Abstract

Health expenditures as a share of GDP and health status vary significantly across countries. While some have argued that the lack of a strong link between the two at the aggregate level is evidence that the marginal returns of health expenditures paid by richest individuals are very small, others have argued that prices and inefficiencies blur the link between health consumption and health. In this paper, we evaluate the contributions of two gaps on the cross-country differences in the GDP share of health expenditures and health status: (i) the TFP gaps measuring the relative economic development (called efficiency gap), and (ii) the price gaps capturing the inefficiencies on the health service market (called the health gap). To this end, we extend a general equilibrium framework à la Aiyagari (1994) by including health production (Grossman, 1972). We estimate its structural parameters as well as the country-specific structural gaps using a method of moments approach on aggregate and micro data from eight countries. We perform counterfactual experiments to quantify the relative role of price and TFP in explaining observed cross-country heterogeneity in health expenditures and health status. Our main finding is that dispersion in health price seems to be the main cause for cross-country differences in health.
1 Introduction

Large differences in health spending are observed across countries. In 2005, the U.S. spent 14.6% of its GDP on health while Germany spent 10.3% and Sweden 8.7% (OECD Health Data, 2005). Yet, health spending does not appear to be strongly associated with health outcome despite compelling evidence that health care improves health. For example, Americans have been repeatedly found to be in worse health than other countries (Banks et al., 2006) and have higher incidence rates for various health conditions (Solé-Auró et al., 2015). This has prompted some to argue that the additional health expenditures is poorly productive and that the health production function is relatively flat (Fuchs, 2004). Thus, the most developed country must devote an increasing share of its income to health expenditures for low returns on the health status. But, a large literature propose an alternative explanation by focussing on health care prices differences and underlying health differences. This paper presents an economic framework that can fit observed patterns in order to understand the role played by these different explanations and their welfare implications.

Since health expenditures is the product of health price and the demand for health services, one potential suspect for such differences is that prices are different. There is compelling evidence of substantial variation in prices for the same services which cannot be explained by differences in quality: we refer to these distortions as wedge on the health service market. For example, we report in Table 1 variation in costs for various products and services from the International Federation of Health Plans in 2013 (IFHP, 2013). The cost of an angiogram in the U.S. was 3.1 times that in Spain while 4.8 times for Bypass surgery. Similar evidence can be found for physician compensation and administrative costs (Anderson et al., 2003; Cutler and Ly, 2011). Cutler and Ly (2011) argue that much of these differences in costs come from the administrative burden of managing a complex reimbursement system while the relationship between providers and payers (insurers) may lead to important wedges due to asymmetric information. Higher prices may lower the consumption of health services but evidence on the price elasticity of health services suggest a relatively inelastic demand curve (Manning et al., 1987). Hence, higher prices have the potential of leading to a higher share of income devoted to health. We interpret these price gaps as the health wedge induced by inefficiencies on the health service market.

The quantity of health services consumed in each country may also from one country to the other because of differences in economic resources. The earlier literature on differences in health expenditures has identified income as a key source of differences. Nevertheless, Gerdtham and Jonsson (2000) conclude that the income elasticity of health expenditures is
close to one which would suggest, as Newhouse (1992) points out, that income differences cannot explain large variation of the income share of health expenditures. However, Hall and Jones (2007) estimate a life-cycle model which generates much higher income elasticities capable of explaining the rise in health expenditures in the US which suggest that the income elasticity may have been underestimated in previous studies. We will take advantage to the estimation of a general equilibrium model in order to evaluate the impact of efficiency wedge measured by the TFP gaps, on the health expenditures as a share of GDP and on health status.

As far as we know, there is no general equilibrium model recognizing the endogeneity of health expenditures, health and economic resources that allows to quantify the source and welfare implications of these heterogenous wedges (health or/and efficiency) across countries. In a simple illustrative model, we show that the relative magnitudes of the various effects depend in part on values of preferences of consumers for consumption and health as well the shape of the health production function: the structural parameters of the economy determine the endogenous responses to price and TFP changes. But, if preferences and technologies are rather similar across countries, what drives cross-country disparity in health would then be the size of the wedges (health and efficiency). Hence, in this paper, we take up the challenge to estimate simultaneously income and price elasticities as well as the size of the price and TFP gaps. We build on the framework developed by Aiyagari (1994), augmented to incorporate health production following Grossman (1972), to estimate structural parameters and health and efficiency wedges using a Method of Simulated Moments (MSM) approach exploiting variation across country in economic resources, health expenditures and health outcomes. One additional advantage of our structural approach is to provide estimates of price and TFP elasticities at the partial and the general equilibrium: the feedback effect of all market adjustments can be measured.

Many others factor can explain the cross country differences in health and health expenditures. The health insurance system can transfer some health services from the richest to the poorest, thus improving the aggregate health status. We thus introduce in our estimation the observed heterogeneity across country of co-insurance rate. While higher

Chari et al. (2007) use a simple macroeconomic model to show that this efficiency wedge can be generated by frictions that cause factor inputs to be used inefficiently. This inefficient factor utilization maps into efficiency wedges and thus a lower TFP.

Our approach is in the spirit of Chari et al. (2007)’s work. They estimate the standard RBC model in order to measure the contributions of various wedges (efficiency, labor) on the variations across time of macroeconomic aggregates. This method is also used by Ohanian et al. (2008) to explain the cross country differences in long-term changes in hours worked.
expenditures may lead to better health, the causality may also run in the opposite direction. The rapid growth of obesity in the U.S. relative to other countries may also explain part of the differences in health expenditures across countries (Thorpe et al., 2004, 2007). According to Cutler et al. (2003), part of the differences in obesity between the U.S. and Europe could originate from differences in food production technology and regulation which lead to higher relative price of less healthy food choices. We also take into account these heterogeneous risky behaviors. Finally, we also introduce heterogeneity across country coming from labor market incomes. It is well known that US earning risks is larger in European labor markets. If we show that a large earning risk may evict health expenditure because agents need to insure themselves against consumption fluctuations (using precautionary savings), the effect at the aggregate level is ambiguous because capital accumulation increases output and thus average earnings, which affects the demand for health services and the GDP share of health expenditures. Hence, in order to purge our estimate of the price and TFP gaps from these observed cross country differences, our estimation takes these other sources of heterogeneity into account.

A joint theory of health and economic resources must also be able to explain the fact that the distribution of health within country is very different, in particular across groups with different economic resources. Avendano et al. (2009) provide evidence that the health gradient with wealth, the relationship between health status and wealth, is larger in the U.S. than in Europe for most health conditions. Smith (1999) reviews a host of factors which may explain the gradient: the effect of health shocks on wealth, for example because of incomplete markets, the effect of wealth on health, for example through demand effects (e.g. income effects) and unobserved third factors (genes, etc). Hence, beyond the aggregate economic indicators, our estimation also take as targeted moments the health gradient in each countries. This allows us to control for the health price and TFP elasticities estimated for each country. Our estimation is performed on 8 countries (the US, Sweden, Denmark, the Netherlands, Germany, France, Italy, and Spain).

We find that the US are characterized by the highest health price of our sample and lies among the highest-TFP countries. In addition, using counterfactuals, we find that inefficiencies on the health market dominate the high technological efficiency when we focus on health indicators. It is also the case for European countries. Therefore, dispersion in health price seems to be the main cause for cross-country differences in GDP share of health expenditure and percentage of population in good health. When we consider welfare, rather than health indicators, the conclusion is reversed (as the utility gain from good health can be low compared to consumption utility): TFP gaps matter more than price gaps. Health price gap (the health wedge) is more than compensated by TFP gap
(the efficiency wedge) in the US. Let us notice that the welfare costs of the price gap are found to be large: while the health sector accounts for only 14% of GDP in the US, the welfare costs of the price gap is only twice lower than the welfare gains induced by the TFP gap. This underlines the strong distortions induced by inefficiencies on the US health market.

The paper is structured as follows. In section 2, we present a simple illustrative model to explain differences in health expenditures and health across countries. In section 3, we expose the general equilibrium model that will be used to fit the data. In section 4, we present the data and estimation method we use and report estimates of the model and its predictive performance. In section 5, using counterfactual simulations, we quantify the importance of elasticities versus structural gaps in understanding cross-country differences in health indicators. We then explore the economics mechanisms (Section 6) and welfare issues (Section 7). Finally, section 8 concludes.

2 Illustrative Model

In this section, we present theoretical foundations for the idea that prices reflect inefficiencies on the health market. We relate the health wedge to informational inefficiencies and administrative costs in the supply of health services. We also show that it is possible to find a large class of demand for health services and health production function such that the GDP share of health spending is increasing in price and in income, as these predictions would be consistent with the US case for instance.

2.1 Setup

Denote by $s_g = \frac{p_g m_g}{y_g}$ the GDP share of health expenditures in country $g$ where $p_g$ is the price of health services relative to other goods, $m_g$ the quantity of health services and $y_g$ denotes income. Similarly, denote by $\pi(m_g)$ the fraction of the population in good health. Agents choose $m$ to maximize expected utility. This results in a demand function for medical services $m_g = M(y_g, p_g)$ with $M'_y > 0$ and $M'_p < 0$.

Given that $y_{US} > y_g$, Americans spend more on health care than any other country, $s_{US} > s_g$ for $g \neq US$ which either means that $p_{US} > p_g$ and/or that $m_{US} > m_g$. But the latter implies that Americans would be in better health than those from other countries, given that $\pi(m_g)$ is increasing in $m$. However, this is not what is observed. Differences in $y$ are unlikely to explain this pattern because income is higher in the U.S. and the income
elasticity is positive. Thus, we have to consider price differences. However, the extent of price differences will likely create a demand response which will lower demand for health services \((m)\). Hence, a theory that aims to focus on prices must explain the existence of health wedge, i.e. the determinant of the price gaps on the health market services, but also determine restrictions on the demand side of the health services to yield both higher share of health expenditure but lower fraction of population in good health.

2.2 A Stylized Model

We first abstract from general equilibrium effects and present a simple model of supply and demand for health services that leads to the observed facts. Price differentials may exist for various reasons. We focus on the inefficiency induced by the information gaps between providers and payers of health services, as well as the administrative costs induced by the size of the sector. These frictions induce an inefficient allocation of input specific to each country, generated by what we call the health wedge measured by a country-specific price gap. Given these frictions, we then build the demand side using a two-period model of the demand for health services and investigate equilibrium properties and restrictions on fundamentals allowing to match the share of the health expenditures in the total income and health levels across countries.

2.2.1 The Supply of Health Services: the health wedge

We focus on two key differences across countries which may explain differences in prices as suggested by Cutler and Ly (2011). First, we introduce informational frictions in the health sector: the quality of services offered by all providers are not perfectly observed by the payers (the insurance system). We assume that the quality can take two values \(q \in \{0, 1\}\). The organization cost allowing to provide \(q = 1\) is \(p_r > 0\). The payers can detect providers’ shirking behavior with probability \(\zeta \in [0; 1]\). Hence, the optimal contract for the payer is a price \(p_p\) for the quality of the service such that \(p_p = \frac{1}{\zeta} - p_r\), which ensures that \(q = 1\) at the equilibrium.\(^4\) The production function of the provider is thus

\(^3\)We use a highly stylized model in order to present the main arguments explaining why the price of health service is not a competitive price. For more detailed discussions on this point, see the surveys of Newhouse (1996), Dranove and Satterthwaite (2000) and Gaynor and Vogt (2000), or Gaynor and Town (2012).

\(^4\)As usual in the contact theory, this equilibrium price \(p_p\) is deduced from the equalization of the value of the honest provider (here \(p_p - p_r\) with \(q = 1\)) and the ones of the shirking provider (here \((1 - \zeta)p_p + \zeta \times 0\) with \(q = 0\)).

\(^5\)Another way to generate a gap between the effective price and the reservation price \(p_r\) (the production cost), is to introduce a bargaining between the payer and the provider. The Nash product is then given
where \( m \) denotes the quantity of health services supplied to the household, \( b \) is the quantity of services used by the providers, and \( z \) is the total factor productivity of the health sector. Using the equilibrium price contract implying \( q(p_p) = 1 \), it becomes \( m = zq(p_p)b = zb \). Second, we allow for administrative cost in the health system. For simplicity, we assume that administrative costs are sunk costs and proportional to the size of the production sector (\( \iota m \)). Health expenditures in units of consumption goods are \( pm \), given that \( p \) is the relative price of health care services. These expenditures are also the resources for the health sector. The profit of the health sector is \( \Pi_h = pm - C(m) \) with the cost function \( C(m) = \left( \frac{p_p}{z} + \iota \right) m \).

