Dynamic effects of health insurance reform

Jan Boone[*]

June 7, 2025

Abstract

In order to find the long term mortality effects of health policy, we estimate a structural dynamic model following age cohorts over time. This allows us to follow the development in the fraction of people with low health status per category defined by region, age, gender and income. The fraction of people that are unhealthy/ill acts as a stock determining the long term effects of health policy. The model is estimated on regional (NUTS 2) Eurostat data and replicates stylized facts like women and people on high income have more quality adjusted life-years (qaly's). Further, we show that an increase in demand-side cost-sharing does reduce qaly's. In contrast to earlier papers, we find that the effect of cost-sharing on low incomes' qaly's is bigger than for high incomes. Our main finding is that the dynamic effect from following an age cohort from 45 till 84 is far bigger than a static estimate (like difference-in-differences or regression discontinuity design) of the mortality effect would suggest.

JEL codes: I11, I13, I18

Keywords: out-of-pocket payments, mortality, health insurance, poverty, unmet medical needs

---

[*]Tilburg University, Department of Economics, Tilec and CEPR, E-mail: *j.boone@uvt.nl*.

1   Introduction

Healthcare policy is long run policy. Policy reforms can affect a population's health in 20 years time. A recent literature analyzes the health effects of demand-side cost-sharing where health is proxied by mortality. To identify the causal effect of demand-side cost-sharing on mortality, these papers focus on a narrow band of years and/or ages around a change in cost-sharing. To illustrate, a number of studies use the Medicaid eligibility expansion under the Affordable Care Act (Borgschulte and Vogler 2020; Miller, Johnson, and Wherry 2021). As this was implemented over time in a number of states, a difference in differences (diff-in-diff) analysis can be used to identify the causal effect. To facilitate the identification only a few years after the policy change are analyzed making it hard to understand long run effects.

We introduce a simple dynamic health model that allows us to follow a cohort from age 45 to 84. The idea is to capture a "health stock" by distinguishing people with high and low health status. We keep track of health status at the age/gender/income/year/region level. That is, we follow the health status of an age category (per gender/income/region) over time as the cohort gets older. For the healthy group there is a probability of falling ill. For the low health status group there are a probability to recover and a probability to die. The recovery probability depends on the probability of getting treatment which in turn depends on the degree of demand-side cost-sharing. With higher out-of-pocket expenditure, fewer people can afford treatment and recovery is less likely.

In such a theoretical model it is straightforward to choose parameter values such that the dynamic effect is far bigger than a static diff-in-diff estimate would suggest. In order to evaluate this difference for reasonably realistic parameter values, we estimate our model on regional (NUTS 2) Eurostat data over an eleven year period. The model allows us to derive the effect of demand-side cost-sharing on mortality and quality adjusted life years (so called qaly's) which give a higher weight to years in full health than years with low health status.

Our estimated model fits the data well and replicates a number of stylized facts. Women expect more qaly's than men and people on high income more than people in poverty. The probability of falling ill increases with age. Although there is uncertainty about the parameter estimates, there is an unambiguous ordering of countries in terms of expected qaly's for 45 year olds; Switzerland has the highest expected qaly's and Hungary, Bulgaria and Lithuania the lowest. Finally, a 10% point increase in demand-side cost-sharing (that is, in the percentage of out-of-pocket over total

healthcare expenditure) reduces qaly's: the average effect is moderate but clearly bounded away from zero.

Two results stand out compared to the literature discussed below. First, the effect of a change in demand-side cost-sharing on low income qaly's is ten times the effect on high income qaly's. Second, the dynamic effects of demand-side cost-sharing for low incomes are twenty times higher than the static estimate that would follow from a diff-in-diff analysis for this group. The latter result suggests a trade off between a clear-cut causal identification –based on, say a diff-in-diff methodology– and a long run effect estimate –if one is willing to put more structure on the data.

We estimate the model as a Bayesian model to easily propagate the uncertainty of the estimated parameters to the results that we present, e.g. in terms of qaly's.

This paper is not the first to examine the impact of demand-side cost-sharing on mortality. There is a collection of recent studies employing innovative methodologies and primarily relying on individual-level data to establish the causal effect of health insurance on health and mortality. Several studies have utilized the Medicaid eligibility expansion under the Affordable Care Act. This expansion was implemented at various times across different states in the US, enabling a diff-in-diff identification strategy. These studies have demonstrated that the Medicaid expansion (resulting in more comprehensive health insurance coverage) has led to a reduction in mortality rates (Borgschulte and Vogler 2020; Miller, Johnson, and Wherry 2021).

However, these studies focus on the static or short run effects of insurance and demand-side cost-sharing. To illustrate, Borgschulte and Vogler (2020) consider a four year period after the Medicaid reform and do not try to identify the dynamic effect. Miller, Johnson, and Wherry (2021) focuses on 55-64 year olds up to three years after the Medicaid reform.

Chandra, Flack, and Obermeyer (2021) analyze the Medicare Part D prescription drug coverage, in which end-of-year pricing displays a non-linear pattern based on expenditure. The primary finding indicates that increases in out-of-pocket (oop) costs for drugs result in reduced drug use, including the use of high-value treatments, subsequently leading to higher mortality rates. There is no dynamic analysis in the sense that the paper only considers 65 year olds and focuses on within year effects.

Finally, Goldin and colleagues conducted an experimental study in which individuals subject to the Affordable Care Act's health insurance mandate were reminded of potential financial penalties for non-compliance. This reminder prompted individuals to opt for health insurance instead of remaining uninsured, and as a result, mortality rates were lower among those who received the reminder compared to the

control group who did not (Goldin, Lurie, and McCubbin 2020). This paper analyzes a two year period and focuses on 45-64 year olds, whereas our paper derives effects across 45-84 year olds.

Our paper adds to this evidence of negative health effects of demand-side cost-sharing in the following dimensions. First, we utilize European instead of US data. European countries tend to have a more homogeneous health insurance system compared to the diverse range of options available within the US. In the US, individuals may have (very generous) employer-sponsored insurance, Medicaid or Medicare coverage, or no insurance at all, making it challenging to detect aggregate-level effects of changes in say, Medicaid coverage. European countries, in contrast, tend to have nationally determined health insurance features, resulting in a higher level of consistency. For instance, the OECD Health Systems Characteristics Survey shows that more than 90% of the population in European countries obtains primary healthcare coverage through automatic or compulsory insurance, with percentages exceeding 99% in most cases. The corresponding figure for the US is less than one third. In this sense, country or region-wide statistics in Europe provide a better representation of the insurance situation for citizens compared to the US, although they may not capture all individual nuances such as the purchase of complementary insurance.

With this data we follow an age/gender/income/region cohort over time and estimate the model parameters capturing the probabilities of falling ill, of recovering from illness and death. Then we can simulate the long term effects of a change in demand-side cost-sharing using our estimated parameters. Using this simulated data we can compare our dynamic results with the static (diff-in-diff) estimates. We find that the (true) dynamic effects are magnitudes higher for low incomes than the static estimates would suggest. We illustrate this by using the estimation method in Finkelstein and McKnight (2008) to explain the difference in detail.

