What Does Insurance Subsidy Do in a Mandate Reform? Evidence from Massachusetts

Hongming Wang*

Abstract

When adverse selection is inherent in insurance pricing and subsidy is tied to premium, universal coverage raises efficiency in both insurance pricing and transfer, and is socially desirable. Subsidy is full for redistributive purposes. Hence there is potential for ACA-like reforms to improve welfare. Actual subsidy, however, depends on the behavioral responses to policy incentives, and how well private valuation and costs align with social counterparts. I quantify the welfare impact of subsidy dollars based on outcome variation in response to subsidy generosity. I find in the Massachusetts case that the social cost of financing new enrollment tends to outweigh the incremental benefit to new enrollees, but accounting for infra-marginal efficiency gain in pricing and transfer, the return to subsidy is positive even at modest risk aversion.

*hongminw@usc.edu, University of Southern California, Department of Economics. I'd like to thank Alice Chen, Jeff Nugent, John Strauss, Geert Ridder, and seminar participants and discussants at USC CESR, WEAI, and ASHECON for helpful comments on the paper. Financial support from the Graduate School and Schaeffer Center is gratefully acknowledged. All errors are mine.

Both the 2010 Patient Protection and Affordable Care Act, and the pilot 2006-2007 reform in Massachusetts, require individuals to purchase health insurance or face a penalty. While employer-sponsored insurance (ESI) covers most of non-disabled working age Americans, public insurance programs play important roles in the health care access of the low-income and vulnerable population. The mandate provision only strengthens the case for insurance transfers: for example, 68% of new enrollees in the first two years of the Massachusetts reform received premium assistance, with 24% enrolling in traditional Medicaid programs and 44% receiving subsidy on the Commonwealth Care.

Despite the clear policy relevance, we still know relatively little about the economic incidence of insurance subsidy in the presence of a mandate. On a very basic level, is it socially desirable to mandate insurance coverage, and direct large amount of resources to assist the insurance purchase of the uninsured? Given the scope of government intervention and the dollars at stake, an economic motivation for universal coverage and insurance transfer is warranted. Practical policy design then balances the benefits and cost of transfer to various members of society, wary of the fact that the policy itself can cause behavioral responses that lower efficiency.

This paper shows that when adverse selection is inherent in insurance pricing, full-subsidy universal coverage is socially optimal. The optimality of universal coverage is due to the efficiency gain in the implicit premium transfer between high and low cost enrollees, *and* the subsidy transfer between high and low income groups. First, universal coverage maximizes the social efficiency of insurance pricing when type-specific contract is not viable: marginal enrollees lower the average cost and the premium payment of all infra-marginal enrollees. As enrollment increases, private willingness to pay of the low-cost types is too low relative to the social value, and coverage too low absent a mandate.

Second, efficient pricing on the insurance market has broader social implications. Because subsidy is tied to premium, more efficient insurance pricing leads to more efficient transfer payments, and both are maximized when coverage is universal. Although subsidizing new enrollees generates additional transfer cost on payers, the new enrollment also lowers premium price and the total transfer cost of all infra-marginal subsidy enrollees. At full-subsidy universal coverage, the net transfer cost is offset by reduction in own premium. It follows that when subsidy flows from a fully insured working popultion to the non-employed, insurance transfer is in fact Pareto.

These results suggest there is scope for ACA-like intervention in the insurance market: mandate and insurance transfers may be effective ways to combat adverse selection and the associated inefficiencies, and improve welfare. Actual transfer policies may deviate from the full-subsidy solution, if private valuation (cost) in decentralized decision making does not align with the social value (cost), or if the policy itself introduces adverse incentive effects on behavior. For the second-best occurrence of the insurance mandate and transfer, I study insurance and labor outcomes in the Massachusetts reform context, and evaluate the practical benefits and cost to various members of society when subsidy becomes marginally more generous.

The characterization of the second-best outcomes is intrinsically empirical: I assume employment decision is optimal given individual type and a set of market prices, but leave the specifics of insurance decisions unspecified. Given the large number of factors relevant for insurance choice, choice sub-optimality, and the selection between public and private options, the empirical approach is reticent on the underlying decision rule, but instead relies on estimated moments and policy-driven change rates to inform welfare. If the interest is in how different transfer generosity affects outcomes and hence welfare, estimated change rates are "sufficient statistics" for the analysis.

In the second-best, there are four ways subsidy dollars improve social welfare. For marginal enrollees, formal insurance provides better risk protection than informal insurance, and the value of the "top-off" is pit against the transfer cost of subsidizing these new premium payments net of any saving from informal insurance transfers. For infra-marginal enrollees, more subsidy dollars are valued for greater premium assistance. Economy wide, new enrollment lowers the (implicit) premium on both the formal and informal market, and in turn raises transfer efficiency when subsidy is tied to pricing. Finally, behavioral responses induced by subsidy pricing and financing create fiscal externality, and the social cost is larger when pre-existing income transfers or other implicit insurance transfers are larger.

To quantify the relative magnitude of the four channels, I estimate the sufficient statistics observing how insurance and labor outcomes respond to varying degrees of subsidy generosity stipulated in the Massachusetts policy schedules. OLS estimates of the change rate are likely biased due to behavioral response to the schedules. Utilizing the fact that the rest of the nation did not undergo the reform, I use a simulated generosity measure from a reference national sample as the instrument.

I assess the validity of the instrument using 1) over-identification tests, where a weaker instrument with no generosity variation over income constrains a stronger instrument that captures the income variation across demographics but is potentially endogenous, 2) Hausman tests (TBD), where the weaker instrument is interpreted as robust but not efficient, and 3), Monte Carlo simulation tests, where random schedules are generated but generosity is coded for the true demographics of potential recipients: if the demographic variation in the stronger instrument is correlated with unobserved determinants of group-specific outcomes, then the correlation tends to bias the pseudo estimates away from zero. Permuting actual Massachusetts generosity across the rest of the 50 states gives similar null results.

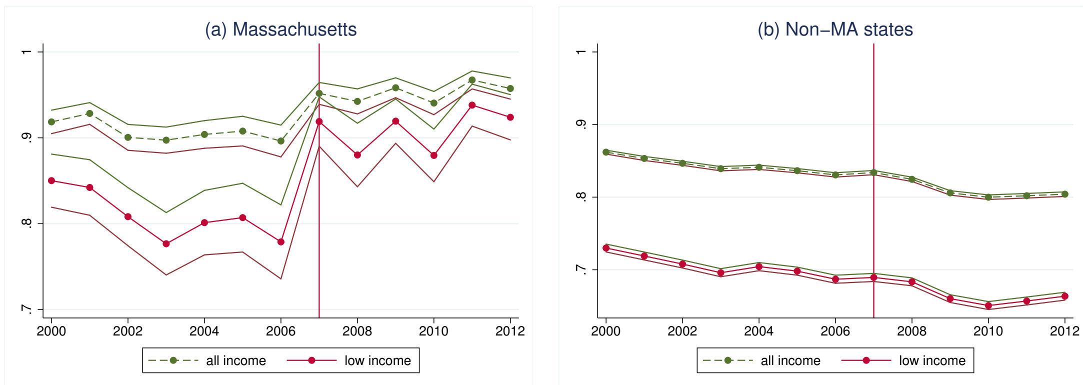
I find modest effect of subsidy on new enrollment: ten percentage point increase in generosity raises coverage by one percentage point. Similar magnitude is found for ACA subsidies. Crowd-out of ESI, on the other hand, is smaller than most estimates based on previous insurance expansions. Labor effect of subsidy is small in the average population, but is significant and large in the near-elderly (55-64) group. Joint insurance-labor outcome suggests the labor reduction associated with insurance selection is also small. I show the IV estimates are consistent with simpler difference-in-difference estimates and graphic evidence following individuals in the SIPP sample over the initiation course of the reform (2005-2007).

Full calibration of the welfare model suggests the social cost of subsidy financing tends to outweigh any benefit to enrollees on the margin, especially when risk aversion is modest and when alternative or informal transfers are large. However, accounting for the social benefit on the infra-margin, such as value of premium assistance to existing enrollees and value of efficient pricing to all insurance and subsidy payers, the overall economic return to subsidy dollars is positive even at modest risk aversion and potentially large implicit transfers. The finding has broad implication for policy designs aimed at balancing cost and coverage.

1 Massachusetts Health Insurance Reform

The Massachusetts comprehensive health reform is signed into law by the then-governor Mitt Romney in April, 2006. Central to the law is the individual mandate requiring all Massachusetts residents over age 17 to acquire affordable health insurance that meets a set of "minimum creditable coverage" standards. Failure to obtain coverage will result in a tax penalty, unless the individual is able to demonstrate economic hardship (for instance, below 150% of FPL) or religious exemption. As evident in Figure 1, Massachusetts has higher pre-reform insurance rate than the nation's average[1], and quickly achieves near universal coverage within first year of the reform.

Figure 1: Insurance coverage trends



Notes. Graphs compare 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. I aggregate micro data for the 27-64 age group in the CPS March supplement, adjusting by insurance weights. 95% confidence intervals are plotted.

The reform coordinates efforts from employers, individuals, private insurers and the government, to reduce uninsurance in the state. Coverage is lower among the low-income population: from Figure 1(a), coverage gain is concentrated among the below 300% FPL group[2]. To cover more low-income families, the state expanded the Medicaid program (MassHealth) and instituted a publicly-subsidized Exchange market known as the Commonwealth Health Insurance Connector. The new MassHealth covers children with family income no greater than 300% FPL, up from the previous 200% cap. Low income population ineligible for Medicaid (for example, non-elderly,

[1] This is attributed to a few factors. First, ESI coverage is high in Massachusetts before the reform: 76% of the 27-64 age group is covered by employers in Massachusetts during 2001-2005, whereas the national average is 70% (author calculation from CPS). Second, regulation of the individual market already implemented guarantee issue and community rating in the state. Also relevant is the fact that Massachusetts merged the risk pool of small group and non-group plans, significantly reducing the premium in the latter group.

[2] Similarly, according to Massachusetts Health Reform Surveys, 2006 and 2007, uninsurance rate in fall, 2006 is 23.8% among the population below 300% FPL, and decreased to 12.9% in fall, 2007. The numbers for the above 300% FPL group are 5.2% in 2006 and 2.9% in 2007, respectively. For more details on coverage gain in the first year, see Long (2008).

non-disabled, childless adults with income above 133% FPL) can obtain coverage from the Connector, a state clearing house bringing together consumers and state-certified individual plans. The Connector has a subsidized Commonwealth Care program and an unsubsidized Commonwealth Choice program. Commonwealth Care is open to eligible individuals whose family income is no greater than 300% of FPL and who are not offered health insurance from their employers. Those not eligible for subsidy can buy from the Commonwealth Choice program.

Affordability and premium subsidy schedules are released in the middle of the previous year. The Connector sets the maximum monthly premium a person in a given income bracket needs to pay towards her coverage, or the affordability. The amount is zero for individuals with family income below 150% FPL. In 2010, for example, out-of-pocket premium cap for an individual plan is $ 39 per month for the 150-200% bracket, $ 77 per month for the 200-250% bracket, and $ 119 for the 250-300% bracket. Individuals are not held accountable for failing to enroll in plans with premium contribution exceeding their affordability threshold.

Subsidy schedule works closely with affordability to ensure most of the low-income population are eligible for a subsidized Commonwealth Care plan. Starting year 2008, after applying subsidy, enrollee contribution falls to 0 for those below 150% FPL, roughly 10% for the 150-200% bracket, 20% for 200-250% and 30% for 250-300%. Table 1 breaks down insurance enrollment in Massachusetts by source of coverage. Commonwealth Care contributed the largest share of new enrollees by June of 2008: of the 442,000 new enrollees, around 40% received subsidy from the program. The second largest source of increase is employer provided group plans (33%). Combining MassHealth and Commonwealth Care, more than half (68%) of the new enrollees in the first two years of the reform received some state assistance in insurance purchase.

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.

Alongside the individual mandate, Massachusetts also implements an employer mandate which requires employers with more than 11 full-time equivalent workers make fair and reasonable contribution towards the premium cost of full-time employees, or pay a fine up to \$ 295 per worker. Employers must also provide its employees a section 125 plan that pays insurance premium on a pre-tax basis. Failing to do so will incur a free rider surcharge. Additionally, employers have Health Insurance Responsibility Disclosure (HIRD) obligation, and must collect signed HIRD forms from employees who decline to enroll in employer-sponsored plans. However, amidst concerns over administrative costs to firms, and in anticipation of the federal reform phasing in, the state repealed all provisions in the employer mandate by June, 2014.

The employer mandate is responsible for the small coverage gain among the high income group above 300% FPL (Appendix Figure 1), where coverage is already near-universal in the baseline. In the low income group, however, the substantial coverage gain is almost completely driven by rising generosity of public assistance programs in Massachusetts, with no significantly different trending in ESI coverage rate compared to the rest of the country. Because employees opting out of ESI are not eligible for subsidy, the employer mandate is valuable in setting a default coverage option for workers, limiting the degree of crowd-out relative to a scenario where workers can freely choose between ESI and subsidized coverage.

A growing literature has looked at the Massachusetts experience and shown generally positive effects on the insurance market. For example, the reform resulted in greater rate coverage (Long, Stockley, and Yemane, 2009), better health care access (Kolstad and Kowalski, 2012) among the state's poor, and lower premium (Graves and Gruber, 2012). Although most studies are positive in nature, Hackmann, Kolstad and Kowalski (2015) characterizes the welfare loss of adverse selection exploiting the individual mandate. Mandate penalty increases the willingness to pay for premium, increases enrollment and lowers equilibrium premium. Recovered cost curve lies below the demand curve for enrollees in the unsubsidized Commonwealth Choice program. In contrast, Finkelstein, Hendren and Shepard (2017) finds that insurance demand is substantially lower than cost for the low-income subsidized enrollees on Commonwealth Choice, and raises issues of implicit insurance and uncompensated care nonetheless provided to the formally uninsured. I address how these issues might affect the argument for subsidy expansion in this paper.

The reform is also found to have broader impacts on the labor market. For example, Heim and Lin (2016) shows increased retirement rate among the near-elderly following the Massachusetts reform. Using tax returns, Heim and Lurie (2015) shows modest increase in mobility among low-income workers. On the more normative side, Kolstad and Kowalski (2016) evaluates the efficiency of employer provision of insurance under the individual mandate. Similar to the Commonweath Choice enrollees, workers that newly obtained ESI have high willingness to pay for the insurance, as evidenced by accepting wages lower by almost the full cost of premium. This again contrasts with the privately uninsured, where government transfers seem to play more prominent roles for either formal or implict coverage.

2 A Model of Social Insurance

Should the government provide health insurance to all members of the society? In private insurance markets, coverage is seldom universal: even at very low prices, take-up is low. Relatedly, revealed willingness to pay (WTP) for insurance is also low, sometimes even below own expected medical cost. Hence based on private WTP alone, an insurance mandate hurts the well-being of the lowest valuation types and is not a welfare-enhancing policy. However, private WTP may not be the ideal metric of welfare, if market failures and behavioral biases suppress the private valuation of insurance below the social value. In these cases, the government has a unique role in correcting the problem with policies such as mandate and subsidy.

In this paper, I focus on two potential sources of market failures: adverse selection in insurance pricing, and moral hazard in insurance transfers. Under adverse selection, there is positive correlation between insurance demand and cost: premium based on higher cost enrollees is above the WTP of low cost types. The WTP of the low cost types is too low in the sense that if they enroll, they check the negative feedback loop on cost and pricing, attract even lower cost types to enroll, and benefit all infra-marginal enrollees who pay less for insurance.

With insurance transfer, the government can expand the WTP of individuals priced out of insurance due to adverse selection. The transfer would have some members of the economy pay for the insurance cost of others, and the moral hazard problem associated with the transfer is smaller, if the insurance pricing is closer to the efficient level, or if coverage is closer to universal in the case of adverse selection. Hence the incidence of insurance subsidy not only opens up new channels of income redistribution, but also strengthens the classic argument of universal coverage in combating inefficiencies in insurance pricing.

In what follows, I present an economy populated by agents heterogenous in risk and productivity[3]. Assuming risk aversion and no additional cost of social insurance, the first-best involves universal coverage where agents pay actuarially fair premium based on own risk type, and income redistribution across productivity types. I then study the planner's allocation where insurance pricing is uniform rather than type-specific; that is, the social implication of adverse selection and insurance transfers are optimized in the employment and insurance outcomes assigned to individuals. I consider it an intermediate benchmark between the full-information first-best outcome and the second-best outcome under uniform pricing. Coverage is still universal, and out of both redistributive and efficiency concerns, features a full-subsidy insurance transfer to the non-employed.

I then study decentralized decision making by utility-maximizing individuals facing uniform insurance pricing. The characterization of the second-best outcomes is intrinsically empirical: I only require individuals to optimize on the extensive margin of the employment decision, but leave the specifics of the insurance selection unspecified. Given the large number of factors relevant for insurance choice, choice sub-optimality, and the selection between public and private options, the empirical approach is reticent on the data generating process, and relies on estimated moments and policy-driven change rates to inform welfare. These estimates serve as the common ground for a variety of model-based decision rules that explain the data. I summarize four practical channels that premium subsidy brings benefits or cost to various members of society. The empirical section of the paper centers around the statistics that quantify these four channels.

I begin the discussion with the environment of the economy, and the first-best allocation.

2.1 Setting and first-best allocation

The economy is populated by a continuum of individuals differing by risk μ and labor productivity ν. $\mu \in [0,1]$ is the probability of staying healthy and not requiring any medical care. The cost of medical treatment is normalized to a constant M. Hence type μ has expected medical cost $(1-\mu)M$.

Individuals who are employed generate output valued at w, and incur a fixed cost of participation $g(\frac{1}{\nu})$ which depends on worker productivity $\nu \in [0,1]$: higher productivity workers spend less effort to produce the

[3] I do not study the case where individuals' perceived risk type differs from the true type, or other behvioral biases that enter utility, which may also drive a wedge between private market outcomes and the social optimal.

output, and have lower disutility from working. I normalize $g(1) = 0$ and $g(+\infty) = +\infty$, so that the highest productivity type always works, and the lowest type never works[4].

I allow for arbitrary correlation between risk and productivity types. The density function $f(\nu, \mu)$ over $[0,1] \times [0,1]$ is assumed to be strictly positive everywhere, and continuosly differential in both arguments.

The social planner assigns each type an employment outcome $e(\nu, \mu) \in \{0, 1\}$, and an insurance outcome $h(\nu, \mu) \in \{0, 1\}$. Consider the insurance provision by the planner as charging individuals their type-specific premium $p(\nu, \mu)$, and the total premium payment from the insured equals the total cost of enrollees receiving medical care. In other words, the insurance is a resource transfer at no additional administrative cost to the planner. The assumption is common in the public economics literature, and plausible when insurance is institutionally arranged by the government.

Moreover, the planner redistributes total output to generate consumption for individuals. Let $t(\nu, \mu)$ denote the transfer type (ν, μ) receives from the planner. In aggregate, $t(\nu, \mu)$ sums up to zero. The desirably of income transfer aross productivity types may depend on the insurance transfer across the risk types, and both considerations are intertwined in the planner's problem.

Individual has von Neumann-Morgenstern utility over consumption in different states of the world. Specifically, $U(\nu, \mu) = \mu u(c_H(\nu, \mu)) + (1 - \mu) u(c_S(\nu, \mu)) - e(\nu, \mu) g(\frac{1}{\nu})$, where consumption in the healthy state $c_H(\nu, \mu) = w e(\nu, \mu) + t(\nu, \mu) - p(\nu, \mu) h(\nu, \mu)$, and consumption in the sick state $c_S(\nu, \mu) = c_H(\nu, \mu) - M + M h(\nu, \mu)$. Agents are risk-averse: $u'(c) > 0$, $u''(c) < 0$. Utility is state-independent and uniform across all agents, although the latter assumption is not essential for first-best insurance allocation: so long as the individual is slightly risk averse, paying actuarially fair premium raises utility and social welfare.

Utility satisfies the Inada condition $u'(0) = +\infty$: marginal utility is infinite when agent nearly spends all income on health-related expenditures and consumes very little. A direct corollary is that the planner will always transfer enough resources to an uninsured and unemployed person so that consumption is positive after paying M. This ensures that any insurance and medical payment is "affordable" given income. The joint nature of income and insurance transfer rules out specifications such as the constant absolute

[4]The key implication of the model holds if productivity has a multiplier effect on output: $w = \nu \bar{w}$, where $\bar{w}$ is the maximum output constrained by capital and technology in the economy. For simplicity, I work with fixed w in the main analysis.

risk aversion (CARA), which would imply an insurance level independent of income.

A welfarist planner maximizes the following social welfare function by choosing allocation $\{e, h, p, t\}_{(\nu,\mu)}$ in the economy:

$$\max_{e,h,p,t} W = \iint_{[0,1]^2} U(\nu,\mu)\, dF(\nu,\mu)$$

subject to the resource constraints

$$\iint_{h(\nu,\mu)} p(\nu,\mu)\, dF(\nu,\mu) = M \iint_{h(\nu,\mu)} (1-\mu)\, dF(\nu,\mu)$$
$$\iint_{[0,1]^2} t(\nu,\mu)\, dF(\nu,\mu) = 0$$

The solution of the program can be summarized as the follows.

Proposition 1. *With risk-averse agents ($u'(c) > 0$, $u''(c) < 0$, $u'(0) = +\infty$) heterogeneous in risk type μ and productivity ν, in the first-best allocation,*

1. *coverage is universal:* $h^{FB}(\nu,\mu) = 1$
2. *individuals are charged expected cost of medical care:* $p^{FB}(\nu,\mu) = (1-\mu)M$
3. *the planner redistributes labor earning with transfer* $t^{FB}(\nu,\mu)$ *such that*
   (a) *average marginal utility equalize between workers and non-workers:* $E[u'(c^{FB}) \mid e^{FB}(\nu,\mu) = 1] = E[u'(c^{FB}) \mid e^{FB}(\nu,\mu) = 0]$
   (b) *transfer is feasible:* $E[t^{FB}] = 0$
4. *the planner chooses employment* $e^{FB}(\nu,\mu)$ *such that* $e(\nu \geq x(\mu), \mu) = 1$, *and that*
   $$E[\Delta U \,|\, (x(\mu),\mu)] = E[u' \,|\, (\nu < x(\mu),\mu)] \cdot E[\Delta t \,|\, (x(\mu),\mu)],$$
   where $\Delta = \lim_{\nu \to x(\mu)^+} - \lim_{\nu \to x(\mu)^-}$ *gives the increases in utility (ΔU) and transfer (Δt) when the marginal types work.*

Proof. *Appendix*

Universal coverage follows directly from risk aversion. Suppose an allocation leaves type (ν,μ) uninsured. At any given degree of risk aversion, the individual is willing to pay for insurance $\pi(\nu,\mu) > (1-\mu)M$ (Jensen inequality). If the planner insures (ν,μ) at actuarially fair rate $(1-\mu)M$, (μ,ν)

is better-off. This new insurance does not require additional transfer from other individuals: because $w\,e(\nu,\mu)+t(\nu,\mu) > M$ for the uninsured (Inada condition), $p(\nu,\mu) = (1-\mu)M$ is always affordable. Hence no one is made worse-off. The planner can always Pareto-improve social welfare till all individuals are covered[5].