**Property 1.** The price gap of health services increases with informational frictions between providers and payers, as well as with the administrative costs.

**Proof.** The zero profit condition leads to equilibrium prices as function of two parameters, informal friction \( \zeta \) and administrative cost \( \iota \):  
\[
p = \frac{1}{z} \frac{p_p}{\zeta} + \iota \equiv \mathcal{P}(\zeta, \iota)
\]  
with \( \mathcal{P}'(\zeta, \iota) > 0 \) and \( \mathcal{P}'(\zeta, \iota) > 0 \).

Property 1 shows that the gap between US price \( p_{US} \) and the European \( p_E \) increases from \( \zeta \approx 0 \) (the extreme case with infinite informational fictions) to \( \zeta \approx 1 \): the larger the providers’ informational rent, the higher the price in economies with informational frictions: the health wedge increases with frictions. Moreover, when administrative cost increases, the price of the health services rise. This can be the case when the number of operators/intermediaries is uselessly large in the market, perhaps due to the administrative burden of handling the insurance reimbursement process. Frictions on the supply side of health services generate the health wedge, implying a price gap.

\[
(p_p - p_r)(R - p_p)^{1-\zeta}
\]  
where \( R \) is the marginal revenues of the payer. In this case, the equilibrium price is \( \zeta R + (1 - \zeta)p_r = p_p \). The larger the provider’s bargaining power (\( \zeta \)), the higher the price. See Gowrisankaran et al. (2015) or Ho and Lee (2017) for a detailed discussion on the bargaining between providers and payers, in a general framework where insurers bargain also with the consumers.

6We interpret these informational frictions as the physicians’ effort at work that is not perfectly observed by the hospital manager. Then, the larger the physicians’ informational rent, the higher the price. This can be consistent with the findings of (Cutler and Ly, 2011) underlining that specialist U.S. physicians earn 5.8 times what the average worker does, compared to the non-U.S. average of 4.3 times.

7In the case where the markup price is determined by a bargaining between payers and providers, two cases can arise: the US system where the provider’s bargaining power is large in a decentralized market, and the European case where, in all countries, a public system reduces the provider’s bargaining power, by setting the price at its lowest level.
2.2.2 The Demand for Health Services

The household’s behavior is summarized by an intertemporal choice of consumption (two-periods model), where the uncertainty comes from both the income level and the health status in the second period of life.

- **Assumption 1.** The probability of being in good health is \( \pi(m) \), with \( \pi' > 0 \) and \( \pi'' < 0 \). The utility benefit of good health is \( \phi \).

- **Assumption 2.** Risk: the second-period income is drawn in the lotteries s.t. \( \tilde{y} \in \{0, \frac{y}{1-\varpi} \} \) with probabilities \( (\varpi; 1-\varpi) \Rightarrow E[\tilde{y}] = y \) and \( \sigma_{\tilde{y}}^2 = 2(\varpi y)^2 \).

- **Assumption 3.** Using the endowment \( y \), each household produces inputs \( b \) used in the health sector. The production cost of \( b \) is given by \( C(b) = \frac{1}{\vartheta} b^\vartheta \), with \( \vartheta > 1 \). This inputs is sold to the health sector at price \( p_p \). The net cost from this activity is \( F = \max_b \{ p_p b - C(b) \} \). Hence, we have \( F = \frac{\vartheta-1}{\vartheta} pm \) at the equilibrium.

- **Assumption 4.** Financial arrangements: One risk-free asset \( a \), yielding the interest rate \( r \). Financial constraint holds: \( a \geq 0 \).

The problem of the household is defined by

\[
\mathcal{U} = \max_{m,a \geq 0} \{ u(c) + \beta E[u(d) + \pi(m)\phi] \}
\]

s.c. \[
\begin{align*}
\quad & c = (1-\tau)y + F - a - \mu pm \\
\quad & d = (1+r)a + \tilde{y}
\end{align*}
\]

where \( c \) and \( d \) are consumptions of the first and second period, \( \mu \) is the co-insurance rate, \( r \) the interest rate and \( \tau \) is the tax used to fund the health insurance system.\(^8\)

\(^8\)The FOC of the program are provided in Appendix B.
2.2.3 Equilibrium: the impact of health and efficiency wedges

Using the budgetary constraint of the Health insurance system, \( \tau y = (1 - \mu) pm \), the equilibrium is defined by

\[
\begin{align*}
    u'(y - a - \frac{1}{\vartheta} pm) &= \beta(1 + r) E[u'(1 + r)a + \bar{y})] \\
    \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pu'(y - a - \frac{1}{\vartheta} pm) &= \beta \pi'(m) \phi \\
    \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pu'(y - \frac{1}{\vartheta} pm) &= \beta \pi'(m) \phi
\end{align*}
\]

where (1) and (2) are the FOC w.r.t. \( a \) and \( m \) for the households having \( a > 0 \), whereas (3) is the FOC w.r.t. \( m \) for the households having \( a = 0 \).

Assume that the uncertainty and the incomplete financial market lead to \( a > 0 \). For simplicity, we also assume that the efficiency wedge can be viewed as a reduction of the average income \( y \). Equation (2) gives \( m = M(p, y, a) \), whereas equation (1) gives \( a = A(p, m, y, \sigma_{\bar{y}}) \). Given these decision rules, we can deduce the properties of the shares of health expenditure \( s = \frac{pm}{y} \).

**Property 2.** If \( |\varepsilon_{\pi'}| > 1 \), then, when \( a > 0 \), \( \frac{\partial s}{\partial p} > 0 \) and \( \frac{\partial s}{\partial y} > 0 \). Moreover, we always have \( \frac{\partial s}{\sigma_{\bar{y}}} < 0 \).

**Proof.** See Appendix B. When \( a = 0 \), the condition is given by equation (21).

Property 2 implies that the model can predict that \( s \) increases when efficiency wedge declines (\( y \) increases) or when health wedge rises (\( p \) increases). The high efficiency wedge in the US (\( y_{US} > y_E \)) or/and large health wedge (\( p_{US} > p_E \)) can explain \( s_{US} > y_E \). We deduce that the relative importance of these wedges can be identified through their impact on the GDP share of medical expenditures that determines the fraction of the population in good health (\( \pi(m) \)): this is a priori indeterminate, even in this stylized model.

Notice that \( s \) decreases with income risk. Hence, the impact of higher income (\( y_{US} > y_E \) implies \( s_{US} > s_E \)) can also be compensated by higher income risk, as in the US economy, (\( \sigma_{\bar{y},US}^2 > \sigma_{\bar{y},E}^2 \) implies \( s_{US} < s_E \)). Nevertheless, if the efficiency in the health production function of a marginal unit of health expenditures is sufficiently high (\( |\varepsilon_{\pi'}| > 1 \)), then Property 2 shows that high price (large health wedge) can help match \( s_{US} > s_E \). Indeed, the forces that drive the share of the health expenditures in total income to be

\[ The estimation of the model will remove this indeterminacy. \]

9
higher in the US than in Europe (a high average income, and high prices) must simply dominate the need for insurance coming from the precautionary saving motive, acting in opposite direction. Given that Property 2 yields increasing demand for health services as income expands, a corollary is that rising income inequality inside a country (say a mean-preserving spread) will lead to rising health inequalities.

This equilibrium analysis without aggregate macro effects shows that more savings can lead to less health expenditures at the individual level. Nevertheless, we need to look at general equilibrium effects because this additional saving leads to more capital, and thus, more aggregate production, leading to higher incomes. This feedback effect of the general equilibrium framework can then reinforce the demand for health. The crowding out effect, driven by the precautionary saving, can be compensated by an income effect driven by a more capital-intensive economy. The general equilibrium analysis can also highlight feedback effects from tax endogenous changes.

3 General Equilibrium Model

3.1 Households

We have a continuum of agents who are heterogeneous with respect to their productivity level $e$, health status $h$ and asset holding $a$. The productivity levels $e$ are determined by an exogenous stochastic process. In contrast, health status and asset are endogenous outcomes of the model.

**Preferences.** Households value both their consumption and their health status. Households’ preferences can be described by the following standard expected discounted utility

$$
\sum_{t=0}^{\infty} \beta^t \sum_{e'} \sum_{h'} p(e'|e)p(h'|h,m)u(c,h)
$$

where $0 < \beta < 1$ is the time discount factor, $c \geq 0$ is consumption and $h$ current health status. Next period’s variables are denoted with a prime. As in DeNardi et al. (2010), health can be either good ($h = 1$) or bad ($h = 0$). The object $p(h'|h,m)$ denotes the probability of being in health status $h'$ next period, given the current health status $h$ and health services $m$. This transition is determined by a health production function that

10Earning risk leads to a positive precautionary saving: savings ($a > 0$) then crowds out health expenditures.
depends on health services \( m \). The exogenous transition probability from the current productivity level \( e \) to next period’s level \( e' \) is denoted \( p(e'|e) \). In contrast, \( p(h'|h,m) \) is endogenous in the model. We assume that the instantaneous utility is additive in consumption \( c \) and health \( h \):

\[
u(c, h) = \frac{c^{1-\sigma}}{1-\sigma} + \phi h. \tag{5}
\]

with \( \phi > 0 \) the utility benefit of good health. Hence, the rate at which the marginal value of health increases with income depends on how fast the marginal utility of consumption is declining.

**Health Production.** Each agent can spend his resources on consumption \( c \) and health services \( m \). Health services \( m \) improve the probability of being in good health next period. In addition, we assume that the function that maps health services in health status is

\[
p(h' = 1|h, m) = 1 - \exp(-(\alpha_0 m + \alpha_{1h} + \eta r_b)) \tag{6}
\]

with \( \alpha_0 > 0, \alpha_{1h} > 0 \). With \( \alpha_{11} > \alpha_{10} \), the probability to be in good health next period is higher for agents who are currently in good health. Finally, we assume that the individual probability to be in good health depends on the country specific baseline health, denoted \( r_b \) for "risky behaviors", for example obesity or other environmental factors.

**Resource Constraint.** Labor income is affected by an idiosyncratic stochastic process that determines the value of efficient labor \( e \). \( e \) is the sum of an AR(1) permanent shock with parameters \( (\rho_e, \sigma_e) \) and a transitory shock with standard deviation \( \sigma_v \). Market incompleteness prevents agents from insuring against the idiosyncratic risk. In addition to labor income, agents collect capital income from asset holding \( a \), with risk-free return \( r \) rate. Next period’s asset \( a' \) is then

\[
a' = a(1 + r) + we(1 - \tau) - c - \mu pm \tag{7}
\]

Income is spent on labor income taxes (with flat-tax rate \( \tau \)), consumption \( c \) and health services \( m \). The variable \( p \) refers to the relative price of health services with respect to consumption good. \( \mu \) refers to the co-insurance rate. This captures the fraction of out-of-pocket expenditures in total health expenditures. In addition, assets have to satisfy a borrowing constraint

\[
a \geq 0. \tag{8}
\]
Endogenous Demand for Health Services and Savings. As far as each agent is concerned, the state variables are the realizations of the household-specific shock, $e$, the stock of wealth, $a$, and health status $h$. The dynamic program solved by an individual whose state is $(a, h, e)$ is

$$
V(a, h, e) = \max_{m, c} \left\{ \frac{c^{1-\sigma}}{1-\sigma} + \phi h + \beta \sum_{e'} \sum_{h'} p(e'|e)p(h'|h,m) V(a', h', e') \right\} 
$$

subject to equations (7) and (8). $V$ denotes the individual’s value function. The solution that solve this problem is a set of decision rules that map the individual state into choices for consumption and health services. We denote these rules by $\{c(a, h, e), m(a, h, e)\}$.