Our finding that demand-side cost-sharing affects health and mortality of people on low income aligns with previous research indicating that healthcare utilization is influenced by individuals' liquidity constraints. Individuals with lower incomes tend to defer or forgo valuable treatments when these are expensive (Gross, Layton, and Prinz 2020; Nyman 2003). Individuals with higher incomes may also forgo necessary treatments (Brot-Goldberg et al. 2017; Chandra, Flack, and Obermeyer 2021). However, in the latter case, the decision to forgo treatment is more likely driven by factors other than liquidity issues and we find a far smaller effect.

Further, Eurostat variables derived from the EU-SILC survey enable us to concentrate on causal mechanisms. The survey includes questions about unmet medical

needs in the past months and the reasons for these unmet needs. One of the reasons cited is the out-of-pocket cost, which leads individuals to postpone or forgo treatment because it is too expensive.

The next section presents a dynamic model of health status and mortality. Then we describe the Eurostat data that we use. We explain the empirical model that we estimate. Estimation results are presented and we conclude with a discussion of the policy implications. The appendix contains more details on our data and estimation. The online appenix is the html version of this paper which includes –per section– the python code that is used in each section's analysis.[1] This is an advantage of using data at the regional level. The repository contains the python code that gets the data from Eurostat so that each step of this analysis can be replicated.

2 THEORY

The idea of our dynamic model is to follow a cohort aged a at time t over time to age $a + 1$ in year $t + 1$. We capture this dynamic with a straightforward Markov model. The model that we estimate is illustrated in Figure 1. We follow individuals from age 45 to age 85. Let's denote the age at which we start following an individual by $a = 0$ at time $t = 0$. If this agent is healthy, there is a probability π that the agent falls ill and moves to state ill at age $a = 1$ in period $t = 1$. With probability $1 - \pi$ the agent stays healthy also at age $a = 1$. For an agent in the unhealthy state, there is a probability σ that she is cured and becomes healthy at age $a = 1$. With probability δ she dies and with probability $1 - \sigma - \delta$ she remains in the unhealthy state at $a = 1$. Finally, death is an absorbing state.
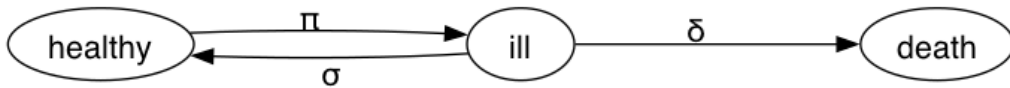

Figure 1: Markov chain model

Instead of deriving the steady state outcome of the model, we estimate parameters using the underlying difference equation. Focusing just on age for the moment to ease notation, the model implies the following difference equation. Let ι_a denote the fraction of people aged a who are ill and $1 - \iota_a$ the fraction who are healthy,

---

[1]See the github repository: https://github.com/janboone/dynamic-effects-of-health-insurance-reform.

then ι_a evolves over age (and time) as:

$$\iota_{a+1} = ((1 - \sigma - \delta)\iota_a + \pi_a(1 - \iota_a))/(1 - \delta\iota_a) \tag{1}$$

and $\delta\iota_a$ people die at age a. In words, the fraction of people ill aged $a + 1$ equals the fraction ill aged a who do neither recover to the healthy state nor die plus the fraction of healthy people who fall ill. We normalize by $1 - \delta\iota_a$ to ensure that the fractions ill and healthy add up to one.

We differentiate the fraction ill, ι_a, by region/year/gender/income. If we add all the sub/superscripts we get ι_{atr}^{ij} for age a in year t and region r among gender $i \in \{f, m\}$ and low/high income $j \in \{l, h\}$. Let α_{tr} denote the fraction of people on low income in region r and year t which is available in our data. We observe in our data mortality μ_{atr}^i, the fraction of females/males i who die at age a in year t in region r. In terms of our model this is given by

$$\mu_{atr}^i = \alpha_{tr}\delta\iota_{atr}^{il} + (1 - \alpha_{tr})\delta\iota_{atr}^{ih} \tag{2}$$

As we do not observe ι directly in our data, we have to estimate the starting points of the fraction of people initially ill, ι_0, in order to solve the difference equations over time. Our estimates of ι_0 vary by age/gender/region/income. The reason ι_0 varies by age is illustrated in Table 1. Differentiated by gender/region/income we follow the cohorts aged 45-74 over time until they are aged 55-84. The data section below motivates this choice. Hence we estimate the initial values ι_0 for these ages in the year 2008. Then we apply equation (1) to determine the development of ι_a over time as cohorts get older.

The remainder of the section explains how π, δ, σ vary with age/year/gender/region/income. First, the probability of falling ill π varies by age/gender/country/income. We know that older people tend to be less healthy than young people, women tend to live longer than men, longevity varies by country and people on low income tend to have lower health status than people on high income. We allow π to vary along these dimensions to capture these stylized facts.

Second, once people are in state "ill" we assume that their probability of death δ is the same across age/gender/income. We do allow δ_c to vary by country as some healthcare systems may be better in prolonging life for the ill than systems in other countries.

A final source of variation is the probability σ with which people recover to full health. This probability σ is determined by the probability of treatment and the quality of treatment conditional on being treated. We have data on unmet medical

Table 1: We follow cohorts aged 45-74 over time from 2008 until they are 55-84 in 2018.

year	age											
2008	45	46	47	48	...	68	69	70	71	72	73	74
2009	46	47	48	49	...	69	70	71	72	73	74	75
2010	47	48	49	50	...	70	71	72	73	74	75	76
2011	48	49	50	51	...	71	72	73	74	75	76	77
2012	49	50	51	52	...	72	73	74	75	76	77	78
2013	50	51	52	53	...	73	74	75	76	77	78	79
2014	51	52	53	54	...	74	75	76	77	78	79	80
2015	52	53	54	55	...	75	76	77	78	79	80	81
2016	53	54	55	56	...	76	77	78	79	80	81	82
2017	54	55	56	57	...	77	78	79	80	81	82	83
2018	55	56	57	58	...	78	79	80	81	82	83	84

needs. Reasons why people forgo treatment –that would be beneficial– include waiting lists, hospital too far away, afraid of treatment. The motivation that we focus on is that the treatment is skipped because it is too expensive. Let $v \in [0, 1]$ denote the fraction of people who have unmet medical needs. We write this as the sum of the fraction of people who indicate that treatment is too expensive, τ, and the fraction of people who give other reasons, v_0.

$$v = v_0 + \tau$$

We expect that τ is increasing in the fraction of healthcare expenditures that people pay out-of-pocket, oop.

$$\tau = \zeta oop$$

where oop is measured as expenditure paid out-of-pocket over total healthcare expenditure. We interpret this as capturing the generosity of a country's health insurance system. If healthcare is completely free at point-of-service, $oop = 0$ and $\tau = 0$: with free healthcare, no one forgoes treatment because it is too expensive. At the other extreme, no health insurance at all: everything is paid out-of-pocket: $oop = 1$. Then a fraction $\zeta \in [0, 1]$ will indicate they forgo treatment because it is too expensive. We allow ζ to depend on income with $\zeta^l \geq \zeta^h$. We expect people on low income to react more strongly to an increase in oop than people on high income. Also people on low income are more likely to avoid treatment for other reasons: $v_0^l \geq u_0^h$. To

illustrate, people on low income may be less skilled, say, in navigating waiting lists or find it more difficult to travel to a hospital that is further away (or even abroad).

