### Abstract

This contribution starts from the observation that the past decades have seen a substantial change in the division of labor between private (PI) social insurance (SI), to the advantage of the latter. The efficiency view of SI (to be expounded in Sect. 37.2) explains the existence of SI with the market failures of PI, namely moral hazard and adverse selection. In Sect. 37.3, a benevolent government is introduced that seeks to determine the optimal division of labor between PI and SI. However, moral hazard effects are found to plague SI at least as much as PI, while the empirical relevance of the adverse selection argument has recently been challenged. In Sect. 37.4, the exposition therefore turns to the public choice view, which emphasizes the interests of risk-averse voters even with below-average wealth in redistribution through SI. This view predicts a crowding out of PI by SI also in markets without adverse selection, which has been observed. Section 37.5 turns to normative issues by proposing a test that indicates whether and in which lines of insurance the division of labor between PI and SI could be improved. The final section, Sect. 37.6, offers concluding remarks and an outlook on future challenges confronting both PI and SI.
The Division of Labor Between Private and Social Insurance

Peter Zweifel

Abstract

This contribution starts from the observation that the past decades have seen a substantial change in the division of labor between private (PI) social insurance (SI), to the advantage of the latter. The efficiency view of SI (to be expounded in Sect. 37.2) explains the existence of SI with the market failures of PI, namely moral hazard and adverse selection. In Sect. 37.3, a benevolent government is introduced that seeks to determine the optimal division of labor between PI and SI. However, moral hazard effects are found to plague SI at least as much as PI, while the empirical relevance of the adverse selection argument has recently been challenged. In Sect. 37.4, the exposition therefore turns to the public choice view, which emphasizes the interests of risk-averse voters even with below-average wealth in redistribution through SI. This view predicts a crowding out of PI by SI also in markets without adverse selection, which has been observed. Section 37.5 turns to normative issues by proposing a test that indicates whether and in which lines of insurance the division of labor between PI and SI could be improved. The final section, Sect. 37.6, offers concluding remarks and an outlook on future challenges confronting both PI and SI.
Table 37.1 Social (SI) and private (PI) insurance in some OECD countries in percent of GDP

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>SI 22.1</td>
<td>21.7</td>
<td>26.6</td>
<td>25.2</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>2.4</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td>France</td>
<td>SI 20.8</td>
<td>24.9</td>
<td>27.7</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>3.4</td>
<td>6.7</td>
<td>7.4</td>
</tr>
<tr>
<td>UK</td>
<td>SI 10.5</td>
<td>16.8</td>
<td>18.6</td>
<td>20.5</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>6.8</td>
<td>12.8</td>
<td>10.4</td>
</tr>
<tr>
<td>Italy</td>
<td>SI 18.0</td>
<td>20.0</td>
<td>23.3</td>
<td>24.9</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>0.6</td>
<td>3.5</td>
<td>8.4</td>
</tr>
<tr>
<td>Japan</td>
<td>SI 10.4</td>
<td>11.3</td>
<td>16.5</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>7.0</td>
<td>8.1</td>
<td>7.5</td>
</tr>
<tr>
<td>USA</td>
<td>SI 13.2</td>
<td>13.5</td>
<td>14.5</td>
<td>10.2</td>
</tr>
<tr>
<td></td>
<td>PI n.a.</td>
<td>4.1</td>
<td>4.5</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Note: Putting labor’s share at two-thirds of GDP, one arrives at a combined PI and SI share in workers’ pay of 42.7% (25.2 + 3.4)/0.67) for Germany and 19.5% (10.2 + 2.9)/0.67 for the USA (as of 2009).

Sources: OECD Social Expenditure Database, several editions; Sigma of Swiss Re, several editions.

SI: Includes benefits (e.g., for housing) that are part of public welfare

2007 for SI

Note: Putting labor’s share at two-thirds of GDP, one arrives at a combined PI and SI share in workers’ pay of 42.7% (25.2 + 3.4)/0.67) for Germany and 19.5% (10.2 + 2.9)/0.67 for the USA (as of 2009).

Sources: OECD Social Expenditure Database, several editions; Sigma of Swiss Re, several editions.