3.2 Good-Producing Firm

Production $Y$ is characterized by constant returns to scale using aggregate capital $K$ and labor $N$ as inputs:\footnote{Aggregate employment is exogenous and result from the Markovian representation of the AR(1) process on productivity $e$.}

$$
Y = AK^\alpha N^{1-\alpha}
$$

$A$ captures technological factor productivity and $0 < \alpha < 1$ the capital share in GDP. The representative firm operates under perfect competition such that profit maximization leads to

$$
\begin{align*}
    r &= \alpha A \left( \frac{N}{K} \right)^{1-\alpha} - \delta_k \\
    w &= (1 - \alpha) A \left( \frac{K}{N} \right)^\alpha
\end{align*}
$$

with $w$ the wage rate, and $\delta_k$ capital annual depreciation rate\footnote{In the model, household’s utility is not a function of leisure. As a result, employment level is independent of factor prices $w$ and $r$, distribution of health and asset and the aggregate capital stock. Aggregate employment can be directly computed from the Markov process governing the evolution of labor productivity $e$.}.
3.3 Health Insurance System

The government provides health insurance. Health insurance reimburses medical expenditures using proportional taxes on labor income:

$$\tau wN = (1 - \mu)p \sum_e \sum_h \sum_a m(a, h, e)\lambda(a, h, e)$$  \hspace{1cm} (13)

where $\lambda(a, h, e)$ is the stationary distribution of individuals across individual states $(a, h, e)$. Given the co-insurance rate $\mu$, the tax rate $\tau$ must finance expenditures. Using Equation (12), leading to $wN = (1 - \alpha)Y$, we deduce from Equation (13) that the tax rate is proportional to the GDP share of health expenditures.

The supply of health services is the same as in section 2.2. Hence, households transform a part of their consumption goods in inputs supplied to the health sector. This sector does not generate profit. The price of health services depends on physicians' informational rent, given the administrative costs.

3.4 Definition of Equilibrium

A steady-state equilibrium for this economy is a household value function, $V(a, h, e)$; a household policy, $\{c(a, h, e), m(a, h, e)\}$; a Health insurance system, $\tau$; a stationary probability measure of households, $\lambda$; factor prices, $(r, w)$; and macroeconomic aggregates, $K, N$, such that the following conditions hold:

(a.) Factor inputs, tax revenues, and transfers are obtained aggregating over households:

$$K = \sum_e \sum_h \sum_a a\lambda(a, h, e), \quad N = \sum_j e_jN_j$$

(b.) Given $K$ and $N$, factor prices $r$ and $w$ are factor marginal productivities (Equations (11) and (12)).

(c.) Given $r, w, \tau, \mu, p$, the household policy solves the households' decision problem described in (9).

(d.) Tax rate $\tau$ adjusts such that health insurance budget constraint (13) is satisfied.

\textsuperscript{13}In Equation (7), these costs do not appear (see Assumption 3 in section 2.2). Indeed, without any information on these costs, we assume $\vartheta \to 1$ for the estimation procedure of the model.
(e.) The goods market clears: \( Y = \sum_e \sum_h \sum_a c(a, h, e) + pm(a, h, e)\lambda(a, h, e) + \delta_k K. \)

(f.) The measure of households \( \lambda(a, h, e) \) is stationary.

## 4 Data and Estimation

We estimate the model on eight countries: the US, Sweden, Denmark, the Netherlands, Germany, France, Italy, and Spain. We chose these countries because both OECD data and comparable longitudinal micro data was sufficiently available to estimate parameters. There is considerable heterogeneity in both outcomes and institutions across these countries. Due to the existence of common parameters, we estimate simultaneously all countries.

We use a two-step strategy to estimate/calibrate parameters of the model for \( g = 1, \ldots, G \) countries. Firstly, we describe how the auxiliary parameters are set using external information (Micro and Macro datasets). In a second step, we use a method of simulated moments approach to estimate remaining parameters.

### 4.1 Auxiliary Parameters

We use different sources of data to obtain auxiliary parameters\(^{14}\).

**Income Risk.** Estimating income processes requires panel data. For the United States, we use eight years of the Panel Study of Income Dynamics (PSID) data (1990 to 1997). Data after 1997 is collected every two years, complicating the estimation of the income process. For European countries, we use eight years of the European Community Household Panel (ECHP) from 1994 to 2001. We first net out the effect of age from income by regressing an household’s total net income on a flexible age polynomial and obtain residuals. We use after-tax household income as it allows for differences across countries in social programs that may mitigate income risk. For the error component, we assume the following process

\[ \eta_t = e_t + u_t \quad \text{with} \quad e_t = \rho e_{t-1} + \nu_t \]

where \( \nu_t \) is the innovation to the persistent component, distributed \( N(0, \sigma_e^2) \), whereas the transitory component \( u_t \) is distributed \( N(0, \sigma_u^2) \). Table\(^2\) shows the estimates of the

\(^{14}\)Details on data sources to construct variables can be found in Appendix A.
income process. Overall, the variances of the transitory components are close enough. The estimates of the stationary variance of the permanent component are larger in the US than in European countries. We find considerable persistence in income, with autocorrelation coefficients ranging from 0.9697 (Netherland) to 0.9798 (Spain). The main source of the difference in income risk is the scale of the innovation to permanent income in the U.S. The variance of the permanent shock is twice as large in the U.S. compared to Europe.

**Risky Health Behaviors.** To calibrate the variable \( r_{b,g} \), we use the Study of Health, Ageing and Retirement in Europe (SHARE) and the Health and Retirement Study (HRS) for the U.S for the year 2004. We use an indicator of obesity for a body mass index larger than 30 and compute the average obesity rate of those between the ages of 50 and 75. Obesity is a good summary measure of past physical activity as well as eating habits. It is unlikely to be the result of past health services or depend on parameters of the health insurance system. Body mass peaks around those ages which justifies looking at this age range as a summary measure. In Figure 1, we can observe that Americans have the largest share of the population which are obese: more than 30% of this population has a BMI over 30, followed of Spain, with 25%, while other European countries tend to have less than 20% of their population with a BMI over 30.

**Co-insurance Rates.** We use aggregate data from OECD Health Data to compute the co-insurance rate \( \mu \) across countries for 2005. We define the co-insurance rate as private out-of-pocket household expenses as percentage total expenditure on health. This data is not available for Italy. We use the World Bank data for this country with a similar definition. OECD numbers for the other countries are very similar to the World Bank numbers. Figure 2 shows the differences across countries. Spain and Italy have large share of out-of pockets over total health expenditures, while France and the Netherlands have the smallest shares. The US ranks in the middle.

**Other Parameters.** We use Penn World Table Feenstra et al. (2015) in order to calibrate the country-specific shares of capital (\( \alpha \)) and the depreciation rates (\( \delta_k \)). The values reported in Table 3 give the estimates for the period 1990-2005. The share of capital in production (\( \alpha \)) is between 0.3399 (Deutschland) to 0.4379 (Italy), the value for the US being 0.3553. In the case of the depreciation rate (\( \delta_k \)), our estimates range between 0.0346 (France) to 0.0483 (Sweden), with a value of 0.0401 for the US.
4.2 Method of Simulated Moments

We have three groups of structural parameters to estimate. The vector of preference parameters is given by \( \{ \beta, \sigma, \phi \} \). Preference parameters are kept constant across countries. The second group consist of parameters of the health production function \( \{ \alpha_0, \alpha_1, \psi, \eta \} \), also constant across countries. Finally, we have two country specific parameters capturing gaps on TFP in producing goods and on the price of health services, \( \{ A_g, p_g \} \). These gaps are estimated by taking as the reference country the US: we thus normalize \( A_{US} = 1 \) and \( p_{US} = 1 \). Hence, the estimation provide measures of the structural gaps across economies. Notice that only cross-country differences can be identified using our data. The structural parameter vector to estimate is given by

\[
\Theta = \{ \beta, \sigma, \phi, \alpha_0, \alpha_1, \psi, \eta, \{ A_g \}_{g \neq \text{US}}, \{ p_g \}_{g \neq \text{US}} \} 
\]

The structural estimation of the model allows to quantify the mechanisms of propagation of the revealed country-specific gaps (price and TFP).

Denote the set of country specific auxiliary parameter \( \chi_g \) and \( \chi = \{ \chi_1, ..., \chi_G \} \). For each country, consider a set of \( M_g \) moments (or targets) denoted

\[
m_g(\Theta, \chi_g) = \{ m_{g,1}(\Theta, \chi_g), ..., m_{g,M_g}(\Theta, \chi_g) \}. \tag{14}
\]

The moment vector over all countries is given by

\[
m(\Theta, \chi) = \{ m_g(\Theta, \chi_g), ..., m_G(\Theta, \chi_G) \}'. \tag{15}
\]

The total number of moments is \( M = \sum_g M_g \). Given \( \Theta \) and \( \chi \), we can simulate these moments from the model using \( S \) draws for income and health. Denote the simulated moment vector, \( \tilde{m}_S(\Theta, \chi) \).

We combine a set of aggregate moments and moments derived from micro data. We convert all monetary amounts to US dollars PPP adjusted. In order to partially pin down risk aversion, we use the ratio of capital to GDP, \( K/Y \). As previously , we use data form Penn World Table [Feenstra et al. (2015)] over the years 1990 to 2005. We also use GDP per capita relative to US, \( \tilde{Y}_g = Y_g/Y_{US} \) to identify TFP gaps. We also use data Penn World Table over the years 1990 to 2005. A third moment involves the share of health expenditures as a fraction of GDP, \( s = \frac{pm}{Y} \). We use information from the OECD Health data from 2003-2007.\footnote{The OECD health data does not include data for Italy. We use the corresponding value from the}
Moments involving micro data help estimate the health production function. We use SHARE and HRS data for 2004 and 2006 to estimate health state transitions. We also use those data to estimate the gradient of health status by levels of wealth (capital). This further strengthens the identification of the health production function. Health transition probabilities depend only on wealth through the choice of medical consumption. Hence, the variation in observed health status by wealth level helps identify the productivity of medical consumption, \( \alpha_0 \). We use self-reported health which is asked in both surveys. Of course, one could be interested in considering multiple dimensions of health but the computational burden of doing this prohibits this possibility. Self-reported health is reliable overall health measure predictive of mortality and use of physician services [Miilumpalo et al. 1997]. Respondents are asked to rate their health from poor to excellent using 5 levels. We convert this measure to a binary indicator where 1 denotes very good or excellent health and zero other cases. It is well-known that self-reported health varies across countries in part due to differences in reporting scales [Jurges 2007, Kapteyn et al. 2007]. We correct for reporting scales by following the strategy proposed by Jurges (2007).

We estimate a logit model relating self-reported health to more objective measures of health and country fixed effects. The country fixed effects reflect different reporting styles. The average predicted probabilities of being in good health is given by \( \tilde{p}(X_g, g) \) where argument \( g \) denotes that the fixed effect of country \( g \) is used in predicting good health in that country and \( X_g \) denote a set of objective health conditions. In order to correct for reporting styles, we set the fixed effect to a base country, say 1, which is Germany in our analysis, and use \( \tilde{p}(X_g, 1) \). We can do something similar for transition rates. Denote by \( \tilde{p}_{jk}(X_g, g) \) the joint probability of being in state \( j \) at time \( t \) and \( k \) at time \( t+1 \). We can predict those probabilities using a logit where the dependent variable equals one when states \( j \) and \( k \) are observed over the two waves and use both set of observed health conditions at \( t \) and \( t+1 \). Using Bayes rule, the transition rate can be computed as \( \tilde{p}_{k|j}(X_g, g) = \frac{\tilde{p}_{k|j}(X_g, g)}{\tilde{p}(X_g, g)} \). Hence, the corrected transition rate is given by \( \tilde{p}_{k|j}(X_g, 1) \). Since transition probabilities sum to one conditional on state \( j \), we use \( \tilde{p}_{1|0}(X_g, 1) \) and \( \tilde{p}_{1|1}(X_g, 1) \) as moments. Given that surveys measure health every two years, we recompute annual transition rates, solving \( \tilde{\Pi}_2 = \tilde{\Pi}_1^2 \) for \( \tilde{\Pi}_1 \) where \( \Pi_q \) is the markov transition matrix for \( q \) year transitions.

