of government: theory and an application to prisons. *The Quarterly Journal of Economics*, **112** (4), 1127–1161.

Heim, B. T., Hunter, G., Isen, A., Lurie, I. Z. and Ramnath, S. P. (2017). Income responses to the affordable care act: Evidence from the premium tax credit notch.

Hendren, N. (2017). Knowledge of future job loss and implications for unemployment insurance. *American Economic Review*, **107** (7), 1778–1823.

Hu, L., Kaestner, R., Mazumder, B., Miller, S. and Wong, A. (2016). *The effect of the patient protection and affordable care act medicaid expansions on financial wellbeing*. Tech. rep., National Bureau of economic research.

Jaffe, S. and Shepard, M. (2018). Price-linked subsidies and imperfect competition in health insurance. *mimeo*.

Keehan, S. P., Sisko, A. M., Truffer, C. J., Poisal, J. A., Cuckler, G. A., Madison, A. J., Lizonitz, J. M. and Smith, S. D. (2011). National health spending projections through 2020: economic recovery and reform drive faster spending growth. *Health Affairs*, **30** (8), 1594–1605.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C. and Wrobel, M. V. (2012). Comparison friction: Experimental evidence from medicare drug plans. *The Quarterly Journal of Economics*, **127** (1), 199–235.

Kucko, K., Rinz, K. and Solow, B. (2018). Labor market effects of the affordable care act: Evidence from a tax notch.

Landais, C., Nekoei, A., Nilsson, P., Seim, D. and Spinnewijn, J. (2017). Risk-based selection in unemployment insurance: evidence and implications.

Lyons, S. (2017). Are employer mandates to offer health insurance effective in reducing subsidized coverage crowd-out of employer-sponsored insurance? *American Journal of Health Economics*, **3** (3), 370–391.

Mahoney, N. (2015). Bankruptcy as implicit health insurance. *American Economic Review*, **105** (2), 710–46.

Miller, K. (2016). Do private Medicare firms face lower costs?, Mimeo. University of Oregon.

Reynolds, L. (2013). The future of the nhs–irreversible privatisation? interview by jill mountford. *BMJ (Clinical research ed)*, **346**, f1848.

Rothschild, M. and Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics*, pp. 629–649.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J. and Sobek, M. (2018). Ipums usa: Version 8.0 [dataset]. *Minneapolis, MN: IPUMS*, *https://doi.org/10.18128/D010.V8.0*.

Saez, E. and Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, **106** (1), 24–45.

Saltzman, E. (2018). Demand for health insurance: Evidence from the california and washington aca exchanges.

Sommers, B. D., Shepard, M. and Hempstead, K. (2018). Why did employer coverage fall in massachusetts after the aca? potential consequences of a changing employer mandate. *Health Affairs*, **37** (7), 1144–1152.

Tebaldi, P. (2017). Estimating equilibrium in health insurance exchanges: Price competition and subsidy design under the aca. *mimeo*.

Vickers, J. and Yarrow, G. (1991). Economic perspectives on privatization. *Journal of Economic Perspectives*, **5** (2), 111–132.

Wilson, R. (1968). The theory of syndicates. *Econometrica: journal of the Econometric Society*, pp. 119–132.

Appendix

A Appendix Proofs

A.1 Risk-based pricing

Consider first the case of risk-based pricing. Insurers observe health type μ, and charge competitive premium $p(\nu,\mu) = (1-\mu)M$. Government in addition observes productivity ν, and varies transfers $t(\nu,\mu)$ and $\Delta t(\nu,\mu)$ across types. Absent a universal mandate, individual take-up decision involves comparing the utility of 1), acquiring formal insurance at cost $(1-\mu)M$ in both health states, and 2), remaining uninsured and receiving uncompensated care support $\Delta t(\nu,\mu)$ in the sick state.

Absent uncompensated care, willingness to pay (WTP) for insurance is above expected cost (risk-based premium), and take-up is complete. With uncompensated care, insurance take-up is generally incomplete. This is because ex-post transfer Δt lowers ex-ante WTP, and because insurance company has less information than the government, risk-based premium cannot adjust for the implicit coverage from uncompensated care.

Given the incomplete take-up, the government can improve welfare with a universal mandate: all individuals purchase competitively priced insurance, and uncompensated care cost to government is zero. To see this, note that for arbitrary employment $e(\nu,\mu)$ and transfers $t(\nu,\mu)$ and $\Delta t(\nu,\mu)$, the uninsured has consumption

$$(w) + t(\nu,\mu) > 0$$

is the healthy state, and

$$(w) + t(\nu,\mu) + \Delta t(\nu,\mu) - M > 0$$

in the unhealthy state. With non-negative consumption (Inada condition), expected transfer $t(\nu,\mu)+(1-\mu)\Delta t(\nu,\mu)$ plus any labor earning is greater than expected cost $(1-\mu)M$, or competitive premium is already affordable given transfer.

To improve welfare, the government instead transfers

$$\bar{t}(\nu,\mu) = t(\nu,\mu) + (1-\mu)\Delta t(\nu,\mu)$$

in both health states, and has the uninsured purchase insurance at expected cost $(1-\mu)M$. The resulting consumption,

$$(w) + t(\nu,\mu) + (1-\mu)\Delta t(\nu,\mu) - (1-\mu)M$$

equals the *expected* consumption in the uninsured case, but gives higher expected *utility* because individual is risk-averse.

Similar argument applies to all uninsured individuals, and a universal mandate improves welfare given any employment and transfer policy. With the mandate, government redistributes income to smooth marginal utility of consumption across employment and risk: low-health individuals with greater cost of insurance receive more generous transfer. Abstracting from the efficiency cost of tax transfers, consumption with risk-based pricing equals

$$c^{RB}(e) = e \cdot w - \bar{p}$$

in the economy, where $e = E[e(\nu,\mu)]$ is the employment size, and $\bar{p} = M(1 - E[\mu])$ is the social cost of insurance. The effectual pooling across risk is socially desirable with risk-based pricing, and is achieved through (risk-adjusted) premium subsidy rather than ex-post uncompensated care. I compare with this benchmark result when more limited information is allowed in premium and transfers.

A.2 Basic trade-off

Consider moving the boundary of social insurance in the risk space to $n = 1$, the ultra-health margin. Social welfare

$$\begin{aligned} W = & \int_0^n \int_{\chi(\mu)}^1 \mathrm{d}F(\nu,\mu) \cdot u(c_e) + \int_n^1 \int_{\chi(\mu)}^1 \mu \, \mathrm{d}F(\nu,\mu) \cdot u(c_e + p) \\ & + \int_n^1 \int_{\chi(\mu)}^1 (1-\mu) \, \mathrm{d}F(\nu,\mu) \cdot u(c_e + p - M) + \int_0^n \int_0^{\chi(\mu)} \mathrm{d}F(\nu,\mu) \cdot u\Big(A - (1-\lambda_p)p\Big) \\ & + \int_n^1 \int_0^{\chi(\mu)} \mu \, \mathrm{d}F(\nu,\mu) \cdot u(A) + \int_n^1 \int_0^{\chi(\mu)} (1-\mu) \, \mathrm{d}F(\nu,\mu) \cdot u(A - M) - \int_0^1 \int_{\chi(\mu)}^1 g\left(\frac{1}{\nu}\right) \mathrm{d}F(\nu,\mu) \end{aligned}$$

where employed enrollees, measured in mass $E[e(\nu,\mu) \cdot hi(\nu,\mu)] = \int_0^n \int_{\chi(\mu)}^1 \mathrm{d}F(\nu,\mu)$, have consumption

$$c_e = w - \frac{1-e}{e} A - \left[\frac{\lambda_p}{e} \int_0^n \int_0^{\chi(\mu)} \mathrm{d}F(\nu,\mu) + 1 \right] p$$

where $\int_0^n \int_0^{\chi(\mu)} \mathrm{d}F(\nu,\mu)$ is the mass of subsidized (non-employed) enrollees, and $e = E[e(\nu,\mu)] = \int_0^1 \int_{\chi(\mu)}^1 \mathrm{d}F(\nu,\mu)$ is the employment size following from arbitrary employment margin $\chi(\mu)$. Employed non-enrollees in $\int_n^1 \int_{\chi(\mu)}^1 \mathrm{d}F(\nu,\mu)$ save premium p if healthy, but pay medical cost M if not.