If someone gets treatment, the probability of recovering to full health is given by λ. Hence, we write the probability of recovery σ as

$$\sigma = \sigma_0 + \lambda(1 - \upsilon) \tag{3}$$

where σ_0 is the probability that someone recovers without treatment. The parameter λ captures the quality of care which we allow to vary by region and year in ways explained below.

We consider two applications of our model. First, we calculate the expected quality adjusted life-years (qaly's) of 45 year olds till they are 85. We normalize the value of a life-year in full health to 1.0 and set the qaly for a year being ill equal to $\theta \in \langle 0, 1]$. The value of a "life-year when dead" is normalized to 0.0. To make this qaly calculation, we keep track of the size of the groups healthy and ill as follows:

$$N_{a+1}^{healthy} = (1 - \pi_a)N_a^{healthy} + \sigma N_a^{ill}$$

and

$$N_{a+1}^{ill} = \pi_a N_a^{healthy} + (1 - \sigma - \delta)N_a^{ill}$$

where we start with 45 year olds at $a = 0$ and calculate up to age A (corresponding to 85 year olds). In contrast to equation (1) we do not assume that $N_a^{ill} + N_a^{healthy} = 1$; indeed, the point is that we lose life-years as age increases. As above $N_a^{healthy}, N_a^{ill}, \pi_a, \sigma, \delta$ vary with gender/region/income.

We then calculate expected quality adjusted life-years from age 1 to A as the undiscounted sum of qaly's:[2]

$$Q = \sum_{a=1}^{A} N_a^{healthy} + \theta N_a^{ill} \tag{4}$$

Second, we consider the effect on qaly's of an increase in demand-side cost-sharing $\Delta oop > 0$. For this, we calculate qaly's in the baseline outcome, Q_0 and compare it with the outcome where the fraction of people with unmet medical needs increases with $\Delta \upsilon = \zeta \Delta oop$. We denote the qaly's in this outcome by Q_1. The loss of qaly's due to increased demand-side cost-sharing is given by:

$$\Delta Q = Q_0 - Q_1$$

---

[2]We use an undiscounted sum here to make it easy to compare to the static version $\tilde{Q}$ discussed below.

In words, suppose there was a reform in year $t = 0$ which increased *oop*. After the reform we follow the cohort aged $a = 0$ over time till age $a = A$ and calculate their qaly's, Q_1. We compare these qaly's with the counterfactual where *oop* was not changed, Q_0. We denote this the dynamic comparison.

We compare this dynamic analysis with the following static one. Instead of following an age cohort over A calendar years, we consider A ages one period after the reform and denote their qaly's by $\tilde{Q}_1$. That is, the equation for $\tilde{Q}$ is comparable to (4) except that in the latter both age a and calendar year t vary over the summation. With $\tilde{Q}$ calendar year is fixed at $t = 1$ (with the reform happening in calendar year $t = 0$). We then compare $\tilde{Q}_1$ with the qaly's these A age groups would have received under the counterfactual where *oop* is unchanged, $\tilde{Q}_0$.

To illustrate with Table 1: we start with ages 45-74 at $t = 0$. We have the reform and calculate $\tilde{Q}_1$ for the second row in the table (ages 46-75). Then we compare this to qaly's for the second row under the counterfactual of no change in *oop*. We argue that this captures the estimate made by an event study like a diff-in-diff analysis, say where the reform was introduced in one region but not in another region. The diff-in-diff approach would then control for year effects like a virus going around in the year after the reform.

This type of analysis can identify the causal effect convincingly but is usually applied for a couple of years only. For instance, because in later years the reform is introduced in other regions as well. Or because over time other factors may change causing a bias in the estimates based on the comparison of the regions.

Theoretically, it is plausible that the dynamic estimate exceeds the static one. In a model where stocks (of healthy and ill) determine the state of the system and its development over time, ignoring these dynamic effects is likely to cause an underestimation of the effect on qaly's or mortality. With our estimated model we can quantify the difference between the static and dynamic estimates.

3   DATA

The data that we use is from Eurostat's regional database and provides for NUTS 2 regions population size and number of deaths per age/gender category. In principle, we have data on 11 countries and 65 regions for the 11 years 2008-2018 and ages 45-84 for women and men. The years 2008-2018 were chosen because, at the time of the analysis, for most variables and regions data was available from 2008 onward till 2018. The age range was chosen with the idea that mortality starts to increase

from 45 years onward and number of observations per region drop significantly from age 85 onward. Finally, NUTS 2 regions were selected where there was at least one observation for unmet medical needs over the years and where the number of deaths during the year does not exceed the population size at the start of the year.

Table 2 shows the summary statistics for our variables. We briefly discuss the main variables, the appendix provides more detail. We have more than 40k observations for population and deaths per age/gender/year/region category.[3] We have missing observations and explain below how we deal with these in our Bayesian setting.

The average population size per region-age-gender category is about 6000 and the average number of deaths 90. Median population size per category equals 5300 and median number of deaths 62. In our data, the percentage of people dying in a NUTS 2/year/age/gender category (`mortality`) equals 2% on average with a maximum of 17.3% for some region and age combination. The material deprivation measure (denoted `deprivation`) comes from the EU statistics on income and living conditions (EU-SILC) survey. It refers to the enforced inability to pay unexpected expenses, afford adequate heating of the home, durable goods like a washing machine etc. This variable captures the idea that people may feel forced to forgo (valuable) treatment because it is too expensive.[4] The mean deprivation rate across regions and years equals 10%. Some regions have (almost) no people in deprivation, while in others 55% of people live in deprivation. The median rate is around 3%.

Also from the EU-SILC survey, we use the variable capturing unmet medical needs because the forgone treatment was too expensive (`too exp`). The variable `unmet` measures percentage of people in need of healthcare that postpone or forgo treatment because it is either too expensive, the hospital is too far away, there is a waiting list for the treatment, the patient hopes that symptoms will disappear without treatment, the patient is afraid of treatment or has no time to visit a physician. The mean fraction of people forgoing treatment because it is too expensive

---

[3]The max. number of observations that we could have is: 65 (regions) ∗ 11 (years) ∗ 30 (ages) ∗ 2 (genders) = 42,900 which we have in first three rows. For the next 5 rows in the table we have data for at max. 11 years and 65 regions which corresponds to 715 obervations and for out-of-pocket we have at max. 11 years and 11 countries corresponding to 121.