To compute the health gradient, we first create bins based on the distribution of net wealth in 2005 PPP adjusted US dollars. We use the bins \([0 - 100k, 100k - 250k, 250k - 450k, 450k +] \). We compute the average adjusted predicted probability of being in good health conditions.
health within each bin, \( \tilde{p}_{q,g}(X_g, 1) \) for \( q = 1, 2, 3, 4 \). We use as moments the relative probability using the first bin as a base: 
\[
\tilde{p}_{q,g}(X_g, 1) = \tilde{p}_{q,g}(X_g, 1) / \tilde{p}_{1,g}(X_g, 1)
\]
for \( q = 2, 3, 4 \). The vector of moments for each country is given by
\[
m_g = \left\{ K_g / Y_g, \tilde{Y}_g, s_g, \tilde{p}_{1|0}(X_g, 1), \tilde{p}_{1|1}(X_g, 1), \tilde{p}_2(X_g, 1), \tilde{p}_3(X_g, 1), \tilde{p}_4(X_g, 1) \right\}
\]
where \( \tilde{Y}_g \) is not included for the U.S.

We report in Figure 3 the computed moments from the data. In terms of capital-output ratio, Italy, Spain and the US have the highest capital to GDP ratio while Sweden has the lowest.\(^{16}\) GDP per capita is in general 40 to 20% lower in European countries relative to the U.S. The U.S. spends 14% of GDP on health while only three countries rise above 10% in Europe (France, the Netherlands and Germany). In terms of transition rates into good health, the U.S. ranks 6 out of 8 countries for transition rates conditional on being healthy and third for transition rates from bad health to good health. Finally, the health gradient by wealth level is much steeper in the U.S. than in any European country. For example, those in the highest wealth level (more than 450,000$ US PPP) had a probability of being in good health 34% higher than those in the lowest wealth group (less than 100,000$ US PPP). Denote by \( m_D(\Theta_0, \chi_0) \) the vector of moments computed from the data, where \( \Theta_0 \) and \( \chi_0 \) are the true values of the structural and auxiliary parameters.

The difference between simulated moments and moments from the data is given by 
\[
\tilde{g}_S(\Theta, \chi) = [\tilde{m}_S(\Theta, \chi) - m_D(\Theta_0, \chi_0)]
\]
The MSM estimator \( \hat{\Theta} \) is the solution to the minimization problem
\[
\min_{\Theta} \tilde{g}_S(\Theta, \chi)'W_N\tilde{g}_S(\Theta, \chi) \tag{17}
\]
where \( W_N \) is a positive definite weighting matrix which depends on the data. We choose a diagonal matrix with elements equal to the inverse of the variance of each moment as a weighting matrix. For moments involving microdata, we use the bootstrap to find the variance while we use the time-series variation to compute the variance for aggregate moments.

Because the function is not a smooth function of the parameters we follow the method proposed by Chernozhukov and Hong (2003). Instead of minimizing directly the objective

\(^{16}\)Overall, the shares of housing capital are larger in European countries than in the US (more than 45% in European countries vs. 39% in the US). These shares of housing capital are especially high in Italy and Spain (48% and 46.5% respectively). See Kamps (2006) and Backus et al. (2008) for more details on this decomposition of capital.
function, we construct a Monte Carlo Markov Chain which converges to a stationary process with a distribution whose mode is equivalent to the MSM estimator\(^{17}\).

### 4.3 Estimation Results

#### 4.3.1 Estimation results

Estimation results are reported in Table 4. There are parameters common to all countries \(\{\sigma, \beta, \phi, \psi, \delta_{h1}, \delta_{h2}, \eta\}\) that summarize the preferences and health technology. The other set of parameters such as price and TFP gaps are country specific.

The coefficient of relative risk aversion, \(\sigma\), is estimated at 3.12 which is within the range of estimates found in the literature for precautionary saving models. The discount factor \(\beta\) is estimated to be 0.823 which implies an annual discount rate of 21\%. This level of impatience is large compared to existing estimates. Such a high discount rate is needed in order to match simultaneously the capital-to-GDP ratio and the health gradient. With risk aversion at 3, the capital-to-GDP ratio would have been too large for standard discount rates (above 5). Similarly, the health gradient with wealth would not be steep enough at values of risk aversion below 2 as the curvature of the utility function has a positive effect on how much health spending is responsive to income/wealth. The marginal utility of being in good health is found to be 0.086 which is large, given that the curvature of the utility function implies low values of utility from consumption. The marginal productivity of health investment \(\alpha_0\) is found to be large, at 2.339. This value is reminiscent of Property 2: large marginal benefits from investing in health is necessary to allow the model to replicate a data-consistent GDP share of health expenditures.\(^{18}\)

In order to gauge the plausibility of our parameter estimates, we compute US elasticities generated by the model. The US coinsurance elasticity of health expenditures \(pm\) is 0.4 (in partial equilibrium), which is larger than estimates from the RAND Health Insurance Experiment (at 0.2). This might be due to the scope of the experiment: in the RAND experiment, out-of-pocket expenditures were subject to a limit, which is not the case in our coinsurance rate counterfactual. Our US income elasticity is 1.53 (in partial equilibrium), which lie between the macro estimates (close to 1) and Hall and Jones (2007)’s\(^{18}\).

---

\(^{17}\)See Appendix D for more details.

\(^{18}\)In addition, the stylized model (equation (21) in Appendix B) suggests that, in simple framework without savings, the model would replicate a data-consistent GDP share of health expenditures with a price elasticity lower than 1 in absolute value and an income elasticity larger than 1. Interestingly, we will show in Section 5 that these conditions also hold in our model.
findings (higher than 2). We estimate strong state-dependence in health transition probabilities with the probability of being in good health next period being much larger if one is already in good health than if it is not ($\alpha_{11} > \alpha_{10}$). Finally, we estimate a negative effect of the obesity rate on the probability of being in good health.

**Estimated (in)efficiencies across countries: the structural gaps.** The estimation procedure allows to measure the cross-country (in)efficiencies in terms of health prices and TFP: the price and TFP gaps.

As for the price of health, price gaps are large. All countries are characterized by lower values than the US economy (normalized to one): the estimation reveals an inefficiency gap for the US. France, the Netherlands and Denmark display the cheapest health prices relative to the US (around .3 - .4). Germany, Italy, Sweden and Spain constitute a second group with health prices hovering around .6 relative to the US. Our results are consistent with the health price dispersion in Table 1. As for TFP gap, it captures the heterogeneity in economic development across countries. The US, Denmark and Sweden appear as the most advanced economies while Spain and Italy lie at the other end of the TFP distribution. The estimated inefficiencies suggest that countries such as the US, with high price of health and high TFP, could use their advanced technological development (TFP efficiency) to pay for expensive price services (health market inefficiency). Other countries have high TFP and cheap price of health, such as Denmark. Estimated gaps show that price dispersion appear very large, even larger than gaps in TFP, thereby suggesting that our focus on cross-country differences in inefficiency of the health sector is supported by the data. Nevertheless, even if these structural gaps are large, it is only the product of elasticities with respect to each structural gap by the estimated size of these gaps that explain the differences between countries. Hence counterfactual experiments must be performed to evaluate the contribution of each gap at the equilibrium.

4.3.2 Model fit

Figure 4 shows that the model succeeds in fitting the share of health expenditures ($p_g m_g / y_g$, the Spearman correlation is 0.9839). The model slightly underestimates the transition from bad to good health and overestimates the transition from good to good health.

---

19 We compute this income elasticity using the TFP counterfactual in Section 5 and make sure to compute the average response of $p m$ to a 1% increase in total income (including labor earnings and financial income).
the US. The model fit is worse for the transition from bad to good health (the Spearman correlations are -0.6720 for \( p_{1|0}(X_g, 1) \) and 0.7611 for \( p_{1|1}(X_g, 1) \)). Nevertheless, the model reproduces the empirical evidence that the share of health expenditures is the highest in the US but with a low percentage of people in good health (only Italy and Spain perform worse) as well as in terms of health inequalities (the highest in the US). At the opposite, Denmark performs much better: for the low share of health expenditures (only Italy and Spain have lower \( s \)), the percentage of people in good health is the largest, and the health inequalities are the lowest. Italy and Spain seems to be the worse country, with high health inequality in Spain. Finally, Germany, France, Netherland and Sweden have similar outcomes, with higher shares of health expenditure in Germany and France: the model fits this feature. Figure 4 shows that the simulated moments are very close to their empirical counterparts: the capital/output ratio and the GDP differences are well reproduced (the Spearman correlations are respectively 0.4871 and 0.9702). Finally, health inequalities, measured by the health gradient are in accordance with the main features of the data (even if the Spearman correlation are low, respectively 0.1233, 0.1966 and 0.45 but always positive), showing than the US is the country where inequalities are the largest, whereas Nordic countries are those where they are the lowest.

5 In search of the main cause for cross-country differences in health: price versus TFP

The main challenge of this paper is to evaluate the respective contributions of gaps in prices and TFP on cross-country health performance. For an individual agent who buys health services, the price can be affected by the health insurance system, here summarized by the coinsurance rate \( \mu \). Hence, in addition to the price and TFP gaps, we will also evaluate the contribution of the gap induced by dispersion in coinsurance rates on health performances. These exogenous changes are indexed by \( x \in \{ p, A, \mu \} \).

We perform counterfactual experiments in which the model predicts what would have happened in one country when faced with the price of health, TFP or coinsurance rate from another country.

---

\[ \frac{\partial y}{\partial x} = \frac{y_1 - y_0}{(y_1 + y_0)/2} \] and \[ dx/x = \frac{x_1 - x_0}{(x_1 + x_0)/2} \]
The leading role of health price gap in the US

We focus below on the two indicators put forward in the introduction: the GDP share of health expenditure and the fraction of the population in good health. We first show below that only price differences can help close the cross-country gap along both dimensions.

This conclusion is based on counterfactual experiments. For the purpose of clarity, we focus first on the case in which the US is characterized by French health prices, TFP or coinsurance rate. These different experiences mean that the US experience (i) a decline in their inefficiency gap on the health service market when the price is set to the one of French market ($p_{US} > p_{FR}$), or (ii) a decline in their efficiency gap in technology when TFP is set to French level ($A_{US} > A_{FR}$), or (iii) an increase of the inefficiency gap induced by a lower coinsurance rate when the French rate is applied in the US ($\mu_{US} > \mu_{FR}$). Figure 5 reports the model predictions, where $dx/x < 0$ for $x \in \{p, A, \mu\}$.

In the data, the US display a larger GDP share of health expenditure than the French economy. By imposing the French price in the US economy, the model generates a decline in the share of health expenditure (panel (a) of Figure 5 where $dy/y$ is negative). This predicted decrease in the share of health expenditures after a reduction in price gap explains a large part of the observed differences between US and France ($dy/y$ is also negative in the data). In panel (b) of Figure 5, we consider the fraction of the population in good health. With a fall in the price gap, the model predicts that more Americans are in good health. In the data, Americans are in worse health than French people. Hence, the model successfully accounts for an increase in the fraction of the population in good health when the price gap is low. Hence, the price counterfactual also accounts for the observed differences measured with the data. This suggests that the price gap can explain the US specificities with respect to the share of health expenditure and the fraction of the population in good health. In contrast, neither TFP nor coinsurance counterfactuals are consistent with the data along both dimensions: by switching to the French TFP, the model correctly predicts that the GDP share of health expenditures declines. However, the counterfactual TFP leads the model to move further away from the data on the fraction of the population in good health. For the coinsurance counterfactual, this is the reversed: the French coinsurance rate in the US economy help close the gap with respect to the data on the fraction of people in good health, but widens the gap with the data on

---

22 Among the many possibilities of counterfactual experiments in our sample of 8 countries, we choose this example because our estimation suggests that these countries are very different in terms of price of health, TFP, and coinsurance rate.

23 A low coinsurance rate is interpreted as a large inefficiency gap because it reduces perception of prices by individuals.
the GDP share of health expenditures.