Average risk $r = 1 - \frac{\int_0^n \int_0^1 \mu \, \mathrm{d}F(\nu,\mu)}{\int_0^n \int_0^1 \mathrm{d}F(\nu,\mu)}$ varies with margin n according to

$$\frac{\mathrm{d}r}{\mathrm{d}n} = -\frac{r - (1-n)}{i} \int_0^1 f(\nu,n) \mathrm{d}\nu$$

where $i = E[hi(\nu,\mu)] = \int_0^n \int_0^1 \mathrm{d}F(\nu,\mu)$ is the insurance rate.

Expanding insurance to the ultra-health margin, welfare increases by

$$\frac{\mathrm{d}W}{\mathrm{d}n}\bigg|_{n=1} = \underbrace{\int_{\chi(1)}^{1} f(\nu,1)\,\mathrm{d}\nu \cdot \Big(u(c_e) - u(c_e + p)\Big)}_{\text{marginal utility loss}}$$
$$\underbrace{+\, e \cdot u'(c_e) \cdot \left[-\frac{\lambda_p}{e} p \int_{0}^{\chi(1)} f(\nu,1)\,\mathrm{d}\nu\right]}_{\text{tax payer cost: new enrollee subsidy}}$$
$$\underbrace{+\, e \cdot u'(c_e) \cdot \left[\frac{\lambda_p}{e}(1-e) + 1\right] p \int_{0}^{1} f(\nu,1)\,\mathrm{d}\nu}_{\text{premium saving}}$$
$$\underbrace{+ \int_{0}^{\chi(1)} f(\nu,1)\,\mathrm{d}\nu \cdot \Big[u\Big(A - (1-\lambda_p)p\Big) - u(A)\Big]}_{\text{marginal utility loss, subsidy enrollees}}$$
$$\underbrace{+\, (1-e) \cdot u'\Big(A - (1-\lambda_p)p\Big) \cdot (1-\lambda_p)p \int_{0}^{1} f(\nu,1)\,\mathrm{d}\nu}_{\text{subsidized premium saving}}$$

Given full subsidy $\lambda_p = 1$, the trade-off only affects workers who are premium payers. The first three terms simply to

$$\frac{\mathrm{d}W}{\mathrm{d}n}\bigg|_{n=1,\lambda_p=1} = \int_{\chi(1)}^{1} f(\nu,1)\,\mathrm{d}\nu \cdot \Big[u(c_e) - u(c_e+p) + p \cdot u'(c_e)\Big]$$
$$\propto\ u(c_e) - u(c_e+p) + p \cdot u'(c_e)$$

Concave utility implies risk pooling is socially desirable: infra-margin cost saving $p \cdot u'(c_e)$ is valued more than the utility cost on the ultra-margin. Once a mandate implements the pooling, the government redistributes across income with cash transfer A and subsidy λ_p. Given transfer A, abstracting from behavioral responses to taxes, standard equity argument implies full subsidy is desirable.

A.3 Risk preference heterogeneity

When risk attitude differs across individuals (Cohen and Einav 2007; Barsky *et al.* 1997), WTP for insurance is not monotone in risk, but in addition depends on the correlation with risk preference (Andrews and Miller, 2013). If low-risk individuals are also less risk-tolerant, adverse selection is lessened (Finkelstein and McGarry, 2006). In this case, the risk-tolerant efficiently bear more risk (Diamond 1967; Wilson 1968), lowering the desirability of complete redistribution across risk. Regardless of risk preferences, the highest health margin $\mu = 1$ has zero demand for insurance and remains the last group to enroll. The basic trade-off carries over, but adjusts for the differences in risk preferences between the infra and the ultra margin.

Assume preference exhibits constant relative risk aversion (CRRA), with risk coefficient γ taking two values $\gamma_1 < \gamma_2$ in the economy: individuals with γ_2 are more risk-averse. Let ϕ (Φ) denote the share of high risk-aversion types on the ultra (infra) margin. Adjusted by risk preference, the trade-off becomes[69]

$$\frac{dW}{dn}\bigg|_{n=1} \propto \phi \cdot \Delta u_2 + \Phi \cdot u_2^{'} \cdot p + (1-\phi) \cdot \Delta u_1 + (1-\Phi) \cdot u_1^{'} \cdot p$$

where $\Delta u_k = u(c;\gamma_k) - u(c+p;\gamma_k)$ is the utility cost of premium p for risk type γ_k, $k = 1,2$, and $u_k^{'} = u^{'}(c;\gamma_k)$ the social cost saving.

Argument for a coverage mandate is strengthened, if ultra-margin individuals are on average more risk-averse, or $\Phi < \phi$[70]. When true, marginal utility loss is smaller than in the homogeneous case, and social benefit is greater. Conversely, if risk tolerance increases with health ($\Phi > \phi$), and social benefit is less than the marginal cost, then high-health individuals are efficiently uninsured.

A.4 Labor productivity and non-labor income

Assume now that labor productivity affects earning: type ν produces output νZ at opportunity cost $g(\frac{1}{\nu})$. Workers take home $(1-\tau)\cdot \nu W$ after a linear tax τ. The tax revenue is used to fund the cash benefit and premium subsidy.

Individuals receive non-labor income $y(\nu,\mu)$ independent of employment. The distri-

[69]Let the distribution of risk preference be $\Pr\{\gamma(\nu,\mu) = \gamma_2\} = \rho(\nu,\mu)$. The welfare effect of enrolling the perfect health types with full subsidy to the non-employed is

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}n_e}\bigg|_{n_e=1,\,\lambda_p=1} &= \int_{\chi(1)}^{1} \rho(\nu,1)\,\mathrm{d}F(\nu,1)\cdot\Delta u_2 + \int_{\chi(1)}^{1}\Big(1-\rho(\nu,1)\Big)\mathrm{d}F(\nu,1)\cdot\Delta u_1 \\ &+ \int_0^1\int_{\chi(\mu)}^1 \rho(\nu,\mu)\,\mathrm{d}F(\nu,\mu)\cdot u_2'(c_e)\cdot\frac{p}{e}\int_{\chi(1)}^1 f(\nu,1)\mathrm{d}\nu \\ &+ \int_0^1\int_{\chi(\mu)}^1 \Big(1-\rho(\nu,\mu)\Big)\mathrm{d}F(\nu,\mu)\cdot u_1'(c_e)\cdot\frac{p}{e}\int_{\chi(1)}^1 f(\nu,1)\mathrm{d}\nu\end{aligned}$$

The trade-off is then simplified using $\Phi = \frac{\int_0^1\int_{\chi(\mu)}^1 \rho(\nu,\mu)\mathrm{d}F(\nu,\mu)}{e}$ and $\phi = \frac{\int_{\chi(1)}^1 \rho(\nu,1)\mathrm{d}F(\nu,1)}{\int_{\chi(1)}^1 f(\nu,1)\mathrm{d}\nu}$.

[70]Note that the social benefit can be written as

$$\Phi\cdot u_2^{'}\cdot p + (1-\Phi)\cdot u_1^{'}\cdot p = \underbrace{p\cdot u_1'\cdot(1-\phi)}_{>(1-\phi)\cdot\Delta u_1} + \underbrace{p\cdot u_2'\cdot\phi}_{>\phi\cdot\Delta u_2} + p\cdot\Delta u'\cdot(\phi-\Phi)$$

Therefore social benefit offsets marginal utility loss, if $\phi > \Phi$, or risk aversion is higher on the ultra-margin.

bution of $y(\nu,\mu)$ is assumed continuous but otherwise unrestricted in the economy.[71][72] The trade-off now adjusts for the tax incidence of subsidy across the productivity space:

$$\underbrace{E[u'|e=1]\,p + \overbrace{\frac{Cov[u',\nu|e=1]}{E[\nu|e=1]}\left(1-\frac{e}{e_{\mu=1}}\right)p}^{\text{tax incidence of subsidy}}}_{\text{social benefit}} \;-\; \underbrace{E[\Delta u|e=1,\mu=1]}_{\text{marginal utility loss}}$$

where e outside the conditional probability is employment size $e = E[e(\nu,\mu)]$, and $e_{\mu=1} = E[e(\nu,\mu)|\mu=1]$ is employment on the ultra margin. $Cov[u',\nu|e=1]$ evaluates the tax burden at marginal utility of consumption. If there are more subsidy eligibles on the ultra-margin than on average ($e_{\mu=1} < e$), expanding insurance raises the per capita cost of subsidy. If the extra cost is directed more to high productivity types, then the social benefit increases.