SI: Includes benefits (e.g., for housing) that are part of public welfare

2007 for SI

Important. Both the high GDP share of SI and PI combined and the preponderance of SI are mainly the result of a remarkable expansion of SI. In Germany, for instance, the SI share was already at a comparatively high value of 22.1% in 1980 but continued to increase to 26.6% until the year 2000, becoming roughly stable at 25.2% by 2009. The PI share was at 2.4% in 1990 (the first year with comparable data), nine times lower than the SI figure. By 2009, it has increased to 3.4%, still more than seven times lower than the SI figure. Detail not reported in Table 37.1 but provided by OECD shows that the expansion of SI was led by health, with pensions second; accordingly, the term “social insurance” as used in this chapter is more comprehensive than “social security” in US language.

Much of the economics literature has justified the existence and even the preponderance of SI by its efficiency-enhancing properties, arguing that SI can mend or at least mitigate market failures of private insurance (PI). A typical exponent of this view is Sinn (1996), who claims that SI offers a service PI cannot offer. Risk-averse parents who want their children to live in a world with an income floor for the unlucky must rely on SI because PI cannot deal with this type of inter-generational risk. In a similar vein, Casamatta et al. (2000), examining the political sustainability of SI, theoretically describe conditions that shift the division of labor between SI and PI to the detriment of SI. However, the expansion of SI also has raised concerns. For instance, Feldstein (1995) has estimated that the privatization of US pensions would increase economic welfare by $1.5 per dollar of net social security wealth (SSW). Accordingly, the concern in this camp is that SI may “crowd out” PI over time. An early piece of empirical evidence came from Cutler and Gruber (1996), who found that the 1987–1992 expansion of US Medicaid (a public program providing health insurance coverage to the poor) displaced private health insurance coverage at the rate of 50%.

This contribution seeks to shed light on this debate in four ways. First, it reviews the efficiency reasons supporting the view that there is a need for SI, possibly complemented by PI. The theoretical mainstay is provided by the famous Rothschild-Stiglitz (1976) model that can be combined with the “market for lemons” dynamics described by Akerlof (1970) to predict the possibility of a “death spiral” for PI markets in the presence of insurers’ lack of information...
about consumers’ true risk status. Second, some empirical evidence concerning the alleged “death spiral” and insurer behavior in the face of their lack of information is examined. At least for the time being, the conclusion is that the efficiency view of SI may well fail to explain the secular expansion of SI and the associated crowding out of PI. Third, the alternative public choice view emphasizing the interests of citizens and voters in wealth redistribution through SI is expounded. It accords well with evidence documenting a crowding out of PI by SI. Forth, normative issues are addressed by asking whether and in which lines of insurance the future division of labor between PI and SI could be improved.

### 37.2 The Efficiency View of Social Insurance

Theoretical explanations of the existence of SI usually refer to two main shortcomings of private insurance markets due to asymmetric information, moral hazard, and adverse selection. It will be argued that moral hazard effects plague SI even more than PI; the emphasis of this section therefore will be on adverse selection.

#### 37.2.1 Moral Hazard as a Market Failure of Private Insurance

Moral hazard is the consequence of the fact that the insured agent does not reap the full benefit of preventive effort anymore (since the insurer participates in any reduction of expected loss) while bearing the full cost of this effort. The observable consequence is a positive correlation between the degree of insurance coverage and the probability of loss (in the case of so-called ex ante moral hazard) and the amount of loss (ex post moral hazard). If this behavior is unobservable (“hidden action”), the insurer cannot sanction it by increasing the individual premium. To the extent that the loading contained in insurance premiums increases along with expected loss (e.g., because of increasing cost of administration and risk bearing), the true cost of insurance rises due to moral hazard.\(^1\) Moral hazard effects therefore cause consumers to choose less than full coverage (the first-best solution), which amounts to a welfare loss.