As individuals cover their own expected medical cost, the worse risk types have lower consumption at given income. Redistribution in a fully insured economy is desirable because agents are heterogenous in their income generating ability *and* their resource cost of insuance. Condition 3(a) states that transfer payment equalizes marginal consumption across employment statets. In aggregate, we have full unemployment insurance that compensates for the insurance cost of risk types.

Employment allocation follows a cut-off strategy: at a given risk type, productivity types above a threshold $v^* = x(\mu)$ produce output w, and those below are not employed. Along this margin, putting one more individual to work lowers the utility of the new worker by $E[\Delta U|(x(\mu),\mu)]$, and the planner saves on transfer payment to that worker by $E[\Delta t|(x(\mu),\mu)]$. However, movement on the margin has infra-marginal implications. The output generated by the new worker benefits the consumption of all lower-productivity net transfer recipients ($E[u'|(\nu < x(\mu),\mu)]$), at the same time lowering the average transfer borne by higher-productivity workers ($E[u'|(\nu \geq x(\mu),\mu)]$), with $E[u'|(\nu < x(\mu),\mu)] = E[u'|(\nu \geq x(\mu),\mu)]$ from the transfer condition. Hence the planner balances the net utility loss on the margin with the *social* benefit of transfer to all members of the society.

2.2 Constrained first-best

The first-best case has the planner using type-specific insurance pricing and transfers. In practice, risk and productivity types can be difficult to observe or to contract on, and the government often charges uniform insurance prices, and provide lump-sum cash transfers. When the planner is restricted to using only second-best contracts independent of individual types, what is the optimal size of insurance, employment, and social transfer?

The case of the constrained first-best is interesting, because it provides the best-case coverage rate when adverse selection is inherent in insurance pricing. Consider the lowest cost type $\mu = 1$, who has precisely zero expected cost *and* private WTP for insurance. Average cost, however, is always above

[5]It can be similarly shown that full insurance is preferable to any partial insurance contract that pays less than M in the sick state.

zero. Absent social considerations, the lowest type never takes up insurance under uniform pricing, and coverage never universal. On the other hand, for a given set of transfer contracts, if universal coverage remains a desirable outcome in the constrained first-best, then there is scope for the government to extend coverage to all using similar contracts. The nature of the social value of insurance missing in private WTP, and the policy implications for income and insurance transfer, are the key interests in the constrained first-best.

For the argument of universal coverage, it suffices to consider a cut-off strategy where only risk types below a threshold get coverage. Specifically, I let n_e denote the last risk type to obtain coverage who is employed, and n_{1-e} denote the last enrollee who is not employed. At a given employment allocation, which again follows a cut-off strategy $x(\mu)$, the uniform premium price p is determined by the expected cost of enrollees: $p = M \cdot E[1-\mu | h(\nu,\mu) = 1] = M\Big(1 - \frac{\int_0^{n_e}\int_{x(\mu)}^1 \mu\, dF(\nu,\mu) - \int_0^{n_{1-e}}\int_0^{x(\mu)} \mu\, dF(\nu,\mu)}{\int_0^{n_e}\int_{x(\mu)}^1 dF(\nu,\mu) - \int_0^{n_{1-e}}\int_0^{x(\mu)} dF(\nu,\mu)}\Big)$.

The non-employed receive two forms of transfer from the planner, funded by taxation on labor earning w of workers. First, they receive cash transfer A. The idea of cash transfer supporting a subsistence consumption in the case of high medical expenditure is implicit in the Inada condition. In addition, insurance enrollees receive premium subsidy at rate $\lambda_p \in [0,1]$. Since the amount of insurane subsidy is linked to insurance pricing, enrolling the low cost types not only has market externality on insurance pricing, but further has social externality on all net payers and recipients of subsidy.

In a notational shorthand, I let e denote the size of employment given an allocation $x(\mu)$: $e = \int_0^1 \int_{x(\mu)}^1 dF(\nu,\mu)$. Utility in the second-best case is the follows:

$$U(\nu,\mu) = \begin{cases} u(c_{11}) - g(\frac{1}{\nu}), & \mu \le n_e,\ \nu \ge x(\mu) \\ \mu\, u(c_{10}) + (1-\mu)\, u(c_{10} - M) - g(\frac{1}{\nu}), & \mu > n_e,\ \nu \ge x(\mu) \\ u(c_{01}), & \mu \le n_{1-e},\ \nu < x(\mu) \\ \mu\, u(c_{00}) + (1-\mu)\, u(c_{00} - M), & \mu > n_{1-e},\ \nu < x(\mu) \end{cases}$$

where $c_{11} = w - \frac{1-e}{e}A - \Big[\frac{\lambda_p}{e}\int_0^{u_{1-e}}\int_0^{x(\mu)} dF(\nu,\mu) + 1\Big]p$: that is, the insured workers pay for the cash transfer and the insurance subsidy of the non-employed, in addition to own insurance premium. c_{10} is the consumption of workers who are uninsured, and is greater than c_{11} by premium $p(n_e, n_{1-e})$ in the healthy state, but drops to $c_{10} - M < c_{11}$ in the sick state. The non-employed receive transfer income $c_{00} = A$. If insured, they pay post-subsidy premium out of transfer income: $c_{01} = A - (1-\lambda_p)p$.

The planner chooses insurance coverage n_e, n_{1-e}, subsidy λ_p, cash transfer A, and employment $x(\mu)$ to maximize social welfare $\iint_{[0,1]^2} U(\nu,\mu)\,dF(\nu,\mu)$. The complete solution of the problem is complicated. However, it can be shown that full-subsidy universal coverage is at least a local maximum: at sufficiently high baseline coverage, further expanding coverage till everyone is covered is always welfare-enhancing.

Proposition 2. *With uniform pricing of insurance, cash transfer and premium subsidy financed by lump-sum tax, at any given employment allocation $x(\mu)$, full-subsidy universal coverage is a local maximum:*

1. *given full subsidy and full coverage of non-employed, the planner does not deviate from full coverage of workers:* $\frac{dW}{dn_e}|_{\lambda_p=1,n_e=1,n_{1-e}=1} > 0$
2. *given full coverage of workers, the planner does not deviate from full coverage of non-employed:* $\frac{dW}{dn_{1-e}}|_{n_e=1,n_{1-e}=1} > 0$
3. *given universal coverage of all, the planner does not deviate from full insurance subsidy to non-employed:* $\frac{dW}{d\lambda_p}|_{\lambda_p=1,n_e=1,n_{1-e}=1} = 0$

Proof. *Appendix*

To understand that marginal and social implication of adverse selection, consider enrolling one more worker of risk type n_e. Because the marginal enrollee has lower expected cost of medical care, premium is lower by $\frac{dp}{dn_e} = -M\frac{E[\mu|h(\nu,\mu)=1]+n_e-1}{Pr[h(\nu,\mu)=1]}\int_{x(n_e)}^{1} f(\nu,n_e)\,d\nu$. Social welfare responds in the following way:

$$\frac{dW}{dn_e} = \underbrace{\Delta u\,|_{(\nu\geq x(n_e),n_e)}}_{\text{marginal enrollee}} + \underbrace{u'|_{(\nu\geq x(n_e),\mu\leq n_e)} + u'|_{(\nu\geq x(n_e),\mu>n_e)} + u'|_{(\nu<x(n_e),\mu\leq n_e)}}_{\text{social externality}},$$

where $\Delta u\,|_{(\nu\geq x(n_e),n_e)} = \int_{x(n_e)}^{1} f(\nu,n_e)\,d\nu\cdot\Big(u(c_{11}) - n_e u(c_{10}) - (1-n_e)u(c_{10}-M)\Big)$ is the utility change on the margin. At very low coverage rate, marginal enrollees have higher WTP than the average cost of enrollees, and utility increases with insurance. As n_e approaches 1, marginal enrollees tend to have WTP lower than average cost. For the lowest valuation type $n_e = 1$, insurance strictly decreases utility.

The loss on the margin is countervailed by three social externality of adverse selection. Term $u'|_{(\nu \geq x(n_e), \mu \leq n_e)}$ gives the social benefit of lower premium rate to working enrollees: they save on both own premium payment and subsidy payment to non-employed enrollees. The transfer saving is greater with higher rate of subsidy λ_p and coverage among the non-employed n_{1-e}, whereas saving in own premium is greater with greater coverage n_e . At full subsidy and near-universal coverage, the social externality is large enough to dominate the utility loss on the margin, so that enrolling risk type $n_e = 1$ strictly increases social welfare: $\Delta u\,|_{(\nu \geq x(1),1)} + u'|_{(\nu \geq x(1),\mu \leq 1)} > 0$.

$u'|_{(\nu \geq x(n_e), \mu > n_e)}$ gives the social benefit to uninsured workers: their transfer payment is lower when premium is lower. Because this group is not protected from health expenditure shock, lower transfer payment increases consumption in both health states, and is more beneficial for riskier types where potential consumptin loss is larger. When evaluated at $n_e = 1$, the externality term vanishes.

$u'|_{(\nu < x(n_e), \mu \leq n_e)}$ gives the social benefit to non-employed enrollees: out-of-pocket premium payment is lower when premium is lower. The benefit vanishes at full subsidy. Put together, we have $\frac{dW}{dn_e}|_{\lambda_p = 1, n_e = 1, n_{1-e} = 1} = \Delta u\,|_{(\nu \geq x(1),1)} + u'|_{(\nu \geq x(1),\mu \leq 1)} > 0$.

At full coverage of workers, the welfare impact of enrolling one more non-employed person is

$$\frac{dW}{dn_{1-e}} = \Delta u\,|_{(\nu < x(\mu), n_{1-e})} + u'|_{(\nu < x(\mu), \mu \leq n_{1-e})} + \underbrace{e\,u'(c_{11})\,(AS - MC)}_{u'|_{(\nu \geq x(\mu),\mu)}},$$

where $\Delta u\,|_{(\nu < x(\mu), n_{1-e})} = \int_0^{x(n_{1-e})} f(\nu, n_{1-e})\,d\nu \Big(u(A - (1-\lambda_p)p) - n_{1-e} u(A) - (1 - n_{1-e}) u(A - M) \Big)$ is the utility change of the new enrollee. Although utility on the margin decreases for risk type $n_{1-e} = 1$ (except when $\lambda_p = 1$, where utility always increases on the margin), out-of-pocket premium cost is lower for all infra-marginal enrollees receiving subsidy. This social externality, given by $u'|_{(\nu < x(\mu), \mu \leq n_{1-e})}$, is larger when coverage of the non-employed is larger. At $n_{1-e} = 1$, regardless of subsidy rate, the social benefit is large enough to dominate utility loss on the margin, and the non-employed are strictly better off with full coverage: $\Delta u\,|_{(\nu < x(\mu), n_{1-e})} + u'|_{(\nu < x(\mu), \mu \leq n_{1-e})} > 0$.

The term $e\,u'(c_{11})\,(AS - MC)$ gives the social externality to workers, operating through two channels. First, enrolling one more non-employed person raises subsidy payment by $MC = \int_0^{x(n_{1-e})} f(\nu, n_{1-e})\,d\nu\,\frac{\lambda_p}{e}\,p$ per worker. Second, better risk pool reduces adverse selection, and for each worker, it amounts to a saving of $AS = [\frac{\lambda_p}{e}\int_0^{n_{1-e}}\int_0^{x(\mu)} dF(\nu,\mu)+1]\frac{dp}{dn_{1-e}}$ in subsidy and own premium payment.

Saving in subsidy is greater when coverage of the non-employed n_{1-e} is greater, and saving in own insurance premium is greater when coverage of workers n_e is greater. As n_{1-e} approaches 1, regardless of subsidy rate, it can be shown that benefit from adverse selection (AS) dominates the marginal cost of transfer (MC). It follows that when the working population is already fully insured, further covering the non-employed increases utility for both workers and non-workers. In other words, when social transfer corrects for the pricing inefficiency in a selected market, redistribution may in fact be Pareto: both net recipients *and* net payers of the transfer are better-off.

Full subsidy is justified, if at full coverage, a marginally lower subsidy rate generates more utility loss for non-employed enrollees than utility gain for working enrollees: $-u'(c_{11})+u'(c_{01})>0$. I show below that universal coverage implies the planner equalizes consumption across employment states with full subsidy, so the condition holds with equality locally. Since $c_{11}\geq c_{01}$ more generally, the planner keeps increasing subsidy till coverage is free for the non-employed.

Hence there is strong argument for universal coverage, when adverse selection is inherent in insurance pricing. The desirablity of full coverage in this case is due to the joint force of two social externality missing in private WTP. First and more familiar, the market implication on insurance pricing is not internalized in the WTP of low cost types: lower insurance pricing increases the utility of all infra-marginal enrollees, which tends to suggest an optimal coverage rate higher than the private market outcome. The market implication of insurance pricing alone, however, is not sufficient to justify universal coverage.

Government attempts at fixing adverse selection, such as insurance mandate and subsidy, introduce additional rationales that strength the case for universal coverage. The interventions allow the pricing efficiency on the insurance market to have braoder social implications. For instance, the resource needed to subsidize the insurance take-up of the uninsured is less costly, when the efficiency gain in insurance pricing in turn cuts back on the scale of the transfer. The two externality is intertwined in the constrained first-best: the pricing efficiency on the insurance market and the transfer efficiency involving the whole economy are both optimized with full-subsidy universal coverage.

2.3 Constrained first-best: the hybrid model

The setting considered previously is closest to the single payer system, where government covers all health care cost using public funds collected from workers. The similarity is by construction, given the focus on social planner's allocation. Actual health care systems are often hybrid, featuring both private insurers and publicly-subsidized programs. In the US, for example, employers sponsor health insurance to most of the non-elderly enrollees, and Medicaid and other programs cover low-income families and vulnerable popultions. In hybrid systems, it could happen that insurance pricing in the private (group) and public (non-group) market is different and reflects the underlying risk pools in each market. In the US context, the group rate for the employed is typically lower than the non-group rate for individuals not offered ESI. Conceptually, when insurance pricing is based on separate risk pools divided by employment, is it still desirable to mandate coverage for all? Does the separation in the risk pook in turn alter employment allocation?

Proposition 3 suggests the full coverage is still desirable. Intuitively, within market it still holds that the social benefit of lower premium to all infra-marginal enrollees dominates the utility loss to the last enrollee. Across markets, additional transfer to the last enrollee in the subsidized program is offset by lower total trasfer to existing subsidy recipients. It follows that at relatively high baseline coverage rate, it is welfare-enhancing to further increase coverage in both markets till all are covered.

Proposition 3. *When the insurance risk pool is separate by employment, and workers self-finance insurance and subsidize non-employed enrollees in public insurance programs, universal coverage is still a local optimum:*

1. *the planner does not deviate from covering all workers:* $\frac{dW}{dn_e}|_{n_e=1} > 0$
2. *the planner does not deviate from covering all non-workers:* $\frac{dW}{dn_{1-e}}|_{n_{1-e}=1} > 0$
3. *given universal coverage, the planner does not deviate from full insurance subsidy:* $\frac{dW}{d\lambda_p}|_{n_e=1, n_{1-e}=1, \lambda_p=1} = 0$

When coverage is universal, there is effectual risk pooling in the financing of health insurance. The planner's allocation of labor (e) and income transfer (A) in the hybrid model is the same as in the single payer model: $e^{SP2hybrid} = e^{SP2}$, $A^{SP2hybrid} = A^{SP2}$.

Proof. *Appendix*

Given coverage is universal, consumption equity motivates a full subsidy to the non-employed. It then follows that insurance and labor allocation in the single and hybrid system is in fact identical. Because full-subsidy universal coverage is desirable in both cases, workers end up bearing the expected medical cost of all agents in the economy. From a pure public financing perspective, the risk pooling across the whole population is implicit in the transfer structure. Importantly, separate insurance pricing by employment has *no* bearing on the optimal employment size in the economy, which is easily violated in the decentralized setting where individuals respond to state-specific prices and adjust employment to insurance cost. I study second-best departures from the constrained first-best in the empirical model below. However, the risk pooling result in the constrained first-best is not always guaranteed. For instance, administrative and marketing activities can raise additional social cost of private insurance. More fundamentally, unlike private insurance that protects agents from risks stipulated at the time of insurance purchase, only the government can provide redistribution in response to ex-ante correlated shocks *before* insurance is formally acquired.

2.4 Second-best

2.4.1 The empirical approach

Second-best outcomes in the hybrid model deviate meaningfully from the constrained first-best. On the extensive margin of coverage, since the value of insurance to other enrollees and members of society do not enter private WTP, coverage is not universal even with substantial transfer. On the margin of insurance choice, although differential pricing across employment has no bearing on optimal employment in the constrained first-best, it has significant behavioral implication for insurance selection and associated labor adjustments in the decentralized setting. These behavioral responses to private valuation and costs suggest optimal insurance transfer in the second-best likely differs from the full-subsidy contract in the constrained first-best.

In what follows I develop an *empirical* model of second-best insurance and employment choice in a hybrid insurance model: individuals take prices and government transfers as given, and take actions to maximize own utility. In particular, I only require individuals to optimize on the employment decision, but leave the specifics of the insurance decision unspecified. In other words, the characterization of the insurance domain is completely empirical: insurance take-up, choice, and pricing are constrained only by their empirical moments and change rates to policy parameters as observed in the data, rather than following specific decision rules derived from optimized programs.

An all-fitting decision rule of insurance choice is difficult to formulate. In fact, behavioral biases and cognitive limitations question the standard assumption of optimality in observed choices. For example, WTP may be biased downward due to misperception of risk by over-confident agents (Barseghyan et al. 2013). Cognitive ability matters for choosing the optimal level of insurance (Fang, Keane, and Silverman, 2008), whereas bounded rationality prevents a complete understanding of the insurance contract (Handel and Kolstad, 2015), leading to sub-optimal plan choice (Abaluck and Gruber, 2011; Bhargava, Loewenstein, and Sydnor, 2017), inertia (Handel, 2013), and low insurance take-up (Spinnewijn, 2017).

Empirical estimates of outcomes and how they vary with policy parameters, on the other hand, are data-driven entities invariant to the choice of decision rules. They contain key information in the data that forms the common ground for alternative models of the decision process. In this context, because the individual insurance choice is not necessarily optimized and enters empirically, the aggregate implication on insurance pricing and public-private insurance selection cannot be determined from primitives alone, and must also enter empirically[6]. These outcome-based "sufficient statistics" (Chetty, 2009) are attractive for welfare analysis where choices are not necessarily optimal, or where full specification of decision rules requires strong modeling assumptions.

I then present the details of the second-best economy.

2.4.2 Insurance allocation and pricing

Again consider a unit mass of individuals differentiated by risk type μ and labor productivity ν. Insurance allocation in the economy is summarized in $\lambda_{i,j}$, $i = 0, 1$, $j = 0, 1, 2$, where $i = e$ indicates workers and $i = 1 - e$ indicates non-workers. Within employment cell, $j = 1$ indicates ESI, $j = 2$ indicates non-ESI sources of coverage subsidized based on income, and $j = 0$ the uninsured. Hence $\lambda_{e,1}$ gives the fraction of workers enrolled in ESI plans,

[6]The exact mapping from risk-productivity primatives to insurance selection, over and above the extensive margin choice on employment and coverage, is complicated and difficult to specify. This is another reason why an outcome-based approach relying on aggregate substitution patterns, captured in "crowd-out" estimates, to inform welfare may be advantageous here.

and $\lambda_{1-e,0}$ the fraction of the non-employed who are also uninsured. By construction, $\lambda_{i,2} = 1 - \lambda_{i,0} - \lambda_{i,1}$, $i = e, 1-e$. For reasons noted above, these insurance measures do not come from one specified behavioral model, are not constrained by any particular optimality condition, and are empirical moments perceived by all agents.

Also taken as given are the prices and transfers in the economy. Although all risk types determine the market premium rate, individual risk types do not internalize the pricing externality to others. Assuming the private and public insurances are priced based on a common risk pool[7], premium r is a function of the population uninsurance rate λ_0: $r(\lambda_0) = \frac{E[1-\mu|h(\mu,\nu)=1]}{1-\lambda_0}(M-C)$, and the slope of the relationship reflects the nature of selection in insurance take-up. Adverse selection, for example, implies $r'(\lambda_0) > 0$. C is the average copay required of insured patients.

2.4.3 Subsidy and public transfer

Individuals not enrolled in ESI can purchase public insurance at reduced cost. Let λ_p denote the average subsidy rate, or percent subsidy, on the non-ESI market. Ex-subsidy premium cost, $(1-\lambda_p)r(\lambda_0)$, coincides with "affordability", or the maximum spending on premium before insurance is deemed unaffordable. If individuals forgo coverage, they are subject to a fine at 50% of affordability, or $k(1-\lambda_p)r(\lambda_0)$ with $k = \frac{1}{2}$ in Massachusetts.

The insurance subsidy is financed by a lump-sum tax τ_{pb} on worker's marginal product of labor w, and penalty collected from the uninsured. Let the aggregate employment size be e. The government's budget implies

$$\tau_{pb} = \lambda_p r(\lambda_0)[\lambda_{e,2} + \frac{1-e}{e}\lambda_{1-e,2}] - k(1-\lambda_p)r(\lambda_0)[\lambda_{e,0} + \frac{1-e}{e}\lambda_{1-e,0}] + \frac{1-e}{e}A,$$

where A is unemployment insurance payout to the non-employed, also financed by taxation on workers.