How can we explain this result? The price counterfactual implies the largest effects on the share of health expenditure and on the fraction of the population in good health, compared to TFP and coinsurance. These sizable quantitative results come from the combination of the economic propagation mechanisms in the model (captured by the elasticity) and the size of the gaps in price of health (price inefficiencies), in TFP (productivity efficiency) or in the health insurance system (coinsurance gap) that are uncovered by our estimation procedure. Does the leading role of price come from a strong sensitivity of behaviors to small gaps or a moderate sensitivity of behaviors but large gaps? Figure 5 shows that the price or TFP elasticities are of rather similar magnitude in the US. In addition, the price gaps are larger than TFP gaps: inefficiencies in the health sector are larger than the technological gaps across countries. Hence, the primary driver for cross-country health differences lies in the large price gaps. Notice that the impact induced by the health insurance system (the OOP gap) is larger than the one induced by the TFP gap. In section 5.2, we argue that these conclusions are relevant for other European countries. We explore the economic mechanisms behind these results in section 6.

5.2 The health price gaps play also the central role in Europe

In this section, we present the predicted share of health spending and fraction of the population in good health when European economies are faced with US price, TFP or coinsurance rate. Obviously, the size of gaps are not identical across European countries. This exercise allows us to test the robustness of our conclusions with respect to various estimations of the country-specific structural gaps. In addition, the elasticities of the model are also country-specific because each country has a specific steady state. Hence, these counterfactuals can be viewed as a robustness check of the leading role of the price in the explanation for the cross country differences in health.

We present in Table 5 the predicted share of health spending (s, panel (a)) and fraction of the population in good health (p(H = 1), panel (b)) when European economies are faced with US price (columns (1)-(3)), TFP (columns (4)-(6)) or coinsurance rate (columns (7)-(9)). As in Figure 5, we decompose the total effect of the counterfactual experiment in elasticities (ε, columns (1),(4),(7)), inefficiency gaps (in prices column (2), in TFP column (5), and coinsurance rate column (8)) and predicted effect on share of health spending and fraction of population in good health (columns (3), (6), (9)). Table 5 allows to identify the source of differences across countries. Price elasticities of the shares of health expenditure
lie between 0.25 and 0.4, which is lower than in US (0.55). They are larger for France and Netherland where they are equal to 0.68 and 0.75 respectively. These results underline the large heterogeneity of market adjustment across countries, despite the common structural parameters. Concerning the price gaps there are two groups: the first composed of Spain, Italy, Deutschland and Sweden where the price gaps are between 0.35 and 0.45, and the second composed of Denmark, France and Netherland where the price gaps are larger (0.7, 0.73 and 0.81 respectively). Hence, the combination of a large price elasticity with a large price gap leads France and Netherland to be the countries where the shift to the US price increases the share of health expenditure the most. At the opposite, Italy and Spain combine low price elasticities with small price gaps leading to the smallest changes in the share of health expenditure.

If we compare these results with those obtained for the TFP gap, the main difference lies in the gap sizes: they are smaller than the price gaps. Nevertheless, the magnitudes of the country-specific TFP-elasticities with respect to the GDP share of health expenditure are larger than the price-elasticity. As in the case of the US, the price gap is the main source of difference between European countries and the US. Hence, the conclusion based on the US economy and showing that the health prices lie at the heart of the cross country differences seems to be robust: it is not dependent of the country chosen as the reference, even if the order of magnitude changes because the elasticities are country-specific. Finally, we note that, as for the US case, the impact of the coinsurance gaps is more important than that of the TFP gaps in explaining the cross country differences of the share of health expenditure. Estimates show that behaviors (here synthesized by elasticities) are very sensitive to changes in co-insurance rates. However, apart from Italy and Spain, the heterogeneity of the health insurance system (coinsurance gaps) combined with the sensitivity of the agents to these gaps, has a smaller impact on the GDP share of health expenditure than on health price differences.

Table 5 also allows to assess the impact of these three gaps (price, TFP and coinsurance rate) on the percentage of people in good health. We observe then that for 4 out of 7 countries the price gap dominates the TFP gap. For this group of country (Denmark, France, Netherland, Sweden), the result comes again from the large size of the estimated price gaps. For the three last (Deutschland, Italy and Spain), the large impact of the TFP comes from a high sensitivity to TFP variation for Deutschland (the size of the TFP gap being small for this country), whereas for Italy and Spain the result is explain by the large size of the TFP gap. Finally, the sensitivity of the performance of the health system seems rather insensitive to heterogeneity in the health insurance system (coinsurance gap): this is mainly due to the inertia of the behaviors summarized by the very low
elasticities. Hence, if the coinsurance gaps have a larger impact than the TFP gap on the share of health expenditure, this is not the case on the percentage of population in good health. This underlines that the heterogeneity in price gaps is crucial to explain the two dimensions of health performance, the share of health expenditure and the percentage of population in good health.

6 Exploring the economic mechanisms

We explore the economic mechanisms of the model by looking at predicted elasticities. We take advantage of our structural approach to compare elasticities under partial and general equilibria.

6.1 The health wedge

How would Americans react if their health wedge declines? US price elasticities (Figure 6) give the reaction of the US economy to a decline in the health wedge. Naturally, the same experience can be done for each European country. However, in their case, it will be an increase in the health wedge so that they all have the same price level as in the US (Figure 7).

In partial equilibrium, a reduction (increase) of health wedge in the US (European countries) leads health services $m$ to increase (decline), which rises (lowers) the percentage of the population in good health. In the US, a 1% reduction of the price of health services increases health expenditures by 0.44%, whereas the percentage of the population in good health is -0.457%. The lower (higher) health wedge leaves fewer resources for individual savings. Aggregate capital stock falls, thereby reducing production. In the US, the price elasticity of the stock of capital is -0.016, and the price elasticity of GDP is -0.0056. At the macroeconomic level, the GDP share of health expenditures declines (increases) as the changes in health price dominates the fall in the demand for health services. In the US, the GDP share of health expenditure in GDP falls by 0.62% for 1% price gap reduction, whereas public health and GDP improve. Beyond these aggregate performances, a decrease (increase) in the health wedge modifies health inequalities, summarized by the health gradient. At the bottom of the wealth distribution, the health wedge reduction allows the poorest to access more easily to the health system (this is the opposite when the health wedge increase in European countries). At the top of the wealth distribution, the

\footnote{In our simulations, the price gap is set to its estimated level for the French economy.}
impact of the price variation is negligible on behaviors: rich groups were already healthy prior to the switch to a more (less) efficient health market. Hence, $p_4/p_1$ largely declines in the US (Figure 6) or increase in European countries (Figure 7). For the middle class, the reduction (increase) of the health wedge is accompanied with a wealth increase (decrease) allowing these agents to deviate from the health conditions of the poorest: this explains the increases in $p_2/p_1$ and $p_3/p_1$ in the US (Figure 6), and their declines in European countries (Figure 7).

In general equilibrium, the tax rate adjusts to balance the health insurance system: with lower inefficiency on the health service market, the US tax rate is reduced by 7.5 points in percentage (from 18.75% to 11.25%). In Europe, the introduction of the US health wedge would lead to increases of the tax rates from 1.02pp (Spain) to 8.87pp (Netherland). The variations in the tax rate are proportional to changes in GDP share of health expenditures (see Equation (13)). This additional reduction of distortions in the US leads households to save more, which increases the capital stock in the economy, therefore shifts wages upward and reduces the interest rate. Hence, this tax reduction, accompanied by a wage increase thereby amplify the effect of the price reduction on all variables: general equilibrium magnifies all the price elasticities (Figure 6). In European countries, the same mechanisms are at work, but in the opposite direction: the increase of the health wedge is magnified by the rise in distortions induced by the tax rate increases (Figure 7). The wealth effects induced by health wedge variations, and specific to general equilibrium, are stronger at the bottom of the wealth distribution, such that, (i) the gap between the poorest and the richest ($p_4/p_1$) is smaller (larger) than at partial equilibrium, and (ii) the differences between the middle class and the poorest ($p_2/p_1$ and $p_3/p_1$) tend to increase (to reduce).

The gaps between the measures of the elasticities at partial vs. general equilibrium are significant, even if changes in the price of health service have only indirect impacts on wages and interest rates. In the US, the elasticities of the GDP share of health expenditure is -0.62% at partial equilibrium vs. -0.55% at general equilibrium. For the European countries, the gaps between the elasticities of the share of health expenditure taken at partial or at general equilibrium lie between 0.05pp (Spain) and 0.15pp (Netherland).

6.2 The efficiency wedge

How would Americans react if their efficiency wedge increases? The US TFP elasticities (Figure 6) give the reaction of the US economy to an increase of the health wedge, ie. a
reduction of the TFP in US. The same experience is done for each European countries. However, in their case, it will be an decrease of the efficiency wedge so that they all have the same TFP level as in the US (Figure 8).

At partial equilibrium, this TFP variation changes only wages: they fall by -1.55% for 1% TFP decrease, suggesting than a reduction in productive capital magnifies the negative impact of the TFP change. In European countries, the increases in wage lie between 1.24 (Denmark) to 1.79 (Italy). Our partial equilibrium analysis must then be understood as an equilibrium where the interest and the tax rates remain at their initial values. In the US, this negative income shock leads to a reduction in health expenditures by 2.16% for 1% TFP decrease, and thus reduces the percentage of the population in good health with an elasticity of 0.211%. It also reduces households' savings and thus the stock of productive capital, thereby magnifying the negative impact of TFP on GDP (the elasticity is equal to 1.54%). The GDP share of health expenditure is reduced, driven by the large decline in health expenditure (the elasticity is 0.625). In European countries, the permanent income shock is positive and thus all is reversed. For comparison, the country specific elasticities lie between 1.69 (Netherland) and 2.87 (Italy) for health expenditures, 0.16 (Netherland) and 0.26 (Italy) for the percentage of the population in good health and between 0.22 (Netherland) and 1.41 (Sweden) for the GDP share of health expenditure. Hence, the model is more sensitive to TFP than health price variations (see Figures 6, 7 and 8). As the increase of efficiency wedge (a TFP decline) reduces all wages, this particularly affects total incomes of the 75% poorest agents who reduce their health expenditures: health inequalities decrease across all households, but less with the top 25% richest ($p_4/p_1$ declines by only 0.18% for 1% TFP decrease, whereas $p_2/p_1$ and $p_3/p_1$ are reduced by respectively 10% and 14% for 1% TFP decrease), who are less sensitive to wage changes thanks to their financial incomes. At the opposite, a reduction of efficiency wedge (a TFP increase) in European countries, boosts European income and wealth for all individuals along the wealth distribution, which tends to dampen health inequality. All effects are reversed in Denmark and Sweden as the switch to the US TFP actually results in a lower TFP level.

At general equilibrium, the US tax rate declines as the GDP share of health spending goes down. This creates a positive wealth effect, which tends to dampen the initial fall in TFP. As a result, effects on macroeconomic aggregates at the general equilibrium tend to be dampened with respect to the partial equilibrium. This is not the case for the health gradient, as the richest individuals benefit more from the decline in taxation (see Figure 25 in our simulations, the TFP is set to its estimated level for the French economy.
The same damping effects related to general equilibrium adjustments are observed in the European countries (see Figure 8).

### 6.3 The coinsurance gap

Countries are divided in two groups: (1) By switching to the US coinsurance rate, individuals in Germany, France and the Netherlands need to increase their out-of-pocket spending, for any level of health utilization (In the benchmark calibration, $\mu$ in these countries was smaller than $\mu_{US}$). (2) By switching to the French coinsurance rate, American and the other European countries that switch to the US coinsurance rate (Denmark, Italy, Sweden, Spain) benefit from a larger insurance coverage (In the benchmark calibration, $\mu_{US} > \mu_{FR}$ and $\mu$ in these European countries was larger than $\mu_{US}$). For the purpose of clarity, we present the economic mechanisms for countries in group (1). Obviously, whatever the sign of the change in coinsurance, elasticities display the same sign in all countries.

Under *partial equilibrium*, individuals spend less on health services. Given the large elasticity of health services, this leads to a decline in GDP-share of health spending (See Figures 6 and 9). With reduced investment in health, the share of individuals with good health falls. The reduced health insurance coverage creates a negative wealth effects, as in the price counterfactual. As in the price counterfactual, this creates contrasting effects on health inequality in the middle ($p_2/p_1$ and $p_3/p_1$) and the top of the distribution ($p_4/p_1$).