A.5 Tax penalty

Suppose the government provides uncompensated care to the uninsured, and retro-actively enroll patients in formal insurance at the site of care. The low-income (non-employed) receives full premium subsidy. High-income uninsured pays tax penalty $k \cdot p$.

Subsidy is financed by a linear tax on payroll and the tax penalty. Uncompensated care is financed by surcharges on paying customers. Let t denote the patient cost share. Surcharge on patients is

$$uc^0 = \frac{t}{\lambda_{1,1,0}}\Big[\underbrace{\lambda_{1,0,0}(M-p) + \lambda_{0,0,0}M}_{UC}\Big]$$

where $\lambda_{i,j,k} = \Pr\{e=i, hi=j, g=k\}$, and $k=0$ for patients. UC gives the total uncompensated cost (net of high-income premium). Surcharge on non-patients is $uc^1 = \frac{1-t}{\lambda_{1,1,1}}UC$.

Consumption of premium payers in the healthy state depends on after-tax income $T(\nu,\mu)$, premium p, tax share of subsidy (net of penalty), and uc^1

$$c_{1,1,1} = T(\nu,\mu) - p - \lambda_{0,1}\frac{\nu p}{e\cdot E[\nu|e=1]} + \lambda_{1,0,1}\frac{\nu k p}{E[\nu|e=1]} - uc^1$$

Excess burden $uc^0 > uc^1$ lowers patient consumption $c_{1,1,0} = c_{1,1,1} - (uc^0 - uc^1)$.

[71] Primitives of the distribution include endowment, wealth, and non-labor activities that generate income. $y(\nu,\mu)$ provides a summary proxy for these differences.

[72] Adjusting for the tax burden of UI and subsidy, consumption equals

$$c_e = y(\nu,\mu) + \nu Z - \frac{1-e}{e\cdot E[\nu|e=1]}\nu A - \frac{\int_0^n\int_0^{\chi(\mu)} \mathrm{d}F(\nu,\mu)}{e\cdot E[\nu|e=1]}\nu p - p$$

for workers, where $\frac{\nu}{\nu\cdot E[\nu|e=1]}$ is the tax burden on type ν in the linear tax schedule.

High-income uninsured pay tax penalty if healthy, but are required to purchase insurance in the case of a health event,

$$
\begin{aligned}
c_{1,0,1} &= T(\nu,\mu) - kp - \lambda_{0,1}\frac{\nu p}{e\cdot E[\nu|e=1]} + \lambda_{1,0,1}\frac{\nu k p}{E[\nu|e=1]} \\
c_{1,0,0} &= T(\nu,\mu) - p - \lambda_{0,1}\frac{\nu p}{e\cdot E[\nu|e=1]} + \lambda_{1,0,1}\frac{\nu k p}{E[\nu|e=1]}
\end{aligned}
$$

Non-employed are fully insured (formally or implicitly). Consumption $c_0 = A + y(\nu,\mu)$ does not vary by enrollment or health.

Increase in social welfare

$$
\begin{aligned}
W =& \int_0^n \int_{\chi(\mu)}^1 \mu\, u(c_{1,1,1})\,\mathrm{d}F(\nu,\mu) + \int_0^n \int_{\chi(\mu)}^1 (1-\mu)\, u(c_{1,1,0})\,\mathrm{d}F(\nu,\mu) \\
&+ \int_n^1 \int_{\chi(\mu)}^1 \mu\, u(c_{1,0,1})\,\mathrm{d}F(\nu,\mu) + \int_n^1 \int_{\chi(\mu)}^1 (1-\mu)\, u(c_{1,0,0})\,\mathrm{d}F(\nu,\mu) \\
&+ \int_0^1 \int_0^{\chi(\mu)} u(c_0)\,\mathrm{d}F(\nu,\mu) - \int_0^1 \int_{\chi(\mu)}^1 g(\frac{1}{\nu})\,\mathrm{d}F(\nu,\mu)
\end{aligned}
$$

when extending formal insurance extends to the ultra-margin $n = 1$ equals

$$
\begin{aligned}
\frac{\mathrm{d}W}{\mathrm{d}n}\Big|_{n=1} =& \int_{\chi(1)}^1 \Big(u(c_{1,1,1}) - u(c_{1,0,1})\Big) f(\nu,1)\,\mathrm{d}\nu \\
&+ \sum_{k=0,1} E[u'|e=1, hi=1, g=k]\cdot E\Big[\frac{\mathrm{d}c_{1,1,k}}{\mathrm{d}n}\Big| e=1, hi=1, g=k\Big]\cdot\lambda_{1,1,k} \\
&+ \sum_{k=0,1} \frac{Cov[u',\nu|e=1, hi=1, g=k]\cdot g_e^k}{E[\nu|e=1]}\, p\left[1-k-\frac{e}{e_{\mu=1}}\right]\int_{\chi(1)}^1 f(\nu,1)\,\mathrm{d}\nu
\end{aligned}
$$

where $g_e^k = \Pr\{g=k|e=1, hi=1\}$. Note that

$$
\sum_{k=0,1} E\Big[\frac{\mathrm{d}c_{1,1,k}}{\mathrm{d}n}\Big| e=1, hi=1, g=k\Big]\cdot\lambda_{1,1,k} = p(1-k)\int_{\chi(1)}^1 f(\nu,1)\,\mathrm{d}\nu
$$

and

$$
E\Big[\frac{\mathrm{d}c_{1,1,0}}{\mathrm{d}n}\Big| e=1, hi=1, g=0\Big]\cdot\lambda_{1,1,0} = g_e^0\, p\int_{\chi(1)}^1 f(\nu,1)\,\mathrm{d}\nu\cdot\left[\frac{e}{e_{\mu=1}} + \frac{E[\nu|e=1, g=0]}{E[\nu|e=1]}\left(1-k-\frac{e}{e_{\mu=1}}\right)\right]
$$

Then

$$\sum_{k=0,1} E[u'|e=1, hi=1, g=k] \cdot E\Big[\frac{\mathrm{d}c_{1,\cdot,k}}{\mathrm{d}n}\Big| e=1, hi=1, g=k\Big] \cdot \lambda_{1,1,k}$$

$$= E[u'|e=1, hi=1, g=1]\, p\,(1-k) \int_{\chi(1)}^{1} f(\nu,1)\,\mathrm{d}\nu$$

$$+ \Delta u'_{e=1}(1-g_e)\, p \int_{\chi(1)}^{1} f(\nu,1)\,\mathrm{d}\nu \cdot \left[\frac{e}{e_{\mu=1}} + \frac{E[\nu|e=1, g=0]}{E[\nu|e=1]}\left(1-k-\frac{e}{e_{\mu=1}}\right)\right]$$

where $\Delta u'_{e=1} = E[u'|e=1, hi=1, g=0] - E[u'|e=1, hi=1, g=1]$.

Adjusting for $\int_{\chi(1)}^{1} f(\nu,1)\,\mathrm{d}\nu$, welfare impact is proportional to

$$E[u'|e=1, hi=1, g=1]\Delta c + \Delta u'_{e=1}(1-g_e)\, p \left[\frac{e}{e_{\mu=1}} + \frac{E[\nu|e=1, g=0]}{E[\nu|e=1]}\left(1-k-\frac{e}{e_{\mu=1}}\right)\right]$$
$$+ \sum_{k=0,1} \frac{Cov[u',\nu|e=1, hi=1, g=1]\cdot g_e^k}{E[\nu|e=1]}\, p \left[1-k-\frac{e}{e_{\mu=1}}\right] - E[\Delta u|e=1, hi=1, g=1]$$

where $\Delta u = u(c_{1,1,1}) - u(c_{1,0,1})$ evaluated at $n=1$, and $\Delta c = (1-k)p$ is net premium.