[4]Eurostat also provides the variable "at-risk-of-poverty" per NUTS 2 region. This is a relative poverty measure: the share of people with disposable income after social transfers below a threshold based on the national median disposable income. But the link between this poverty measure and forgoing treatment because it is too expensive is weaker. Further, there are more missing observations for this variable than for the deprivation variable.

equals 1.5%, while the mean fraction of people with unmet medical needs (because it is too expensive or other reasons) equals 6.5%. In some regions the fraction of people forgoing treatment because it is too expensive goes up to almost 20% and the fraction of people with unmet medical needs to almost 30%.

For the model application considering the effect of demand-side cost-sharing on health and mortality, healthcare quality could be a collider. Indeed, if government resources are reduced, there could be both an increase in out-of-pocket payments and a reduction in healthcare quality (by reducing investments in technology and/or reducing the number of physicians). This reduction in quality would also affect health and mortality thereby confounding the healthcare demand effect of out-of-pocket spending. We use two variables to control for the quality of care. First, infant mortality is a well known measure for the quality of a healthcare system. Moreover, infant mortality is not directly related to our measure of mortality which starts at age 45. On average there are almost four dead infants (younger than one year of age at death) per 1000 live births. But for some regions this almost reaches 16 per 1000 live births. The second quality measure is the number of physicians per 100k inhabitants. On average there are 380 physicians per 100k inhabitants but this varies between regions from 180 to 800. We view this as an approximate measure of resources available for healthcare in a region. Finally, we use the fraction of healthcare expenditure paid out-of-pocket as a signal of how generous health insurance is. This captures country wide policy. On average people pay a bit more than 20% out-of-pocket but this varies from 12% in some countries to more than 40% in others.

Table 2: Summary statistics main variables

	count	mean	std	min	median	max
population	42900	5961.0	3788.7	574.0	5286.0	30491.0
deaths	42900	88.9	83.7	0	62.0	867.0
mortality (%)	42900	1.8	1.9	0	1.2	17.3
deprivation (%)	481	9.9	13.3	0	2.8	55.2
too expensive (%)	527	1.5	2.6	0	0.5	18.4
unmet (%)	527	6.4	4.6	0.7	5.1	28.1
infant mortality (‰)	705	3.8	2	0.8	3.4	15.8
physicians per 100k inhab.	594	379.4	108.6	180	363.3	807.9
out-of-pocket (%)	55	22.8	9.5	12.0	18.9	43.2

4 ESTIMATION

We estimate the model using Bayesian methods. Here we present the equations of the model. The online appendix contains the details on the choice of priors and the `pymc` code of the model. Further, we explain how the Bayesian algorithm deals with missing observations.

We present the model in reverse order. We observe the number of deaths m and the population size n per age/gender/region/year dimensions. We model the probability of death μ and m has a Binomial distribution:

$$m_{atr}^{i} = Binomial(n_{atr}^{i}, \mu_{atr}^{i}) \tag{5}$$

where μ_{atr}^{i} is given by equation (2). Poverty α_{tr} per region and time period is observed in the data using the material deprivation measure discussed in the previous section. The fraction of people ill ι_{atr}^{ij} varies by age a, gender i, time t, region r and income level j. This fraction ι follows from difference equation (1). Before we explain how this equation is estimated, recall that the estimate of δ_c is allowed to vary by country c.

To solve differential equation (1), we start by estimating the initial condition ι_{a0r}^{ij} for ages $a = 0 - 30$ (i.e. ages 45-74) in year $t = 0$ (2008) per region r, gender i and income category j.

Next, we need to estimate the probability of falling ill π_{ar}^{ij} which we assume to be additive in age, region, gender and income fixed effects in log-odds space. That is we add together the log-odds effects for age, gender, region and income and apply the inverse logit function $e^x/(1 + e^x)$ to the sum of these log-odds. This gives the probabilities $\pi_{ar}^{ij} \in \langle 0, 1 \rangle$.

The probability $\sigma = \sigma_0 + \lambda(1 - \upsilon)$ of recovering to full health in equation (3) is estimated as follows. We estimate the probability $\sigma_0 \in \langle 0, 1 \rangle$ of recovering to health without medical intervention. Further, we observe the fraction of people who indicate that they have unmet medical needs υ_{tr} at the year/region level. For the ill who do get treatment, $1 - \upsilon$, there is a probability λ of full recovery. The logg-odds of λ are the sum of three effects: (i) year fixed effects, capturing that over time the quality of care improves due to technological progress, (ii) infant mortality and (iii) number of physicians per 100k inhabitants. The latter two capture the variation across time and region in quality of care. Again we use the inverse logit function to turn the log-odds into a probability λ.

We expect both υ_o and τ to be bigger for low incomes than for high incomes.

If your income is higher, you are less likely to skip a treatment because it is too expensive: $\tau^j = \zeta^j oop$ with $\zeta^l \geq \zeta^h$ and oop_{tc} varies with country and time. Further, also with the other reasons for unmet medical needs (like waiting lists) we expect these to be higher for low incomes. Hence, we find that

$$v_{tr} = \alpha_{tr}(\zeta^l oop_{tc} + v^l_{otr}) + (1 - \alpha_{tr})(\zeta^h oop_{tc} + v^h_{otr})$$

where we observe $\tau_{tr} = (\alpha_{tr}\zeta^l + (1 - \alpha_{tr})\zeta^h)oop_{tc}$ and $v_{otr} = \alpha_{tr}v^l_{otr} + (1 - \alpha_{tr})v^h_{otr}$.

Hence we find

$$\sigma^j_{tr} = \sigma_0 + \lambda_{tr}(1 - \zeta^j oop_{tc} - u^j_{otr})$$

Given that we estimate the initial conditions ι_0 and the parameters π_a, σ, δ we can solve differential equation (1) to find ι^i_{atr} and thus calculate the probability of death μ^i_{atr} in equation (2) as input in the Binomial distribution (5).

One of the conceptual advantages of a Bayesian model is that there is no clear distinction between data and parameters. Both are basically distributions. This comes to the fore when dealing with missing values which are prevalent in our data as shown in Table 2. The Markov Chain Monte Carlo (MCMC) algorithm generates the posterior distribution of parameters by drawing from this distribution. For each parameter we will have 4 chains of 2000 draws, that is 8000 draws from the posterior distribution. When an observation is missing, the algorithm will randomly draw the observation from a pre-defined distribution. This implies that each draw is different signalling the uncertainty that surrounds the missing observation. This is better than imputing the value of the observation as the latter suggests that we are as certain about the imputed value as we are about an observed value; which is obviously not the case.

For variables based on summation, it is natural to assume a normal distribution. The expectation (standard deviation) of this distribution is then given by the average value (standard deviation) of the observations that we do have for this region or country.

To improve the efficiency of the algorithm in solving the difference equation while estimating the parameters, the data is not in (two dimensional) dataframe format. Instead we work with multi-dimensional tensors. To illustrate, the tensors with the population and deaths data are 4-dimensional tensors with coordinates age/gender/year/region. The difference equation then iterates over age and calendar year using pytensor's `scan` function.