Under *general equilibrium*, interestingly, the fraction of the population in good health does not evolve in a similar way as the aggregate demand for health services $m$. In group (1), in general equilibrium, more people are in good health in spite of a reduced demand for health services. A closer look at the change in the gradient suggests that the change in coinsurance rate affects individuals differently along the wealth distribution. Indeed, in group (1), with higher out-of-pocket health spending, the demand for medical services falls, which tends to reduce GDP-share of medical spending, hence the tax rate. Individuals are then faced with two contrasting forces on their budget constraint: on the one hand, higher out-of-pocket health spending; on the other hand, lower tax rate, which creates a positive wealth effect. For low-income-low-asset individuals, the demand for health services is very low, even before the switch to the US coinsurance rate. The fall in the tax rate has a sizeable positive effect on their wealth and income, which allows them to start spending more on health services. This effect is at work on $p_2/p_1$ and $p_3/p_1$ that are reduced in the counter-factual experiment. There groups constitute a large share
of the population, so that the average percentage of people in good health goes up. At
the top of the asset distribution, medical spending is high, even before the switch to the
US coinsurance rate system. Medical expenditure is a normal good for rich individuals.
Demand for medical services by the rich group slightly declines, but remains high. The
health gradient $p_t / p_0$ increases.

The coinsurance experiment illustrates the difference between aggregate responses to the
economic policy and responses along the wealth distribution, thereby underlining the
interest of modeling health choices in a heterogeneous agent setting. As these composition
effects appear only in general equilibrium, the coinsurance experiment also illustrates the
importance of general equilibrium effects.

7 Welfare

Our counterfactual experiments show that large health wedge (high price of health ser-
vice) tends to lower the utilization of health services, thereby the fraction of people in
good health, hence welfare. In contrast, small efficiency wedge (high TFP) increases in-
come, which allows individuals to spend more on health and reach a higher probability of
good health, thereby increasing welfare. The US is characterized by a high health wedge
and a small efficiency wedge. With respect to the health indicators, we have shown that
the negative impacts of the price gaps are not compensated by those induced by the TFP
gaps. Is this conclusion also relevant when we consider welfare? Indeed, the weight of
health in welfare can be small and thus dominated by other components highly linked to
the TFP gap. Which gap is quantitatively more important on the US welfare?

We aim at answering this question by looking at welfare implications of changes in US
health, efficiency and coinsurance wedges. In these counterfactual experiments, US price
(TFP, coinsurance rate, respectively) is replaced by the French estimated value. Table 6
reports the model’s predictions under general equilibrium. Welfare changes are computed
as $\gamma$ % change in permanent consumption that US individuals would willing to pay to
switch to the French value. The formula is the following. We compute the fictitious
consumption level $c_0$ corresponding to the benchmark welfare such that

$$
\sum_t \sum_e \sum_a \sum_h \beta^t \left( \frac{\bar{c}_0}{\bar{c}_0} + \phi h \right) \lambda_0 (a, h, e) = \sum_e \sum_a \sum_h \nu (a, h, e) \lambda_0 (a, h, e)
$$

We repeat the same calculation in order to get $c_1$ in the counterfactual experiment. The
welfare change is measured as $\gamma = \frac{c_1 - c_0}{(c_1 + c_0)/2}$ which can be decomposed as follow: $\gamma =$
\[
\varepsilon_{\pi|\pi,\mu} \frac{\sigma_1 - \sigma_0}{(x_1 + x_0)/2}, \quad \text{where } \varepsilon_{\pi|\pi,\mu} = \frac{\sigma_1 - \sigma_0}{(x_1 + x_0)/2}, \quad \text{for } x = p, A, \mu.
\]

Table 6 suggests that welfare consequences of health wedge are large. A switch to French health price increases US welfare by more than 8% in terms of permanent consumption. This result comes from the large price gap between US and France. Table 6 also suggests that efficiency wedge matters more than health wedge. Indeed, Americans would need to be compensated by 16.27% of permanent consumption to switch to the French TFP. This result is mainly due to the large impact of the TFP on all the economic mechanism (larger elasticity \(\varepsilon_{\pi|\pi}\)), but not from the estimated size of the TFP gap. Finally, Americans would need to be compensated by 5.76% of permanent consumption to switch to the French coinsurance rate. As for the price, this result mainly comes from the size of the coinsurance gap. Interestingly, even though the GDP share of health spending amounts to 14% of GDP, changes in health variables, such as price of health and coinsurance rate, has large welfare consequences: it represents a half of the impact of the efficiency wedge.

Figure 10 reports welfare changes across the distribution of labor earnings. A large part of the cross individual differences with a country come from the health expenditures. The poorest people spend a modest part of their income on health care costs: they are not very sensitive to price variations in health services and thus benefit from the most modest welfare gains when health market inefficiencies are reduced. The same reasoning applies for a reduction in TFP: a decrease in their incomes (almost only wages) only reduces their consumption, whereas the "quality" margin of their welfare (ie, the state of health) does not change as much, as it was already low before the TFP decline. For the middle class, the impact of a wedge variation is, on the contrary, very important. In fact, these agents, with modest financial wealth, had begun to taste the well-being of their health. Reducing the price of health brings them great welfare gains, while reducing their income through a reduction of the TFP, thereby leading them to give up health expenses. TFP changes tend to affect all income levels equally (due to the direct effect of TFP on the wage rate that applies to all productivity levels). As a result, welfare impact of lower TFP negatively affects US workers in the same way along the income distribution. In contrast, Figure 10 confirms that changes in prices affect more high-income groups than low income groups. Indeed, with lower price of health services, GDP-share of health spending declines, hence the tax rate falls. All income group benefit from lower taxation, especially the high-income group. The richest, meanwhile, are characterized by a high share of health expenditure in their total expenditure, and by a large share of financial income in their total income. It is clear then that they are the primary winners of the reduction of inefficiencies in the health market. For high-income individuals, the welfare
decline induced by the decline in TFP, which is largely explained by the renunciation of health, is negligible because the increase in health expenditure has a marginal return. Finally, the uniform decrease in the coinsurance rate mainly improves the welfare of the middle class. Indeed, the cheaper access to health services makes them spend more on health, whereas, for the rich, this has a very low impact on their wealth and thus improves their welfare by a smaller amount than the one of a middle class individual.

8 Conclusion

Health expenditures as a share of GDP and health status vary significantly across countries. While some have argued that the lack of a strong link between the two at the aggregate level is evidence that the marginal returns of health expenditures paid by richest individuals are very small, others have argued that prices and inefficiencies blur the link between health consumption and health. In this paper, we evaluate the contributions of two gaps on the cross-country differences in the GDP share of health expenditures and health status: (i) the TFP gaps measuring the relative economic development (called efficiency gap), and (ii) the price gaps capturing the inefficiencies on the health service market (called the health gap). To this end, we extend a general equilibrium framework à la Aiyagari (1994) by including health production (Grossman 1972). Using a method of moments approach, we estimate its structural parameters using macro and micro data from the US and seven European countries. Our estimation reveals not only deep parameter values related to preferences and health production but also country-specific structural price and TFP gaps. We perform counterfactual experiments to quantify the relative role of price, TFP and coinsurance rates in explaining observed cross-country heterogeneity in health expenditures and health status.

We find that the US are characterized by the highest health price of our sample and lies among the highest-TFP countries. In addition, using counterfactuals, we find that inefficiencies on the health market dominate the high technological efficiency when we focus on health indicators. It is also the case for European countries. Therefore, dispersion in health price seems to be the main cause for cross-country differences in GDP share of health expenditure and percentage of population in good health. When we consider welfare, rather than health indicators, the conclusion is reversed (as the utility gain from good health can be low compared to consumption utility): TFP gaps matter more than price gaps. Health price gap (the health wedge) is more than compensated by TFP gap (the efficiency wedge) in the US. Let us notice that the welfare costs of the price gap
are found to be large: while the health sector accounts for only 14% of GDP in the US, the welfare costs of the price gap is only twice lower than the welfare gains induced by the TFP gap. This underlines the strong distortions induced by inefficiencies on the US health market.

References


A Data Sources

We use four longitudinal surveys to construct our auxiliary estimations and the distribution of some moments. Our main data sources are the Health and Retirement Study (HRS) for the U.S., the Study of Health, Ageing and Retirement in Europe (SHARE), the Panel Study of Income Dynamics (PSID) for the U.S, and the European Community Household Panel (ECHP) for Europe.

We first use the two aging surveys to construct estimates of health status and health transitions. HRS and SHARE are longitudinal surveys of the over-50 population and conducted every two years surveys. The HRS, and SHARE cover an equally broad range of topics, including demographics (age, gender and education), labor supply, income, pension benefits, wealth, and health, and they contain identical question wording wherever possible. We use data from 2004. In particular the demographic variables that we use in HRS and SHARE for the health transitions rates are age, gender and education. The education variable is college versus non college. The chronic variables used are hypertension, stroke, lunge problems and cancer. We also build two limited physical indicators, the first is defined as self-reported difficulties with activities of daily living (ADLs) as well as instrumental ADLs (IADLs). For limitations in ADLs, questions were asked in all surveys about difficulties in five basic activities: bathing, dressing, eating, getting in and out of bed, and walking across a room. Individuals were classified as having any ADL limitation if they reported limitations with one or more of the five activities. Limitations in IADLs were measured by questions about difficulties in the following five activities: making meals, shopping, making phone calls, taking medications and managing money. Those who reported having some difficulty with any of the five activities were classified as having any IADL limitation. For the quintiles of wealth variables across countries, we use the net assets variable (all assets minus debt including housing). Finally, we use self-reported height and weight to construct our measure of obesity. Survey weights are used when computing statistics.

To estimate income processes, we use the PSID and ECHP. We use the PSID data, years 1990 till 1997, for the U.S., and ECHP, years 1994 till 2001. Both data sets have extensive information of income variables. We define income as total household income minus taxes and transfers and restrict the age range between 21 and 85 years old for estimating income processes.
B Basic two-periods model analysis

B.1 The FOC of the household program

The FOC are

\[ u' \left( (1 - \tau) y - a - \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pm \right) = \beta (1 + r) E [u'(1 + r)a + \tilde{y}] \quad (18) \]

\[ \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pu' \left( (1 - \tau) y - a - \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pm \right) = \beta \pi'(m) \phi \quad (19) \]

\[ \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pu' \left( (1 - \tau) y - \left( \mu - \frac{\vartheta - 1}{\vartheta} \right) pm \right) = \beta \pi'(m) \phi \quad (20) \]

where (18) and (19) are the FOC w.r.t. \( a \) and \( m \) if \( a > 0 \), whereas (20) is the FOC w.r.t. \( m \) when \( a = 0 \).

B.2 Proof of Property 2

The elasticities of \( s \) when \( a > 0 \). The properties of \( M \) and \( A \) are determined using the log-linearization of equations (1) and (2). We first Log-linearize equation (1), leading to

\[ \hat{a} = \frac{1}{2} \frac{(\nu_1 + \nu_2)\nu_5\nu_3}{-\nu_4[\nu_1 + (1 + r)(\nu_1 + \nu_2)]} \sigma_y^2 + \frac{-\nu_1\nu_3}{-\nu_4[\nu_1 + (1 + r)(\nu_1 + \nu_2)]} \hat{y} + \frac{\nu_2(\nu_1 + 1)}{-\nu_4[\nu_1 + (1 + r)(\nu_1 + \nu_2)]} \hat{p} \]

where \( \nu_1 = \frac{\pi''}{\pi'} m < 0, \nu_2 = \frac{1}{y} u'' \frac{m}{w} < 0, \nu_3 = u'' \frac{y}{w} < 0, \nu_4 = u'' \frac{a}{w} < 0 \) and \( \nu_5 = \frac{u''' y}{w} \). We assume that \( u'' > 0 \) in order to have precautionary saving, leading to \( \nu_5 < 0 \). Moreover, we consider the case where \( \beta R \to 1 \) in order to have \( a = 0 \) in an hypothetical economy without risk. If \( |\varepsilon| > 1 \), then \( \Gamma_3 < 0 \) because this leads to \( \nu_1 + 1 < 0 \).