2.4.4 Uncompensated care and private transfer

Private transfers occur because, 1) not all enrollees in ESI contribute to the financing of the coverage, and 2), informal or implicit coverage may nonetheless be provided to the uninsured under some circumstances. In

[7]The uniform pricing assumption is plausible in Massachusetts because of the 2005 merger of the small group and non-group risk pools. It simplifies analysis below because selection by insurance types has only fiscal externality but no pricing externality.

the first case, when ESI are self-financed by workers[8], the expected medical cost of non-working enrollees are paid by working enrollees through private transfer.

Furthermore, the possibility of implicit or informal insurance imposes additional tax on working enrollees. Because uninsured patients can still receive emergency care, charity care providers such as hospitals bear the cost of treating the uninsured (Garthwaite, Gross, and Notowidigdo, 2017). The existence of informal insurance is consistent with the low willingness to pay for formal insurance among the low-income (Finkelstein, Hendren, and Luttmer, 2015; Finkelstein, Hendren, and Shepard, 2017).

It is less clear if charity care providers then distribute the cost to other members of society. For example, hospitals may increase the billing and pass the cost to formal insurance enrollees, or receive special funds from private donation or the state that compensates the care to the uninsured[9]. Alternatively, hospitals may absorb the cost and endure lower profit. In the main model, the privately insured bear the medical cost of the uninsured. I essentially assume uncompensated care compresses wages of ESI enrollees either through premium increase or profit loss of businesses. In the robustness check, I examine the case where the cost of charity care affects hopsital profit that enters social welfare function separately, without being passed on to other members of society.

Specifically, I assume with probability $j \cdot (1-\mu)$, the uninsured encounters a health condition that qualifies them for use of uncompensated care. Probability j may depend on the urgency and acuteness of the illness, and on hospital finance. In this case, the uninsured pay the same copay as the insured, and the remaining cost $M - C$ is paid by the privately insured. That is, fraction j of the uninsured receive informal insurance at premium rate $ri(\lambda_0)$, where $ri(\lambda_0) = \frac{E[1-\mu|h(\mu,\nu)=0]}{\lambda_0}(M-C)$ is the expected medical cost per uninsured, and again depends on the nature of risk selection in $h(\mu,\nu)$.

Put together, workers enrolled in ESI pay the premium for all enrollees, and with probability j, the expected medical cost of the uninsured. Per person private transfer τ_{pv} follows from the budget constraint

[8] or when wage rate received by workers is lower than marginal product of labor by almost the full cost of ESI (Kolstad and Kowalski, 2016)

[9] Massachusetts, for example, has a special fund that covers the state's uninsured high risk pool. Since the pool is funded by all paying customers of hospital, including a tax on insurance premium, it is plausible the privately insured contribute significantly to the cost of the uninsured.

$$\lambda_{e,1}\tau_{pv} = [\lambda_{e,1} + \frac{1-e}{e}\lambda_{1-e,1}]r(\lambda_0) + j[\lambda_{e,0} + \frac{1-e}{e}\lambda_{1-e,0}]ri(\lambda_0).$$

2.4.5 Consumption, utility, and employment

Taking prices, transfers and insurance allocation as given, average consumption of a worker in good health is

$$c_{10} = w - \tau_{pb} - \lambda_{e,1}\tau_{pv} - (1-\lambda_p)\overline{\lambda_{e,2}}r(\lambda_0),$$

and in bad health is

$$c_{11} = c_{10} - (1-\lambda_{e,0})C - \lambda_{e,0}jC - \lambda_{e,0}(1-j)M = c_{10} - C - \lambda_{e,0}(1-j)(M-C),$$

where $\overline{\lambda_{e,2}} = 1-\lambda_{e,1}-(1-k)\lambda_{e,0}$ is the mandate-adjusted enrollment in public insurance. It reflects the fact that a higher penalty allows subsidy to be financed more by the uninsured *within* the employment cell[10], lowering the moral hazard cost that occurs *across* cells. $\overline{\lambda_{e,2}}$ is therefore the relevant measure for evaluating transfer efficiency. Uninsured patients on average pay $(1-j)(M-C)$ more than insured patients. As j approaches one, difference between informal and formal insurance contracts vanishes, and at $j = 1$, coverage is universal implicitly.

Consumption of a non-employed person in good health is

$$c_{00} = A - (1-\lambda_p)\overline{\lambda_{1-e,2}}r(\lambda_0),$$

and in bad health is

$$c_{01} = c_{00} - C - \lambda_{1-e,0}(1-j)(M-C),$$

where $\overline{\lambda_{1-e,2}} = 1-\lambda_{1-e,1}-(1-k)\lambda_{1-e,0}$.

Individuals maximize expected utility choosing employment, taking all prices and transfers as given. Let $e(\nu,\mu) \in \{0,1\}$ denote optimized employment decisions. Expected utility

$$U(\nu,\mu) = e(\nu,\mu)E_\mu u(c_{1\cdot}) + [1-e(\nu,\mu)]E_\mu u(c_{0\cdot}) - e(\nu,\mu)g(\frac{1}{\nu}),$$

[10]Recall that k is the ratio of mandate penalty over affordability. Of the $\lambda_{e,0}$ who are uninsured, $k\lambda_{e,0}$ are paying affordability. Although $k\lambda_{e,0}$ do not actually possess insurance, from a pure transfer perspective, they are self-subsidized public insurance enrollees. The mandate-adjusted measure reflects the self-subsidy by incorporating $k\lambda_{e,0}$ at no additional transfer cost.

where $E_\mu u(c_{i\cdot}) = \mu u(c_{i0}) + (1-\mu)u(c_{i1})$, $i = e, 1-e$. It is easy to see that each risk type μ follows a cut-off decision rule in employment: more productive types with $\nu \geq x(\mu) = \frac{1}{g^{-1}(E_\mu u(c_{1\cdot}) - E_\mu u(c_{0\cdot}))}$ work, and less productive types do not. Employment size in the decentralized setting $e = \int_0^1 \int_{x(\mu)}^1 dF(\nu, \mu)$.

Summing over individuals, social welfare can be expressed as

$$\begin{aligned} W &= \iint_{[0,1]^2} U(\nu, \mu) dF(\nu, \mu) \\ &= e E_{\bar{\mu}_e} u(c_{1\cdot}) + (1-e) E_{\bar{\mu}_{1-e}} u(c_{0\cdot}) - G(e), \end{aligned} \tag{1}$$

where $E_{\bar{\mu}_e} u(c_{1\cdot}) = \bar{\mu}_e u(c_{10}) + (1-\bar{\mu}_e)u(c_{11})$, and $\bar{\mu}_e = E[\mu | \nu \geq x(\mu)]$ is the mean risk type among workers. Similarly, $\bar{\mu}_{1-e} = E[\mu | \nu < x(\mu)]$ is the mean risk type among the non-employed. $G(e) = \int_0^1 \int_{x(\mu)}^1 g(\frac{1}{\nu}) dF(\nu, \mu)$ is the total disutility from employment in the economy.

When the government changes transfer payments by increasing λ_p or A, individual employment decision responds optimally. However, as decentralized employment is only optimal up to a given set of prices and transfers, effects of employment adjustment on equilibrium prices and transfers are not internalized in individual decision. In other words, as policy changes, envelopment theorem applies to quantities outside the utility function, but prices, and hence consumption, vary endogenously to policy inside utility.

2.5 Government choice of λ_p and the welfare formula

Compared to the constrained first-best, individuals in second-best fail to internalize two social externality. First, efficiency gain in insurance pricing benefits all enrollees. Second, between net payers and recipients of subsidy, insurance transfer is more efficient when insurance pricing is more efficient. Both externality is necessary to motivate the full-subsidy full coverage in the constrained first-best.

The government approach the constrained first-best to some extent by judicially choosing subsidy rate λ_p. The full-subsidy soluation, however, is not directly applicable to the second-best. Since employment decision is only optimal up to a set of prices and transfers, a very high subsidy discourages employment either through selection into subsidized insurance or the moral hazard effect of taxation. The fiscal externality of funding a growing subsidy program from a diminishing tax base is a realistic second-best constraint, which might suggest optimal subsidy is less than full. On the whole, whether the social benefits of coverage expansion outweigh the social cost of transfer financing is ultimately an empirical question, and particularly so in this case given the empirical nature of the second-best.

Formally, the welfare impact of a marginal increase in subsidy rate is given by $\frac{dW}{d\lambda_p} = eE_{\bar{\mu}_e}[u'(c_{1\cdot})\frac{dc_{1\cdot}}{d\lambda_p}] + (1-e)E_{\bar{\mu}_{1-e}}[u'(c_{0\cdot})\frac{dc_{0\cdot}}{d\lambda_p}]$. Envelope theorem applies to $eE_{\bar{\mu}_e}$, $(1-e)E_{\bar{\mu}_{1-e}}$, and $G(e)$: employment responds optimally to new prices, and conveys no direct impact on welfare. Within utility, however, insurance allocation, pricing and transfers all respond endogenously to policy parameters. In particular, transfer responses $\frac{d\tau_{pb}}{d\lambda_p}$ and $\frac{d\lambda_{e,1}\tau_{pv}}{d\lambda_p}$ reflect the equilibrium response of insurance $\lambda_{i,j}$, pricing r and ri, and employment e to subsidy generosity λ_p, and are total derivatives of λ_p.

For a monetized metric of welfare, define dollar subsidy $\widetilde{\lambda_p} = \lambda_p r(\lambda_0)$. A dollar increase in subsidy raises utility per subsidy eligible by $\frac{dW}{d\tilde{\lambda}_p}/\lambda_2$, where $\lambda_2 = e\lambda_{e,2} + (1-e)\lambda_{1-e,2}$. Normalized by worker's utility gain of a dollar increase in wage, the dollar change in welfare following a dollar increase in subsidy is given by $WM = \frac{\frac{dW}{d\tilde{\lambda}_p}/\lambda_2}{\frac{dW}{dw}/e}$. Proposition 4 characterizes the metric analytically.

Proposition 4. *When insurance choices are potentially sub-optimal, and employment decisions are optimal taking prices and transfers as given, insurance subsidy affects social welfare in the following ways:*

1. *to all enrollees, there is efficiency gain in insurance pricing $P(\lambda_p; j, k)$: greater coverage lowers insurance premium and transfer:*

$$P(\lambda_p; j,k) = -\frac{1-\lambda_0}{\lambda_2}\frac{\epsilon_{r,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p} - j\frac{\lambda_0}{\lambda_2}\frac{ri(\lambda_0)}{r(\lambda_0)}\frac{\epsilon_{ri,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p} \\ - \frac{1-e}{\lambda_2}\overline{\lambda_{1-e,2}}\left[\frac{E_{\bar{\mu}_{1-e}}u'(c_{0\cdot})}{E_{\bar{\mu}_e}u'(c_{1\cdot})} - 1\right](1-\lambda_p)\frac{\epsilon_{r,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p}$$

2. *to existing enrollees, there is premium assistance value $\Delta W(\lambda_p; j,k)$ that allows enrollees to stay covered despite adverse income shock:*

$$\Delta W(\lambda_p; j,k) = \frac{1-e}{\lambda_2}\overline{\lambda_{1-e,2}}\left[\frac{E_{\bar{\mu}_{1-e}}u'(c_{0\cdot})}{E_{\bar{\mu}_e}u'(c_{1\cdot})} - 1\right]$$

3. *to the newly insured,*

(a) *there is risk protection value of insurance $I'(\lambda_p;j)$ that reduces consumption loss in the bad health state:*

$$I'(\lambda_p;j) = -\frac{1-j}{\lambda_2}\frac{M-C}{r(\lambda_0)}\cdot \left[e(1-\bar{\mu}_e)\frac{u'(c_{11})}{E_{\bar{\mu}_e}u'(c_{1\cdot})}\frac{d\lambda_{e,0}}{d\lambda_p} + (1-e)(1-\bar{\mu}_{1-e})\frac{u'(c_{01})}{E_{\bar{\mu}_e}u'(c_{1\cdot})}\frac{d\lambda_{1-e,0}}{d\lambda_p}\right]$$

(b) *new insurance on the margin raises additional premium cost $P'(\lambda_p;j,k)$, paid by both workers and subsidy recipients:*

$$P'(\lambda_p;j,k) = \frac{e}{\lambda_2}\left[1-j\frac{ri(\lambda_0)}{r(\lambda_0)}\right]\frac{d\lambda_{e,0}}{d\lambda_p} + \frac{1-e}{\lambda_2}\left[1-j\frac{ri(\lambda_0)}{r(\lambda_0)}\right]\frac{d\lambda_{1-e,0}}{d\lambda_p} - \frac{1-e}{\lambda_2}\left[\frac{E_{\bar{\mu}_{1-e}}u'(c_{0\cdot})}{E_{\bar{\mu}_e}u'(c_{1\cdot})}-1\right](1-\lambda_p)\frac{d\overline{\lambda_{1-e,2}}}{d\lambda_p}$$

(c) *Welfare change on the margin equals*

$$W'(\lambda_p;j,k) = I'(\lambda_p;j) + P'(\lambda_p;j,k)$$

4. *to workers, subsidy incurs additional moral hazard cost on labor, and the cost is greater with greater pre-existing insurance transfers:*

$$MH(\lambda_p;j,k) = -\frac{1-e}{\lambda_2}\left[\overline{\lambda_{1-e,2}} + \frac{\frac{A}{r(\lambda_0)} + \lambda_{1-e,1} + \left[j\frac{ri(\lambda_0)}{r(\lambda_0)}-k\right]\lambda_{1-e,0}}{\lambda_p}\right]\frac{\epsilon_{1-e,\lambda_p}}{e}$$

5. *Total welfare change from a dollar increase in premium subsidy equals*

$$WM(\widetilde{\lambda_p};j,k) = \frac{\frac{dW}{d\widetilde{\lambda_p}}/\lambda_2}{\frac{dW}{dw}/e} = P(\lambda_p;j,k)+\Delta W(\lambda_p;j,k)+W'(\lambda_p;j,k)+MH(\lambda_p;j,k).$$

Insurance subsidy affects both marginal and infra-marginal enrollees. Economy wide, when insurance take-up is adversely selected, new enrollees lower the premium rate faced by all enrollees, and in particular benefits net payers of insurance subsidy who save on both own premium and transfer payments. In the case where workers are the single payer of formal and informal insurance in the economy, total saving from premium pricing in workers' utiles equals $-\frac{1-\lambda_0}{\lambda_2}\frac{\epsilon_{r,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p} - j\frac{\lambda_0}{\lambda_2}\frac{ri(\lambda_0)}{r(\lambda_0)}\frac{\epsilon_{ri,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p}$. Since subsidy recipients $(1-e)\overline{\lambda_{1-e,2}}$ still pay $1-\lambda_p$ of the premium out of pocket, the utility gain from pricing efficiency is greater for the non-employed by the factor $\frac{E_{\bar{\mu}_{1-e}} u'(c_{0\cdot})}{E_{\bar{\mu}_e} u'(c_{1\cdot})} > 1$. Total efficiency gain in insurance pricing $P(\lambda_p; j, k)$ accounts for the fact that value of premium saving is greater for the non-employed.

To existing enrollees, subsidy provides a form of insurance for insurance affordability against adverse income shocks. A dollar increase in subsidy transfers resources from the employed to the non-employed state, so that enrollees have means of continuing coverage when income is low. The rick protection against insurance unaffordability is valued at $\frac{1-e}{\lambda_2} \overline{\lambda_{1-e,2}} \frac{E_{\bar{\mu}_{1-e}} u'(c_{0\cdot})}{E_{\bar{\mu}_e} u'(c_{1\cdot})}$ by recipients. The net benefit of premium assistance, after accounting for increased payment in the employed state, is given in $\Delta W(\lambda_p; j, k)$.

To new enrollees who are previously uninsured, there is additional risk protection value $I'(\lambda_p; j)$ from formal insurance, and this value is decreasing in j: when $j = 1$, informal insurance perfectly substitutes formal insurance, and universal coverage is implicit without recourse to subsidy programs.

The risk protection gain $I'(\lambda_p; j)$ is measured against new premium payments on the margin, summarized in $P'(\lambda_p; j, k)$. Net of the informal coverage already provided, new enrollees incur additional premium cost $\frac{e}{\lambda_2}\left[1 - j\frac{ri(\lambda_0)}{r(\lambda_0)}\right]\frac{d\lambda_{e,0}}{d\lambda_p} + \frac{1-e}{\lambda_2}\left[1 - j\frac{ri(\lambda_0)}{r(\lambda_0)}\right]\frac{d\lambda_{1-e,0}}{d\lambda_p}$, where $1 - j\frac{ri(\lambda_0)}{r(\lambda_0)}$ is the actuarial "top-off" of formal insurance. This cost is not completely borne by workers, as subsidy recipients contribute towards own premium. The cost sharing reduces utility in the non-employed state by $(1-e)\frac{d\overline{\lambda_{1-e,2}}}{d\lambda_p}(1-\lambda_p) E_{\bar{\mu}_{1-e}} u'(c_{0\cdot})$. Term $-\frac{1-e}{\lambda_2}\left[\frac{E_{\bar{\mu}_{1-e}} u'(c_{0\cdot})}{E_{\bar{\mu}_e} u'(c_{1\cdot})} - 1\right](1-\lambda_p)\frac{d\overline{\lambda_{1-e,2}}}{d\lambda_p}$ accounts for the cost sharing and the fact that consumption value is greater in the non-employed state. Subsidizing new enrollees is welfare-enhancing on the margin if $W'(\lambda_p; j, k) = I'(\lambda_p; k) + P'(\lambda_p; j, k) > 0$, which is more likely to hold when j is smaller.

Lastly, as workers do not internalize the social implication of transfers, the incidence of private and public transfer incurs moral hazard cost $MH(\lambda_p; j, k)$ on labor supply. This cost is increasing in the size of subsidy recipients $\overline{\lambda_{1-e,2}}$, and is further compounded by pre-existing insurance transfers in the economy. To the extent that private insurance provided to $\lambda_{1-e,1}$ is of value $\frac{1}{\lambda_p}$ per subsidy, and informal insurance to $\lambda_{1-e,0}$ is of actuarial value $j\frac{ri(\lambda_0)}{r(\lambda_0)}/\lambda_p$, moral hazard cost of public expansion is greater by the order of $-\frac{1-e}{\lambda_2}\frac{\frac{A}{r(\lambda_0)} + \lambda_{1-e,1} + \left[j\frac{ri(\lambda_0)}{r(\lambda_0)} - k\right]\lambda_{1-e,0}}{\lambda_p}$. Other welfare programs that buffer income shocks ($\frac{A}{r(\lambda_0)}$) similarly weaken the argument for subsidy expansion, whereas greater self-financing through mandate penalty k lowers welfare cost. The exact magnitude of moral hazard is pinned down by the empirical elasticity of labor reduction following a unit increase in subsidy, given in ϵ_{1-e,λ_p}.

Therefore selection from pre-existing insurance arrangements worsens the moral hazard cost of labor. When ESI enrollees switch into subsidized programs, there is no additional value of risk protection, but an increase in premium cost $-\frac{1-e}{\lambda_2}\left[\frac{E_{\tilde{\mu}_{1-e}} u'(c_{0\cdot})}{E_{\tilde{\mu}_e} u'(c_{1\cdot})} - 1\right](1-\lambda_p)\frac{d\overline{\lambda_{1-e,2}}}{d\lambda_p}$ from $P'(\lambda_p; j,k)$. If the substitution is accompanied by labor reduction, or $\epsilon_{1-e,\lambda_p} > 0$, then the moral hazard cost $MH(\lambda_p; j,k)$ also applies. In this case, subsidy payment to a growing body of recipients is financed from a smaller tax base. The fiscal externality then magnifies the total welfare cost of insurance selection.

2.6 Sufficient statistics

Subsidy rate

The welfare formula $WM(\widetilde{\lambda_p}; j,k) = P(\lambda_p; j,k) + \Delta W(\lambda_p; j,k) + W'(\lambda_p; j,k) + MH(\lambda_p; j,k)$ expresses the *dollar* effect of subsidy in terms of marginal changes in *percent* subsidy, or the subsidy rate. Why is the relative price measure useful for characterizing behavioral responses to subsidy?

Firstly, it provides a standardizing measure of program *generosity*. If enrollees value a dollar subsidy more (less) when facing lower (higher) market premium rate, then a dollar subsidy is not directly comparable across markets. Subsidy rate, on the other hand, reflects the relative generosity across markets. Furthermore, the same generosity also captures the opportunity cost of paying full, private insurance premium as opposed to subsidized, public insurance premium, and is therefore closely related to selection incentives on the insurance and labor market. It is then reasonable to expect that these behavioral responses are better characterized as a result of relative price changes, rather than dollar price changes alone.

Insurance-labor selection

Behavioral responses on the insurance and labor market are summarized in the level and marginal changes in conditional insurance allocation $\lambda_{i,j}$, $i = e, 1-e$, $j = 0,1,2$. Selection from ESI to subsidized insurance with employment reduction, for example, increases (reduces) the denominator of $\lambda_{1-e,2}$ ($\lambda_{e,1}$), and reduces (increases) the numerator of $\lambda_{e,1}$ ($\lambda_{1-e,2}$). Analytically, $\frac{\lambda_{i,j}}{d\lambda_p} = \left[\frac{d(i\lambda_{i,j})}{d\lambda_p} - \lambda_{i,j}\frac{di}{d\lambda_p}\right]/i$, so that both insurance and labor selections as a result of subsidy are captured in these statistics.

These statistics are completely empirical quantities unconstrained by optimality conditions. This is desirable because employment responses affect welfare only through transfer externality not internalized in individual's optimal employment choice, and transfer externality further depends on pricing externality on the insurance market where choices can be difficult to rationalize with optimization models. The empirical statistics are "sufficient" in the sense that welfare evaluation is still feasible without strong modeling assumption of the underlying decision processes.

Moral hazard
When insurance subsidy is financed by taxes, absent insurance selection responses, higher subsidy raises the tax burden on workers and lowers employment size. The pure tax effect is compounded by insurance selection, if some ESI enrollees reduce employment and switch to subsidized coverage. The selection results in even larger tax burden on workers. The joint effect of tax and selection responses is summarized in ϵ_{1-e,λ_p}, the elasticity of employment reduction following a unit increase in subsidy.

Insurance pricing I quantify the effect of adverse selection on insurance pricing using estimates from Finkelstein, Shepard, and Hendren (2017). At income thresholds where subsidy decreases and price paid by consumers increases, Finkelstein, Shepard, and Hendren (2017) finds that lower cost enrollees in the Massachusetts Exchange drop out, and average cost to insurers is higher at higher income bands. The regression discontinuity estimates are used to infer the extend of pricing efficiency gain when the remaining uninsured take up formal insurance.