However, one may note at once that SI is subject to the same moral hazard effects as PI unless one is willing to assume that institutions of SI can observe preventive behavior. In the case of health, accidents, disability, and old age, this is clearly not true, to the contrary. Whereas private insurers usually tailor the parameters of their contracts to individual behavior reflecting (the control of) moral hazard (e.g., by granting rebates for no claims), SI is strongly bound to the solidarity principle which requires equal ex ante benefits for equal (rates of) contributions. Therefore, the argument that SI has an efficiency advantage over PI due to its lower loading holds with regard to the cost of administration at best. But administrative expense may well be endogenous, reflecting effort at controlling moral hazard. For instance, it is lower in Canadian health insurance (which continues to rely on fee-for-service) than in US health insurance (which has created managed care alternatives to combat moral hazard) (Danzon 1992). One might argue that compared to PI, SI can better count on physicians as agents for the verification of health claims. However, a study by Dionne and St. Michel (1991) finds evidence suggesting that physicians helped workers to benefit from an increase in generosity in the public workplace accident scheme of Quebec. According to the authors, this effect did not depend on the severity of the condition but rather on its observability. It was marked when the diagnosis was ambiguous but absent when easily checked by another physician. Therefore, at least in this case it cannot be said that SI was able to enlist the support of physicians as intermediaries for reining in moral hazard.

---

\(^1\)Recall that the major part of an insurance premium is redistribution between those who do not suffer a loss and those who do. The true cost of insurance amounts to the loading in excess of the expected loss. It is for this reason that full coverage is predicted for riskaverse consumers who are charged a fair premium (i.e., a premium without a loading).
The one remaining component of the loading is acquisition expense, which is negligible in the case of SI because SI constitutes a monopoly. The cost advantage of SI therefore has to be weighed against the imposed uniformity of promised benefits which entails an efficiency loss as soon as preferences with regard to insurance differ within the population.

37.2.2 Adverse Selection: The Crucial Market Failure of Private Insurance?

The crucial efficiency drawback of PI has to do with the other consequence of asymmetric information, adverse selection. The theoretical mainstay is provided by the Rothschild-Stiglitz (1976) model, which can be combined with the dynamics of Akerlof’s (1970) “market for lemons,” where good quality is driven out by bad quality. The model assumes that private insurers do not (and never will) know the true type of consumers, reflected by the probability of loss (“hidden type”). Therefore, a pooling contract reflecting the average probability in the population is the best PI can come up with. Since the trade-off between premium and coverage is valued differently by high and low risks, a challenger can always offer a contract with less coverage but a lower premium that appeals to the favorable but not the high risks. Note that this again induces a positive correlation between the amount of insurance coverage and the probability of loss, very much the same as in the case of (ex ante) moral hazard. However, contrary to the case of moral hazard, this time there is an incumbent insurer who is stuck with its high risks and forced to increase the pooling premium to maintain financial equilibrium. This only strengthens the incentive of favorable risks to leave the incumbent (who offers “good” quality in Akerlof’s terms), triggering a “death spiral.” Since the challenger still writes a pooling contract, it can in turn be attacked by a third competitor; the death spiral may therefore even continue to wipe out the entire PI market.

Before discussing the potential of SI to at least mitigate this allegedly crucial market failure of PI, it may be worthwhile to consider some empirical evidence. As noted above, a positive correlation between the amount of coverage and probability of loss is a sign of both types of asymmetric information. Puelz and Snow (1994) presented evidence suggesting that US automobile insurance was indeed characterized by adverse selection. However, as shown by Dionne et al. (2001) using data from Quebec, this conclusion cannot be upheld as soon as nonlinear terms of explanatory variables are included. Based on French automobile insurance data and conditioning on a large set of potential confounding variables, Chiappori and Sâlanic (2000) again find evidence of adverse selection. Cawley and Philip- son (1999) test the model’s prediction that unit premiums increase with the amount of coverage in order to compensate for the positive correlation between coverage and probability of loss. Their empirical analysis of term life insurance premiums shows that to the contrary, unit premiums fall with coverage. Moreover, low risks hold more coverage than high ones, contradicting the notion of a competing insurer siphoning off low risks with a low-premium, low-coverage alternative.