5 RESULTS

In this section we present the results of the estimation of the model. Before presenting the outcome of our estimation, we present some (graphical) checks of our model.

5.1 Model fit

A first check of the model is to see whether it fits the data. Figure 2 shows for all our observations the observed number of deaths on the horizontal axis and the predicted number on the vertical axis. Further, for each prediction we add a dashed vertical line representing the 95% prediction interval (of the posterior predictive distribution). The predictions clearly follow the 45-degree line and almost all 95% prediction intervals intersect the 45-degree line. In this sense the model fit seems quite good.
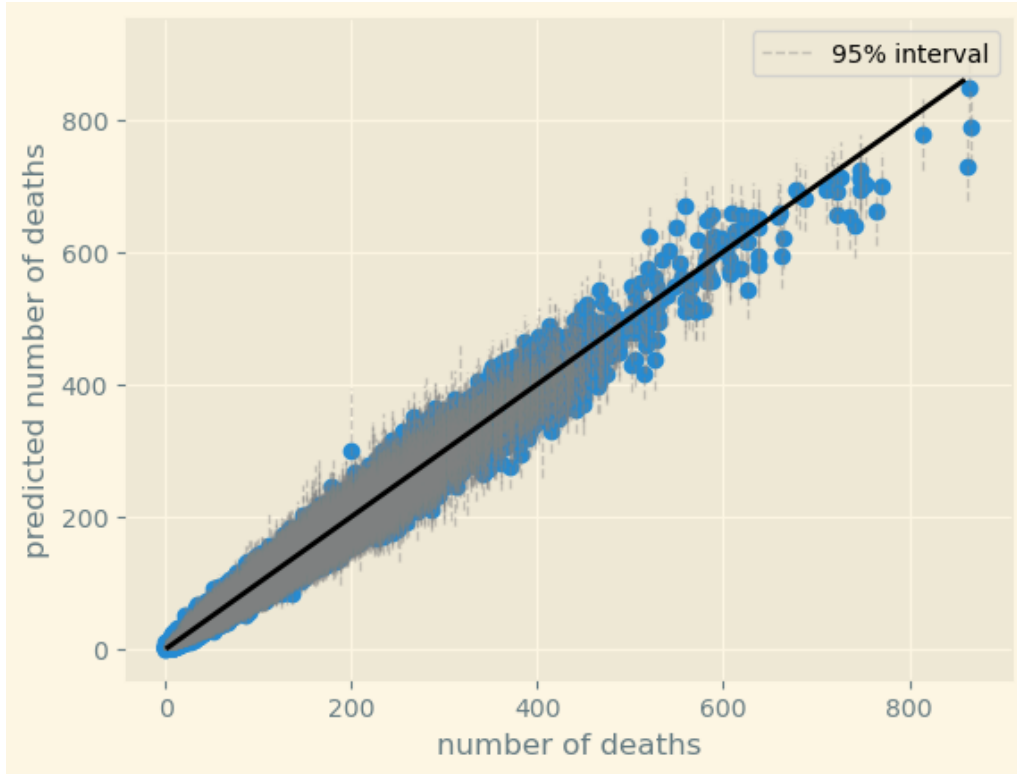

Figure 2: Observed vs predicted number of deaths (per age/gender/region/year) together with the 95% prediction intervals.

Figure 6 in the appendix presents the trace plots and Table 5 the values for r-hat for estimated parameters. The r-hat values are close to one and the traceplots

satisfy the following three criteria. First, the plots in the right column of the figure should be stationary; that is, not trending upward or downward. This implies that the posterior mean of the coefficient is (more or less) constant as we sample. Second, there should be good mixing which translates in condensed zig-zagging. In other words, the algorithm manages to draw values across the whole domain of the posterior quickly one after the other. Finally, the four chains cover the same regions. This is most easily checked in the left column of the trace plot. All three features are satisfied in the figure.

To gain an intuition on how the model works, Figure 3 shows a number of estimated outcomes. The top-left panel shows the probability of falling ill, π, as a function of age for women and men.[5] This probability is quite low around age 45 and increases to 40% for men in their eighties. As shown in the panel below this one, the probability of death –conditional on being ill– is above 10% per year. In other words, falling ill in our model does not refer to a regular cough or breaking a leg. The observation that the average probability of falling ill is higher for men than for women is in line with the observation that women tend to live longer than men. The 95% posterior probability interval shows that there is quite some variation in the probability of falling ill at older ages but that uncertainty is small till age 75. In order not to crowd the figure we do not report the 95% intervals for males but they are of similar magnitude.

The top-right panel shows the country offsets (in log-odds terms) for four countries. There is a clear dichotomy between Switzerland and Norway on the one hand and Lithuania and Hungary on the other. But even between these countries there is hardly overlap between the distributions. Of these four countries, people in Switzerland are least likely to fall ill (at any age) and in Hungary most likely.

Similarly, the probability of dying conditional on being ill, δ_c, varies by country. Of the four countries presented, it is lowest for Denmark and Finland and highest for Hungary and Bulgaria.

Finally, the bottom-right panel shows the fraction of people ill for four categories: people aged 55 and 60 with low/high income. The fraction of people who are ill is higher for low than for high incomes, as one would expect. Also the fraction increases with age. The posterior distributions for these fractions do not overlap.