Secondly, we Log-linearize equation (2), leading to

\[ \hat{m} = \frac{1 - \nu_2}{\nu_1 + \nu_2} \hat{p} + \frac{\nu_3}{\nu_1 + \nu_2} \hat{y} - \frac{\nu_4}{\nu_1 + \nu_2} \hat{a} \]

where \( \nu_1 = \frac{\pi'}{\pi'} s < 0, \nu_2 = \frac{1}{y} u'' \frac{s}{w} < 0, \nu_3 = u'' \frac{s}{w} < 0, \nu_4 = u'' \frac{a}{w} < 0 \) and \( \nu_5 = \frac{u''' s}{w} \). We assume that \( u'' > 0 \) in order to have precautionary saving, leading to \( \nu_5 < 0 \). Moreover, we consider the case where \( \beta R \to 1 \) in order to have \( a = 0 \) in an hypothetical economy without risk. If \( |\varepsilon| > 1 \), then \( \Gamma_3 < 0 \) because this leads to \( \nu_1 + 1 < 0 \).
Hence, we deduce

\[ \hat{m} = \varepsilon_y \hat{y} - \varepsilon_a \hat{a} + \varepsilon_p \hat{p} \]
\[ \hat{a} = \Gamma_1 \sigma_y^2 + \Gamma_2 \hat{y} + \Gamma_3 \hat{p} \]

Given this solution for \( \hat{a} \), we obtain:

\[ \hat{m} = (\varepsilon_y - \varepsilon_a \Gamma_2) \hat{y} + (\varepsilon_p - \varepsilon_a \Gamma_3) \hat{p} - \varepsilon_a \Gamma_1 \sigma_y^2 \]

and, using \( \hat{s} = \hat{p} + \hat{m} - \hat{y} \), we deduce that

\[ \hat{s} = (\varepsilon_y - 1 - \varepsilon_a \Gamma_2) \hat{y} + (1 + \varepsilon_p - \varepsilon_a \Gamma_3) \hat{p} - \varepsilon_a \Gamma_1 \sigma_y^2 \]

If \( |\varepsilon_{\pi'}| > 1 \), then we have \( \frac{\partial s}{\partial y} > 0 \) and \( \frac{\partial s}{\partial p} > 0 \). Indeed, the elasticities of \( s \) w.r.t. \( y \) and \( p \) satisfy the following properties.

- For the elasticity w.r.t. \( y \), we have:

\[ \varepsilon_y \Gamma_2 = \frac{-\nu_1 \nu_3 \nu_4}{\nu_1 (\nu_1 + \nu_2) \nu_3 (\nu_3 + (1 + r) (\nu_1 + \nu_2))} \]
\[ = \frac{(\pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'')}{\pi'' m \pi'' + (1 + r) \left( \pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'\right)} \]
\[ \varepsilon_y - 1 = \frac{\nu_3}{\nu_1 + \nu_2} - 1 = \frac{\pi'' y}{\pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'} \]
\[ = \frac{1}{\pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'} \left[ \pi'' y \pi' - \pi'' m \pi' - \frac{1}{p} \pi'' pm \pi' - \frac{\pi'' m \pi'' \pi' y}{\pi'' m \pi'' + (1 + r) \left( \pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'\right)} \right] \]

Hence, we have \( \varepsilon_y - 1 - \varepsilon_a \Gamma_2 > 0 \) iff

\[ u'' y \pi' - \pi'' m \pi' - \frac{1}{p} \pi'' pm \pi' < 0 \]
\[ \Leftrightarrow u'' y - \frac{1}{p} pm \pi' - \pi'' m \pi' \left( 1 + \frac{u'' y \pi'}{\pi'' m \pi'' + (1 + r) \left( \pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'\right)} \right) < 0 \]

A sufficient condition is:

\[ 1 + \frac{u'' y \pi'}{\pi'' m \pi'' + (1 + r) \left( \pi'' m \pi'' + \frac{1}{p} \pi'' pm \pi'\right)} < 0 \Leftrightarrow \frac{1}{1 + (1 + r) \frac{u'' y - \frac{1}{p} pm \pi'}{u'}} < -\pi'' m \pi' - \frac{1}{p} \pi'' pm \pi' \]

which is always satisfied because the LHS is negative \( (u'' < 0) \) whereas the RHS is
positive \((-\pi''_{\pi m} - \frac{1}{\varrho} u''_{\pi m w} > 0)\).

- For the elasticity w.r.t \(p\), we have:

\[
\varepsilon_a\Gamma_3 = \frac{\nu_4\nu_2(\nu_1 + 1)}{-\nu_4(\nu_1 + \nu_2)[\nu_1 + (1 + r)(\nu_1 + \nu_2)]} = \frac{\frac{1}{\varrho} u''_{\pi m w} (\pi''_{\pi m} + 1)}{-\left(\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}\right)\left[\pi''_{\pi m} + (1 + r)\left(\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}\right)\right]} \times [1 + \epsilon_p] = \frac{\pi''_{\pi m} + 1}{\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}} \left[1 + \frac{1}{\varrho} u''_{\pi m w} \left(\pi''_{\pi m} + (1 + r)\left(\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}\right)\right)\right]
\]

where \(\frac{\pi''_{\pi m} + 1}{\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}} > 0\) if \(\pi''_{\pi m} + 1 < 0\), i.e. \(|\varepsilon_{\pi'|} > 1\). We assume that this last restriction is satisfied. Hence, \(|\varepsilon_{\pi'|} > 1\) is a sufficient condition for \(1 + \epsilon_p - \varepsilon_a\Gamma_3 > 0\) because

\[
\frac{1}{\varrho} u''_{\pi m w} \left(\pi''_{\pi m} + (1 + r)\left(\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}\right)\right) > 0 \iff 1 + \frac{1}{\varrho} u''_{\pi m w} \left(\pi''_{\pi m} + (1 + r)\left(\pi''_{\pi m} + \frac{1}{\varrho} u''_{\pi m w}\right)\right) > 0
\]

The elasticities of \(s\) when \(a = 0\). Without risk, the agent problem becomes static: the optimal choice is reduced to the optimal health expenditures. We have \(a = 0, c = y - pm\) and \(d = y\). In order to analyze the other properties of the model, we log-linearize the equation (3):

\[
\nu_1 \hat{m} = \hat{p} - \nu_2 (\hat{m} + \hat{p}) + \nu_3 \hat{y} \Rightarrow \hat{m} = \frac{\nu_3}{\nu_1 + \nu_2} \hat{y} + \frac{1 - \nu_2}{\nu_1 + \nu_2} \hat{p}
\]

where \(\nu_1 = \pi''_{\pi m} < 0, \nu_2 = \frac{1}{\varrho} u''_{\pi m w} < 0\) and \(\nu_3 = u''_{\pi w} < 0\), implying that \(\frac{\nu_3}{\nu_1 + \nu_2} > 0\) and \(1 - \nu_2 > 0\). For all functional forms, i.e. \(\forall\{\pi(\cdot), u(\cdot)\}\), the equilibrium health expenditures \((m^*)\) behave as a normal goods, i.e. \(\frac{\partial m^*}{\partial p} < 0\) (because \(\frac{\nu_3}{\nu_1 + \nu_2} < 0\)) and increase with the income, i.e. \(\frac{\partial m^*}{\partial y} > 0\) (because \(\frac{\nu_3}{\nu_1 + \nu_2} > 0\)). This implies that \(\varepsilon_p < 0\) and \(\varepsilon_y > 0\).

Given that \(\hat{s} = \hat{p} + \hat{m} - \hat{y}\), we deduce that

\[
\hat{s} = \frac{\nu_3 - (\nu_1 + \nu_2)}{\nu_1 + \nu_2} \hat{y} + \frac{1 + \nu_2}{\nu_1 + \nu_2} \hat{p}
\]

\[
\hat{s} = (\varepsilon_y - 1) \hat{y} + (\varepsilon_p + 1) \hat{p}
\]
with \( \frac{1+\nu_1}{\nu_1+\nu_2} \geq 0 \) and \( \frac{\nu_3-(\nu_1+\nu_2)}{\nu_1+\nu_2} \geq 0 \). Let us denote \( \varepsilon_{\pi'} = \pi'' \frac{x}{x'} < 0 \) and \( \varepsilon_{u'} = u'' \frac{c}{w} < 0 \). Using these notations, we have (i) \( \frac{1+\nu_1}{\nu_1+\nu_2} > 0 \) iff \( 1 + \nu_1 < 0 \). This leads to \( |\varepsilon_{\pi'}| > 1 \). (ii) the restriction \( \eta \left[ \frac{\nu_3-(\nu_1+\nu_2)}{\nu_1+\nu_2} \right] + (1+\nu_1) \frac{\nu_1+\nu_2}{\nu_1+\nu_2} > 0 \) is equivalent to

\[
\eta u'' \frac{y}{u'} + 1 + \nu'' \frac{m}{\pi'} < \eta \pi'' \frac{m}{\pi'} + \eta u'' \frac{pm}{w'} \quad \Leftrightarrow \quad \varepsilon_{u'} < \varepsilon_{\pi'} \left( 1 - \frac{pm}{y} \right) + \varepsilon_{u'} \frac{pm}{y} - \frac{1}{\eta} \left( 1 + \varepsilon_{\pi'} \right) \frac{c}{y}
\]

For (i) satisfied, ie. \( |\varepsilon_{\pi'}| > 1 \), we have \( 1 + \varepsilon_{\pi'} < 0 \). Hence, a sufficient condition on elasticities ensuring that the previous inequality is satisfied is \( \varepsilon_{u'} < \varepsilon_{\pi'} \left( 1 - \frac{pm}{y} \right) + \varepsilon_{u'} \frac{pm}{y} \). Given that \( \frac{pm}{y} \in (0,1) \), this is true iff \( \varepsilon_{u'} < \varepsilon_{\pi'} \). Hence, the share of health expenditures in the total income

(i) can be increasing with the price of the health sector if \( |\varepsilon_{\pi'}| > 1 \),

(ii) can be increasing with the total income if \( \varepsilon_{u'} < \varepsilon_{\pi'} \).
C Solving the General Equilibrium Model

Step 1: Households’ decision rules. In step 1, we compute the household policy. Given \( r, w, \tau, \mu, p \), we determine, for each state \((a, h, e)\), consumption, savings and medical expenditures \(\{c(a, h, e), a'(a, h, e), m(a, h, e)\}\) that solve the households’ decision problem described in (9). We rely on a discrete approximation of the state space. \( h \) takes 2 values (good or bad), the number of \( e \) ability level is \( N_e \) and the asset grid is captured by a discrete set of points \( N_k \). We then compute \(2\times N_e \times N_k \) value functions. Let us make several comments on the asset grid. First, we use piecewise linear interpolation, so that next period’s asset choice can lie outside the initial grid on asset. Secondly, as it is standard in the literature (Castaneda et al. (2003)), the asset grid is not equally spaced. For very low values of asset holdings, the distance between grid points is small. This is done to allow financially constrained individuals to increase their savings by small increments.

With respect to Aiyagari (1994)’s model, the complexity lies in the computation of two optimal choices \( c \) and \( m \) (\( a' \) being determined by the household’s budget constraint) that are related through a dynamic first-order condition. We rely on value function iteration. Starting from a guess on optimal choices of \( c \) and \( m \), for a given state \((a, h, e)\), using Nelder-Mead method, we compute values of \( c \) and \( m \) that maximize the value function (9), using a guess on next period’s value function. The new values for \( V, c \) and \( m \) are compared to the initial guess. If they are not close, replace the guess by the new values of \( c, m, V \) and repeat the optimization procedure. If they are close enough, the household’s policy was found for the given state \((a, h, e)\). We then repeat the whole process for all possible values of state \((a, h, e)\).

Step 2: Stationary distribution. We compute the invariant wealth and health distribution by Monte Carlo simulations. We use simulated paths to generate an approximation of the distribution. We start with an individual agent to whom we assign an asset level, labor efficiency and health status. Using the policy rule computed in step 1, we can infer the individual optimal choices, then the probability of future health status. We draw a new productivity level and repeat the procedure next period. We track the individual’s choices and realization of idiosyncratic shock over a very large number of periods. The stationary distribution is obtained by counting the number of times the individual happens to be in each state of the space \((a, h, e)\). We check that the Monte Carlo distribution converges to the stationary distribution.
**Step 3: General Equilibrium.** We compute the general equilibrium factor prices \( r \) mentioned in (b.) in Section 3.4 then \( w \) is inferred from equation (12) and the equilibrium tax rate \( \tau \) (mentioned in (d.) in Section 3.4). As a result, Steps 1 and 2 must be repeated until the interest rate \( r \) clears the asset market and the tax rate \( \tau \) ensures that health insurance budget constraint is satisfied.