Risk protection
The risk protection value of insurance is greater, if the consumption loss without insurance is greater in the bad health state, or when $u'(c_{11})$ and $u'(c_{01})$ are greater. The premium assistance value of subsidy is greater, if income loss in the non-employment state is greater, or when $\frac{E_{\bar{\mu}_{1-e}} u'(c_{0\cdot})}{E_{\bar{\mu}_e} u'(c_{1\cdot})}$ is greater. Both values increase with the curvature in the utility function, or when agents are more risk averse.

I estimate statistics related to insurance-labor selection and moral hazard on labor, but calibrate risk preference and values of risk protection and insurance pricing in Section X. I assess the sensitivity of welfare evaluation to different calibrated values.

3 Data

3.1 Sample summary

I estimate insurance-labor statistics using a sample of 27-64 year olds living in Massachusetts in 2008-2011 waves of the American Community Survey (ACS). ACS samples over 3.5 million addresses per year, and is representative of sub-state geographical areas with a minimum population of 65,000[11]. I end the sample period in year 2011 because the same sub-state areas cannot be followed due to a change in census map in 2012.

Health insurance variables first appeared in the survey in 2008, and have since become the primary data for insurance coverage estimates published by the Census Bureau[12]. Questions ask about coverage from 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.

Labor questions distinguish between active employment and participation, which also includes the phase of unemployed job search. Specifically, respondents are asked if they were working on or looking for any job in the past week. Those who reply yes to either are in the labor force; those reporting having a job are coded employed. Respondents are also asked if they worked at all over the past 12 months. I compare results using different labor measures in the analysis below.

I exclude the elderly who already have subsidized insurance from Medicare, and focus on non-elderly adults above the age 27. Younger adults may be eligible for dependent coverage from parental ESI according to state and federal laws[13]. Of the 135,223 individuals in the 27-64 age group, 2,863 are inmates in "group quarter" households. Because information is missing on their family members, their subsidy eligibility is hard to determine. I hence drop these observations from my analysis.

The final sample has 132,360 individuals, of which 30,389 do not have

[11] ACS also has a fairly low non-response rate under 5% during the study period. By contrast, CPS has a non-response rate around 12% in the basic monthly survey, and 15% in the Annual Social and Economic Supplement (ASEC).

[12] Fortuitously, 2008 is also the first year the individual mandate, the employer mandate, and premium subsidy all took effect in Massachusetts after an initiation period in 2006-2007.

[13] For instance, Massachusetts enacted special provisions in 2007 that reduced uninsurance among young adults (Long, Yemane, and Stockley, 2010). The 2010 ACA dependent mandate reduced uninsurance in the 19-26 age group nation-wide (Akosa Antwi, Moriya, and Simon, 2013).

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: Full sample includes non-institutionalized Massachusetts residents aged 27-64 in ACS between year 2008 and 2011. ACS sampling weights are applied. 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.

ESI coverage; the latter group is entitled to subsidy based on family income[14]. I detail the the determination of subsidy exposure in the following section. Table 1 summarizes demographics and insurance-labor outcomes. In general, those without ESI are younger, less likely to be married, and from a lower socio-economic standing: they are less educated, more likely to be ethnic minorities, and have lower income. Average subsidy rate (in percent of relevant market premium rate) is 29%, but is much higher at 68% among potential recipients not already enrolled in ESI.

3.2 Subsidy rate

I generate measures of subsidy generosity using information in Schedule HC Worksheets and Tables. The document is the official guideline for determining mandate penalty, affordability of ESI and subsidized coverage from the Commonwealth Care. Figure 2 shows a screen shot of relevant tables in the 2010 Worksheets and Tables. Tables on the left determine affordability, the maximum amount a person needs to pay towards own insurance premium. It rises with the person's family income, measured in percent federal poverty line (FPL).

The affordability level limits the price faced by potential enrollees, and the difference between consumer price and pre-subsidy market premium rate is the subsidy implicit in the regulation. In practice, for individuals not enrolled in ESI, there exists plans on the Commonwealth Care that require enrollees to pay exactly the affordability amount. Hence affordability is in fact the (lowest) price faced by consumers. I therefore construct a measure of subsidy *generosity*, which gives the potential cost saved by the subsidy as opposed to paying the full price, in $subs = 1 - \frac{affordability}{market\_rate}$, and call this measure subsidy *rate* throughout the text. I discuss the measurement of the numerator and the denominator below.

3.2.1 Numerator: affordability

Affordability is a mapping from percentage federal poverty line, which in turn depends on family income and family size. For example, single tax filers with income less than 150% FPL ($16,248 in 2010) are fully subsidized, and so are couples who file jointly with income less than $21,864, or 150% FPL for a family of two. Whether married couples file jointly or separately has no

[14]More precisely, they are not required to pay more than an "affordability" amount for insurance premium. The difference from the full market premium rate is the subsidy.

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	$16,248	$ 0
$16,249	$21,660	$ 39
$21,661	$27,084	$ 77
$27,085	$32,496	$116
$32,497	$39,000	$175
$39,001	$44,200	$235
$44,201	$54,600	$354
$54,601	Any individual with an annual income over $54,600 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	$21,864	$ 0
$21,865	$29,148	$ 78
$29,149	$36,432	$154
$36,433	$43,716	$232
$43,717	$54,600	$315
$54,601	$65,000	$422
$65,001	$85,800	$589
$85,801	Any couple with an annual income over $85,800 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	$ 27,468	$ 0
$27,469	$ 36,624	$ 78
$36,625	$ 45,780	$154
$45,781	$ 54,936	$232
$54,937	$ 72,800	$373
$72,801	$ 93,600	$586
$93,601	$114,400	$849
$114,401	Any family with an annual income over $114,400 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	$124	$248	$ 732
27–29	$206	$412	$ 732
30–34	$206	$412	$ 760
35–39	$218	$436	$ 774
40–44	$250	$500	$ 774
45–49	$280	$560	$ 834
50–54	$372	$744	$ 910
55+	$412	$824	$1,066

Region 2. Bristol, Essex, Hampden, Middlesex, Norfolk, Suffolk and Worcester Counties			
Age	Individual[1]	Married couple[2] (no dependents)	Family[3]
0–26	$156	$312	$ 672
27–29	$223	$446	$ 672
30–34	$224	$448	$ 774
35–39	$227	$454	$ 788
40–44	$259	$518	$ 788
45–49	$285	$570	$ 850
50–54	$338	$676	$ 927
55+	$445	$890	$1,085

Region 3. Barnstable, Dukes, Nantucket and Plymouth Counties			
Age	Individual[1]	Married couple[2] (no dependents)	Family[3]
0–26	$153	$306	$ 662
27–29	$214	$428	$ 662
30–34	$216	$432	$ 835
35–39	$216	$432	$ 863
40–44	$271	$542	$ 874
45–49	$271	$542	$ 906
50–54	$321	$642	$1,030
55+	$427	$854	$1,280

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).

Note: screen print of page 3 in 2010 Schedule HC Worksheets and Tables, available at http://www.mass.gov/dor/docs/dor/forms/inctax10/f1-nrpypdfs/form-1-nrpy-worksheets.pdf bearing on affordability, which is always determined using combined income and has the same value for both spouses[15]. At higher income, affordability changes discretely at 200%, 250% and 300% FPL, beyond which there is no further subsidy.

I transform ACS income variables into percentage FPL in the following way. First, I construct nuclear families for tax purposes using intra-household relationship pointers developed by Ruggles et al. (2015). I split multi-generational households into nuclear families where only children below age 18 can be claimed as tax dependents, not adult children living with parents. The procedure also excludes grand-parents from the tax-filing unit of adult children. Next, to measure federal adjusted gross income, which is the base income for determining affordability, I use total personal income in ACS. The variable gives respondent's total pre-tax income and losses from all sources in the preceding 12 months.

I sum over personal income within nuclear families to derive family income for tax purposes. This family income is smaller than the raw family income variable in ACS, which also includes adult children and grandparents. I then apply yearly poverty guidelines published by Department of Health and Human Services to generate poverty level that differs by family size. Measured against this poverty level, I transform family income into percent FPL. Lastly, I assign the corresponding affordability level to all members in the family.

3.2.2 Denominator: market premium rate

Market premium rate differs by year, location and age band. In year 2010, for example, premium rate is lowest in the region of Berkshire, Franklin and Hampshire counties, and within region, premium is twice as large for the near-elderly (55+) as it is for young adults (27-29). Family plan premium for married couple is a simple sum of individual premium rates. Therefore I assign premium rate at the individual level based on year, location and age band.

I match regions, which are collections of counties, to public use micro-data area (PUMA) in ACS. PUMAs are geographically contiguous units build on census tracts and counties and do not cross state borders. There are 52 PUMAs in Massachusetts superimposed on 14 counties and 3 rating regions,

[15] For example, in the 2010 Worksheet, the section determining eligibility for government-subsidized health insurance has the following instruction: "If married filing separately and living in the same household, each spouse must combine their income figures from their separate U.S. returns when completing this worksheet." (p. WS-2).

with population within PUMA ranging from 100,000 to 200,000. I map each PUMA to one of the rating regions, and calculate subsidy rate according to formula $subs_{itp} = 1 - \frac{affordability_{it}}{market\_rate_{itp}}$, where subscript i denotes individual (age and income), t year and p PUMA.

For PUMAs straddling two regions, true market rate is not known. I calculate average premium rate in the PUMA, averaging two regional rates adjusted by the share of population in each region. I apply this average market rate to all residents in the PUMA[16]. This is the strategy adopted in Frean, Gruber and Sommers (2017). In the Appendix I show results are not affected if I average calculated subsidy rate (rather than premium) over regions, assign the PUMA to region with larger population share, or simply drop these PUMAs from the sample.

3.2.3 Comparing with administrative records

How does the calculated subsidy compare with actual subsidy to the eligible low-income population? According to Independent State Auditor's Report on Certain Activities of the Commonwealth Health Insurance Connector Authority[17], from July to November, 2009, total insurance premium to Commonwealth Care insurance plans was \$ 808,729,633, out of which enrollees contributed \$ 41,404,805. The implied subsidy rate is 95%. In my data, average subsidy rate in year 2009 among enrollees in a non-ESI coverage is 70%. Further limited to eligible recipients with income below 300% FPL, the mean is 93%. Therefore calculated subsidy rate in the ACS data is comparable to actual subsidy generosity to program enrollees.

4 Empirical Strategy

I harness variation in both the numerator and denominator of subsidy rate to identify the effect of subsidy generosity on behavior. As stated previously, in each year, affordability depends only on income, whereas market premium rate varies across age band and location. Taken together, apart from the temporal variation over years, three sources of variation jointly determine subsidy generosity for eligibles:

1. for a given income (hence affordability) and location, individuals in older age groups are subsidized more;

[16] This generates 7 additional rating regions in the data. They account for about 14% of the Massachusetts population.

[17] available at http://www.mass.gov/auditor/docs/audits/2010/201014673a.pdf

2. for a given income (hence affordability) and age group, residents in lower cost area are subsidized less;

3. for a given age group and location, higher income individuals face higher affordability and are subsidized less.

In other words, the effect of subsidy generosity can be identified comparing across location, age and income groups within year, and the "triple difference" strategy is implied in the construction of subsidy rate discussed above. However, unlike most difference-in-difference analyses where the treatment status is either pre-existing or exogenous, in this case, behavioral responses to schedules can inflate certain groups' exposure to subsidy, causing biases of reserve causality. Omitted variables correlated with group-specific subsidy exposure also bias OLS estimates regressing outcomes on subsidy. I discuss in detail potential sources of biases and ways to overcome them below.

Firstly, subsidy rate defined using reported family income in ACS is a biased measure of subsidy *generosity* available to these households, and the bias is due either to behavioral responses to subsidy schedule, or to measurement errors[18] when behavioral response is minimal. For example, if generous schedules induce marginal enrollees to reduce labor effort, then subsidy exposure is *larger* than program generosity. Alternatively, if workers qualifying for subsidy sort out of ESI, and if ESI premium has previously been passed on to wages, then subsidy exposure can be *lower* than program generosity for workers switching insurance types. Because these behavioral selections are driven by program rules, but in turn affect actual subsidy exposure of enrollees, they cause reverse causality biases in OLS estimates on subsidy exposure. Given some of these behavioral responses are precisely the interest of the empirical exercise, it is necessary to address the challenges

[18]I introduce measurement error in the numerator of subsidy rate when I approximate federal adjusted gross income using reported pre-tax income. I introduce measurement error in the denominator when I average market premium rate in PUMAs straddling two regions using population weights.

they pose to causal identification[19][20].

I address reverse causality biases exploiting the fact that the rest of the nation did not undergo the reform, and that calculated subsidy rate from the reference national sample does not contain behavioral responses endogenous to subsidy take-up in Massachusetts. The simulated instrument strategy (Currie and Gruber 1996a, 1996b; Cutler and Gruber, 1996) has been widely applied in the literature. In this context, the instrument parametrizes the generosity of a policy schedule for different income groups in a population where individuals are kept unaware of the policy change and do not respond to the reform. The instrument then corrects for measurement error and reverse causality biases in the OLS estimates.

Secondly, omitted factors that jointly determine subsidy exposure and outcomes bias OLS estimates. For example, place-specific factors in the health care industry may affect both insurance pricing and local labor and insurance outcomes. Cohort differences in life cycle employment patterns may be correlated with differential subsidy exposure across age groups and year. Finally, risk preference can potentially explain the positive correlation between income and insurance take-up. To address these concerns, I control for the main effects of the known factors that determine subsidy generosity: income, year, age band, and location. In addition, I include three-way interaction terms of these factors to control for any *unknown* factors relevant for outcome that do not vary at the same level as subsidy schedules.

Could there be confounding factors that vary at exactly the same level as the subsidy schedule? In light of the concurrent recession that particularly affects year 2008-2009, I allow for economic shocks to have differential impact across year, location, age and income groups. Specifically, I interact

[19]The behavioral responses do not necessarily imply that individuals target income to a small range just below certain threshold. Endogenous subsidy rate can carry behavioral responses to policy even without much deliberate sophistication on the part of agents. For example, a near-elderly person may choose to retire early, and the lower retirement income will qualify her for subsidized coverage. In this case, her actual subsidy exposure is larger than generosity measured from a reference sample that did not respond to subsidy schedules, and the selection need not occur locally around policy cut-offs to still bias OLS estimates.

[20]The separate question whether individuals indeed manipulate income around policy thresholds meets somewhat mixed evidence. On the intensive margin of generosity, Exchange enrollees in Massachusetts do not appear more likely to report income just below thresholds, suggesting minimal manipulation (Finkelstein, Shepard, and Hendren, 2017). On the extensive margin of eligibility, there is some evidence of bunching at the threshold where subsidy fades out: Shi (2015) finds bunching before 300% FPL in Massachusetts from the ACS data, and Heim et al. (2016) finds bunching before the 400% FPL threshold in ACA from tax return data.

annual Massachusetts age-specific unemployment rate, obtained from Bureau of Labor Statistics, with a host of location-year dummies, demographic dummies, and in some specification, a full set of location-year-demographics dummies. This battery of controls, together with the fact that simulated subsidy rate comes from a pre-recession national sample, suggests local economic shocks are unlikely to be a major confound to the estimated effects of subsidy generosity.

I then turn to the construction of simulated instrument in the Massachusetts context.

4.1 Simulated subsidy rate

I apply Massachusetts affordability and premium schedules to a fixed national sample of households in year 2005-2006[21], assuming the entire simulation sample lived in one particular policy region and year. Similar to Mahoney (2015), I calculate simulated subsidy rate $subiv$ using the following equation:

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

That is, I apply the location(p)-year(t) schedule to every individual in the simulation sample. I assign market premium rate using age a, and affordability using individual's family income as percentage FPL. Having calculated subsidy rate for each individual, I average within demographic-age groups to compute group-specific generosity of a given location-year schedule[22]. $\mathbb{N}_{da}$ denotes the set of individuals of age a and demographics d. The resulting simulated rate $subiv_{dapt}$ instruments for the subsidy exposure of type da individuals in Massachusetts subject to pt schedules.

Although schedules vary across location, year and age bands, the instrument additionally vary by 144 demographic cells in d, generated by interacting gender, race (White, Black, other), Hispanic origin, education levels (high school drop-out, high school, and some college), marital status and presence of dependent children below 18. What is the value-added of the demographic variation? For reference, consider the lean instrument

[21] I construct the simulation sample from the 48 contiguous states and the Washington D.C., pooling over the 2005 and 2006 wave of ACS. In this sample, none of the individuals have experienced the 2007 Massachusetts reform or have reacted to the subsidy schedules. I choose the pre-recession period to avoid the confounding effects of the economic downturn.

[22] I use the STATA command *collapse* to generate cell means, adjusted by ACS sampling weights.

$sublean_{apt}$ varying only by location, year and age bands, constructed as the follows:

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

The instrument characterizes subsidy generosity across location-year-age-band cells, taking into account any baseline income differences over the life cycle. Within cell, however, there is no further variation for individuals across the income distribution. This is in contrast to the design of affordability schedules, which implies subsidy is more generous at lower income for all age groups. This income gradient in generosity, while important for outcome, is not included in the lean instrument as a potential source of identifying variation.

Because income distribution within Massachusetts is likely endogenous to the program, I generate income variation across race/ethnicity groups, education levels and family types from the simulation sample. That economic outcomes differ along the socio-economic gradient is well-established in the literature (). Demographic groups of lower socio-economic status (SES) are predicted to face higher subsidy generosity due to lower baseline income. This ex-ante association allows for a parametrization of generosity over income that does not contain any behavioral response to the program that may be differential across demographics[23]. Assuming the same SES gradient applies to Massachusetts, instrument $subiv_{dapt}$ is able to capture the variation in generosity *within* location-year-age-band cells, in addition to variation across cells.

The demographic variation in generosity proves valuable for identification. First, the variation is large and meaningful. Consider the 2010 schedule in Region 1 (Berkshire, Franklin and Hampshire Counties) for the 30-34 age group, where market rate is \$206. For a college-educated, non-Hispanic White male married but childless, average income is 743% FPL in the national sample, implying a 7.2% subsidy rate. On the other hand, a non-Hispanic Black single mother and high-school drop-out has income below the poverty line (63% FPL), and a near full subsidy rate (97%). Table 3 shows that subsidy is substantially more generous for minority groups, the lower educated groups, and single persons[24]. Because the policy targets

[23]The demographic association may introduce omitted variable biases, if demographic-specific factors correlated with subsidy generosity also determine group-specific outcomes. To address this issue, I always control for main effects of demographics in the regression. I also conduct over-identification tests where the demographic variation provides additional information on instrument validity.

[24]Alternatively, Appendix Figure shows the geographical variation in subsidy exposure

disadvantaged population less likely to have private insurance, this variation in generosity is meaningful.

Second, the demographic variation provides the additional information needed for over-identification tests to assess the exogeneity of policy variation across location, year, age *and* income groups. Conceptually, one might believe the location-year-age-band variation in the lean instrument $sublean_{apt}$ is plausibly exogenous, but the demographic variation in $subiv_{dapt}$ may be correlated with unobserved differences in demographics that are determinants of group-specific outcomes. Such concern is alleviated, when over-identification tests powered by demographic variation suggest both instruments contain valid sources of exogenous variation. Given this, $subiv_{dapt}$ is the preferred instrument because first-stage explanatory power is higher when generosity also varies by income, although similar qualitative effects hold using either instrument.

For a given demographic group, generosity variation across location, year, and age band is also large. For purely illustrative purposes, I focus on two groups corresponding to the margin of insurance selection (74th percentile in subsidy generosity) and new insurance take-up (95th percentile) in Massachusetts[25]. With 74% of the sample enrolled in ESI, the 74th percentile approximates marginal groups likely to select into subsidized plans at higher subsidy. With a 95% total insurance rate, the 95th percentile approximates marginal groups who are newly insured.

The 74th percentile is a single Hispanic woman of "other" race with college education and no children. Subsidy generosity for this group ranges from 41.83% (year 2008, Region 2, age 35-39) to 58.89% (year 2011, Region 1, age 55-64), with a mean of 48.94% in the estimation sample. The 95th percentile is a White Hispanic single male who did not finish high school and have no children. Generosity ranges from 71.26% (year 2008, Region 2, age 30-34) to 81.61% (year 2011, Region 1, age 55-64), with a mean of 75.99%.

(Panel A) and generosity (Panel B). The difference reflects the extent of income selection in actual exposure of local residents. Overall, there is similar distribution of higher subsidy areas in either measure. In terms of sample mean, simulated rate strongly predicts actual susbidy exposure, but distribution wise, the simulated rate is less concentrated at zero or full subsidy (Appendix Figure). Individuals receiving zero subsidy nonetheless enjoy small amount of generosity, and high subsidy recipients (> 80%) tend to have somewhat lower generosity. Wilcoxon matched sign test rejects the null of equal distribution with $p < e-4$.

[25] I first average out the year, location, and age-band variation in simulated rate $subiv_{dapt}$, and then rank by demographics to identify the 74th and 95th percentile.

Table 3: Demographic variation in subsidy rate

		subsidy rate		simulated subsidy rate	
	Observation	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: In this table I show average subsidy rate by demographic categories in Massachusetts, adjusted by ACS sampling weights. Individuals in Massachusetts are assigned two subsidy rates. The endogenous rate is calculated based on reported sub-family income applying the relevant PUMA-year-age-band schedule. The simulated rate is derived from applying the said schedule to a fixed national sample of individuals of similar demographic characteristics. Details are in the main text.

4.2 Econometric model

I use simulated subsidy rate $subiv_{dpt}$ to instrument for endogenous subsidy exposure $subs_{ipt}$. In the reduced form, I have the following specification:

$$
\begin{aligned}
y_{iapt} = \theta \cdot subiv_{d(i)apt} + \chi_0 \cdot incb_{d(i)} + \rho_a + \phi_p + \tau_t \\
+ \gamma_0 \cdot X_{d(i)} + f(UE_{at}; \gamma^{UE}_{0,pt}, \gamma^{UE}_{0,d(i)}) + \mu_{iapt},
\end{aligned} \tag{2}
$$

where the unit of observation is individual i of age a living in PUMA p sampled in year t. I control for the mains effects of policy variation, namely the location-year-age group variation in premium pricing, and affordability by income, with a full set of dummies and an exogenous measure of income as percentage FPL $incb_{d(i)}$, constructed from the same national sample that generates the instrument. Variable $incb_{d(i)}$ hence controls for any baseline differences in generosity across the income distribution, and admits a direct income effect on outcomes[26]. I control for baseline differences in demographics with indicators of gender, race/ethnicity, education levels and family composition in $X_{d(i)}$. In more complete specifications, I further include triple interaction terms between income and dummies of location, year and age group, to weed out any confounding variation that does not differentially affect all four margins as does the subsidy schedule[27].