---

[5]These probabilities are calculated by averaging π_{ar}^{ij} over region r and income j.
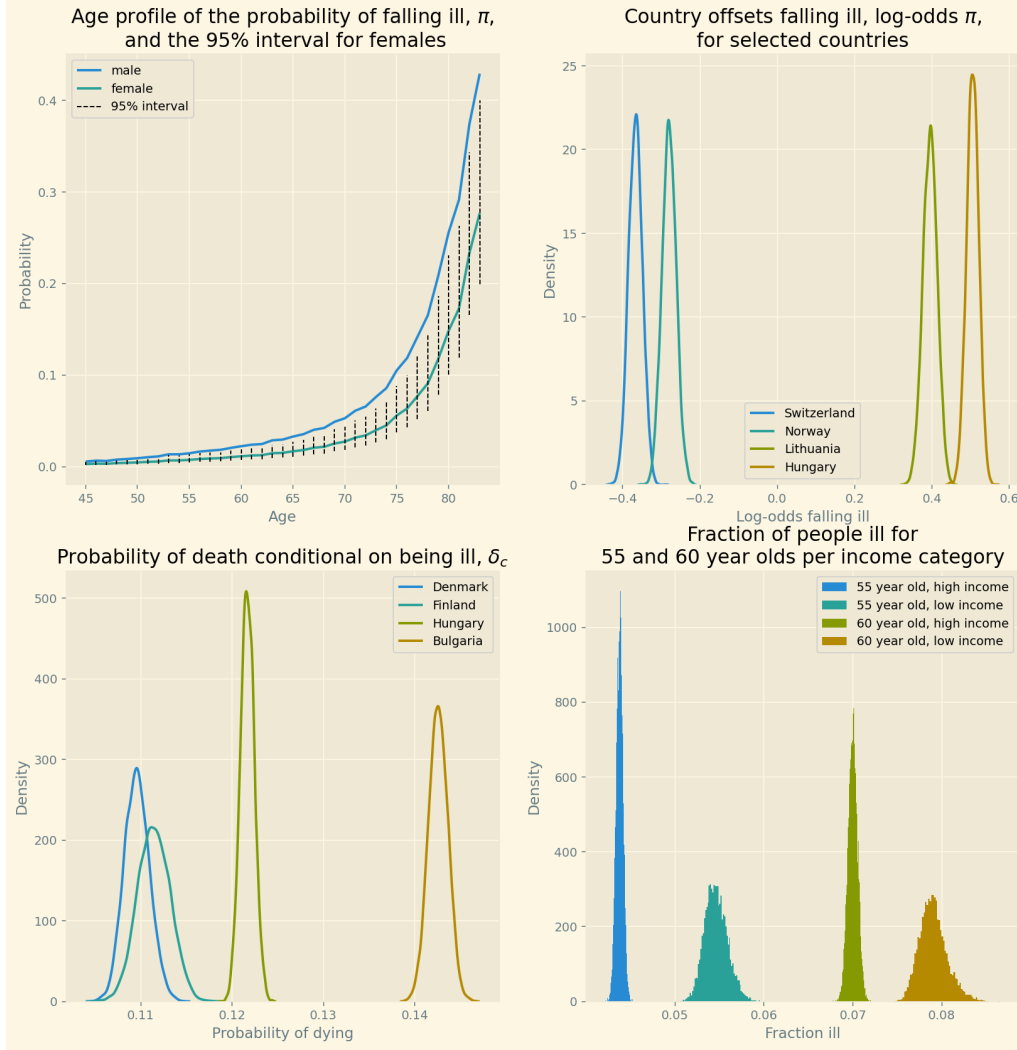
Figure 3: Summary of selected features of the model

5.2 Analysis

The analysis in this section makes the following points. First, the expected qaly's per person vary a lot between countries and within countries between gender and income categories. Second, when we increase demand-side cost-sharing by ten percentage points ($\Delta oop = 0.1$), there is a clear but modest reduction in qaly's. Third, the effect is almost zero for people on high income but far higher for people on low income. In this sense, effect heterogeneity is high and considering the average effect (across incomes) is not informative. Finally, instead of comparing qaly's lost over, say, a thirty year time period (for which one needs a dynamic model), we can compare the loss in qaly's for the same age groups from one year before the increase in *oop* to one year after. The latter is close to a (static) measure used in most event study analyses. We show that the dynamic effect is an order of magnitude bigger than the
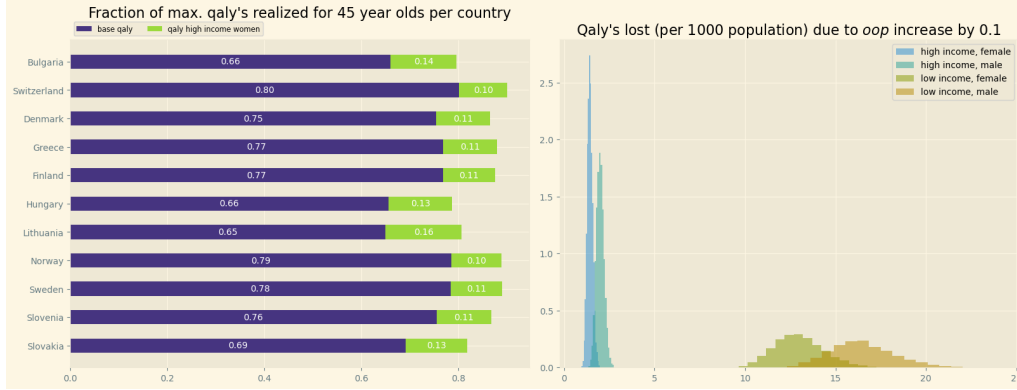
estimated static effect.

Figure 4: Summarizing two applications of the model: left panel shows the fraction of the maximum qaly's realized by country and the right panel shows qaly's lost due to 0.1 increase in *oop*.

Figure 4 summarizes the two applications that we analyse with the model. First, since our estimation method identifies the fraction of people who have low health status, we can derive qaly's using equation (4). In the figure we use $\theta = 0.6$: the quality of life for someone who is ill (and has a 10% probability of dying per year) is 60% of the quality of life for someone in full health. This is an arbitrary choice and obviously different choices can be made here. Different values for θ will change the precise numerical values but not the main (qualitative) conclusions of the analysis.

The figure was made as follows. For each group (female/male, low/high income) of 45 year olds we create a population of 1000 individuals. At 45 we use our estimates of ι_0 to split the groups into low/high health status. With our estimated parameters π, σ, δ we follow these individuals as they age over time till they are 85. Hence, for each group there are at max. 40,000 life-years (in full health) to be had over this period. For each subgroup gender/income we calculate the realized quality adjusted life-years and divide this by 40,000. The left panel of Figure 4 shows the lowest qaly's realized (base qaly) and the highest qaly's. The latter are always realized by women with high income. The former are (almost) always by low income males.[6] The panel shows that in, say Switzerland, low income males get 80% of the max. number of qaly's and high income females get 10% of 40,000 qaly's more. That is, the latter get 12.5% (0.10/0.80) more qaly's than the former. This gender/income gap ranges all the way up to 20% in Hungary, Bulgaria and Lithuania.

---

[6]The two exceptions are Bulgaria and Hungary where high income males' qaly's were just below the low income males' qaly's.

The panel on the right shows what happens if *oop* is increased by 0.1: people pay an additional 10% of their healthcare expenditure out-of-pocket for each country in our data. The average effect is around 3.0: among 1000 people, a region/gender/income group loses 3 life-years for 45-85 year olds due to $\Delta oop = 0.1$. This effect is not huge (people do not just pass away because cost-sharing is increased), but it is clearly bounded away from 0.

What is more surprising is the effect heterogeneity for high and low incomes. The effect is almost ten times bigger for people on low income. People on high income hardly react in terms of healthcare expenditure to the increase in *oop* as they can afford it anyway. But for low incomes, the increase reduces their demand for treatment. This raises unmet medical needs because it is too expensive and hence, for this group, reduces the probability σ of recovering to the healthy state. This has two effects on their qaly's. First, the quality of life in the ill state is lower than in the healthy state. Second, staying longer in the ill state, increases the probability of death (compared to being in the healthy state).

Finally, we compare the dynamic (across years) qaly effect derived above with a year-on-year effect for the same ages. To avoid making additional assumptions we do this for the ages 45-74 in the first row of Table 1. This is the cohort of ages that we use to estimate the model. Hence we have a well defined ι_0 fraction of people that are ill per age/gender/region/income category in the year before the reform. Then we use the equations for $\tilde{Q}_0, \tilde{Q}_1$ defined in the theory section to find the static effect on qaly's. We compare this to the dynamic estimate in equation (4) for the same ages till $A = 75$.