Negative correlation between risk and productivity implies $\frac{E[\nu|e=1,g=0]}{E[\nu|e=1]} < 1$, and uncompensated care strengthens the social benefit of formal insurance. When $Cov(\nu,\mu) \geq 0$, then $\sum_{k=0,1} \frac{Cov[u',\nu|e=1,hi=1,g=1]\cdot g_e^k}{E[\nu|e=1]} \leq \frac{Cov[u',\nu|e=1]}{E[\nu|e=1]} < 0$. Trade-off evaluated at $\frac{Cov[u',\nu|e=1]}{E[\nu|e=1]}$ understates the social benefit of tax-financed subsidy. As penalty increases, net effect on welfare depends on $-Cov[u',\nu|e=1]$ relative to $\Delta u'_{e=1}(1-g_e)E[\nu|e=1, g=0]$. When consumption variation implies greater value of redistribution across productivity, and the scope of redistribution is smaller between patients, the tax penalty strengths the welfare argument for a mandate.

A.6 Uncompensated care

Suppose the government enrolls risk type μ_{1-e} who otherwise receives uncompensated care.[73] Welfare effect is given by

$$\frac{\mathrm{d}W}{\mathrm{d}\mu_{1-e}}\Big|_{n_e=1} = \sum_{k=0,1} E[u'|e=1, hi=1, g=k] \cdot E\Big[\frac{\mathrm{d}c_{1,1,k}}{\mathrm{d}\mu_{1-e}}\Big| e=1, hi=1, g=k\Big] \cdot \lambda_{1,1,k}$$
$$- \sum_{k=0,1} \frac{Cov[u',\nu|e=1, hi=1, g=k]\cdot g_e^k}{E[\nu|e=1]} \frac{pe + MC\cdot i_{1-e}}{i} \int_0^{\chi(1)} f(\nu,1)\,\mathrm{d}\nu$$

[73]For simplicity I assume workers are already enrolled in formal insurance with the tax penalty. I hence focus on the implication of replacing uncompensated care with premium subsidy in the low-income.

From resource neutrality,

$$\sum_{k=0,1} E\Big[\frac{\mathrm{d}c_{1,1,k}}{\mathrm{d}\mu_{1-e}}\Big| e=1, hi=1, g=k\Big]\cdot\lambda_{1,1,k} = 0$$

Note that

$$E\Big[\frac{\mathrm{d}c_{1,1,0}}{\mathrm{d}\mu_{1-e}}\Big| e=1, hi=1, g=0\Big]\cdot\lambda_{1,1,0} = \\ (1-g_e)\left[\frac{e\cdot p + MC\cdot i_{1-e}}{i}\left(1-\frac{E[\nu|e=1,g=0]}{E[\nu|e=1]}\right)+MC\left(\frac{t}{1-g_e}-1\right)\right]\int_0^{\chi(1)} f(\nu,1)\mathrm{d}\nu$$

Then the welfare effect is proportional to

$$\Delta u'_{e=1}\left[\frac{e\cdot p + MC\cdot i_{1-e}}{i}\left(1-\frac{E[\nu|e=1,g=0]}{E[\nu|e=1]}\right)+MC\left(\frac{t}{1-g_e}-1\right)\right](1-g_e) \\ -\sum_{k=0,1}\frac{Cov[u',\nu|e=1,hi=1,g=k]\cdot g_e^k}{E[\nu|e=1]}\,\frac{e\cdot p+MC\cdot i_{1-e}}{i}$$

Tax-financed subsidy lowers patient surcharge burden, when subsidy cost is directed to the high productivity negatively correlated with risk. In this case, both the tax finance and uncompensated saving improve welfare. When $Cov(\nu,\mu)\geq 0$, $\sum_{k=0,1}\frac{Cov[u',\nu|e=1,hi=1,g=k]\cdot g_e^k}{E[\nu|e=1]} \leq \frac{Cov[u',\nu|e=1]}{E[\nu|e=1]} < 0$. If $-Cov[u',\nu|e=1] > \Delta u'_{e=1}(1-g_e)E[\nu|e=1,g=0]$, or when tax transfer has greater scope of redistribution and social value, subsidized formal insurance dominates uncompensated care. Subsidized universal insurance maximizes welfare.

A.7 Welfare formula

For Equation 3, first derive the pricing externality on state utility $u_t(\omega_t)$. The key externalities include the cost composition effect on premium p, and the burden of uncompensated care uc_p. In addition, if individuals do not internalize the penalty on credit rating, policy that induces take-up has external benefit on interest charges. Given state distribution, the utility change from the pricing externalities on consumption equals

$$\frac{\mathrm{d}V}{\mathrm{d}\mathbf{K}}\Big|_{Pr^{i,j}} \approx -\frac{\mathrm{d}\tau_{pb}}{\mathrm{d}\mathbf{K}}\cdot e\cdot u'(c_{1..}) - \frac{\mathrm{d}\tau_{pr}}{\mathrm{d}\mathbf{K}}\cdot\lambda_1^e\cdot u'(c_{11.}) \\ -\frac{\mathrm{d}(1-\lambda_p)p}{\mathrm{d}\mathbf{K}}\cdot\lambda_2\cdot u'(c_{.2.}) - \frac{\mathrm{d}kp}{\mathrm{d}\mathbf{K}}\cdot\lambda_0\cdot u'(c_{.0.}) \\ -\frac{\mathrm{d}uc_p}{\mathrm{d}\mathbf{K}}\cdot\lambda_{>0}^0\cdot u'(c_{.>0,0}) + \Pi\left(\frac{\mathrm{d}\pi}{\mathrm{d}\mathbf{K}}\right) \tag{20}$$

where c_{ijg} is average consumption for individuals with employment i, insurance j, and health state g. The first-order approximation occurs when moving expectation inside the state utility using Taylor expansion, ignoring second-order effects on $u''(c)$. π is the interest rate, or cost of borrowing charged to individuals, negatively affected by medical debts from uninsured health events. Saving in interest payments when policy increases take-up is given by $\Pi\left(\frac{d\pi}{d\mathbf{K}}\right)$. Effect on tax burden $\frac{d\tau_{pb}}{d\mathbf{K}}$ comes from total differentiation of *BC*.1, on private transfer $\frac{d\tau_{pr}}{d\mathbf{K}}$ from *BC*.2, on survice surcharge $\frac{duc_p}{d\mathbf{K}}$ from *BC*.3, and on premium $\frac{dp}{d\mathbf{K}}$ from *BC*.4.

Then, to focus on changes in state distribution, I Taylor-expand utility at the mean consumption of employment i and insurance j

$$U \approx \int_0^1 \sum_{i=0,1} \sum_{j=0,1,2} \rho_t^{ij} \left[u(c_t^{ij}) + \frac{1}{2} u''(c_t^{ij})(\sigma_t^{ij})^2 \right] dt \tag{21}$$

Ignoring second-order terms and variance, and averaging over life cycle, consumer welfare $V = \sum_i \sum_j \rho^{ij} u(c^{ij})$ responds to policy through distribution ρ^{ij} and the utility change of movers. Absent friction, movers choose outcomes optimally, and the behavior has no direct impact on welfare.

With friction, observe outcomes are not always rationalized with the decision model, and the deviation may respond to policy. In the case of insurance, take-up pattern not predicted by the decision model is attributed to the taste shock η_t. When a policy change induces more take-up than the model prediction, η_t responds positively to policy in generating over-demand for insurance and sub-optimal utility loss. The unintended effects on utility enter the welfare calculation in $\frac{dV}{d\mathbf{K}}\Big|_{u'}$. Continuing the insurance example, for a total take-up effect $\frac{d\lambda^0}{d\mathbf{K}}$ observed in the data, if fraction $\theta^0(\mathbf{K})$ are over-insured resulting in utility loss $\Delta u\Big|_{\theta^0(\mathbf{K})}$, the friction component $\frac{dV}{d\mathbf{K}}\Big|_{u'} = \theta^0(\mathbf{K})\frac{d\lambda^0}{d\mathbf{K}} \cdot \Delta u\Big|_{\theta^0(\mathbf{K})}$.