Main effects and interaction terms are sufficient defense against confounding factors, if these factors do not vary at the same level as the instrument. One concern in this context is differential economic shocks affecting income groups across location, year and age bands. For this possibility, I generate variation in unemployment rate at the same level of the subsidy schedule. Specifically, I interact annual age-group unemployment rate in Massachusetts UE_{at}, with location dummies to allow for differential shocks by year, location and age. Moreover, I interact UE_{at} with demographic controls in $X_{d(i)}$. I include these additional controls, summarized in $f(UE_{at}; \gamma^{UE}_{0,pt}, \gamma^{UE}_{0,d(i)})$, in the regression. In the full specification, I interact

[26] In particular, this income measure does not suffer from potential behavioral response to the subsidy schedule. It captures the main effect of income since generosity is parametrized implicitly using this information on income. Results are similar if I use demographic level baseline income, demographic-age-band baseline income, a spline function of income with knots at 150% FPL, 200% FPL, 250% FPL, 300% FPL (all policy thresholds), or dummies for intervals created by these knots.

[27] For interactions, I use 10 policy regions instead of 52 PUMAs. The resulting interaction terms include year-region-age-band dummies, which absorb any level effect of market premium rate, differential income effects by year-region, by year-age-bands, and by region-age-bands.

UE_{at} with a complete set of location-year-demographic dummies, generating exactly the same level of variation as the instrument. I compare results with different controls below.

Two-stage-least-square (2SLS) estimates give the local average treatment effect (LATE), or changes in behavior from those receiving a small increase in generosity. These estimates are the empirical counterparts of sufficient statistics relevant for policy impact and welfare. As is clear from the reduced form, in this context, the small policy change is induced by moving an individual from one year, age, location or income group to another, or by permuting schedules along these margins. The first-stage equation is

$$\begin{aligned} subs_{iapt} = &\alpha \cdot subiv_{d(i)apt} + \chi_1 \cdot incb_{d(i)} + \rho_a + \phi_p + \tau_t \\ &+ \gamma_1 \cdot X_{d(i)} + f(UE_{at}; \gamma^{UE}_{1,pt}, \gamma^{UE}_{1,d(i)}) + \nu_{iapt}. \end{aligned} \tag{3}$$

Using the predicted value $\widehat{subs_{d(i)apt}}$, the second stage is

$$\begin{aligned} y_{iapt} = &\beta \cdot \widehat{subs_{d(i)apt}} + \chi_2 \cdot incb_{d(i)} + \rho_a + \phi_p + \tau_t \\ &+ \gamma_2 \cdot X_{d(i)} + f(UE_{at}; \gamma^{UE}_{2,pt}, \gamma^{UE}_{2,d(i)}) + \omega_{iapt}. \end{aligned} \tag{4}$$

Standard results have that $\hat{\beta}_{2SLS} = \frac{\hat{\theta}_{OLS}}{\hat{\alpha}_{OLS}}$. Following Frean, Gruber, and Sommers (2017), I cluster standard error at the PUMA level. Results are similar if I instead cluster by region, year and age band, or by region and age band.

4.3 Assessing instrument validity

Although the demographic variation in policy eligibility has previously been exploited to study insurance-related outcomes (Mahoney, 2015), in this paper I formally assess the validity of this source of identification. The central concern with the instrument is not whether it parametrizes group-specific program generosity free from any endogenous response (or reverse causality), but rather whether key determinants of group-specific outcomes have been controlled for in the model, so that netting out these determinants, differences in the instrument across demographics are not correlated with omitted factors of outcome.

I take multiple steps to assess the extent of omitted variable bias in this context. First, as a model specification test, I use the lean instrument varying across location, year and age band, joint with the main instrument which additionally varies across demographics, to construct over-identification tests. The additional demographic variation in the main instrument allows the model to be over-identified. For almost all outcomes across different specifications, I cannot reject the null that both instruments are exogenous to the error term. The result suggests estimates using the main instrument are unlikely to contain significant biases from omitted variables.

Second, to directly assess the size of potential bias, I generate random schedules across location, year, and age band, but keep the demographic characteristics of potential recipients unaltered. Absent omitted variable bias, the OLS estimates of outcome on random program generosity should be precisely zero. Any systematic deviation from zero indicates the extent of correlation between demographic variation in the instrument and omitted factors in the error term. Results indicate the size of bias in the main estimates is likely very small. In lieu of artificial schedules, I permute actual subsidy generosity in Massachusetts across the rest of the 50 states that did not undergo the reform, and find similar null results. I show more details in Section .

5 Results

5.1 First stage

Table 4 shows first-stage correlation between the endogenous variable and the instrument. Column (1) corresponds to the basic specification where I include PUMA, year, and integer age fixed effects, region-age-band fixed effects, demographic level baseline income, and main effects of demographics. Column (2) in addition allows for differential impact of age-specific economic shocks across demographics, with interaction terms between age-band unemployment rate and demographic controls. Column (3) is the main specification. In addition to PUMA and integer age dummies, it has a full set of triple interaction terms over region, year, age band, and income, and controls for economic shocks at the same level as the subsidy schedule. Specifically, I estimate the following first-stage regression in column (3):

$$\begin{aligned} subs_{iapt} = {} & \alpha \cdot subiv_{d(i)apt} + \chi_1 \cdot incb_{d(i)} + \rho_a + \phi_p + \tau_t + \rho_{b(a)} \cdot \phi_{r(p)} \cdot \tau_t \\ & + \rho_{b(a)} \cdot \phi_{r(p)} \cdot incb_{d(i)} + \phi_{r(p)} \cdot \tau_t \cdot incb_{d(i)} + \rho_{b(a)} \cdot \tau_t \cdot incb_{d(i)} \\ & + \phi_{r(p)} \cdot \tau_t \cdot X_{d(i)} + \gamma_0 \cdot UE_{b(a)t} + \phi_{r(p)} \cdot \tau_t \cdot X_{d(i)} \cdot UE_{b(a)t} + \nu_{iapt}. \end{aligned}$$

That is, I control for region-year-age-band determinants of outcome, and differential income effect across age-region, region-year, and age-year, with triple interaction terms[28]. In $\phi_{r(p)} \cdot \tau_t \cdot X_{d(i)}$, I interact indicators of demographic categories with region-year dummies to control for unobserved factors relevant for demographic differences in outcome. I then further interact $\phi_{r(p)} \cdot \tau_t \cdot X_{d(i)}$ with age-band unemployment rate $UE_{b(a)t}$ so that recession may affect outcome at at the same level as subsidy schedules.

On average, simulated subsidy rate strongly predicts endogenous subsidy exposure: first-stage coefficient is nearly one in all specifications. F-statistic suggests the instrument is strong, although the explanatory power decreases somewhat when controls at the same level as the instrument are included.

5.2 Insurance outcome

5.2.1 Coverage gain

Table 5 examines the effect of subsidy on insurance outcomes. The first three columns focus on having any health insurance. Panel A shows OLS estimates on endogenous subsidy exposure $subs$. According to these estimates, a ten percentage point increase in subsidy exposure is associated with a significant *decrease* in coverage by 0.7 percentage point. This association possibly reflects the fact that coverage is lower among lower-income population who has higher subsidy eligibility, but this correlation does not recover the causal impact of subsidy.

Results using simulated subsidy rate, on the other hand, show more generous subsidies significantly increase insurance take-up in the reduced form (Panel B) and the second stage (Panel C). The difference suggests that even controlling for a large number of group level controls, individual subsidy exposure is still endogenous, possibly due to reverse causality and unobserved individual characteristics. Hence a simulated instrument varying by group level generosity can potentially weed out sources of endogeneity in actual subsidy exposure, and restore unbiasedness in the second-stage.

[28] Interaction $\rho_{b(a)} \cdot \phi_{r(p)} \cdot incb_{d(i)}$ indicates coefficient before the continuous income variable is differentiated by age band and region. Year fixed effects are then τ_t are swept out with the interaction terms.

Table 4: First stage: endogenous subsidy on simulated instrument

	(1)	(2)	(3)
subiv	1.00***	1.00***	0.99***
	(0.038)	(0.039)	(0.042)
region-year FE	Y	Y	
region-year-age-band FE			Y
recession controls		Y	Y
R^2	0.29	0.29	0.29
F-statistic on instrument	687.32	669.50	549.49

Notes: Table shows the ordinary least squares (OLS) estimate in the first-stage regression of endogenous subsidy rate on the simulated instrument. In all specifications I include demographic controls, baseline income at the demographic level, PUMA, year, and region-year fixed effects. Column 2 additionally include age-band unemployment rate interacted with demographic controls. Column 3 includes all three-way interactions between age band, region, year, and income, as well as unemployment rate varying at the same level as the instrument. See main text for details. Robust standard errors clustered at the level of PUMA are in the parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 5: Effect of subsidy on insurance outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	have any insurance			have ESI		
	Panel A: OLS					
$subs$	-0.071***	-0.071***	-0.071***	-0.50***	-0.50***	-0.50***
	(0.0034)	(0.0033)	(0.0032)	(0.0087)	(0.0087)	(0.0090)
R^2	0.066	0.066	0.083	0.34	0.34	0.34
	Panel B: reduced form					
$subiv$	0.061**	0.062**	0.10***	-0.67***	-0.67***	-0.59***
	(0.025)	(0.026)	(0.025)	(0.047)	(0.046)	(0.050)
R^2	0.053	0.053	0.071	0.18	0.18	0.19
	Panel C: 2-stage least squares					
$\widehat{subs}$	0.061**	0.062**	0.10***	-0.67***	-0.66***	-0.60***
	(0.025)	(0.026)	(0.025)	(0.030)	(0.029)	(0.034)
R^2	0.021	-0.035	-0.077	0.32	0.32	0.18
region-year FE	Y	Y		Y	Y	
region-year-age-band FE			Y			Y
recession controls		Y	Y		Y	Y

Notes: Table shows the OLS estimates regressing insurance outcome on endogenous subsidy rate $subs$ in Panel A, OLS estimates regressing outcome on simulated subsidy rate $subiv$ (the reduced form) in Panel B, and 2-stage least square estimates instrumenting $subs$ with $subiv$ in Panel C. I experiment with different controls for location-year-age-band confounds, in particular, recession, as indicated in the lower half of the table. Robust standard errors clustered at the level of PUMA in the parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

2SLS estimates suggest that for a ten percentage point increase in subsidy generosity, there is a one percentage point increase in new insurance take-up. To put the effect into perspective, I infer the counterfactual coverage gain if all regions followed the most generous schedule in the data, the one in region 1 and year 2011 (Appendix Table). The new schedule would increase subsidy generosity in Massachusetts by 1 percentage point, which would then increase coverage by 0.1 percentage point. Hence subsidy generosity can explain about 14% of the location variation in insurance rate[29].

The previous exercise uses only within-sample variation. A bolder counterfactual is to infer the subsidy rate needed for universal coverage. Since none of the sample regions achieved 100% insurance rate, assuming a constant effect on take-up, subsidy should increase by 50 percentage points to cover the remaining 5% uninsured. Currently, subsidy among eligible population not covered by ESI is 68%. The implication is that even with full subsidy, coverage is not necessarily universal[30].

Furthermore, assuming effects in Massachusetts are externally valid for the national reform of ACA, how do estimates compare across the two experiments? Frean, Gruber, and Sommers (2017) estimates that a ten percentage point increase in subsidy rate lowers uninsurance by 0.98 percentage point in 2015[31], very similar to the average effect in Massachusetts.

5.2.2 Exit from ESI

A related question is, did the subsidy expansion lower coverage by ESI? In principle, ESI coverage does not respond to subsidy, since firms are required to continue coverage for eligible workers and families, and ESI eligibles are not entitled to subsidy. In practice, workers may select out of ESI coverage, and the incentive is greater with greater subsidy generosity. A large body of evidence suggests insurance selection does occur after public expansion, lowering net coverage gain by 50% to 60% (Gruber and Kosali, 2008), and it

[29] Specifically, insurance rate is 95.54% in region 1 and year 2011, and 94.81% in other regions. Out of the 0.7 percentage point difference in coverage, about 0.1 percentage point is due to subsidy generosity.

[30] The modest effect of subsidy on insurance take-up is consistent with the low demand for formal insurance uncovered in Finkelstein, Sheppard, and Hendren (2017): in-sample willingness to pay for formal insurance lies completely below expected cost to insurers, limiting the effect of subsidy on take-up.

[31] Frean, Gruber, and Sommers (2017) exploits insurance rating variation and implied subsidy differences across regions (county), year and income group differentiated by demographics (household type) to identify effect on insurance outcomes. Similar variation, along with age band rating, is present in Massachusetts.

often involves salient employment adjustments on the labor market (Gruber and Madrian, 2002). In Massachusetts, the selection is likely concentrated among workers who find it difficult to match with an ESI-sponsoring job, using subsidy to fulfill the individual mandate, and those planning on leaving the labor force.

Column (4)-(6) looks at reduction in ESI: the main specification shows a one percentage point increase in generosity reduces ESI coverage by 0.60 percentage point. The effect is comparable to the 50%-60% range suggested in Gruber and Kosali (2008). However, adjusting for the extensive margin of subsidy eligibility, crowd-out in Massachusetts is possibly smaller than in previous experiments. For example, consider the case where a marginal ESI drop-out receives subsidy at rate 70%. Treatment increases from 0 to 0.70 for the individual. In full-subsidy Medicaid programs, treatment for new enrollees always increases discretely from 0 to 1. For the comparison to reflect the same margin of enrollment, one needs to multiply Massachusetts estimate by 0.7.

I focus on two marginal groups: the 74th percentile in the subsidy distribution, and an average person who is not enrolled in ESI. Since the marginal switcher must forgo ESI coverage, with 74% enrolled in ESI, the 74th percentile (subsidy rate 0.65) is the potential "next in line" to select into subsidy. The implied crowd-out rate for this group is 0.39 ($= 0.65 * 0.60$), smaller than previous estimates. Focusing on the subgroup not enrolled in ESI and potentially eligible for subsidy, average subsidy is 68%, and the implied crowd-out is 0.41 ($= 0.68 * 0.60$).

5.2.3 Take-up of subsidized insurance

Identifying actual and potential recipients of subsidy in ACS is difficult: government assisted insurance coverage can either be reported as Medicaid, or private purchase from insurance companies. Outcomes of having any insurance and ESI coverage, on the other hand, are well measured, and can be used to infer the incentive effect on subsidy take-up: assuming public program attracts both new enrollees and ESI drop-outs, the incentive effect on subsidy take-up is around 0.70 (=0.60+0.10).

How well does the estimate describe actual enrollment in Medicaid and Commonwealth Care? I first construct the pool of public insurance enrollees in ACS. The complementary nature of Medicaid and Commonwealth Care implies individuals below 150% FPL not covered by ESI receive full subsidy, and a higher-income group up to 300% FPL receive partial subsidy. Therefore I identify as subsidy recipients individuals with qualifying income ($\leq 300\%$

FPL) who are not enrolled in ESI and who have either Medicaid or individual coverage purchased from insurance companies[32].

The resulting share of subsidized insurance enrollment is 18% in the ACS data for the non-elderly, which is close to if not somewhat *lower* than the administrative record. Based on program reports to Massachusetts Division of Health Care Finance and Policy, total insurance enrollment among the Massachusetts non-elderly population is 5,545,447 by the end of 2010 (Key Indicators, June 2011[33]), of which about 20% enrolled in subsidized insurance[34].

Appendix Table shows the effect on subsidized insurance take-up: with one percentage point increase in subsidy, subsidized insurance take-up increases by 0.76 percentage point. The effect is similar to the inferred estimates based on better measured outcomes. For the quantification of sufficient statistics, I similarly circumvent the measurement problem by writing out public insurance coverage as function of any insurance and ESI: $\lambda_{i,2} = 1 - \lambda_{i,0} - \lambda_{i,1}$, $i = e, 1-e$, and $\frac{d\lambda_{i,2}}{d\lambda_p} = -\frac{d\lambda_{i,0}}{d\lambda_p} - \frac{d\lambda_{i,1}}{d\lambda_p}$.

5.3 Labor outcome

Whether means-tested public insurance programs affect labor supply has been the interest of a long line of empirical research. As crowd-out is more likely to occur among higher income groups where ESI coverage is more common, labor reduction is possibly larger following policies that expanded eligibility based on income. In these cases, the government needs to finance subsidy to a larger number of recipients from a smaller tax base, and the fiscal externality lowers the efficiency of subsidy transfer. The related welfare cost is proportional to the elasticity of labor reduction to program generosity, ϵ_{1-e,λ_p}. Here I focus on the empirical characterization of marginal change

[32]I assume commercial plan enrollees buy from Commonwealth Care whenever their income qualifies for subsidy.

[33]accessible at http://archives.lib.state.ma.us/bitstream/handle/2452/118563/ocn232606916-2011-06.pdf?sequence=1

[34]The subsidized category combines administrative counts from four programs: MassHealth, Commonwealth Care, Medical Security Program, and Commonwealth Care Bridge. Commonwealth Care Bridge provides continued premium assistance to previous CommCare enrollees who lost eligibility due to changes in state law. Medical Security Program is tied with unemployment insurance and subsidizes the premium of existing COBRA coverage or coverage from the program. Apart from Medical Security Program, which accounts for 0.7% of total enrollment, the combined subsidy schedule of these programs are similar to ones described in the main text.

rate $\frac{d(1-e)}{d\lambda_p}$.

Table 6: Effect of subsidy on non-participation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	not in labor force			not employed			last 12 months did not work		
					Panel A: OLS				
$subs$	0.30***	0.30***	0.30***	0.41***	0.40***	0.41***	0.35***	0.35***	0.35***
	(0.0065)	(0.0065)	(0.0066)	(0.0070)	(0.0070)	(0.0071)	(0.0079)	(0.0079)	(0.0080)
R^2	0.18	0.18	0.18	0.20	0.20	0.21	0.20	0.20	0.21
					Panel B: reduced form				
$subiv$	0.081**	0.075*	0.056	0.12**	0.11**	0.055	0.12***	0.11**	0.077*
	(0.040)	(0.040)	(0.044)	(0.048)	(0.048)	(0.053)	(0.042)	(0.042)	(0.045)
R^2	0.094	0.095	0.10	0.083	0.084	0.090	0.092	0.092	0.099
					Panel C: 2-stage least squares				
$\widehat{subs}$	0.081**	0.075*	0.057	0.12**	0.11**	0.056	0.11***	0.11***	0.077*
	(0.038)	(0.038)	(0.043)	(0.044)	(0.045)	(0.052)	(0.039)	(0.040)	(0.044)
R^2	0.13	0.13	0.023	0.14	0.14	0.024	0.15	0.063	0.039
region-year FE	Y	Y		Y	Y		Y	Y	
region-year-age-band FE			Y			Y			Y
recession controls		Y	Y		Y	Y		Y	Y

Notes: Table shows the OLS estimates regressing labor outcome on endogenous subsidy rate $subs$ in Panel A, OLS estimates regressing outcome on simulated subsidy rate $subiv$ (the reduced form) in Panel B, and 2-stage least square estimates instrumenting $subs$ with $subiv$ in Panel C. Column (1)-(3) show effects on whether individual is in labor force last week. Column (4)-(6) show effects on whether individual is employed last week. Column (7)-(9) show effects on whether individual worked in the past 12 months. I experiment with different controls for location-year-age-band confounds, in particular, recession, as indicated in the lower half of the table. Robust standard errors clustered at the level of PUMA in the parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

I inspect three labor outcomes in Table 6: whether the individual reported not in labor force last week in column (1)-(3), not employed in column (4)-(6), and not having worked in the past 12 months in column (7)-(9). Including a host of unemployment controls at the same level as subsidy renders the effect on labor smaller and insignificant: for participation and employment, a one percentage point increase in subsidy results in 0.057 percentage point decrease in labor effort, and for past year working experience, a marginally significant reduction of 0.077 percentage point. Overall, effect on labor in the full population cannot be distinguished from zero, although specifications with less aggressive controls give somewhat larger and more significant effects. In permutation tests, estimates with full controls are larger than 95% of the pseudo estimates from the rest of the nation.

The small effect on labor in Masschusetts is consistent with growing evidence that the labor effect of ACA is also small. Most papers rely on the state variation in Medicaid expansion, and find little impact on labor (Leung and Mas, 2017; Kaestner et al., 2016). By comparison, subsidy on the Exchange seems to receive much less consideration. Duggan, Goda and Jackson (2017) shows differential labor response across the income distribution. Areas with more uninsured with income too low to qualify for Exchange subsidy increased labor supply, whereas those with more uninsured eligible for Exchange subsidy decreased labor supply. Specifically, a ten pecentage point difference in local Exchange eligibles predicts a one percentage point difference in labor force participation.

How does the location-based evidence on labor compare with the micro level evidence in this paper? Conceptually, subsidy generosity within a region depends on the product of two statistics, the size of eligible population, or in the case of ACA Exchange, $Pr(138\% \leq income \leq 399\%)$, and the average subsidy generosity for eligibles, $E(subs|138\% \leq income \leq 399\%)$. The micro estimate β_m recovers the relationship between generoisty $subiv$ and outcome y: $y \propto \beta_m \cdot E(subs|138\% \leq income \leq 399\%) \cdot Pr(138\% \leq income \leq 399\%)$. The location-based estimate β_l recovers $y \propto \beta_l \cdot Pr(138\% \leq income \leq 399\%)$. Potentially β_m and β_l can be related through $\beta_l = \beta_m \cdot E(subs|138\% \leq income \leq 399\%)$, and the conditional mean is observable in the data.

In practice the relationship need not hold. In measuring potential eligibility, for good reasons, Duggan, Goda and Jackson (2017) uses pre-existing income distribution within region. The ex-ante measure would be exact, if there were no behavioral responses that altered the income and subsidy eligibility of potential enrollees. When such response does occur, the ex-ante measure departs from actual subsidy, making it more difficult to compare estimates. For example, also studying the ACA reform, Frean, Gruber and Sommers (2017) finds the micro estimate on any insurance coverage is around 0.098 in 2015. The location-based estimate in Duggan, Goda and Jackson (2017) is 0.25. To generate this pattern, *actual* subsidy to recipients need to be substantially greater than predicted values from pre-existing income distributions[35].