Einav et al. (2010a) argue that the more telling test of adverse selection is whether the cost to the health insurer (measured by healthcare expenditure) increases with exogenously varying premiums. Exogenous premium variation is crucial as it precludes reverse causation running from insurer’s cost to premium in response to moral hazard effects. The authors benefit from such exogeneity because in 2004, Alcoa let the presidents of some forty business units set the contribution their employees had to pay for health insurance, without having access to information about their past healthcare expenditure. On the one hand, the authors find that employee’s choices of contract have no recognizable relationship with a number of demographics. On the other hand, they do find that the insurer’s marginal and average cost increases significantly with the premium (and hence decreases with the quantity demanded), pointing to adverse selection. Their data also allow them to estimate a demand curve.
and to determine the efficient equilibrium point where price equals marginal cost. The divergence caused by the break-even condition given adverse selection, price = average cost, results in a “Harberger triangle” indicating a welfare loss of some US$10 per employee. Relative to the total welfare attainable from purchasing health insurance (the aggregated excess of marginal willingness to pay over marginal cost), this amounts to some 3%. For all the theoretical emphasis on adverse selection as a source of inefficiency of PI, this is a remarkably small figure; however, there may of course be other instances and lines of insurance where the SI alternative could reduce or even avoid a more substantial welfare loss.

Einav et al. (2011) derive another estimate of the welfare loss caused by adverse selection, this time in the UK annuity market. Between 1988 and 1994, they observe some 9,000 consumers who were mandated to purchase annuities but had the choice of a so-called guarantee period ranging from 0 to 5 to 10 years. During the guarantee period, annuitants would receive a fixed nominal benefit; in return, the longer this period, the lower the payment at death. An adverse selection effect arises since individuals with a high remaining life expectancy are predicted to opt for the 10-year period. The authors can test this prediction because they observe annuitants’ mortality up to the end of 2005. Calibrating parameters such as the coefficient of relative risk aversion and the rate of time preference and assuming that consumers can correctly predict their remaining life expectancy, they determine optimal choices. However, they also account for another source of heterogeneity, namely the utility value of consumption during the guarantee period relative to the utility value of wealth at death (e.g., for bequests). They find that if the government were to neglect this second source of heterogeneity by mandating the 5-year period chosen by the great majority, it would forgo a possible efficiency gain over the observed market outcome amounting to $423 per new annuitant. Turned the other way around, this is the welfare loss due to adverse selection effects that could be avoided by an optimal government mandate of 10 years; however, being no more than an estimated 2% of total annuitized wealth, this loss is small.

Findings of this type leave open one important question, whether a competitive market for PI would exist at all in view of the threat of “death spirals” occurring. A case study by Cutler and Reber (1998) suggests this to be a possibility. When Harvard University employees had to come up with a much higher personal contribution to health insurance, those with a favorable cost record migrated from the contract with comprehensive coverage to a more restrictive but cheaper alternative, while those with an unfavorable record kept their comprehensive policies. Within 2 years, the more generous contracts had to be withdrawn. However, it is not clear that the insurers writing them approached insolvency; rather, the evidence suggests that they withdrew loss-making contracts to find a new equilibrium, as theoretically described by Wilson (1977) and Miyazaki (1977). A similar shift towards managed care type health insurance was observed by Buchmueller and DiNardo (2002) in the state of New York, where strict community rating was imposed in 1993, in contradistinction to Pennsylvania and Connecticut. Community rating prohibits insurers from grading premiums according to risk and therefore enforces pooling contracts that allegedly triggers the death spiral. The authors did not find any sign of an increasing number of individuals without coverage in New York, which would be the consequence of a death spiral ultimately leading to the disappearance of private health insurance.

37.2.3 Separating Contracts and Efficiency Enhancement by Social Insurance

Rothschild and Stiglitz (1976) already pointed to separating contracts as a possible response of private insurers to the adverse selection challenge. The inconclusive empirical evidence presented in Sect. 37.2.2 suggests that they may often be successful in this endeavor. The intuition is that through appropriate contract design, consumers
Fig. 37.1 Potentially Pareto-improving social insurance

of a certain risk type can be made to opt for the contract appropriate for them. However, it will be shown that this subjects the low risks to a rationing of coverage. Compulsory SI serves to relax this rationing and therefore acts like a complementary element enhancing the efficiency of PI (Dahlby 1981).