B Appendix Tables

Table B1: Incentive effect of subsidy, by age

	(I) any insurance	(II) employed	(III) ESI	(IV) ESI + employed	(V) ESI + not employed
27-29	0.18**	-0.15	-0.87***	-0.57***	-0.31***
	(0.081)	(0.13)	(0.11)	(0.13)	(0.045)
	[0.91]	[0.20]	[0.67]	[0.61]	[0.059]
30-34	0.076	-0.062	-0.65***	-0.29***	-0.36***
	(0.053)	(0.081)	(0.097)	(0.11)	(0.037)
	[0.92]	[0.19]	[0.71]	[0.64]	[0.070]
35-39	0.056	0.13	-0.57***	-0.17*	-0.40***
	(0.048)	(0.085)	(0.089)	(0.098)	(0.044)
	[0.94]	[0.19]	[0.74]	[0.66]	[0.087]
40-44	0.13***	0.11	-0.48***	-0.072	-0.41***
	(0.047)	(0.075)	(0.063)	(0.066)	(0.040)
	[0.95]	[0.20]	[0.75]	[0.66]	[0.090]
45-49	0.099***	0.0023	-0.53***	-0.27***	-0.26***
	(0.029)	(0.073)	(0.068)	(0.076)	(0.032)
	[0.95]	[0.20]	[0.76]	[0.68]	[0.086]
50-54	0.11***	-0.12*	-0.55***	-0.33***	-0.22***
	(0.029)	(0.062)	(0.061)	(0.061)	(0.040)
	[0.96]	[0.22]	[0.76]	[0.66]	[0.092]
55-64	0.087***	-0.24***	-0.66***	-0.26***	-0.40***
	(0.027)	(0.069)	(0.059)	(0.065)	(0.035)
	[0.97]	[0.33]	[0.74]	[0.58]	[0.16]
R^2	0.071	0.091	0.19	0.13	0.054
y mean	0.95	0.77	0.83	0.64	0.10

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table shows reduced-form estimates of subsidy incentive *subiv* interacted with full set of age band indicators, based on the main specification with full interaction terms. Robust standard error clustered at the level of PUMA in the parenthesis. Group mean of dependent variable in the square bracket.

Table B2: Subsidy effect on ESI and alternative employment measures

	(I) worked in 12 months	(II) worked in 5 years	(III) ESI + not in LBF	(IV) ESI + 12 month UE	(V) ESI + 5 year UE
			Panel A: OLS		
subs	-0.35***	-0.21***	0.034***	0.048***	0.024***
	(0.0080)	(0.0067)	(0.0034)	(0.0034)	(0.0025)
R^2	0.21	0.15	0.063	0.055	0.037
			Panel B: reduced form		
subiv	-0.077*	-0.061*	-0.30***	-0.28***	-0.19***
	(0.045)	(0.031)	(0.021)	(0.020)	(0.018)
R^2	0.099	0.090	0.064	0.053	0.037
2008	-0.043	-0.020	-0.33***	-0.28***	-0.18***
	(0.076)	(0.061)	(0.042)	(0.033)	(0.031)
2009	-0.048	-0.087**	-0.32***	-0.33***	-0.19***
	(0.066)	(0.042)	(0.032)	(0.036)	(0.026)
2010	-0.12**	-0.11**	-0.27***	-0.25***	-0.17***
	(0.059)	(0.048)	(0.030)	(0.033)	(0.028)
2011	-0.079	-0.020	-0.29***	-0.27***	-0.21***
	(0.070)	(0.052)	(0.033)	(0.031)	(0.029)
			Panel C: over-identified 2SLS		
$\widehat{subs}$	-0.075	-0.012	-0.26***	-0.22***	-0.17***
	(0.049)	(0.034)	(0.028)	(0.028)	(0.023)
F-stat	227.99	227.99	227.99	227.99	227.99
p-value	0.82	0.48	0.36	0.15	0.32
y mean	0.83	0.90	0.077	0.071	0.039

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table estimates the effect of subsidy on alternative measures of employment (ever worked in 12 months and in 5 years), and interactive outcomes with ESI: ESI coverage while not in labor force (LBF) last week in column 3, ESI coverage while unemployed in 12 months in column 4, and ESI coverage while unemployed in 5 years in column 5. Panel A shows OLS estimates using endogenous subsidy rate *subs*. Panel B shows reduced-form effect of simulated generosity *subiv*, and year-specific effects interacting *subiv* with year dummies from a separate regression. Panel A and B are based on the main specification with full interaction terms. Panel C shows 2-stage least square estimates instrumenting *subs* with *subiv* and *sublean*, based on a specification with main effects of PUMA, age, year, income, and demographic variables interacted with unemployment rate. P-value from Hansen over-identification test is reported. Robust standard errors clustered at the level of PUMA in the parenthesis.

Table B3: Incentive effect of subsidy, basic controls

	(I) any insurance	(II) employed	(III) in labor force	(IV) ESI + employed	(V) ESI + not employed
	Panel A: OLS, endogenous subsidy exposure				
$subs$	-0.071***	-0.41***	-0.31***	-0.55***	0.048***
	(0.0033)	(0.0068)	(0.0064)	(0.0068)	(0.0046)
R^2	0.068	0.20	0.18	0.29	0.053
	Panel B: 2SLS estimates, instrument varying by location, year, age, and demographics				
$\widehat{subs}$	0.16***	-0.067	-0.053	-0.33***	-0.26***
	(0.036)	(0.057)	(0.047)	(0.047)	(0.029)
F-stat	456.33	456.33	456.33	456.33	456.33
	Panel C: 2SLS estimates, instrument varying by location, year, and age				
$\widehat{subs}$	0.18	-0.62	-0.26	-1.18**	-0.28
	(0.29)	(0.42)	(0.37)	(0.57)	(0.36)
F-stat	6.06	6.06	6.06	6.06	6.06
	Panel D: over-identified 2SLS estimates				
$\widehat{subs}$	0.16***	-0.068	-0.054	-0.33***	-0.26***
	(0.036)	(0.057)	(0.047)	(0.047)	(0.029)
F-stat	228.36	228.36	228.36	228.36	228.36
p-value	0.93	0.17	0.57	0.048	0.95
y mean	0.83	0.90	0.077	0.071	0.039

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table estimates the incentive effect of subsidy on main outcomes, from a basic specification controlling for main effects of location, year, age, and demographic variables, and region-year fixed effects. Panel A shows OLS estimates on endogenous subsidy $subs$. Panel B shows 2SLS estimates instrumented by $subiv$ varying at location, year, age, and demographics. Panel C shows 2SLS estimates instrumented by $sublean$ varying at location, year, and age. Panel D uses both instruments, and reports p-value from the Hansen over-identification test. Robust standard errors clustered at the level of PUMA in the parenthesis.

Table B4: Incentive effect of subsidy, second instrument simulated from Massachusetts

	(1)	(2)	(3)	(4)	(5)
	any insurance	employed	in labor force	ESI + employed	ESI + not employed
			Panel A: instrument *sublean* and *sublean_ma*		
$\widehat{subs}$	0.20	-0.42	-0.35	-0.52	-0.19
	(0.29)	(0.34)	(0.33)	(0.37)	(0.29)
F-stat	3.03	3.03	3.03	3.03	3.03
p-value	0.88	0.39	0.77	0.061	0.60
			Panel B: instrument *subiv* and *sublean_ma*		
$\widehat{subs}$	0.10***	-0.055	-0.057	-0.084	-0.43***
	(0.025)	(0.052)	(0.043)	(0.078)	(0.043)
F-stat	261.84	261.84	261.84	261.84	261.84
p-value	0.17	0.38	0.42	0.79	0.35
y mean	0.95	0.77	0.83	0.64	0.10

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table estimates the incentive effect of subsidy on main outcomes, using different simulated generosity as instruments. Panel A estimates the basic specification that controls for location, year, age, and demographics, and region-year fixed effects. I instrument endogenous subsidy *subs* with *sublean* from the national sample and *sublean_ma* from the Massachusetts sample. Panel B estimates the main specification in Equation 17. Endogenous subsidy *subs* is instrumented by *subiv* from the national sample and *sublean_ma* from the Massachusetts sample. First-stage F statistics and p-value from the Hansen over-identification test are reported in each panel. Robust standard errors clustered at the level of PUMA in the parenthesis.