Table 3: Dynamic vs. static estimates of qaly's lost due to $\Delta oop = 0.1$.

estimate	male low income	female low income	male high income	female high income
$Q_0 - Q_1$	8.01	5.23	0.82	0.46
$\tilde{Q}_0 - \tilde{Q}_1$	0.31	0.24	0.12	0.06

As shown in Table 3, the dynamic low income estimates are lower than in Figure 4 (right panel): the dynamic estimates across ages till 75 are lower than over the full range till 85.

Comparing static and dynamic estimates shows that for this type of health model the dynamic estimates of qaly's lost $Q_0 - Q_1$ are a factor 20 higher for low incomes than the static estimates $\tilde{Q}_0 - \tilde{Q}_1$. For high incomes the factor exceeds 7. As noted

in section 2, we expect the dynamic effect to be bigger than the static effect because the former takes the effect on the stock of ill/healthy into account. However, the size of the difference is larger than one might have expected.

This suggests that for health models where an underlying (latent) health variable is important, short-term estimates (as often with a diff-in-diff analysis) severely under-estimate the overall effect of a policy change.

One can see the following trade-off. A diff-in-diff approach leads to a cleaner causal identification. It often allows for a convincing way to keep other factors constant. In contrast, identification in our model (which allows for a dynamic analysis) is based on underlying theory but suffers from potential missing variable bias. We show that for reasonable parameter values, the static estimate under-estimates the effect compared to a dynamic one taking into account the cumulative effects over a life-time.

Figure 5 illustrates how a static estimate like diff-in-diff ignores the dynamics of health. The figure is an abstract representation of the diff-in-diff estimate of Finkelstein and McKnight (2008) analysing the introduction of Medicare in 1966 providing universal public health insurance to individuals aged 65 and over. To find the health effect of Medicare, the paper compares the average mortality rate of 65-74 year olds with the rate for 55-64 year olds both before and after the introduction of Medicare; the years 1952-1975.

The logic of comparing age category 55-64 with 65-74 is the underlying assumption that the two categories do not differ except for the introduction of Medicare. This assumption would be harder to motivate when comparing 55-64 with 82 year olds. The idea of the paper is that if Medicare has substantial health effects one expects that the mortality rate for 65-74 year olds falls compared to 55-64 year olds after 1966. The paper does not find such an effect and the figure illustrates one mechanism why this can be the case.

To place this analysis in the context of our paper, we simulate the effect of a 10% point increase in oop ($\Delta oop = 0.1$) in year 0 for people aged 65 and over. Hence, in our model this implies that mortality will increase for the 65+ category (not decrease as in the Medicare example). We can simulate both the reform and the counterfactual and hence can make the comparison within age category; no need to compare the effect to 55-64 year olds. For this comparison between the diff-and-diff and dynamic estimates we only need the estimated parameters to be reasonable, not perfectly accurate. As shown above, this assumption is likely to be satisfied.

We do the simulations for low income men in one region in Bulgaria aged 65 and

above who face $\Delta oop = 0.1$ in year 0. On the horizontal axis we plot age for this group. The vertical axis shows the years after the introduction in year 0. The color intensity of a point indicates the number of years an agent has faced this higher oop. In the first year everyone aged 65 and above faced the higher out-of-pocket for one year (bottom row in the figure). In the second year, people aged 66 and over faced this for two years and the 65 year olds only for one year etc.

The horizontal black line indicates the comparison group used in the paper (65-74) for year 10 (last year in the Finkelstein and McKnight (2008) data). The figure illustrates how –from a dynamic point of view– the effect gets diluted as the 65 year olds in this group have only experienced the reform for one year, the 66 year olds for two years etc. In contrast the blue diagonal line gives the true dynamic effect when we follow 65 year olds in year 0 till they become 83 with 19 years of increased oop. The diagonal includes points with a darkness intensity that the horizontal line does not capture for any of the years 0-18. In this sense, the static estimate under-estimates the true dynamic effect.

The right panel shows the consequences of this. The vertical axis shows the increase in mortality by age normalized on the average mortality effect for 65-74 year olds. Hence, the static estimate for 65-74 year olds is one by construction. But due to the dynamics of health in our model, the effect is first below 1 (till age 69) and then increases to six times the static estimate for 80 year olds. In order to identify this (far bigger) effect, one has to estimate a model capturing the age effects instead of relying on the diff-in-diff estimate. It will depend on the application whether a static precise causal estimate is preferable or a less precise estimate that takes the dynamics of the problem into account.

6   DISCUSSION AND POLICY IMPLICATIONS

We introduce a simple dynamic model of health status and mortality. We estimate the fraction of people with high and low health status per age/gender/year/region/income category and the flow probabilities between these health states and death. Using this we follow cohorts over time and predict the number of deaths per age/gender/year/region. The fit of the model is quite good and it replicates stylized facts like mortality is higher for men than women, for low than high incomes.

The model can be used to rank countries in terms of expected quality adjusted life years for 45 year olds and estimate the health gap between low income men and high income women. We also use the model to illustrate the quantitative difference
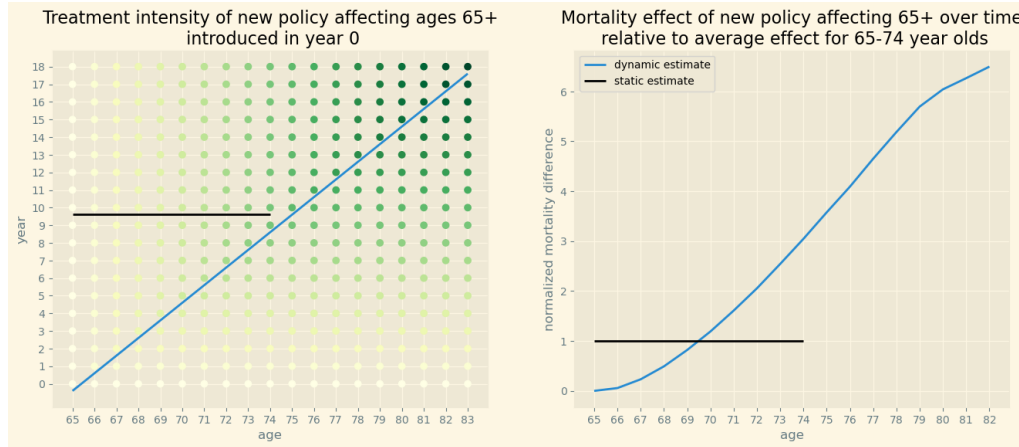
Figure 5: Diff-in-diff vs. dynamic estimates of a 10% point increase in out-of-pocket expenditure for low income men aged 65 and over. The left panel shows the intensity of the effect across ages and years after the reform. The right panel shows the effect when following the 65 year old cohort till they are 83.

between static estimates (like a diff-in-diff) of the mortality effect of a health reform and the dynamic estimate. The former needs fewer assumptions to identify the causal effect. But when health is viewed as a stock variable, it turns out that the dynamic effect is up till six times bigger than the static estimate would suggest.