In Fig. 37.1, the initial situation (no information asymmetry, prior to SI) is depicted as a benchmark. Along the two insurance lines \( P^H \) and \( P^L \), a competitive insurer breaks even, with costs of administration and risk bearing neglected for simplicity [the insurance line \( P^H (\lambda) \) will be explained below]. Premiums therefore simply cover expected loss given by \( \pi^H \cdot \text{LOSS} \) and \( \pi^L \cdot \text{LOSS} \), respectively Indifference curves \( EU^H \) and \( EU^L \) reflect the subjective tradeoffs between additional coverage and higher premium

The slope of the \( EU^H \) curve must be steeper than that of \( EU^L \) because the high risks by definition are confronted with a higher probability of loss than the low risks; therefore they are willing to accept a higher extra premium in return for an increase in coverage \( I \).

Generally the marginal utilities of wealth in the loss and the no-loss state are involved in the definition of such an indifference curve However in the special case where the wealth levels are equal (i.e., where the insurance benefit fully compensates for the loss) the two marginal utilities of wealth are equal as well. For the high risk, eg., the tradeoff at this particular point amounts to the one between paying the additional premium with probability one in return of receiving the additional benefit with probability \( \pi^H \). Therefore the indifference curve \( EU^H \) must have slope \( \pi^H \) where benefit \( I \) equals the loss. This amounts to the optimum for this type of consumer because the tangency condition is satisfied. Hence points \( H^* \) and \( L^* \) denote the first-best optima of the high and the low risk, respectively in the absence of asymmetric information.

Since a pooling contract is not sustainable in the presence of informational asymmetry, the next step is to show how separating contracts can deal with the situation. For simplicity, assume that the insurer launches a high-premium full-coverage contract designed for the high risks, which therefore attain point \( H^* \) of Fig. 37.1. At the same time, it launches a low-premium contract which however would attract the high risks as well if it were to offer full coverage (according to indifference curve \( EU^H \)), a high risk prefers point \( L^* \) to \( H^* \). Indeed, benefits pertaining to this second contract cannot be extended beyond point \( \tilde{L} \), where the indifference curve \( EU^H \) intersects the insurance line labeled \( P^L \). Evidently, separating contracts can avoid unsustainable pooling but at a price. They entail a rationing of coverage for the low risks, although they would be willing to pay the necessary premium for the additional coverage (note that \( L^* \) lies on a higher-valued indifference curve than \( \tilde{L} \)).

In this situation, SI can indeed improve the welfare of both risk types Since SI is mandatory, it amounts to a pooling contract; since it has
The benefit of SI for the low risks is that PI can now offer a complementary benefit up to $L'$ where the high risk’s indifference curve $H^*$ intersects the new insurance line $P^{L'}$. The downside for the low risks is that they must pay the SI contribution up to point $S$, which reflects also the loss probability of the high risks.

The welfare comparison between points $L'$ and $L$ is therefore ambiguous. An indifference curve drawn through $L'$ with strong curvature (indicating marked risk aversion because willingness to pay for additional coverage increases quickly as the consumer moves away from full coverage) would intersect the original $P^L$ insurance line to the right of point $L$, implying that SI with complementary PI affords a welfare gain. Conversely, an almost linear indifference curve through $L'$ (indicating little risk aversion on the part of the low risk) would intersect the $P^L$ insurance line to the left of point $L$. Therefore, SI may but need not be Pareto-improving according to this model.

Coverage provided by mandatory SI can even be strictly Pareto-improving as shown by Besley (1989). The crucial assumption is that SI is able to observe the risk type. However, the amount of coverage provided by PI must fall in order to limit moral hazard effects, an effect that was to be dubbed “crowding out” later (see Sect. 37.4). By way of contrast, Blomqvist and Johansson (1997) argue that in the absence of observability, moral hazard effects caused by PI spill over to SI, making a mixed system strictly less efficient than either the fully private alternative or comprehensive SI.

While this analysis on the whole motivates the existence of SI, it has three shortcomings:

1. It fails to explain why the political debate invariably focuses on the inability of high risks to obtain PI rather than the rationing of coverage imposed on the low ones.