Table B5: Robustness analysis: border PUMAs

	(I) any insurance	(II) employed	(III) in labor force	(IV) ESI + employed	(V) ESI + not employed
			Panel A: main results		
subiv	0.10***	-0.055	-0.056	-0.26***	-0.33***
	(0.025)	(0.053)	(0.044)	(0.058)	(0.023)
R^2	0.071	0.090	0.10	0.13	0.054
			Panel B: average subsidy (instead of price) for border PUMA		
subiv	0.10***	-0.055	-0.056	-0.26***	-0.33***
	(0.025)	(0.053)	(0.044)	(0.058)	(0.023)
R^2	0.071	0.090	0.10	0.13	0.054
			Panel C: assign border PUMA to larger region		
subiv	0.093***	-0.059	-0.061	-0.28***	-0.33***
	(0.024)	(0.053)	(0.044)	(0.059)	(0.024)
R^2	0.061	0.088	0.10	0.13	0.053
			Panel D: cluster by region-age band		
subiv	0.093**	-0.059	-0.061	-0.28***	-0.33***
	(0.033)	(0.062)	(0.060)	(0.070)	(0.036)
R^2	0.061	0.088	0.10	0.13	0.053

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Notes: Table shows reduced-form estimates of subsidy incentive *subiv* from the main specification, treating border PUMAs (intersecting two rating regions) differently. In Panel A, the main result averages premium across regions and then calculates subsidy. Panel B calculates subsidy under each region before averaging. Panel C assigns border PUMAs to region with the greatest population share. Panel D uses the region division in C, but clusters standard error at the level of region and age band. In Panel A-C, cluster is at the level of 52 PUMAs.

Table B6: Consumption, in thousands of dollars

	(I) $c_{1\cdot\cdot}$	(II) $c_{11\cdot}$	(III) $c_{\cdot 2\cdot}$	(IV) $c_{\cdot 0\cdot}$	(V) $c_{\cdot >0,0}$	(VI) $c_{\cdots}$
	Panel A: non-medical consumption					
mean	44.00	46.41	27.84	24.95	33.53	38.82
s.e.	(3.10)	(3.45)	(6.16)	(6.50)	(6.92)	(2.42)
$\frac{\cdot}{c_{1\cdot\cdot}}$	1	1.05	0.63	0.57	0.76	0.88
	Panel B: food consumption					
mean	1.92	1.94	1.78	1.76	1.59	1.87
s.e.	(0.09)	(0.10)	(0.14)	(0.34)	(0.22)	(0.07)
$\frac{\cdot}{c_{1\cdot\cdot}}$	1	1.01	0.93	0.92	0.83	0.97
N	242	203	50	22	18	345
externality	τ_{pb}	τ_{pr}	λ_p	k	uc_p	Π

Notes: Table shows average quarterly non-medical consumption expenditures (in thousands of dollars) in Panel A, and food expenditures in Panel B, for beneficiary groups. Sample includes 27-64 Massachusetts adults in 2011 panel of Consumer Expenditure Survey (CEX). Standard error of mean estimates in the parenthesis.

Table B7: Welfare effect of subsidy, alternative incidence of uncompensated cost

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$			(III) $\frac{dW_\Pi}{d\lambda_p \cdot p}$	(IV) $\frac{dW_C}{d\lambda_p \cdot p}$	(V) (1)+(2)+(3)+(4)			(VI) $\frac{dW_F}{d\lambda_p \cdot p}\Big\|_{u'}$
		$\alpha = 0.32$	$\alpha = 1$	premium tax			$\alpha = 0.32$	$\alpha = 1$	premium tax	
Panel A: non-medical consumption										
$\gamma = 0$	0.232	0.074	0.074	0.077	0.011	-0.338	-0.021	-0.021	-0.018	-0.018
$\gamma = 1$	0.368	0.082	0.097	0.086	0.013	-0.342	0.121	0.136	0.125	-0.028
$\gamma = 2$	0.585	0.092	0.128	0.095	0.014	-0.346	0.345	0.381	0.348	-0.045
$\gamma = 3$	0.928	0.105	0.169	0.106	0.016	-0.350	0.699	0.763	0.700	-0.071
Panel B: food consumption										
$\gamma = 1$	0.249	0.079	0.089	0.079	0.011	-0.339	0	0.010	0	-0.018
$\gamma = 2$	0.268	0.085	0.107	0.080	0.012	-0.339	0.026	0.048	0.021	-0.019
$\gamma = 3$	0.288	0.092	0.129	0.082	0.012	-0.340	0.052	0.089	0.042	-0.021

Notes: Table calculates the welfare effect of premium subsidy under different scenarios of uncompensated care incidence. $\alpha = 0.32$ corresponds to the main analysis where patient surcharge finances 32% of uncompensated cost. Assuming the remaining cost (assessment) is completely passed-through to commercially insured patients, $\alpha = 1$. Alternative pass-through to insurers amounts to a premium tax on enrollees, and subsidy lowers the tax by: $\frac{dW_{UC}}{d\lambda_p \cdot p} = -g \frac{u'(c_{\cdot>0\cdot})}{u'(c_{1\cdot\cdot})} \frac{ri}{p} \left(\frac{1}{1-\lambda_0} + \varepsilon_{ri,\lambda_0} \right) \frac{d\lambda_0}{d\lambda_p}$, where $\frac{u'(c_{\cdot>0\cdot})}{u'(c_{1\cdot\cdot})} = 0.90^{-\gamma}$ for non-medical consumption, and $0.98^{-\gamma}$ for food.

Table B8: Consumption, unsubsidized population > 300% FPL, in thousands of dollars

	(I) $c_{1\cdot\cdot}$	(II) $c_{11\cdot}$	(III) $c_{\cdot 2\cdot}$	(IV) $c_{\cdot 0\cdot}$	(V) $c_{\cdot >0,0}$	(VI) $c_{\cdot\cdot\cdot}$
			Panel A: non-medical consumption			
mean	49.21	49.44	70.84	17.98	37.71	48.54
s.e.	(3.60)	(3.74)	(16.42)	(7.00)	(7.48)	(3.15)
$\frac{\cdot}{c_{1\cdot\cdot}}$	1	1	1.44	0.37	0.77	0.99
			Panel B: food consumption			
mean	2.07	2.03	2.64	2.15	1.82	2.10
s.e.	(0.10)	(0.10)	(0.33)	(1.21)	(0.25)	(0.09)
$\frac{\cdot}{c_{1\cdot\cdot}}$	1	1	1.28	1.04	0.88	1.01
N	189	175	14	6	14	231
externality	τ_{pb}	τ_{pr}	$\lambda_p = 0$	$k = \frac{1}{2}$	uc_p	Π

Notes: Table shows average quarterly non-medical consumption expenditures (in thousands of dollars) in Panel A, and food expenditures in Panel B, for beneficiary groups with income above 300% FPL. Sample includes 27-64 Massachusetts adults in 2011 panel of Consumer Expenditure Survey (CEX). Standard error of mean estimates in the parenthesis.

Table B9: Welfare effect of penalty, alternative incidence of uncompensated cost

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$			(III) $\frac{dW_C}{d\lambda_p \cdot p}$	(IV) (1)+(2)+(3)			(V) $\frac{dW_F}{d\lambda_p \cdot p}\Big\|_{u'}$
		$\alpha = 0.32$	$\alpha = 1$	premium tax		$\alpha = 0.32$	$\alpha = 1$	premium tax	
Panel A: non-medical consumption									
$\gamma = 0$	-0.023	0.011	0.012	0.012	0.003	-0.009	-0.008	-0.008	-0.006
$\gamma = 1$	-0.062	0.013	0.015	0.012	0.003	-0.046	-0.044	-0.047	-0.016
$\gamma = 2$	-0.168	0.014	0.020	0.012	0.003	-0.151	-0.145	-0.153	-0.044
$\gamma = 3$	-0.454	0.016	0.025	0.012	0.003	-0.435	-0.426	-0.439	-0.118
Panel B: food consumption									
$\gamma = 1$	-0.022	0.012	0.013	0.012	0.003	-0.007	-0.006	-0.007	-0.006
$\gamma = 2$	-0.021	0.012	0.015	0.011	0.003	-0.006	-0.003	-0.007	-0.006
$\gamma = 3$	-0.020	0.013	0.017	0.011	0.003	-0.004	0	-0.006	-0.005

Notes: Table calculates the welfare effect of mandate penalty under different scenarios of uncompensated care incidence. I restrict the analysis to the unsubsidized population with income above 300% FPL. $\alpha = 0.32$ corresponds to the main analysis where patient surcharge finances 32% of uncompensated cost. Assuming the remaining cost (assessment) is completely passed-through to commercially insured patients, $\alpha = 1$. Alternative pass-through to insurers amounts to a premium tax on enrollees, and penalty lowers the tax by $\frac{dW_{UC}}{dk \cdot p} = -g \frac{u'(c_{\cdot>0\cdot})}{u'(c_{1\cdot\cdot})} \frac{ri}{p} \left(\frac{1}{1-\lambda_0} + \varepsilon_{ri,\lambda_0} \right) \frac{d\lambda_0}{dk}$, where $\frac{u'(c_{\cdot>0\cdot})}{u'(c_{1\cdot\cdot})} = 1^{-\gamma}$ for non-medical consumption, and $1.01^{-\gamma}$ for food.