The steps of the algorithm are then

i. Compute the stationary level of employment \( N \)

ii. Make an initial guess of the interest rate \( r \) and tax rate \( \tau \)

iii. Compute the wage rate \( w \) using equation (12)

iv. Compute the household’s decision rules (Step 1)

v. Compute the invariant distribution (Step 2)

vi. Calculate aggregate variables using the agents distribution. Check market clearance on the asset market. Check that health budget constraint is satisfied. If these conditions do not hold, update the guess of the interest rate \( r \) and tax rate \( \tau \). If not, go back to ii.

vii. Check for convergence and update the guess
D MCMC Algorithm for MSM estimator

Denote by $Q(\Theta) = -\tilde{g}_S(\Theta, \chi)'W_N\tilde{g}_S(\Theta, \chi)$ the objective function of the MSM estimator. We use the adaptive Metropolis-Hastings algorithm proposed by Haario et al. (2001). The algorithm uses a Gaussian proposal distribution with a covariance matrix which depends on the entire markov chain and shrinks slowly.

A Metropolis-Hastings algorithm is used to simulate a chain that converges to the quasi-posterior distribution. For the initialization of the algorithm, we compute an initial chain where $\Theta \sim N(\Theta_0, \Omega_0)$. We use $n_0 = 100$ draws as a burn-in period with a fixed (and large) covariance matrix $\Omega_0$ (we choose $\Sigma_0 = I_d$). The first element of the chain is drawn in this initial chain. It is denoted $\Theta_0$. Afterwards, we process as follow, for $n = 1, \ldots, N_c$:

- **Step 1.** Start with an evaluation of the objective function $Q$ with a value $\Theta_n$.
- **Step 2.** Draw a proposal $\Theta_n^{*+1}$ in the normal law $N(\theta_n, \Omega_n)$.
- **Step 3.** Compute the acceptance probability $\alpha(\Theta_n^{*+1}|\Theta_n) = \min\{\exp(Q(\Theta_n^{*+1}) - Q(\Theta_n)); 1\}$.
- **Step 4.** Draw a value $\tilde{\eta}$ in the uniform distribution over $[0; 1]$. The new value for the chain is thus
  \[ \Theta_{n+1} = \begin{cases} 
  \Theta_n^{*+1} & \text{if } \tilde{\eta} < \alpha(\Theta_n^{*+1}|\Theta_n) \\
  \Theta_n & \text{otherwise}
  \end{cases} \]
- **Step 5.** Update the covariance matrix $\Omega_{n+1}$ for the next draw:
  \[ \Omega_{n+1} = \frac{n}{n + n_0} \left( \frac{n - 1}{n} \Omega_n + \frac{n}{n - 1} \Sigma_{n-1}' \Sigma_{n-1} - \frac{(n + 1)}{n} \Sigma_n \Sigma_n' + \Sigma_n \Sigma_n' \right) + \frac{n_0}{n + n_0} \Omega_0 \]
  where
  \[ \bar{\Theta}_n = \frac{n + 1}{n + 1 + n_0} \left( \frac{n}{n + 1} \bar{\Theta}_{n-1} + \frac{1}{n + 1} \bar{\Theta}_n \right) + \frac{n_0}{n + 1 + n_0} \bar{\Theta}_0 \]
- **Step 6.** come back to step 1 to compute $\Theta_{n+2}$.

Estimates can be obtained from the chain of length $N_c$ using averages and confidence intervals using empirical quantiles. We use chain of length 10,000 and keep the last 1000 draws to compute estimates. We varied the length of the Markov Chain and found little change in the estimates.
Figures

Figure 1: Obesity Rates in HRS and SHARE

Figure 2: Out-of-pocket as % Health Expenditures
Figure 3: Moments used in Estimation
Figure 4: Model fit
Figure 5: US economy with French price of health \((p)\), TFP \((A)\) or coinsurance rate \((\mu)\). Each counterfactual \(x\) is decomposed into elasticity \(\epsilon\), the ratio of change in exogenous variable \(dx/x\) and change in variable of interest \(dy/y\) with \(y = s, p(H = 1)\). For instance, in panel (a), in the "price" counterfactual, \(dx/x\) is the % change in price, \(ds/s\) % change in share of health expenditure. \(dx/x\) are the same in both graphs. General equilibrium results.
Figure 6: US economy with French characteristics. Arc-elasticities - Partial vs. general equilibrium.
Figure 7: European countries with US health price. Arc-elasticities - Partial vs. general equilibrium.
Figure 8: European countries with US TFP. Arc-elasticities - Partial vs. general equilibrium.
Figure 9: European countries with US coinsurance rate. Arc-elasticities - Partial vs. general equilibrium.
Figure 10: Welfare changes if the US economy were to switch to French values. % change in permanent consumption with respect to benchmark. "p5" 5% percentile on labor earnings, "p50" median labor earnings, "p95" top 5% labor earnings.
### Tables

<table>
<thead>
<tr>
<th></th>
<th>Diagnostics</th>
<th>Drugs</th>
<th>Hospital cost</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angiogram</td>
<td>Gleevec (Cancer)</td>
<td>per day</td>
<td>Bypass surgery</td>
</tr>
<tr>
<td>US</td>
<td>907$</td>
<td>6214$</td>
<td>4293$</td>
<td>75345$</td>
</tr>
<tr>
<td>EU</td>
<td>290$ (SP)</td>
<td>3321$ (NL)</td>
<td>481$ (SP)</td>
<td>15742$ (SP)</td>
</tr>
<tr>
<td>price ratio</td>
<td>3.1</td>
<td>1.9</td>
<td>8.9</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Table 1: Comparison of Prices (IFHP, 2013)

<table>
<thead>
<tr>
<th>Country</th>
<th>$ DE $</th>
<th>$ DK $</th>
<th>$ FR $</th>
<th>$ IT $</th>
<th>$ NL $</th>
<th>$ SE $</th>
<th>$ SP $</th>
<th>$ US $</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ \rho_e $</td>
<td>0.9436</td>
<td>0.9182</td>
<td>0.9588</td>
<td>0.9433</td>
<td>0.9697</td>
<td>0.9182</td>
<td>0.9798</td>
<td>0.959</td>
</tr>
<tr>
<td>$ \sigma_e^2 $</td>
<td>0.0285</td>
<td>0.0150</td>
<td>0.0191</td>
<td>0.0303</td>
<td>0.0108</td>
<td>0.0150</td>
<td>0.0111</td>
<td>0.0396</td>
</tr>
<tr>
<td>$ \sigma_u^2 $</td>
<td>0.0967</td>
<td>0.0751</td>
<td>0.1143</td>
<td>0.0806</td>
<td>0.1192</td>
<td>0.0751</td>
<td>0.1364</td>
<td>0.1257</td>
</tr>
<tr>
<td>$ \sigma_u^2 + \frac{\sigma_e^2}{1-\rho^2} $</td>
<td>0.3567</td>
<td>0.1707</td>
<td>0.3510</td>
<td>0.3556</td>
<td>0.3002</td>
<td>0.1707</td>
<td>0.4140</td>
<td>0.6187</td>
</tr>
</tbody>
</table>

Table 2: Estimates of Income Process

<table>
<thead>
<tr>
<th>Country</th>
<th>$ DE $</th>
<th>$ DK $</th>
<th>$ FR $</th>
<th>$ IT $</th>
<th>$ NL $</th>
<th>$ SE $</th>
<th>$ SP $</th>
<th>$ US $</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ \alpha $</td>
<td>0.373</td>
<td>0.338</td>
<td>0.373</td>
<td>0.456</td>
<td>0.383</td>
<td>0.353</td>
<td>0.348</td>
<td>0.358</td>
</tr>
<tr>
<td>$ \delta $</td>
<td>0.037</td>
<td>0.041</td>
<td>0.035</td>
<td>0.041</td>
<td>0.038</td>
<td>0.048</td>
<td>0.034</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Table 3: Calibration

<table>
<thead>
<tr>
<th>Common Parameters</th>
<th>Country Specific Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimates</td>
<td>Country</td>
</tr>
<tr>
<td>$ \sigma $</td>
<td>3.120</td>
</tr>
<tr>
<td>$ \beta $</td>
<td>0.823</td>
</tr>
<tr>
<td>$ \phi $</td>
<td>0.086</td>
</tr>
<tr>
<td>$ \alpha_0 $</td>
<td>2.339</td>
</tr>
<tr>
<td>$ \alpha_{10} $</td>
<td>0.258</td>
</tr>
<tr>
<td>$ \alpha_{11} $</td>
<td>2.110</td>
</tr>
<tr>
<td>$ \eta $</td>
<td>-0.056</td>
</tr>
<tr>
<td>$ \text{US} $</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4: Estimated Parameters
(a). Share of health expenditures ($s$)

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>TFP</th>
<th>coinsurance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon$</td>
<td>$\frac{dp}{p}$</td>
<td>$\frac{ds}{s}$</td>
</tr>
<tr>
<td>DE</td>
<td>0.3936</td>
<td>0.3960</td>
<td>0.1559</td>
</tr>
<tr>
<td>DK</td>
<td>0.4418</td>
<td>0.6998</td>
<td>0.3092</td>
</tr>
<tr>
<td>FR</td>
<td>0.6818</td>
<td>0.7352</td>
<td>0.5013</td>
</tr>
<tr>
<td>IT</td>
<td>0.2507</td>
<td>0.3735</td>
<td>0.0936</td>
</tr>
<tr>
<td>NL</td>
<td>0.7532</td>
<td>0.8168</td>
<td>0.6152</td>
</tr>
<tr>
<td>SE</td>
<td>0.2613</td>
<td>0.4243</td>
<td>0.1109</td>
</tr>
<tr>
<td>SP</td>
<td>0.3263</td>
<td>0.3641</td>
<td>0.1188</td>
</tr>
</tbody>
</table>

(b). % of population in good health ($p(H=1)$)

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>TFP</th>
<th>coinsurance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\varepsilon$</td>
<td>$\frac{dp}{p}$</td>
<td>$\frac{dp(H=1)}{p(H=1)}$</td>
</tr>
<tr>
<td>DE</td>
<td>-0.0577</td>
<td>0.3960</td>
<td>-0.0229</td>
</tr>
<tr>
<td>DK</td>
<td>-0.0680</td>
<td>0.6998</td>
<td>-0.0476</td>
</tr>
<tr>
<td>FR</td>
<td>-0.0603</td>
<td>0.7352</td>
<td>-0.0443</td>
</tr>
<tr>
<td>IT</td>
<td>-0.0667</td>
<td>0.3735</td>
<td>-0.0249</td>
</tr>
<tr>
<td>NL</td>
<td>-0.0626</td>
<td>0.8168</td>
<td>-0.0511</td>
</tr>
<tr>
<td>SE</td>
<td>-0.0643</td>
<td>0.4243</td>
<td>-0.0273</td>
</tr>
<tr>
<td>SP</td>
<td>-0.0491</td>
<td>0.3641</td>
<td>-0.0179</td>
</tr>
</tbody>
</table>

Table 5: Share of health expenditures and % of individual in good health: Europe with US price, TFP or coinsurance rate.

<table>
<thead>
<tr>
<th></th>
<th>Price</th>
<th>TFP</th>
<th>coinsurance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma$</td>
<td>$\varepsilon_{\gamma</td>
<td>x}$</td>
</tr>
<tr>
<td></td>
<td>0.0839</td>
<td>-0.1627</td>
<td>-0.0576</td>
</tr>
<tr>
<td></td>
<td>0.0805</td>
<td>-1.4000</td>
<td>-0.1059</td>
</tr>
<tr>
<td></td>
<td>-0.0044</td>
<td>-0.0044</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 6: Welfare gains ($\gamma > 0$)/losses ($\gamma < 0$) if the US were to switch to French values.