Along the same margin of selection, what is the micro estimate on labor implied by the location-based estimate in Duggan, Goda and Jackson (2017)? I proxy the magnitude of selection using the ratio of the two estimates on coverage: $select = \frac{\beta_l}{\beta_m}|_{insurance} = 2.5$. $select$ measures the difference bewteen actual subsidy exposure and the ex-ante size of eligible population[36]. Assuming the same selection factor applies to labor outcomes, the implied micro estimate $\beta_m = \frac{\beta_l}{select} = 0.04$, smaller than yet close to the 0.057 estimate in Massachusetts. Although subsidy schedules are not exactly identical, estimates in the ACA case imply similar insurance and labor adjustment to subsidy generosity as found by estimates in Massachusetts.

Finally, previous studies have identified age sub-groups more likely to change labor effort following public insurance expansion. In particular, the near-elderly population may prefer subsidized insurance to the "job-lock" of ESI (Gruber and Madrian, 1995). Some may exit the labor market when switching to public insurance. The retirement effect turns out to be significant in Massachusetts (Heim and Lin, 2017). I next explore any differential effect by age group.

5.4 Heterogeneous effects by age

I fully stratify the age band variation in subsidy generisosity and show age-specific effects in Table 7. I focus on reduced-form estiamtes. In square brackets I show the dependent mean for each age group. New coverage gain is most concentrated among the youngest age group, where enrollment in subsidized program is also most responsive to increased genersosity (column 3), and the probability of acquiring ESI is the lowest (column 2). At older ages, subsidy enrollment and ESI crowd-out are both less responsive to

[35] Both papers used PUMA-level insurance pricing variation, coupled with income-specific affordability schedule, to identify the effect of subsidy geneorsity on coverage. In Duggan, Goda and Jackson (2017), generosity is measured ex-ante by the size of eligible population before the reform, whereas in Frean, Gruber and Sommers (2017), realized subsidy exposure is measured and endogeneity bias is corrected with simulated instruments. Differences in effect size between the two studies partly reflect the degree of endogenous selection into the program.

[36] Specifically, $\frac{\beta_l}{\beta_m} = \frac{E(subs\_post|138\% \leq income\_post \leq 399\%)Pr(138\% \leq income\_post \leq 399\%)}{Pr(138\% \leq income\_pre \leq 399\%)}$.

Table 7: Age-specific effect of subsidy on insurance and labor outcomes

	(1)	(2)	(3)	(4)	(5)
	any insurance	ESI	subsidized insurance	not in labor force	not employed
27-29	0.18**	-0.87***	1.04***	0.042	0.15
	(0.081)	(0.11)	(0.10)	(0.095)	(0.13)
	[0.91]	[0.67]	[0.20]	[0.13]	[0.20]
30-34	0.076	-0.65***	0.83***	-0.010	0.062
	(0.053)	(0.097)	(0.088)	(0.069)	(0.081)
	[0.92]	[0.71]	[0.17]	[0.13]	[0.19]
35-39	0.056	-0.57***	0.67***	-0.18**	-0.13
	(0.048)	(0.089)	(0.086)	(0.076)	(0.085)
	[0.94]	[0.74]	[0.15]	[0.13]	[0.19]
40-44	0.13***	-0.48***	0.71***	-0.084	-0.11
	(0.047)	(0.063)	(0.065)	(0.066)	(0.075)
	[0.95]	[0.75]	[0.15]	[0.14]	[0.20]
45-49	0.099***	-0.53***	0.68***	0.0012	-0.0023
	(0.029)	(0.068)	(0.060)	(0.059)	(0.073)
	[0.95]	[0.76]	[0.14]	[0.14]	[0.20]
50-54	0.11***	-0.55***	0.67***	0.17***	0.12*
	(0.029)	(0.061)	(0.056)	(0.050)	(0.062)
	[0.96]	[0.76]	[0.13]	[0.16]	[0.22]
55-64	0.087***	-0.66***	0.80***	0.30***	0.24***
	(0.027)	(0.059)	(0.061)	(0.054)	(0.069)
	[0.97]	[0.74]	[0.14]	[0.28]	[0.33]
R^2	0.071	0.19	0.18	0.10	0.091

Notes: Table shows reduced-form estimates regressing insurance outcomes on age-stratified susbidy generosity. I include in the regression all three-way interaction terms across region, year, age band, and income, and unemployment rate controls varying at the same level as susbidy schedules. Robust standard errors clustered at the level of PUMA in the parenthesis. Age group mean of the dependent variable in the square bracket. Subsidized insurance enrollment is constructed from ACS based on reported Medicaid/private insuance purchase and income eligibility for subsidy. See main text for details.

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

subsidy. Relative to more experienced workers, younger workers rely more on subsidy to obtain insurance, as opposed to demanding insurance from potential employers.

An uptick appears for both subsidy enrollment and ESI crowd-out in the near-elderly group (55-64): compared to 45-54 year olds, they are more likely to drop ESI and enroll in subsidized insurance. Net insurance gain is modest and similar to other older age groups. The insurance selection is accompanied by a significant reduction in labor: participation drops markedly starting age 50 (and especially after 55), and non-employment effect is strongest among the 55-64 group. Similar effects among the younger groups, on the other hand, are largely missing.

Overall, evidence points to a modest coverage gain concentrated among the young, and increased sorting into subsidized coverage among the near-elderly, who may also leave ESI jobs or the labor force in general. The evidence is hence consistent with a large literature on "job-lock", and complementary to the discussion of age rating regulation on the Exchange, where market performance improves if more generous subsidy differentially attracts healthier, younger enrollees (Ericson and Starc, 2015; Tebaldi, 2017; Aizawa, 2016).

Appendix Figure shows year-specific age profiles of outcomes in Table 7: I stratify *subiv* by both age band and year, and plot coefficients. The year profiles are informative in that if age differences are driven by differential economic shocks, then one might expect systematically bigger effects on behavior during worse-hit years in 2009-2010. However, estimated crowd-out and the labor reduction among the near-elderly appear fairly constant over years, and so is subsidy enrollment among the young. Hence recession is unlikely to explain the behavioral pattern in age sub-groups.

5.5 Insurance-labor outcome

Table 8 shows the effect of subsidy on different combination of labor and insurance outcomes. Column (1)-(4) looks at insurance selection and labor, and column (5)-(6) looks at new coverage and labor. First notice that effects on interactive outcomes recover effects on either insurance or labor: for example, summing up coefficients in column (1) and (2), the reduction in employment is about 0.06 percentage point per subsidy, similar to the main estimates. Coefficients in column (1) and (3) also recover the main effect on ESI coverage.

Table 8: Effect of subsidy on insurance-labor outcome

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	in labor force + ESI	in labor force + no ESI	not in labor force + ESI	not in labor force + no ESI	in labor force + insured	in labor force + not insured	not in labor force + insured	not in labor force + not insured
				Panel A: OLS				
$subs$	-0.54***	0.23***	0.034***	0.27***	-0.35***	0.045***	0.28***	0.026***
	(0.0077)	(0.0067)	(0.0034)	(0.0071)	(0.0064)	(0.0031)	(0.0064)	(0.0015)
R^2	0.30	0.14	0.063	0.23	0.19	0.073	0.17	0.033
				Panel B: reduced form				
$subiv$	-0.30***	0.24***	-0.30***	0.36***	0.048	-0.10***	0.054	0.0028
	(0.058)	(0.038)	(0.021)	(0.037)	(0.039)	(0.024)	(0.041)	(0.0085)
R^2	0.14	0.098	0.064	0.12	0.10	0.068	0.099	0.024
				Panel C: 2-stage least squares				
$\widehat{subs}$	-0.30***	0.24***	-0.30***	0.36***	0.049	-0.11***	0.054	0.0029
	(0.050)	(0.035)	(0.025)	(0.031)	(0.040)	(0.024)	(0.040)	(0.0085)
R^2	0.14	0.045	-0.22	0.11	-0.030	-0.067	0.028	-0.0064

Notes: Table shows the OLS estimates regressing combinations of insurance-labor outcomes on endogenous subsidy $subs$ in Panel A, on simulated generosity $subiv$ (the reduced form) in Panel B, and 2-stage least square estimates instrumenting $subs$ with $subiv$ in Panel C. Estimates are from the preferred specification with three-way interaction terms across region, year, age band, and income, as well as unemployment rate controls varying at the same level as the subsidy schedule. Robust standard errors clustered at the level of PUMA in the parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

What new information do joint outcomes generate? Consider the insurance selection and labor outcomes in column (1)-(4). If we assume that individuals not in the labor force do not subsequently enter labor force upon acquiring subsidized insurance, then the 0.30 reduction in column (3) is absorbed in the 0.36 increase in non-participants on subsidy in column (4), and the remaining 0.06 increase comes from new exits from the labor force. The same labor reduction due to insurance selection is also evident in column (1)-(2): inability to get ESI accounts for about 24 percentage points of the 30 percentage point reduction in the share of ESI-covered labor, implying a net change of 6 percentage points along the participation margin.

Therefore estimates on joint outcomes reveal how much of the insurance crowd-out involves changes in employment status. Based on the previous calculation, of the 60 percentage points that dropped out of ESI, around 6 percentage points exited the labor force, 30 percent points were already outside the labor force, and the remaining 24 percentage points remained active labor. The break-down puts into perspective the small LATE estimate on labor: only a small fraction of insurance switchers altered labor supply.

Appendix Table shows age-specific effects on joint outcomes. In the 27-29 group, although selection into subsidy is large, there is virtually no effect on labor supply: only 5.8% (= $\frac{0.31-0.26}{0.61+0.26}$) of ESI drop-outs also exited labor force, and 65.5% (= $\frac{0.57}{0.61+0.26}$) remained active labor. In the 55-64 group, on the other hand, as many as 46.2% (= $\frac{0.64-0.34}{0.31+0.34}$) of ESI drop-outs exited labor force, and only 2.5% (= $\frac{0.016}{0.31+0.34}$) remained active labor. In terms of new coverage, only 28.8% (= $\frac{0.12-0.035}{0.26+0.035}$) of labor exists in the 55-64 group are new enrollees, whereas only 12.3% (= $\frac{0.022}{0.20-0.021}$) of new enrollees in the 27-29 group reduced labor effort.

5.6 Robustness and falsification tests

I conduct a number of robustness tests on the key estimates that inform sufficient statistics. Table 9 lists these estimates: in principle, only estimates on labor and joint outcomes are needed for welfare analysis, but I nonetheless include estimates on coverage and ESI in column (1)-(2), because of interest in these estimates on their own. In Panel A, I go over the main reduced-form results presented before.

Table 9: Robustness tests on main estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	any insurance	ESI	not in labor force	in labor force + insured	not in labor force + insured	not in labor force + ESI
			Panel A: main results			
subiv	0.10***	-0.59***	0.056	0.048	0.054	-0.30***
	(0.025)	(0.050)	(0.044)	(0.039)	(0.041)	(0.021)
R^2	0.071	0.19	0.10	0.10	0.099	0.064
			Panel B: average subsidy, not prices, for border PUMA			
subiv	0.10***	-0.59***	0.056	0.048	0.054	-0.30***
	(0.025)	(0.050)	(0.044)	(0.039)	(0.041)	(0.021)
R^2	0.071	0.19	0.10	0.10	0.099	0.064
			Panel C: assign border PUMA to region with larger share			
subiv	0.093***	-0.61***	0.062	0.034	0.059	-0.30***
	(0.024)	(0.050)	(0.044)	(0.040)	(0.042)	(0.022)
R^2	0.061	0.18	0.10	0.097	0.097	0.062
			Panel D: cluster by region-age band			
subiv	0.093**	-0.61***	0.062	0.034	0.059	-0.30***
	(0.033)	(0.053)	(0.060)	(0.071)	(0.060)	(0.037)
R^2	0.061	0.18	0.10	0.097	0.097	0.062

Notes: Table shows the main reduced-form estimates on key outcomes in Panel A. Panel B shows estimates where simulated generosity is constructed for border PUMA (straddling two policy regions) by averaging calculated subsidy rate using population weights, as opposed to averaging prices and then calculating subsidy rate. Panel C shows estimates where border PUMAs are absorbed into the region with greater population share. Panel D shows estimates that are identical to those in Panel C, but with robust standard errors clustered at the level of region-age band, or 21 (3 regions * 7 age bands) units. Clustering by region-year-age band (84 units) gives very similar standard errors as those in the parenthesis in Panel A, B and C, which are clustered at the level of PUMA (52 units).

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

Panel B and C show robustness of estimates to different treatment of border PUMAs straddling two rating regions. In the main analysis, I generate average premium pricing in the PUMA, weighted by population share in each region. Alternatively, I could have calculated subsidy rate in different regions, before weighting over regions overlaying the border PUMA. Results are shown in Panel B. In Panel C, I assign border PUMAs to the region with larger population share.

In Panel D, I cluster standard error at the level of region and age band, following the same treatment of border PUMAs as in Panel C. Number of cluster units is 21 (3 regions and 7 age bands) in Panel D, as opposed to 52 (PUMAs) in Panel A-C. While location-based clustering allows for potential correlation over year and age, neither the original rating variation across 3 regions nor the constructed variation across 10 regions is sufficient for credible inference. Following Frean, Gruber and Sommers (2017), I cluster at the level of 52 PUMAs in the main analysis. Panel D shows that clustering at the level of original rating regions and age band gives somewhat larger standard errors. In results not shown, clustered standard errors at the level of region, age band *and* year are very similar to those at the level of PUMAs.

5.6.1 Assessing instrument validity

The main instrument is valid, if the demographic variation coded in simulated generosity is uncorrelated with unobserved determinants of outcome differences across demographics. In the preferred specification I use a large number of controls including triple interaction terms and unemployment rate at the level of location, year, age band *and* demographics to limit the bias from unobserved factors. The assumption is that, partialing out these controls, there is no correlation between the instrument and unmodelled facors in the error term. I first test the plausibility of this assumption using over-identification tests.

Relative to a lean instrument that varies only by location, year and age band, the main instrument additionally varies by demographics. The demographic variation can be used to test the exogeneity of both instruments. Table 10 shows estimates using only the main instrument in Panel A, only the lean instrument in Panel B, and over-identified estimates using both instruments in Panel C[37]. The lean instrument turns out to be weak: first-stage statistic is 4.78, and the resulting estimates are qualitatively similar to the main estimates but larger in magnitude and insignificant. The main

[37] To avoid soaking up the variation in the lean instrument, I do not include triple interaction terms in specifications using both instruments.

instrument is strong. Antoine and Renault (2017) shows that the GMM over-identification test is valid in the presence of weak instruments, and the power of the test improves with greater variance in instrument strength[38].

For all outcomes, I cannot reject the null that both instruments are valid. Furthermore, results are similar when I vary the set of controls in the over-identified regression, hence varying the strength of the weaker instrument: Appendix Table shows instruments remain valid across different specifications for most outcomes.

The over-identification tests suggest the main instrument is unlikely to introduce substantial omitted variable bias. To directly assess the size of any potential bias, I construct pseudo subsidy schedules based on random variation in premium pricing across region and age bands, random affordability schedule over income[39], and random variation of both schedules over period 2008-2011.

I apply 100 sets of pseudo schedules to the estimation sample, coding the generosity of each from the simulation sample. I keep the demographic information in both samples unchanged: given any schedule, a higher-income demographic group receives smaller subsidy generosity than a lower-income group. The baseline income differences help quantify the within-schedule generosity over income, but because schedules are random, the true effect of subsidy is precisely zero. Across the 100 estimates from the pseudo scheules, any systematic deviation from zero indicates the extent of bias due to demographic variation in the instrument that is not randomized. In this case, the main estimates likely capture both the true effect of subsidy on outcome, *and* any omitted factors related to income differences across demographics.

[38]In the notation of Antoine and Renault (2017), the moment conditions for the instruments are

$$\begin{bmatrix} \omega_i(S_i - \Sigma_{SW}\Sigma_{WW}^{-1}W_i) \\ \omega_i W_i \end{bmatrix} = 0,$$

where ω_i is error term in the structural equation, and S is the set of strong instrument, in this case $subiv$, and W the set of weak instruments, or $sublean$. $S_i - \Sigma_{SW}\Sigma_{WW}^{-1}W_i$ captures the generosity variation by demographics. While estimation is driven by the stronger and potentially invalid instrument, in the testing stage, the estimates are required to be compatible with information in the weaker but plausibly more valid instrument. Heteroeneity in the strength of instruments allows distinct information to be used in the estimation and testing stage, and appears to improve test power.

[39]I randomly generate 4 income thresholds. I generate the first threshold h_1 from uniform $[0,1]$, and then generate an increment ih also from uniform $[0,1]$, so that subsidy fades out at $h_1 + 3ih$ of poverty line. Below h_1, affordability a_1 is generated from uniform $[0,100]$, and increases by a fixed amount ia, also generated from uniform $[0,100]$, at each new income threshold.

Table 10: 2SLS estimates using main, lean, or both instruments

	(1)	(2)	(3)	(4)	(5)	(6)
	any insurance	ESI	not in labor force	in labor force + insured	not in labor force + insured	not in labor force + ESI
	Panel A: 2SLS estimates, instrument varying by location, year, age band, and demographics					
$\widehat{subs}$	0.10***	-0.60***	0.057	0.049	0.054	-0.30***
	(0.025)	(0.034)	(0.043)	(0.040)	(0.040)	(0.025)
R^2	-0.077	0.18	0.023	-0.039	0.020	-0.22
First stage F statistic	549.49	549.49	549.49	549.49	549.49	549.49
	Panel B: 2SLS estimates, instrument varying by location, year, and age band					
$\widehat{subs}$	0.23	-1.52***	0.28	-0.078	0.30	-0.39
	(0.33)	(0.67)	(0.40)	(0.39)	(0.39)	(0.39)
R^2	-0.16	-0.34	0.18	0.13	0.16	-0.026
First stage F statistic	4.78	4.78	4.78	4.78	4.78	4.78
	Panel C: over-identified 2SLS estimates					
$\widehat{subs}$	0.10***	-0.60***	0.057	0.046	0.053	-0.30***
	(0.025)	(0.034)	(0.043)	(0.040)	(0.040)	(0.025)
R^2	-0.073	0.18	0.024	-0.036	0.020	0.062
First stage F statistic	275.47	275.47	275.47	275.47	275.47	275.47
Hansen test p-value	0.18	0.16	0.46	0.88	0.30	0.35

Notes: Table shows 2-stage-least-square estimates using different instruments for endogenous subsidy. Panel A estimates the preferred specification with full three-way fixed effects and unemployment rate at the level of subsidy schedule using the main instrument. Panel B estimates a basic specification with region-year fixed effects and demographic controls (first stage corresponding to column (1) in Table 4) using the lean instrument. Panel C estimates the preferred specification but without three-way fixed effects using both instruments. I show first-stage F-statistic on instrument(s) in each case, and in Panel C, the p-value of Hansen J over-identification test. Robust standard errors clusterd at the level of PUMAs are in the parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 11: Effect on randomly constructed schedules

	(1)	(2)	(3)	(4)	(5)	(6)
	any insurance	ESI	not in labor force	in labor force + insured	not in labor force + insured	not in labor force + ESI
			Panel A: main estimates			
subiv	0.10***	-0.59***	0.056	0.048	0.054	-0.30***
	(0.025)	(0.050)	(0.044)	(0.039)	(0.041)	(0.021)
R^2	0.071	0.19	0.10	0.10	0.099	0.064
			Panel B: pseudo estimates			
subiv	-3.4E-06	-8.7E-06	8.6E-05	-9.6E-05	9.3E-05	7.5E-06
	(0.00066)	(0.0012)	(0.0012)	(0.0013)	(0.0012)	(0.00068)
	[-0.0014, 0.0022]	[-0.0028, 0.0025]	[-0.0038, 0.0027]	[-0.0037, 0.0030]	[-0.0033, 0.0022]	[-0.0024, 0.0014]

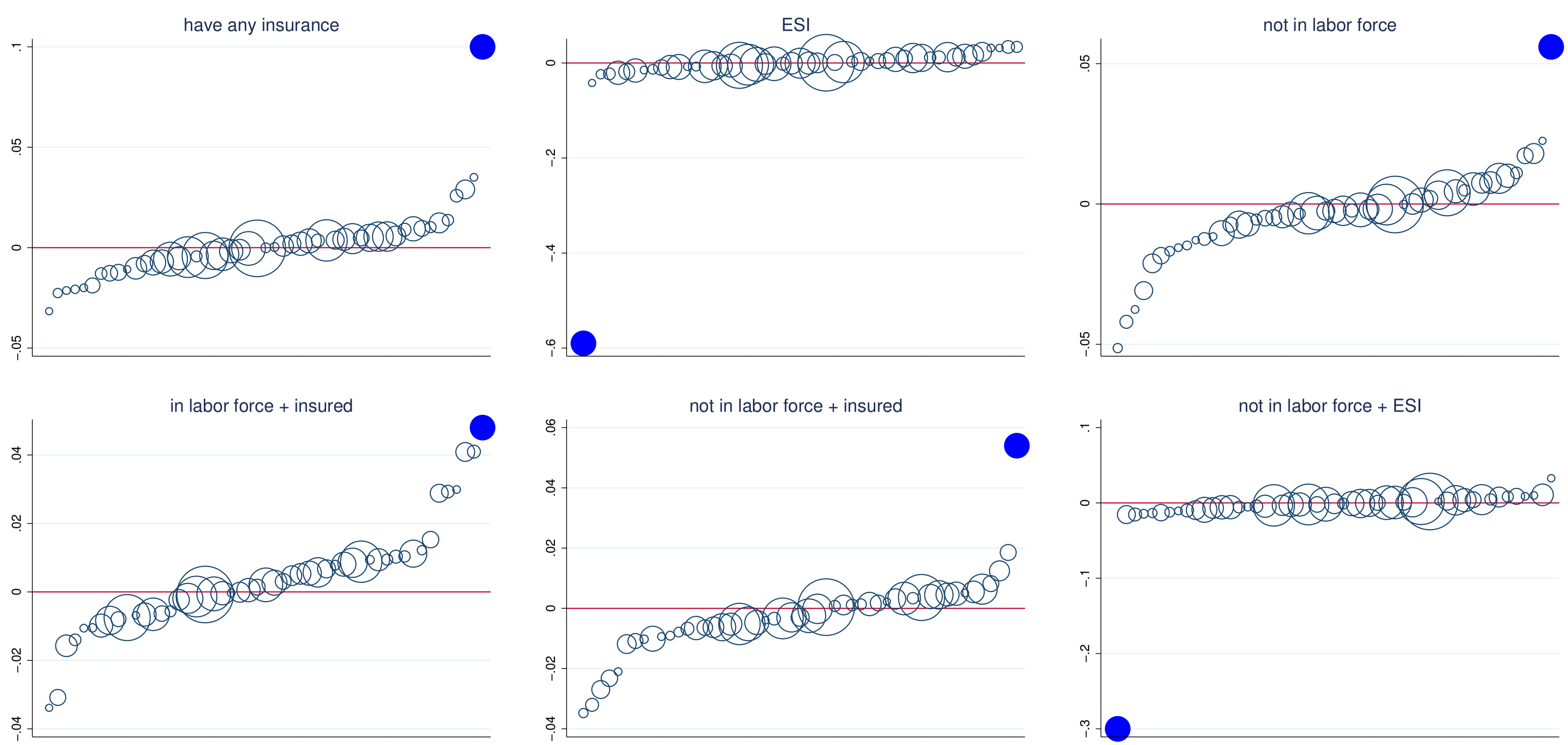
Notes: Table shows reduced-form estimates on outcomes using true policy scheules (Panel A) and the 100 constructed random schedules (Panel B). I show mean statistic of the 100 pseudo estimates, followed by standard deviation in the parenthesis, and the spread between the second smallest and the second largest estimate in square brackets. The construction of random schedules are explained in the main text.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 11 shows pseudo estimates are tightly centered around zero. It follows that after controlling for main effects of demographics and baseline income differences, the scope of omitted factors correlated with group-specifc within-schedule generosity is small.