7 BIBLIOGRAPHY

Borgschulte, Mark, and Jacob Vogler. 2020. "Did the Aca Medicaid Expansion Save Lives?" *Journal of Health Economics* 72 (July):102333. `https://doi.org/10.1016/j.jhealeco.2020.102333`.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What Does a Deductible Do? the Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics* 132 (3): 1261–1318. `https://doi.org/10.1093/qje/qjx013`.

Chandra, Amitabh, Evan Flack, and Ziad Obermeyer. 2021. "The Health Costs of Cost-Sharing," February. `https://doi.org/10.3386/w28439`.

Finkelstein, Amy, and Robin McKnight. 2008. "What Did Medicare Do? the Initial Impact of Medicare on Mortality and out of Pocket Medical Spending." *Journal of Public Economics* 92 (7): 1644–68. `https://doi.org/10.1016/j.jpubeco.2007.10.005`.

Goldin, Jacob, Ithai Z. Lurie, and Janet McCubbin. 2020. "Health Insurance

and Mortality: Experimental Evidence from Taxpayer Outreach." *The Quarterly Journal of Economics* 136 (1): 1–49. `https://doi.org/10.1093/qje/qjaa029`.

Gross, Tal, Timothy Layton, and Daniel Prinz. 2020. "The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments," October. `https://doi.org/10.3386/w27977`.

Miller, Sarah, Norman Johnson, and Laura R Wherry. 2021. "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data." *The Quarterly Journal of Economics*, January. `https://doi.org/10.1093/qje/qjab004`.

Nyman, J.A. 2003. *The Theory of Demand for Health Insurance*. Stanford University Press.

A Data

All our variables come from Eurostat. Table 4 shows the dimensions over which our variables vary: country, NUTS 2, calendar year, age and sex. We also present a clickable link to the variable on the Eurostat website for ease of reference. The file `./getting_data.org` presents the code to download the Eurostat data.[7]

Table 4: Variables and the dimensions over which they vary.

variable	country	NUTS 2	time	age	sex	reference
population		x	x	x	x	link
deaths		x	x	x	x	link
deprivation		x	x			link
too expensive		x	x			link
unmet		x	x			link
infant mortality		x	x			link
number of physicians		x	x			link
out-of-pocket	x		x			link

The variables on deprivation and access to care (unmet and too expensive) come from the EU statistics on income and living conditions (EU-SILC) survey.

From the Eurostat Glossary: "Material deprivation refers to a state of economic strain and durables, defined as the enforced inability (rather than the choice not to do so) to pay unexpected expenses, afford a one-week annual holiday away from home, a meal involving meat, chicken or fish every second day, the adequate heating of a dwelling, durable goods like a washing machine, colour television, telephone or car, being confronted with payment arrears (mortgage or rent, utility bills, hire purchase instalments or other loan payments)." Our variable "material deprivation" equals the share of people in a NUTS 2 region in material deprivation.

Fraction of people with self-reported unmet needs for medical examination is based on the same survey. In particular, the definition of this item is "Self-reported unmet needs for health care: Proportion of people in need of health care reporting to have experienced delay in getting health care in the previous 12 months for reasons of financial barriers, long waiting lists, distance or transportation problems." We use both the general definition of unmet needs and the specific reason that treatment was too expensive.

---

[7]This file can be found in the github repository: https://github.com/janboone/dynamic-effects-of-health-insurance-reform.

Infant mortality measures the number of deaths of infants per 1000 live births in a year at the NUTS 2 region level.

We measure the number of physicians per 100k inhabitants per NUTS 2 region and calendar year. It gives an idea of the resources spent on healthcare in a region. Together with infant mortality and unmet medical needs it gives an idea of healthcare quality in a region.

We characterize how generous a health insurance system is using the variable `OOP` in our analysis. This variable is derived from data on health care expenditure by financing scheme at the country level. For our `OOP` measure we focus on household out-of-pocket payment (`out-of-pocket`) as share of total current health expenditure.

B ESTIMATION

Table 5 shows the details of the parameters in the model that do not have too many indices. To illustrate we do not present the results for `logodds_age_`π as this would entail 40 rows in the table.

The value of `r_hat` is equal or close to one for all variables. This suggests that the NUTS algorithm converged in drawing samples from the posterior distribution. The probability of falling ill is clearly lower for women than for men. The increase in the fraction of people forgoing treatment because it is too expensive, ζ, is higher for low incomes than for high incomes. The ν parameter of the Beta distribution is sometimes referred to as sample size. For the distribution of the fraction of people reporting unmet medical needs due to reasons other than the treatment being too expensive, ν_{U_o} is above 100. For the fraction of people reporting unmet medical needs (for any reason) ν_U is close to 300. The variable ε captures the extent to which unmet medical needs for other reasons than too expensive originates with low incomes compared to high incomes. Although the posterior expectation of ε is positive, zero is part of the 94% credibility interval. The uncertainty of this parameter, including that ε equals zero, is propagated through our results and simulated outcomes. Finally, λ_1 equals the (negative) effect of healthcare quality as measured by infant mortality on the log odds of λ, the probability of recovery conditional on being treated. When measuring quality with the number of physicians per 100k inhabitants, λ_2 captures the effect on the log odds of λ.

$$\log(\lambda/(1-\lambda)) = \lambda_0 - \lambda_1 M_{infant} + \lambda_2 N_{physician}$$

where λ_0 varies by year to capture technical progress, M_{infant} measures infant mortality and $N_{physician}$ the number of physicians. The table shows that λ_1 is positive in expectation but can be zero. For λ_2 there is clearer evidence that the effect of physicians on recovery is positive.

Table 5: Estimated parameters of the main model

	mean	sd	hdi_3%	hdi_97%	ess_bulk	r_hat
logodds_sex_π[F]	-0.19	0.03	-0.25	-0.13	349.00	1.01
logodds_sex_π[M]	0.50	0.03	0.44	0.56	354.00	1.01
ζ[y_l]	0.27	0.01	0.25	0.30	8429.00	1.00
ζ[y_h]	0.03	0.00	0.03	0.04	6048.00	1.00
nu_U	299.81	18.94	264.86	336.40	7950.00	1.00
nu_U_o	109.35	6.59	97.78	122.34	5584.00	1.00
ε	0.07	0.07	0.00	0.19	10331.00	1.00
lambda_1	0.01	0.01	0.00	0.02	6389.00	1.00
lambda_2	0.06	0.01	0.04	0.07	1704.00	1.01

Figure 6 shows the traceplots for the main parameters of the model. We check the following three characteristics of the plots. First, the plot should be stationary; that is, not trending upward or downward. This implies that the posterior mean of the coefficient is (more or less) constant as we sample. Second, there should be good mixing which translates in condensed zig-zagging. In other words, the algorithm manages to draw values across the whole domain of the posterior quickly one after the other. Finally, the four chains cover the same regions. All three features are satisfied for the coefficients in the right panel of the figure.

Figure 6: Traceplot for the main model