2. It does not determine the optimal amount of SI (note that point $S$ of Fig. 37.1 was chosen arbitrarily).

3. It does not explain the expansion of SI over time (from a few percent of GDP prior to World War II to around 15% by 1980 and some 25% at present in a typical industrial country (OECD 2007; see also Table 37.1).

The first shortcoming can be remedied by assuming risk-specific transaction costs for PI that are reduced by SI [the argument is a modification of Newhouse (1996); see also Zweifel and Eisen (2012) Ch. 9214]. In Fig. 37.1 let these transaction costs give rise to a proportional loading $\lambda$ so large as to cause the high risks to go without any PI coverage at all, while the low risks continue to pay the actuarially fair premium ($\lambda = 0$). Graphically, the dashed indifference curve $EU^H[0]$ through the origin indicates a higher expected utility than any point on the dashed insurance line $P^H(\lambda)$. In addition, let this lading reflect the price of risk bearing by PI, which is substantial in the case of high risks because, eg., their losses not only have high expected value but also high variance and (positive) skewness.

In this situation, SI, by imposing coverage up to point $S$, can serve to reduce the conditional variance and skewness of losses to be borne by PI. For simplicity, assume the loading to become zero, making the insurance line $P^H$ originating from $S$ the relevant one, as before. The high risk is now predicted to move from zero coverage to point $H^{**}$ of Fig. 37.1 entailing an even greater welfare gain than in the case discussed initially.

In this sense, SI can be said to solve the problem of high risks not obtaining coverage.

The other two shortcomings are addressed in turn in Sects. 37.3 and 37.4, respectively.

### 37.3 The Optimal Amount of Social Insurance

As noted towards the end of Sect. 37.2.3, the efficiency view of SI leaves the optimal amount
of coverage to be provided by SI relative to PI undetermined and with it, the optimal division of labor between SI and PI. There have been two strands of literature addressing this issue. One is to introduce a benevolent government as the agent who optimizes SI on behalf of a representative citizen. The other recognizes that the implementation of SI ultimately requires a majority support by voters.

### 37.3.1 A Benevolent Government

**Optimizing the Amount of Social Insurance**

Health insurance constitutes a leading case where the government determines the division of labor between PI and SI, in particular by permitting citizens to purchase supplementary coverage offered by private insurers. This is the focus of Petretto (1999), whose model comprises three stages. First, the government introduces SI with a degree of coverage \( \alpha (0 < \alpha < 1) \). Next, consumers select their preferred amount of PI, which is partial in response to the loading contained in the premium (in Fig. 37.1, the high risks choose zero coverage because of a very high loading). This leaves them with some positive net expense in the loss state, which depends on the consumer’s decision with regard to healthcare expenditure and labor supply (and hence available income). While the model is couched in health insurance terms its three-layer structure is rather general. With regard to unemployment, for instance, SI imposes a rate of income replacement.

Frequently, the employer acts as a private insurer by providing a severance package (the premium being a wage rate below the marginal product of labor). In the case of provision for old age, SI provides for a first layer in the guise of a (typically tax-financed) public pension. A second layer comes from PI either purchased individually in the guise of a whole life insurance contract or again provided by the employer and financed by a wage deduction.

The amount of loss falling on PI is not exogenous anymore (contrary to the exposition in Sect. 37.2.1) but now subject to an ex post moral hazard effect. In the case of a health loss, this is the extra healthcare expenditure that would not be incurred without insurance coverage. In the case of unemployment, the worker may exert less search effort to find a new job. Similarly, retirees may seek to maximize their remaining life expectancy in response to a pension that is topped up by a PI contract. Here, the out-of-pocket component could be equated to the difference between preretirement consumption and achievable consumption over the life span remaining.

In keeping with the three layers, optimization occurs in three stages. First, the government sets an optimal uniform degree of coverage \( \alpha \) of the possible loss by maximizing the sum of citizens’ utilities. It is subject to a budget constraint that involves a uniform rate of taxation \( t \) on average labor income that depends on labor supply in the loss and the no-loss state. In the second stage, individuals select their (nonuniform) degree of private insurance coverage \( 1 - k_i \), with \( k_i \) denoting the rate of cost sharing. To simplify the analysis (and contrary to Sect. 37.2), the loss probabilities are assumed to be known; therefore, SI cannot be credited with relaxing the rationing constraint imposed on low risks by separating contracts. In the third stage, individuals decide about their amount of healthcare expenditure (or more generally, the amount of loss falling on the private insurer).