Table B10: Welfare effect of subsidy, alternative calculation of fiscal cost

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$	(III) $\frac{dW_\Pi}{d\lambda_p \cdot p}$	(IV) $\frac{dW_C}{d\lambda_p \cdot p}$						(V) $\frac{dW_B}{d\lambda_p \cdot p} + \frac{dW_C}{d\lambda_p \cdot p} + \frac{dW_\Pi}{d\lambda_p \cdot p} + \frac{dW_C}{d\lambda_p \cdot p}$						(VI) $\frac{dW_F}{d\lambda_p \cdot p}\Big\|_{u'}$
				$\frac{de}{d\lambda_p} = 0$	$= -0.08$	$= -0.16$	$\frac{d\lambda_1^e}{d\lambda_p} = 0$	$\frac{d\lambda_1^{1-e}}{d\lambda_p} = 0$	$\frac{d\lambda_1}{d\lambda_p} = 0$	$\frac{de}{d\lambda_p} = 0$	$= -0.08$	$= -0.16$	$\frac{d\lambda_1^e}{d\lambda_p} = 0$	$\frac{d\lambda_1^{1-e}}{d\lambda_p} = 0$	$\frac{d\lambda_1}{d\lambda_p} = 0$	
Panel A: non-medical consumption																
$\gamma = 0$	0.242	0.071	0.010	-0.358	-0.423	-0.487	-0.311	-0.462	-0.381	-0.035	-0.100	-0.164	0.012	-0.139	-0.058	-0.017
$\gamma = 1$	0.390	0.076	0.011	-0.364	-0.429	-0.493	-0.318	-0.460	-0.381	0.113	0.049	-0.016	0.159	0.017	0.096	-0.027
$\gamma = 2$	0.630	0.083	0.013	-0.369	-0.434	-0.498	-0.324	-0.459	-0.381	0.357	0.292	0.228	0.402	0.267	0.345	-0.044
$\gamma = 3$	1.015	0.091	0.015	-0.374	-0.439	-0.503	-0.330	-0.458	-0.381	0.747	0.682	0.618	0.791	0.663	0.740	-0.071
Panel B: food consumption																
$\gamma = 1$	0.260	0.076	0.010	-0.360	-0.425	-0.489	-0.312	-0.423	-0.381	-0.014	-0.079	-0.143	0.034	-0.077	-0.035	-0.018
$\gamma = 2$	0.280	0.081	0.011	-0.360	-0.425	-0.489	-0.313	-0.423	-0.381	0.012	-0.053	-0.117	0.059	-0.051	-0.009	-0.019
$\gamma = 3$	0.300	0.088	0.011	-0.361	-0.426	-0.490	-0.314	-0.422	-0.381	0.038	-0.027	-0.091	0.085	-0.023	0.018	-0.021

Notes: Table calculates the welfare effect of premium subsidy under different calibration of fiscal costs. Sample includes Massachusetts adults aged 19-64 in Massachusetts. $\frac{de}{d\lambda_p} = 0$ corresponds to the baseline calibration for this sample, where $\frac{d\lambda_1^2}{d\lambda_p} = -0.25$, $\frac{d\lambda_1^{1-e}}{d\lambda_p} = -0.27$, and $\frac{d\lambda_1}{d\lambda_p} = -0.52$. I vary the labor response to subsidy to equal either half or the full size of take-up (-0.08 and -0.16), and shut down ESI selection from workers, non-workers, or both, when employment effect is zero. Incidence of uncompensated cost follows the main analysis, with $\alpha = 32\%$ born by patient surcharge finances 32% of uncompensated cost. Welfare weights (based on consumption) differ only slightly from the 27-64 sample in the main analysis.

Table B11: Welfare effect of subsidy, alternative moral hazard on formal insurance spending

	(I) $\frac{dW_B}{d\lambda_p \cdot p}$	(II) $\frac{dW_{UC}}{d\lambda_p \cdot p}$			(III) $\frac{dW_{\Pi}}{d\lambda_p \cdot p}$	(IV) $\frac{dW_C}{d\lambda_p \cdot p}$	(V) (1)+(2)+(3)+(4)			(VI) $\frac{dW_F}{d\lambda_p \cdot p}\Big\|_{u'}$
		$g = 1$	$g = 0.8$	$g = 0.7$			$g = 1$	$g = 0.8$	$g = 0.7$	
Panel A: non-medical consumption										
$\gamma = 0$	0.232	0.089	0.074	0.062	0.011	-0.338	-0.006	-0.021	-0.033	-0.018
$\gamma = 1$	0.368	0.095	0.082	0.067	0.013	-0.342	0.134	0.121	0.106	-0.028
$\gamma = 2$	0.585	0.104	0.092	0.073	0.014	-0.346	0.357	0.345	0.326	-0.045
$\gamma = 3$	0.928	0.114	0.105	0.080	0.016	-0.350	0.708	0.699	0.674	-0.071
Panel B: food consumption										
$\gamma = 1$	0.249	0.095	0.079	0.066	0.011	-0.339	0.016	0	-0.013	-0.018
$\gamma = 2$	0.268	0.102	0.085	0.071	0.012	-0.339	0.043	0.026	0.012	-0.019
$\gamma = 3$	0.288	0.110	0.092	0.077	0.012	-0.340	0.070	0.052	0.037	-0.021

Notes: Table calculates the welfare effect of premium subsidy for the 19-64 population under different moral hazard on formal insurance spending, modeled as relative generosity of informal insurance g. I re-calculated uncompensated cost saving assuming formal insurance does not increase spending ($g = 1$), increases spending by 25% ($g = 0.8$), or increases spending by 43% ($g = 0.7$). I assume $\alpha = 0.32$ of uncompensated cost is financed by patient surcharge, as in the main analysis.