Finally, since the rest of the nation did not undergo the reform, the remaining 50 states provide naturally occurring placebo samples. Instead of constructing pseudo schedules for each state, I permute actual subsidy generosity in Massachusetts across region-year-age-demographic cells[40]. How likely will one observe the effect size found in Massachusetts in states where no similar reform took place? Figure 3 suggests the probability is fairly small. Even in cases where asymptotic inference suggests insignificant policy impact, the permutation results suggest significant effect at 95% level. Similarly, although estimates using the lean instrument are insignificant, placebo estimates permuted over location, year and age band suggest significant policy impact in Massachusetts (Appendix Figure).

[40] Within a state and year, premium in one of the Massachusetts region and age band is randomly assigned to each PUMA-age-band in placebo states, and affordability is randomly permuted over the 144 demographic groups. That is, affordability schedule in placebo states need not bend at the same income thresholds as in Massachusetts, or increase in income. Nor does insurance premium necessarily increase in age. Similar results hold if I construct random schedules for each state and simulate generosity measures for each schedule, as with the pseudo estimates.

Figure 3: Permutation tests across non-MA states



Notes. Graphs plot placebo estimates from non-MA states, where actual Massachusetts generosity is permuted across location-year-age-demographic cells. Hollow circles show placebo estimates, and solid circles show estimates in Massachusetts. The size of circles reflects the size of 27-64 population in each state. For all outcomes, Massachusetts estimates are the largest in magnitude.

6 Supplementary evidence from program roll-out

To complement main findings based on simulated instruments, this section shows simple difference-in-difference estimates and graphic evidence over the initiation course of the reform. I follow Massachusetts individuals in the Survey of Income and Program Participation (SIPP) since wave 6 (which generally starts in October 2005) till the end of 2007, and use individuals in other Northeastern states as controls. Although the panel does not extend to year 2008 and beyond, and hence has no overlap with the study period in the main analysis, the behavioral response to the initial roll-out is qualitatively similar to the responses found in the following years.

Different provisions of the reform are phased in during 2006-2007. The reform law was passed in April, 2006. In July, 2006, MassHealth expanded coverage to more children. In October, 2006, enrollment in Commonwealth Care began for those below 100% FPL. Starting 2007, the program expanded to the 100%-300% group, who received tiered premium subsidy[41]. Employer mandate took effect in July, 2007. The individual mandate was not effective until 2008: individuals not insured by December 31st, 2007 are subject to a small tax penalty. A larger penalty based on 2008 coverage status became effective in 2009.

Matching the monthly variation in program roll-out, SIPP tracks the monthly employment and insurance outcome of sampled respondents. I use the 2004-2007 panel. Households are sampled every 4 months, at which point they report outcomes in the current month and the preceding three months. The sample rotates and replenishes itself every 4 months, called a wave. A total of 12 waves are in the 2004-2007 panel. I apply longitudinal sampling weights in all analysis of the SIPP sample[42].

Table 12 shows estimates on insurance coverage and ESI based on the following specification:

$$y_{istwr} = \beta_0 + \beta_1 \cdot MA \cdot I\{t \geq July, 2006\} + \alpha_s + \tau_t + \theta_i + \eta_w + \phi_r + \epsilon_{istwr},$$

where time $t = 1, ..., 28$ covers September, 2005 till December, 2007. I construct the post period as after July, 2006, the time the first expansion under

[41] In year 2007, the 100%-150% FPL group is partially subsidized. Since 2008, all individuals below 150% FPL are fully subsidzied.

[42] I use longitudinal weights to maintain constant composition throughout the study period. These weights are positive only for those who stay in the survey throughout the 2004-2007 period, and are designed to reflect any differential attrition predicted by individual characteristics. The weighting is appropriate given the interest in within-individual variation over time, and a wave 9 budget cut (October, 2006-January, 2007) that lost 53% of the original sample.

the law occurred. I include state (α_s), time (τ_t), and individual (θ_i) fixed effects, and integer age dummies in the regression. I use wave (η_w) and reference month (ϕ_r) fixed effects to deal with seam bias across waves. I block bootstrap standard error clustered at the level of state[43], following the procedure in Bertrand, Duflo, and Mullainathan (2004).

Table 12: Difference-in-difference estimates on insurance coverage and ESI, by age

	(1)	(2)	(3)	(4)
	any insurance	ESI	any insurance	ESI
$MA \cdot post$	0.012*	0.0040	-0.0027	-0.017*
	(0.0065)	(0.0071)	(0.0077)	(0.0094)
age group	27-49	27-49	50-64	50-64
R^2	0.0080	0.0064	0.0096	0.0095
N	42,260	42,260	28,806	28,806
# individuals	1,627	1,627	1,099	1,099
# MA individuals	191	191	124	124

Notes: Table shows difference-in-difference estimates on insurance outcomes. All regression includes individual, year-month, state, wave and reference month fixed effects. The *post* variable takes value 1 after July, 2006, the month the first expansion under the reform (MassHealth) took place. Column 1-2 shows effect on the younger group between age 27-49 in the baseline (wave 6), and column 3-4 shows effect on the near-elderly group between age 50-64 in the baseline. I show cross sectional sample size of individuals in the bottom rows. I cluster standard error at the level of state, using the block bootstrap procedure by Bertrand, Duflo, and Mullainathan (2004). Standard errors in the parenthesis come from 500 replication samples. All regressions are weighted by longitudinal sampling weights.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

The small cross-sectional sample size in SIPP implies the analysis is probably under-powered to detect meaningful changes in behavior. Still, there is a marginally significant effect on the insurance take-up in the younger group (age 27-49 in wave 6), and a marginally significant reduction in ESI coverage in the near-elderly[44] (age 50-64 in the baseline). Figure 4 shows coverage of the young increased towards the end of 2007, and was *not* driven

[43]Control states are Connecticut, Delaware, District of Columbia, Maine, Maryland, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. Including Massachusetts, 12 states are represented in the sample.

[44]The effect is not driven by transition into Medicare: restricting to the baseline 50-62 group shows similar results.

by the employer mandate since ESI coverage saw no similar surge over the same period. Rather, the impending mandate penalty seems to have more younger individuals enrolling in subsidized rather than employer-sponsored insurance.

Similar pattern holds for the baseline young and uninsured group (bottom row, Figure 4), where the small increase in ESI appears to account for less than half of the increase in any insurance since July, 2007. However, possibly because the rotating panels thin out eventually, the difference-in-difference estimate (Column 2, Table 13) in fact shows a marginally significant *decrease* in ESI for this group, driven by a jump in ESI coverage after the Commonwealth Care started in October, 2006. Total insurance coverage stayed constant over the same period before the penalty set in. Given the small cross-section sample size, the evidence is inconclusive, but appears to indicate some ESI crowd-out after public expansion, and a smaller ESI crowd-in after the mandate. Once the mandates are in effect, the initial ESI crowd-in fades out. The incentive effect on subsidy enrollment, on the other hand, continues to operate after the mandate, implying a net ESI crowd-out in subsequent years.

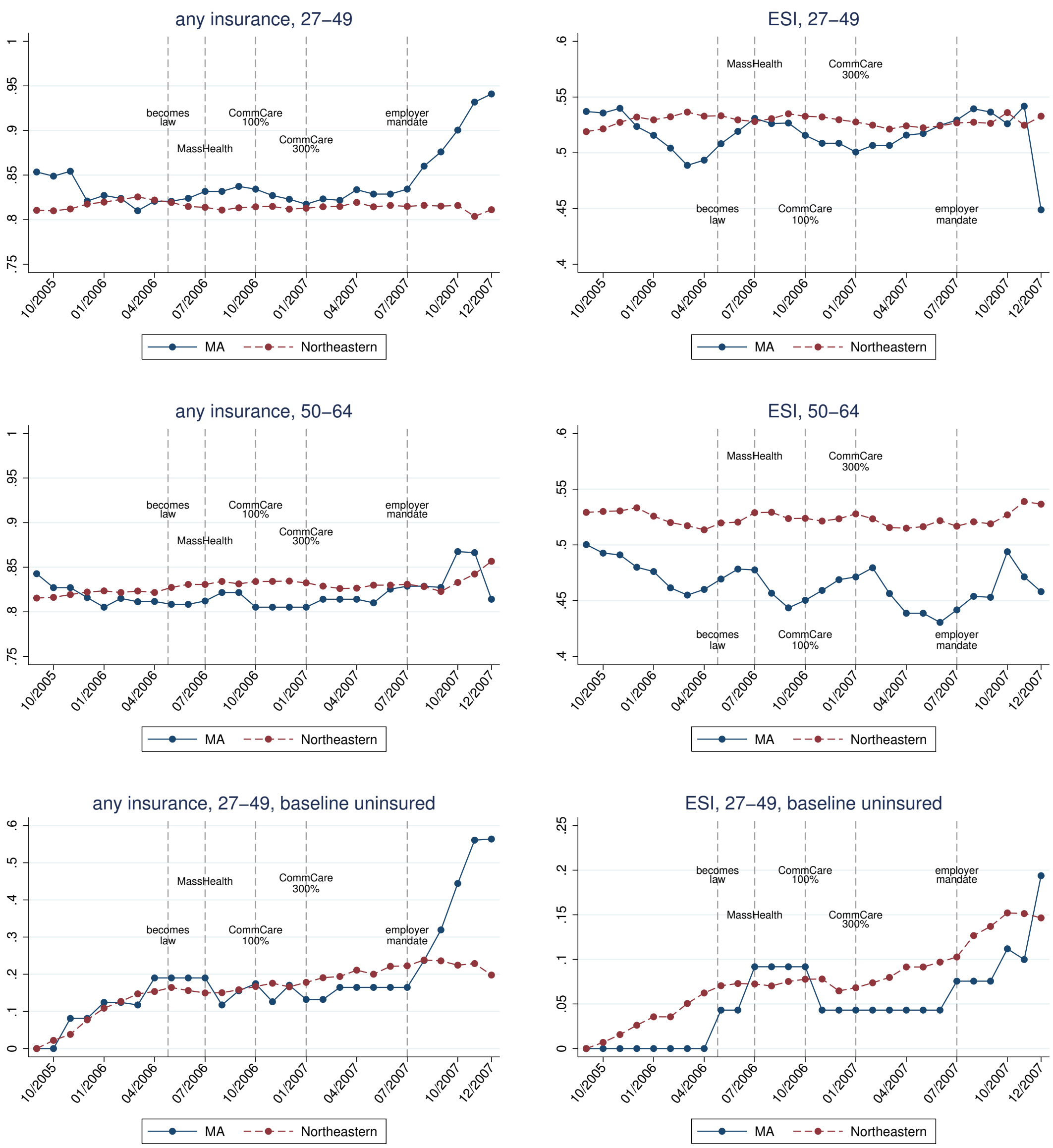
Table 13: Difference-in-difference estimates on insurance and labor, by baseline coverage and age

	(1)	(2)	(3)	(4)	(5)	(6)
	any insurance	ESI	in labor force	any insurance	ESI	in labor force
$MA \cdot post$	0.0013	-0.030*	0.011	0.045***	-0.052***	-0.052***
	(0.025)	(0.016)	(0.028)	(0.012)	(0.016)	(0.012)
age group		27-49			50-64	
baseline		not insured			own-name ESI	
R^2	0.062	0.070	0.062	0.044	0.055	0.039
N	7,504	7,504	7,504	10,876	10,876	10,876
# individuals	289	289	289	414	414	414
# MA individuals	28	28	28	50	50	50

Notes: Table shows difference-in-difference estimates on insurance and labor outcomes, for sub-samples defined by baseline (wave 6) age and insurance status. All regression includes individual, year-month, state, wave and reference month fixed effects. The *post* variable takes value 1 after July, 2006, the month the first expansion under the reform (MassHealth) took place. Column 1-2 looks at the baseline uninsured younger group, and column 3-4 looks at the older group covered by ESI in one's own name in the baseline. I show cross sectional sample size of individuals in the bottom rows. I cluster standard error at the level of state, using the block bootstrap procedure by Bertrand, Duflo, and Mullainathan (2004). Standard errors in the parenthesis come from 500 replication samples. All regressions are weighted by longitudinal sampling weights.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

The marginally significant crowd-out in the near-elderly group is concentrated among workers dropping out of own-name ESI *and* the labor force (Column 4-6, Table 13). I recode the SIPP monthly labor status variable *rmesr* to generate a smooth measure of labor attachment taking eight values

Figure 4: Insurance in Massachusetts and Northeastern states, SIPP trend



Notes. Graphs plot raw trends of insurance coverage and ESI from Sep. 2005 to Dec. 2007 in SIPP. The top row looks at the baseline younger age group. The middle row looks at the baseline near-elderly group. The bottom row looks at the baseline *uninsured* younger group. Longitudinal weights applied in all averages.

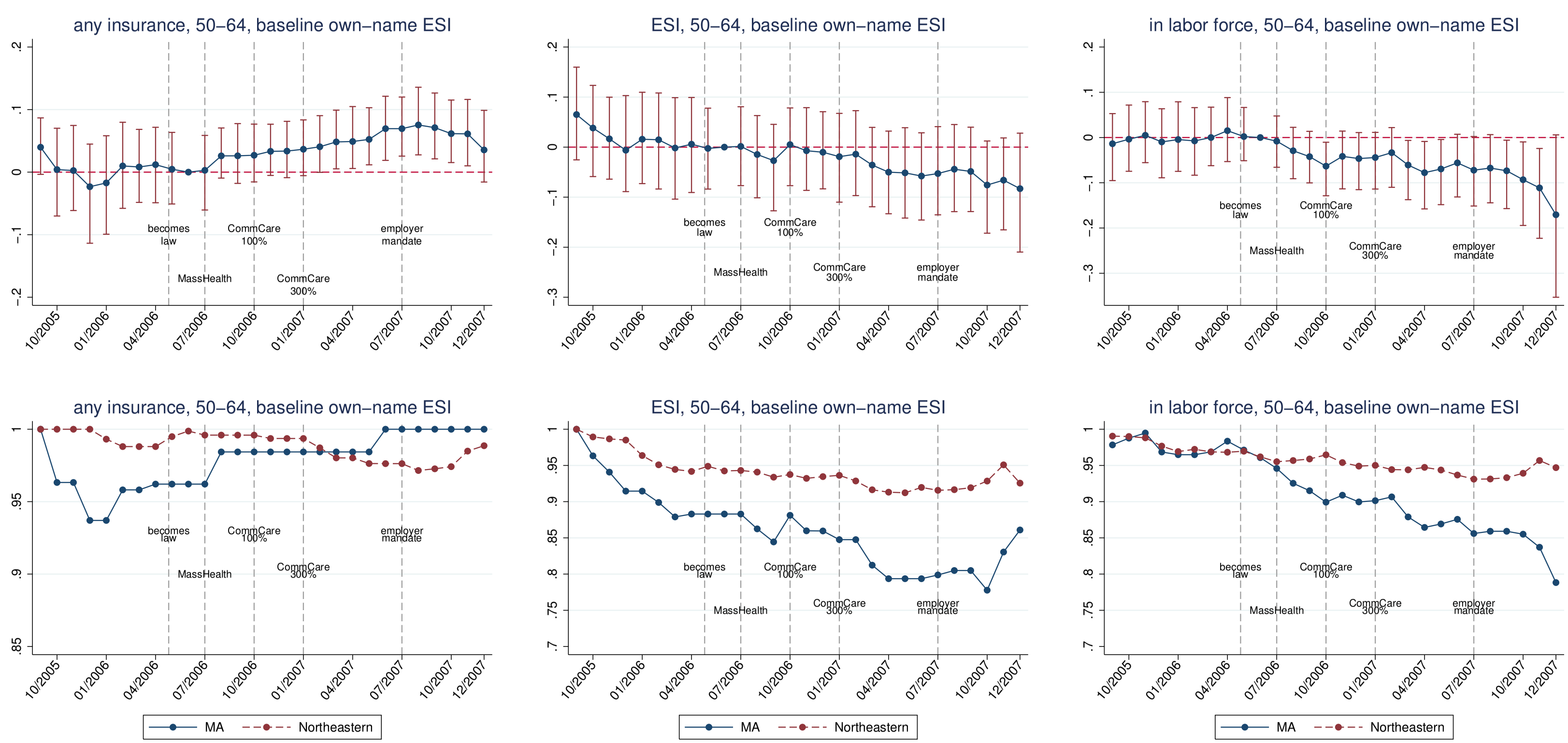
between 0 and 1. 0 applies to individuals with "no job all month, no time on layoff and no time looking for work", and 1 applies to those "with a job entire month, worked all weeks". The six intermediate levels are defined at increment $\frac{1}{7}$. The vast majority of those covered with own-name ESI are employed in the baseline, but trend begins to deviate in Massachusetts after the reform, in particular, after Commonwealth Care started enrollment of the below 100% FPL group, and later of the below 300% FPL group (Figure 5).

Insurance outcomes are noisier in the beginning: although all started out insured by definition, the Massachusetts sample seemed to quickly lose the coverage in the next few months, and remained on a lower parallel trend. After the reform, there was a clear decrease in ESI coverage, and a smaller increase in any insurance, and both effects became stronger after CommCare expanded eligibility to 300% FPL. Absent the small-sample irregularity in the beginning, where ESI already dropped, the implied crowd-out tends to be larger, and the coverage gain smaller.

How does the difference-in-difference estimate based on program roll-out match IV estimates in the main text? Qualitative implication is similar: a significant fraction of ESI drop-outs in the near-elderly also exited labor force. Quantitatively, over the roll-out period, sampled Massachusetts families decreased monthly earned income from \$ 7811 (or annualized 739% FPL adjusting for sub-family size) in June, 2006, to \$ 7528 (691% FPL) in June 2007. Subsidy exposure at the June, 2007 income level was 19.41% according to the 2007 CommCare schedule, and Medicaid covered 1.75% of families with dependent children whose June, 2006 income fell below 133% FPL. The increase in subsidy exposure is 17.66% for this sample. The IV estimates then predict a 3.0%-5.3% reduction in labor[45], and the difference-in-difference estimate of a 4.8% extensive margin exit falls within this range.

[45] Applying age-stratified estimates, the reduction in the 50-54 group is around $0.17 * 17.66\% = 3.00\%$, and in the 55-64 group is around $0.30 * 17.66\% = 5.30\%$.

Figure 5: Insurance selection and labor of near-elderly with own-name ESI, event study estimates



Notes. Graphs plot event study estimates of insurance and labor outcome for the baseline older group covered by ESI in own name. 95% confidence intervals based on 500 replications of block bootstrapped samples are plotted. The resampling occurs at the level of state, following the procedure in Bertrand, Duflo, and Mullainathan (2004). I normalize the effect in June, 2006 to zero. The bottom row shows raw trends of these outcomes. All estimates are adjusted by SIPP longitudinal weights.

ESI reduction, on the other hand, is predicted to be around 9.7%-11.3% according to IV estimates. In the difference-in-difference setup, since all individuals are covered in own name in the baseline, if one interprets the 4.5% coverage gain in Massachusetts as reflecting the bias from ESI loss shortly after the baseline but before the reform, then the adjusted ESI crowd-out rounds up to around 10%, again within the range of difference-in-difference estimates. Using the adjusted estimate, around 50% of ESI drop-outs in the 50-64 group exited labor force. Inferred rate from IV estimates ranges from 31% in the 50-54 group to 46% in the 55-64 group.

7 Calibration

The welfare calculation additionally depends on measures of utility over consumption ($c_{i,j}$ and risk preference parameter γ), health care utilization rate ($\bar{\mu}_e$, $\bar{\mu}_{1-e}$), and pricing variation due to new insurance take-up (ϵ_{r,λ_0}, ϵ_{ri,λ_0}). The census data do not contain relevant information for these measures. Instead, I resort to alternative data sources and extrapolate prior estimates to calibrate these statistics.

Consumption I use Consumption and Activity Mail-out Survey (CAMS) of the Health and Retirement Study (HRS) to calibrate consumption by employment and health care utilization status. CAMS contains detailed consumption expenditures for households over the age 50 who are also sampled in the main HRS surveys where labor and health care utilization information is recorded. Direct measurement of consumption pattern for the full population is less available. To proceed, I assume that younger groups vary consumption similarly as older groups, but differ in the utilization rate of health care, which I later characterize from the National Health Interview Survey (NHIS).