As usual, the model is solved backwards. As for the optimal degree of PI coverage, Petretto (1999) obtains

\[
\frac{1 - k_i^*}{k_i^*} = \frac{u'[\text{Loss}]}{u_i(1 + \lambda)} \cdot \frac{1}{e_i}, \quad (37.1)
\]

with \( e_i > 0 \) an elasticity (in absolute value) relating the loss borne by PI to the net cost borne by the insured.

This result is quite intuitive. The optimal degree of coverage in the PI layer [the right-hand side of (37.1)] is higher:

- The higher the degree of risk aversion of individual \( i \). The marginal utility of income in the loss state, \( u'[\text{Loss}] \), is higher than the weighted average \( \overline{u}_i \) over the loss and the no-loss states.
The marginal benefit associated with an increase of $\alpha$ is shown on the left-hand side. It consists of two components:

- The first term on the left-hand side amounts to a social risk-sharing gain. It is high if over the $n$ individuals considered, there is a marked covariance between their marginal utility of wealth in the loss state and the amount of net loss they have to bear because PI is only partial. Here, one has to take into account that PI coverage demanded typically decreases when SI coverage expands [this “crowding-out” effect is taken into account in the contribution by Chetty and Saez (2009), to be discussed in Sect. 37.3.2 below]. Generally, the amount of covariance is high if the marginal utility of income is high in the loss state, i.e., if individuals exhibit strong risk aversion.

- The second term on the left-hand side of (37.2) is the social redistribution gain. Its main term is a covariance between the individual’s relative average marginal utility and his or her relative net loss. The two quantities are relative because they relate the individual to a mean marginal utility in the population ($\bar{u}^*$) and the mean loss striking the population (NetLoss), respectively. The individuals considered have a strong interest in redistribution through SI if this covariance is high, i.e., if an individual suffers from an above-average net loss simultaneously is characterized by an above-average marginal utility of income (which can be taken to reflect below-average income or wealth). The marginal utility of income prevailing in the population $\bar{u}^*$ is used to translate this benefit into utility terms. Together, these two marginal benefits have to equal marginal cost at the optimum. Marginal cost consists of three multiplicative components:

- The last factor is the basic trigger. It shows how strongly the expected net loss associated with individual $i$ reacts to an expansion of SI. This is the ex post moral hazard effect originating with SI (the probability of loss is observable by assumption, thus precluding ex ante moral hazard).

- The second factor shows that the degree of both SI and PI coverage acts as amplifiers. The main impulse comes from $\alpha^*$ itself. In addition, however, PI covers the remainder $(1 - \alpha^*)$ to the tune of $(1 - k_i^*)$.

- Finally, the first factor transforms these effects into a utility value, this time using the marginal utility in the loss state because ex post moral hazard occurs in the loss state by definition. Equations (37.1) and (37.2) combined define an optimal division of labor between PI and SI.

\[
\sum_i^n \text{Cov}(u'_i \text{[Loss]}_i, \text{NetLoss}_i) + n[\bar{u}' \text{[Loss]} \cdot \text{Cov}(\bar{u}' / \bar{u}', \text{NetLoss}_i / \text{NetLoss}) - 1] \cdot \text{NetLoss}
\]

\[
= \sum_i^n u'_i \text{[Loss]} [(1 - \alpha^*)(1 - k_i^*) + \alpha^*] \frac{\partial \text{NetLoss}_i}{\partial \alpha}
\]
37.3.2 Taking into Account Adverse Selection and Crowding Out