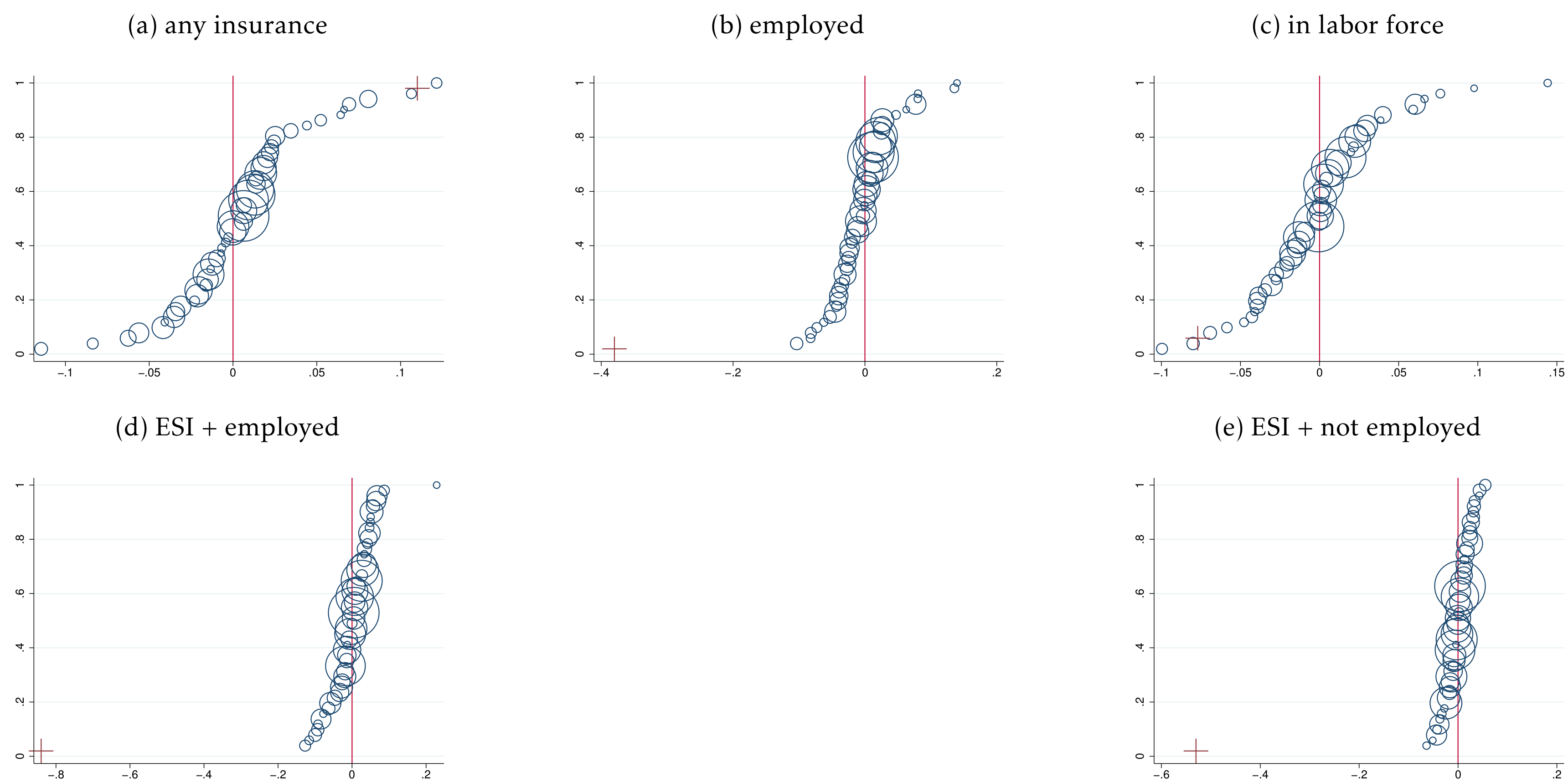
Table B12: Welfare effect of penalty, alternative moral hazard on formal insurance spending

	(I) $\frac{dW_B}{dk \cdot p}$	(II) $\frac{dW_{UC}}{dk \cdot p}$			(III) $\frac{dW_C}{dk \cdot p}$	(IV) (1)+(2)+(3)			(V) $\frac{dW_F}{dk \cdot p}\Big\vert_{u'}$
		$g=1$	$g=0.8$	$g=0.7$		$g=1$	$g=0.8$	$g=0.7$	
Panel A: non-medical consumption									
$\gamma=0$	-0.023	0.014	0.011	0.010	0.003	-0.006	-0.009	-0.010	-0.006
$\gamma=1$	-0.062	0.015	0.013	0.011	0.003	-0.044	-0.046	-0.048	-0.016
$\gamma=2$	-0.168	0.017	0.014	0.013	0.003	-0.148	-0.151	-0.152	-0.044
$\gamma=3$	-0.454	0.019	0.016	0.014	0.003	-0.432	-0.435	-0.437	-0.118
Panel B: food consumption									
$\gamma=1$	-0.022	0.014	0.012	0.011	0.003	-0.005	-0.007	-0.008	-0.006
$\gamma=2$	-0.021	0.015	0.012	0.011	0.003	-0.003	-0.006	-0.007	-0.006
$\gamma=3$	-0.020	0.016	0.013	0.012	0.003	-0.001	-0.004	-0.005	-0.005

Notes: Table calculates the welfare effect of mandate penalty for the 19-64 population above 300% FPL under different moral hazard on formal insurance spending, modeled as relative generosity of informal insurance g. I re-calculated uncompensated cost saving assuming formal insurance does not increase spending ($g = 1$), increases spending by 25% ($g = 0.8$), or increases spending by 43% ($g = 0.7$). I assume $\alpha = 0.32$ of uncompensated cost is financed by patient surcharge, as in the main analysis.

C Appendix Figures

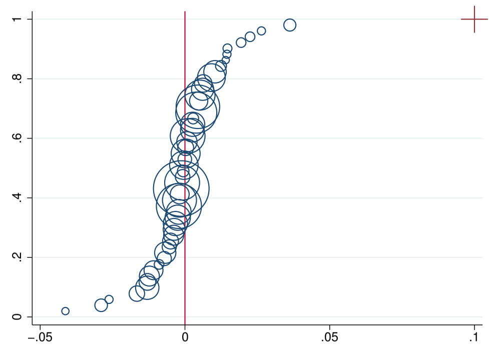
Figure C1: Permutation test: random rating communities in non-MA states



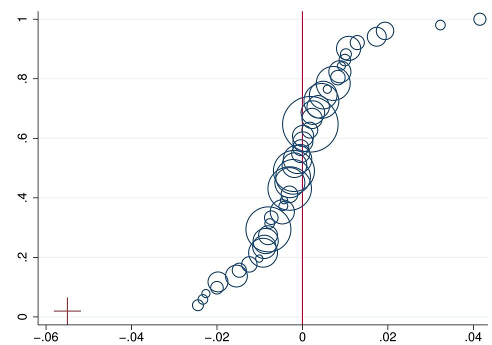
Notes. Graphs plot the empirical cumulative distribution of pseudo estimates from non-MA states and Massachusetts (marked with a plus). I permute actual Massachusetts community rating across location, year and age band cells in control states, and estimate pseudo policy effect using simulated generosity *sublean*, based on the reduced-form specification with main effects of age, PUMA, year, and demographic variables interacted with unemployment rate. Each hollow circle represents estimate from a non-MA state, with the size of the circle corresponding to the number of clustering unit (PUMA) in the state.

Figure C2: Permutation test: random rating communities in non-MA states

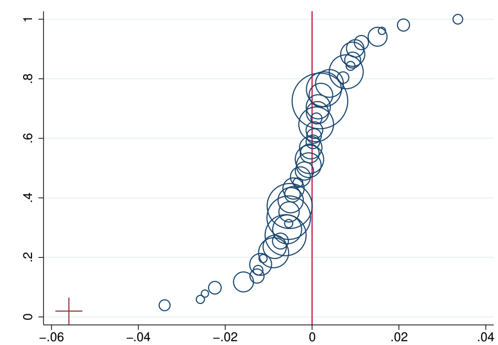
(a) any insurance



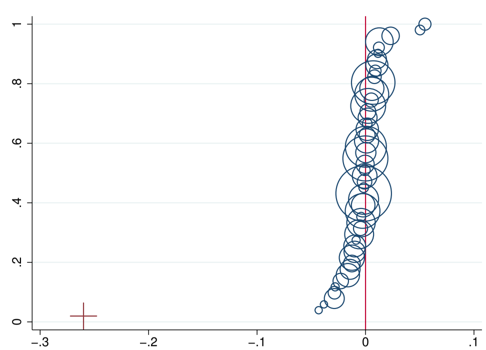
(b) employed



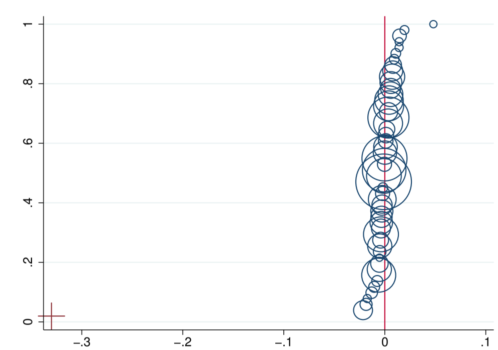
(c) in labor force



(d) ESI + employed



(e) ESI + not employed



Notes. Graphs plot the empirical cumulative distribution of pseudo estimates from non-MA states and Massachusetts (marked with a plus). I permute actual Massachusetts community rating across location, year and age band cells in control states, and Massachusetts affordability across income (demographic) groups. For each state I estimate pseudo policy effect using simulated generosity *subiv*, based on the reduced-form specification with full interaction terms. Each hollow circle represents estimate from a non-MA state, with the size of the circle corresponding to the number of clustering unit (PUMA) in the state.