In HRS-CAMS, I focus on a sample of 55-64 year old households with no co-residing children, and divide consumption evenly between spouses. Most of these households are covered by health insurance. To better approach the high coverage rate under a mandate, I further focus on households where all members are covered by insurance. A household is working if either spouse is working full-time, and not working if neither is employed when sampled. To determine utilization status, I look at number of nights household members spent in hospital or nursing home in the past two years. About 10 percent of the households in HRS have at least 4 institution nights over a two year period, and I assign utilization status 1 to this sub-sample[46].

[46]The cut-off is based on the calculation that for couple households, the implied per

Table 14: Consumption calibration

Non-Durable Consumption:					
mean	no utiliz.	utiliz.	median	no utiliz.	utiliz.
work	13,609	9,798	work	11,250	8,948
no work	10,224	8,832	no work	9,048	7,996
Food Consumption:					
mean	no utiliz.	utiliz.	median	no utiliz.	utiliz.
work	4,365	3,392	work	3,868	2,987
no work	3,430	3,344	no work	3,207	2,630

Notes: Author calculation from HRS-CAMS. All dollars are in 2005 value. Details are in the main text.

I measure consumption with expenditures on non-durable goods and housekeeping activities such as laundry. Similar patterns hold for food consumption only. Table 14 shows mean and median comsumption across employment-utilization cells. All dollars are indexed to 2005 value. Median consumption is lower than the mean, but similar patterns across cells appear. To avoid excessively large consumption in the right tail, I focus on medians. From either non-durable consumption or food consumption, it follows that $\frac{c_{1,0}}{c_{1,1}} = \frac{1}{0.8}$, $\frac{c_{1,0}}{c_{0,0}} = \frac{1}{0.8}$, and $\frac{c_{1,0}}{c_{0,1}} = \frac{1}{0.7}$. I assume this consumption pattern applies to the full population.

Risk preference Following standard practice, I assume CRRA utility over consumption, calibrating the risk coefficient γ at 2. There is no general consensus on the appropriate range of the risk coefficient in the retirement context, with the estimated value close to 1 in Rust and Phelan (1997) and around 5 in French and Jones (2011). Chetty (2006b) places an upper bound at $\gamma = 2$, beyond which the expected utility model may not be consistent with observed substitution patterns between wage and leisure. Both larger curvature in the utility function and consumption disparity between states increase the welfare gain from social transfer. I later assess the sensitivity of welfare calculation to alternative values of risk preference and consumption patterns.

person per year institution night is at least 1. Individual data in NHIS similarly give a 10.3% hospitalization rate on the extensive margin for the 55-64 group in the 2005-2011 national sample. It is also close to the 10.6% inpatient usage rate for the Massachusetts 45-64 age group in 2012 (CHIA).

Utilization rate I characterize utilization pattern by employment in the average population (27-64) from the NHIS, 2005-2011. I measure utilization by the extensive margin of hospitalization in the past 12 months. Utilization rate for the 55-64 age group in NHIS is 10.3%, nearly identical to the rate derived from HRS. I determine employment status over the same recall period combining labor outcomes in the past 2 weeks and in the past 12 months[47]. Average utilization rate is 8% in the 27-64 group. Therefore $\frac{M-C}{r(\lambda_0)} = \frac{1}{0.08}$. Subsetting by employment, utilization is 6.2% among workers, and 15.1% among non-workers: $\bar{\mu}_e = 0.062$, $\bar{\mu}_{1-e} = 0.151$.

Pricing variation and adverse selection I assume insurance pricing is equal to the expected cost of enrollees, and different risk types choose between formal and informal coverage in response to the subsidy schedule. When the formal insurance market is adversely selected, marginal new enrollees in subsidized program have lower expected cost than existing enrollees, lowering the aggregate premium of formal insurance. On the contrary, a marginal disenrollee from formal insurance has larger expected cost of uncompensated care, increasing the implicit premium of informal insurance. Hence with adverse selection, both ϵ_{r,λ_0} and ϵ_{ri,λ_0} are signed positive.

To pin down the magnitude of the pricing elasticity, one needs information on the cost type of marginal and infra-marginal enrollees in formal and informal insurance. The cost of marginal enrollees is estimated in Finkelstein, Hendrean and Shepard (2017) based on the subsidy varation across income thresholds on the CommonwealthCare program[48]. I then determine how much of the average cost of risk pool is driven by movements on the margin.

For example, in year 2011, at the 150% FPL threshold, average cost is lower to the left by \$47 (\$333 to the left and \$380 to the right), and enrollment is higher to the left by 24 percentage points (94% of the eligible population enrolled on the left, versus 70% on the right). Because total cost of infra-marginal enrollees stays constant on either side, the average cost of marginal enrollees who dropped out at lower subsidy is recovered as $\frac{\$333*0.94-\$380*0.70}{0.94-0.70} = \$195.92$. 1054 persons disenrolled, corresponding to an increase in total uninsurance rate of 0.024%[49]. Average cost of formal

[47]The two variables recover employment history in the past 2 weeks to 12 months. I assign non-employment status to individuals who have not been employed in the past 12 months.

[48]I assume extensive margin coverage loss only involves exit from the subsidized program, which is plausible given the last-resort nature of the program for the insured.

[49]This is measured against the full population count for the Massachusetts 19-64 group, estimated to be 4,302,131 from ACS. Of this group, insurance rate in 2011 is 94%.

insurance enrollees is higher by $\frac{94*\$400-0.024*\$195.92}{94-0.024} - \$400 = \$0.052$, since 94% of the population is insured at average cost $400. The implied elasticity based on disenrollment at this threshold is given by $\epsilon_{r,\lambda_0}^{150\%FPL} = \frac{\$0.052}{0.024}\frac{6}{\$400} = 0.033$.

Similar calculation around alternative thresholds shows even smaller elasticity: $\epsilon_{r,\lambda_0}^{200\%FPL} = 0.021$, $\epsilon_{r,\lambda_0}^{250\%FPL} = 0.019$. Averaging over the three thresholds, I quantify $\epsilon_{r,\lambda_0} = 0.024$. Because exit on the margin is very small compared to the bulk of infra-marginal enrollees, its impact on formal insurance pricing is also negligible.

Formal insurance exits matter more for informal insurance pricing. In my model, informal insurance requires the same copay C as formal insurance, but differs in viability: it only covers a fraction j of all medical incidences. The assumption of a copay reflects the fact that the uninsured typically pays only 20% of the total medical cost, which is approximately the co-insurance rate in formal insurance. To quantify j, I assume that marginal enrollees into formal insurance tend to have higher expected cost compared to those who remain uninsured, but because the pool of informal insurance is small (6% in 2011), average cost does not substantially differ from marginal cost. Observed average cost in the uncompensated care pool is $136[50]. Adjusting for the fact that we only observe fraction j of the cost, true average cost $\frac{\$136}{j}$ should be close to and smaller than the marginal cost recovered in Finkelstein, Hendrean and Shepard (2017). I quantify $j = 0.8$.

At $j = 0.8$, the uninsured has average cost $170, smaller than the cost of the marginal uninsured ($196 – $281). Essentially, I apply the same model of uncompensated care cost as in Finkelstein, Hendrean and Shepard (2017), where the uninsured is assumed to pay 20% of the total cost incurred, and reduce utilization by 20% compared to the formally insured. Alternative calibration, such as $j = 0.7$, also satisfies the requirement that average cost stays below marginal cost, but the difference is smaller. I examine the sensitivity of welfare analysis to different calibration of uncompensated care usage.

At 150% FPL, CommCare exits inflate the implicit premium of informal insurance by $\frac{6*\$170+0.024*\$195.92}{6+0.024} - \$170 = \$0.10$. The resulting elasticity is $\epsilon_{ri,\lambda_o}^{150\%FPL} = \frac{\$0.10}{0.024}\frac{6}{\$170} = 0.15$. Similarly, $\epsilon_{ri,\lambda_o}^{200\%FPL} = 0.24$, and $\epsilon_{ri,\lambda_o}^{250\%FPL} = 0.65$. On average, $\epsilon_{ri,\lambda_o} = 0.35$.

[50] Total program cost in the Health Safety Net, formerly known as the Uncompensated Care Pool, is $420 million in 2011. Distributed over 12 months and $0.06 * 4,302,131 = 258,128$ uninsured, average recorded cost per uninsured is $136.

8 Welfare

8.1 Turning IV estimates to sufficient statistics

Some of the sufficient statistics in the welfare formula are insurance choice probabilities conditional on an employment outcome. As individuals adjust employment to a given set of prices and probabilities, the equilibrium distribution of insurance by employment also changes in response to policy. I use estimated statistics on employment and the joint insurance-employment outcome to quantify change rate in the conditional probability: $\frac{d\lambda_{1-e,i}}{d\lambda_p} = [\frac{d(1-e)\lambda_{1-e,i}}{d\lambda_p} - \lambda_{1-e,i}\frac{d(1-e)}{d\lambda_p}]/(1-e) = -[\frac{d(1-e)\lambda_{1-e,1-i}}{d\lambda_p} - \lambda_{1-e,1-i}\frac{d(1-e)}{d\lambda_p}]/(1-e)$, $i = 0, 1$. Both equations give very similar results.

Moreover, labor elasticity ϵ_{1-e,λ_p} depends on the marginal change rate in employment, *and* the average subsidy generosity facing marginal enrollees. Who are the marginal enrollees to receive subsidy in this context? Since the privately insured are not entitled to subsidy, the size of the eligible population is given by $1 - \lambda_1$. I hence quantify the generosity parameter based on the eligible population in Massachusetts: $\lambda_p = 0.68$.

When an ESI enrollee selects into subsidized insurance, he faces an average subsidy generosity of 68%[51]. The implied reduction in labor is given by $\frac{d(1-e)}{d\lambda_p} \cdot \lambda_p = 3.88\%$. It represents a 23% increase from the 17% baseline non-participation rate, or a 17% increase from the 23% baseline non-employment rate. Therefore the labor elasiticity changes slightly with the margin of employment considered, even though the change rate $\frac{d(1-e)}{d\lambda_p}$ is almost identical on both margins. In the main analysis I focus on the participation margin, and check the robustness of results on the employment margin.

Appendix Table lists all sufficient statistics and their quantified values. Based on these values, I calculate the social cost and benefits of a dollar expansion of subsidy as specified in the welfare formula.

8.2 Welfare calculation

Table 15 shows the efficiency gain in insurance pricing: new enrollment in formal insurance lowers premium on both the formal and informal market, which amouts to a welfare gain of \$0.27 per dollar subsidy. About \$0.21

[51] In principle, new subsidy enrollees come from either the ESI pool or the uninsurance pool. However, as net coverage gain is small, and the uninsurance pool is also small, average subsidy mostly reflects generosity to potential ESI drop-outs.

comes from the formal insurance market (column 1), which is around 19 times larger than the informal market. The welfare calculation partly reflects this size difference. Column (3) normalizes the premium payment made by the non-employed into units of worker wage, the adopted dollar denomination in the welfare metric. Quantitatively, the adjustment is inconsequential at the range of risk coefficients considered.

Table 16 summarizes welfare gain for marginal and infra-marginal enrollees. Because formal insurance provides additional risk protection relative to informal insuance, it is valued by marginal enrollees at $I'(\lambda_p; j) = \$1.81$ per dollar subsidy ($\gamma = 2$), and the value increases with risk aversion. New insurance incurs new payment either by enrollees themselves or by net payers of subsidy. The asociated premium cost $P'(\lambda_p; j, k)$ is small compared to the risk protection value, resulting in a net welfare gain on the margin of $W'(\lambda_p; j, k) = \$1.18$ per dollar subsidy.

Table 15: Efficiency gain in insurance pricing

	(1)	(2)	(3)	(4)
	formal insurance	informal insurance	transfer adjustment	$P(\lambda_p; j, k)$
γ				
1	0.21	0.057	0.0016	0.27
2	0.21	0.057	0.0036	0.27
3	0.21	0.057	0.0061	0.27

Notes: Table shows the dollar value of efficiency gain in insurance pricing, charactered in $P(\lambda_p; j, k)$ in the welfare formula. Column (1) characterizes the efficiency gain in formal insurance pricing, or $-\frac{1-\lambda_0}{\lambda_2}\frac{\epsilon_{r,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p}$. Column (2) characterizes the efficiency gain in the implicit pricing of informal insurance, or $-j\frac{\lambda_0}{\lambda_2}\frac{ri(\lambda_0)}{r(\lambda_0)}\frac{\epsilon_{ri,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p}$. Column (3) characterizes the adjustment that normalizes own premium payment made by the non-employed into dollars of worker wage, the adopted denomination in the welfare metric. Specifically, column (3) quantifies $-\frac{1-e}{\lambda_2}\overline{\lambda_{1-e,2}}\left[\frac{E_{\tilde{\mu}_{1-e}}u'(c_{0\cdot})}{E_{\tilde{\mu}_e}u'(c_{1\cdot})} - 1\right](1-\lambda_p)\frac{\epsilon_{r,\lambda_0}}{\lambda_0}\frac{d\lambda_0}{d\lambda_p}$. I show results separately for relative risk aversion coefficients ranging from 1 to 3.

Infra-marginal enrollees benefit from subsidy expansion, because insurance is made more affordable in the low-income state. Subsidy to the non-employed therefore provides some financial protection for insurance affordability against income shocks, allowing recipients to remain insured against large medical expenditure shocks. The continued risk protection in the low-income state is valued by subsidy enrollees at $\Delta W(\lambda_p; j, k) = \$0.25$. Combining new and existing subsidy enrollees, a dollar subsidy is valued at \$1.43.

Finally, Table 17 shows the social cost of subsidy financing (column 4) and the net return of subsidy dollars (column 5), summing over benefits (column 1-3) and cost. The social cost is mostly driven by pre-existing unemployment insurance (UI) that already transfers a fair amount of wage earning to the non-employed state[52]. Essentially, the UI payment broadly represents many other transfer payments from safety net programs which support a consumption floor for the non-employed. In particualr, because of the Inada condition, when charity care is not available, which is relevant for $(1-j)M = \$1,000$ of the medical cost, transfer from alternative programs and implicit insurance should recover comsumption from the medical loss. This would suggest an A of at least $\$1,000$[53]. A large substitution ratio $\frac{A}{r(\lambda_0)} = 3.97$ then magnifies the moral hazard cost.

Another way to interpret $\frac{A}{r(\lambda_0)}$ is to consider the possibility that *effectual* uninsurance in the low-income population may be smaller than reported. Enrollment in Medicaid and other public insurance programs is incomplete. Instead, health care providers may enroll eligible patients in programs that pay for the cost incurred. In these cases, coverage is effectual and retrospective, and no uncompensated care is generated once the eligible individual is enrolled. The new subsidy dollar needed for reported coverage gain may be partly replacing existing dollars already implicitly insuring eligible individuals[54]. Shifting focus to *effectual* coverage tends to lower A and hence the transfer cost, although how to measure effectual coverage and its change rate to subsidy is not entirely clear.

Nonetheless, even if one counts effectual enrollees as "new" enrollees, and hence by construction lowers transfer efficiency, the net return on a dollar subsidy expansion is still positive at modest risk aversion ($\gamma = 2$): return to one more dollar of subsidy is valued at \$0.14. The vast majority of the return accrues to new enrollees on the margin, and the crowding out of resources already implicitly insuring the same people worsens the social cost of transfer financing. Other social considerations, however, more than compensate for the cost. Efficiency gain in insurance pricing benefits all enrollee, and premium assistance to the low-income benefits all subsidy enrollees. Put together, subsidy as is currently administered improves social

[52] According to UI claims data in the Department of Labor Unemployment Insurance Financial Data Handbook, weekly benefit averages \$397 in Massachusetts during 2008-2011, which is over 30% of wage earning.

[53] I calibrate $A = \$1589 = 4 * \$397$, or the monthly average UI benefit.

[54] Although a mandate tends to incentivize program enrollment, in Massachusetts, individuals with income below 150% FPL are exempt from uninsurance penalty. Although almost all qualify for public insurance, reported coverage is only 88% in the 27-64 group.

welfare, and at higher risk aversion, there is scope for further expansion of the program.

Table 16: Benefits to enrollees, marginal and infra-marginal

	(1)	(2)	(3)	(4)	(5)
	$I'(\lambda_p; j)$	$P'(\lambda_p; j,k)$	$W'(\lambda_p; j,k)$ $= I'(\lambda_p; j) + P'(\lambda_p; j,k)$	$\Delta W(\lambda_p; j,k)$	$W' + \Delta W$
γ					
1	1.47	-0.46	1.01	0.11	1.12
2	1.81	-0.62	1.18	0.25	1.43
3	2.21	-0.82	1.38	0.42	1.80

Table 17: Welfare benefits and cost of subsidy

	(1)	(2)	(3)	(4)	(5)
	$P(\lambda_p; j,k)$	$W'(\lambda_p; j,k)$	$\Delta W(\lambda_p; j,k)$	$MH(\lambda_p; j,k)$	$WM(\widetilde{\lambda_p}; j,k)$ $= P + W' + \Delta W + MH$
γ					
1	0.27	1.01	0.11	-1.56	-0.17
2	0.27	1.18	0.25	-1.56	0.14
3	0.27	1.38	0.42	-1.56	0.51

9 Conclusion

References

Abaluck, Jason, and Jonathan Gruber. 2011. "Heterogeneity in Choice Inconsistencies among the Elderly: Evidence from Prescription Drug Plan Choice." American Economic Review 101 (3): 377-81.

Aizawa, Naoki. 2016. "Health Insurance Exchange Design in an Empirical Equilibrium Labor Market Model." https://sites.google.com/site/naokiaizawaweb/research.

Aizawa, Naoki, and Hanming Fang. 2015. "Equilibrium Labor Market Search and Health Insurance Reform." https://sites.google.com/site/naokiaizawaweb/research.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." Journal of Economic Perspectives 27 (1): 197-222.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." Quarterly Journal of Economics 130 (4): 1623-67.

Baily, Martin Neil. 1978. "Some Aspects of Optimal Unemployment Insurance." Journal of Public Economics 10 (3): 379-402.

Chetty, Raj. 2006a. "A General Formula for the Optimal Level of Social Insurance." Journal of Public Economics 90 (10): 1879-1901.

Chetty, Raj. 2006b. "A New Method of Estimating Risk Aversion." The American Economic Review 96 (5): 1821-1834.

Chetty, Raj, and Amy Finkelstein. 2013. "Social Insurance: Connecting Theory to Data." Handbook of Public Economics 5: 111.

Chetty, Raj, and Emmanuel Saez. 2010. "Optimal Taxation and Social Insurance with Endogenous Private Insurance." American Economic Journal: Economic Policy 2 (2): 85-114.

Currie, Janet, and Jonathan Gruber. 1996a. "Health Insurance Eligibility and Child Health: Lessons from Recent Expansions of the Medicaid Program." Quarterly Journal of Economics 111 (2): 431-466.

Currie, Janet, and Jonathan Gruber. 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, David M., and Jonathan Gruber. 1996. "Does Public Insurance Crowd Out Private Insurance?" Quarterly Journal of Economics 111 (2): 391-430.

Cochrane, John H. 1991. "A Simple Test of Consumption Insurance." Journal of Political Economy 99 (5): 957-76.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2010. "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the UK Annuity Market." Econometrica 78 (3): 1031-92.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." American Economic Review 103 (1): 178-219.

Ericson, Keith M. Marzilli, and Amanda Starc. 2015. "Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange." Review of Economics and Statistics 97 (3): 667-82.

Finkelstein, A., N. Hendren, and M. Shepard (2017): "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," NBER Working Paper No. 23668

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers. "Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act." Journal of Health Economics 53 (2017): 72-86.

French, Eric, and John Bailey Jones. 2011. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." Econometrica 79 (3): 693-732.

Graves, John A., and Jonathan Gruber. 2012. "How Did Health Care Reform in Massachusetts Impact Insurance Premiums?." The American Economic Review 102 (3): 508-513.

Gruber, Jonathan. 1997. "The Consumption Smoothing Benefits of Unemployment Insurance." American Economic Review 87 (1): 192-205.

Gruber, Jonathan, and Brigitte C. Madrian. 1997. "Employment Separation and Health Insurance Coverage." Journal of Public Economics 66 (3): 349-82.

Gruber, Jonathan, and Kosali Simon. 2008. "Crowd-out 10 Years Later: Have Recent Public Insurance Expansions Crowded Out Private Health Insurance?" Journal of Health Economics 27 (2): 201-17.

Hackmann, Martin B., Jonathan T. Kolstad, and Amanda E. Kowalski. 2015. "Adverse Selection and an Individual Mandate: When Theory Meets Practice." American Economic Review 105 (3): 1030-66.

Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." American Economic Review 103 (7): 2643-82.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." American Economic Review 105 (8): 2449-2500.

Heim, Bradley T., Gillian Hunter, Adam Isen, Ithai Lurie, and Shanthi Ramnath. 2016. "Income Responses to the Affordable Care Act: Evidence from the Premium Tax Credit Cliff." https://ashecon.confex.com/ashecon/2016/webprogram/Paper4556.html

Heim, Bradley T., and Ithai Z. Lurie. 2015. "The Impact of Health Reform on Job Mobility: Evidence from Massachusetts." American Journal of Health Economics 1 (3): 374-398.

Kolstad, Jonathan T., and Amanda E. Kowalski. 2012. "The Impact of Health Care Reform on Hospital and Preventive Care: Evidence from Massachusetts." Journal of Public Economics 96 (11): 909-929.

Kolstad, Jonathan T., and Amanda E. Kowalski. 2016. "Mandate-based Health Reform and the Labor Market: Evidence from the Massachusetts Reform." Journal of Health Economics 47: 81-106.

Long, Sharon K., Karen Stockley, and Alshadye Yemane. 2009. "Another Look at the Impacts of Health Reform in Massachusetts: Evidence Using New Data and a Stronger Model." The American Economic Review 99 (2): 508-511.

Madrian, Brigitte C. 1994. "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?" Quarterly Journal of Economics 109 (1): 27-54.

Mahoney, Neale. 2015. "Bankruptcy as Implicit Health Insurance." American Economic Review 105 (2): 710-46.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." American Economic Review 77 (3): 251-77.

Rust, John, and Christopher Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets." Econometrica 65 (4): 781-831.

Shi, Julie. 2016. "Income Responses to Health Insurance Subsidies: Evidence from Massachusetts." American Journal of Health Economics 2 (1): 96-124.

Summers, Lawrence H. 1989. "Some Simple Economics of Mandated Benefits." The American Economic Review 79 (2): 177-183.

Tebaldi, Pietro. 2015. "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA." http://dx.doi.org/10.2139/ssrn.2688894.

A Appendix

A.1 TBD