In view of the results presented in Sect. 37.2, a major challenge for determining the optimal division of labor between PI and SI is to take into account problems associated with informational asymmetry but also the fact that an expansion of SI does not increase total insurance coverage in step because PI is displaced to some degree (the "crowding-out" phenomenon, which will be the topic of Sect. 37.4). This challenge has been taken up by Chetty and Saez (2009), who use unemployment insurance (UI) as their leading example. One might argue that this type of insurance is exclusively provided by SI. While the division of labor between PI and SI is indeed much more debated in the context of health insurance, there is a PI involvement in UI to the extent that employers provide severance pay. Both components of UI are financed by a payroll tax, one explicit (\( \tau \)) levied by the government, the other, implicit (\( t_k \)) by the employer, who can observe the wage income \( W_k \) of type \( k \). As in the preceding section, there is a benevolent government who seeks to maximize citizens’ expected utilities. Through setting \( \tau \), it also determines the benefit rate \( \alpha \) of the preceding section if the SI scheme is to maintain its financial equilibrium. Chetty and Saez (2009) show that the optimum explicit contribution rate is given by

$$\frac{\tau}{1 - \tau} = \frac{1}{e_{W,1-r}^{\text{Cov}(\overline{u}_k, \overline{W}_k)}} + \frac{1}{e_{W,1-r}^{u \cdot W}} \cdot \sum_k S_k (1 - r_k) e_k (t_k^* - t_k) \cdot \frac{\overline{u} \cdot \overline{W}_k}{u \cdot W}. \tag{37.3}$$

Note that the left-hand side of (37.7) is increasing in \( \tau \). The three effects of interest are represented as follows:

- **Moral hazard:** This is reflected by two parameters. The first is the elasticity \( e_{W,1-r}^{\text{Cov}(\overline{u}_k, \overline{W}_k)} \) which indicates how strongly (labor) income averaged over the loss and the no-loss state reacts to the amount that the worker retains per dollar earned and hence to the payroll tax \( \tau \) levied by SI. The stronger this response, the more the benefits of SI in terms of income smoothing (see below) have to be adjusted downwards. It corresponds to the elasticity \( e_i \) in (37.1) as in Petretto (1999), who however does not yet distinguish between moral hazard effects induced by the SI and PI layers. This is achieved here by the second parameter \( e_k \), an elasticity indicating the response of \( \overline{W} \) to the overall retention per dollar earned, given by \((1 - \tau)(1 - t_k)\). The stronger this group-specific effect of PI (relative to SI), the more the scaling down of SI occasioned by \( e_{W,1-r}^{u \cdot W} \) has to be corrected in favor of SI again.

- **Adverse selection:** This is represented by \( (t_k^* - t_k) \), the difference between the PI contribution rate (levied by the employer) in the absence of the self-selection constraint (\( t_k^* \)) and in its presence (\( t_k \)). This difference is positive due to the rationing effect of adverse selection (see Fig. 37.1 again), indicating a benefit favoring SI.

- **Crowding out of PI:** This is represented by the group-specific parameter \( r_k > 0 \) which is the elasticity of the PI retention rate \((1 - t_k)\) with respect to the SI retention rate \((1 - \tau)\). It represents what Chetty and Saez (2009) call a fiscal externality, which would be absent from informal insurance provided e.g., among members of the same household. Since both components of unemployment insurance are subject to budget balance, this elasticity indirectly indicates how much of employer-provided coverage is crowded out by SI (however, for simplicity the model does not contain a module representing optimizing behavior of either employers or employees in this respect).

The greater the \( r_k \), the smaller should be the optimal extent of SI, ceteris paribus. As will be shown in Sects. 37.4.2 and 37.4.3 below, crowding-out effects are indeed substantial, serving to reduce the social benefit that can be expected from a (continued) expansion of SI. The remaining elements of (37.3) can be interpreted as follows: The \( \text{Cov} \) term represents the
consumption smoothing benefit of SI and PI combined. For each group \( k \) of the population, their marginal utility of income averaged over the loss and the no-loss state \( \bar{u}_k \) is compared with that of the population of the whole \( \bar{u} \). This ratio is high (due to the concavity of the risk utility function) when relative income \( (\bar{W}_k/\bar{W}) \) is low, resulting in a negative covariance, very much as in (37.2) above. The group-specific covariance values have to be averaged to obtain \( \bar{\text{Cov}}(\cdot) \); together with the negative sign, this amounts to a benefit. It has to be qualified in terms of the moral hazard effect \( e_{\bar{W},1-r} \) discussed above. The summation in the second term is over the population shares \( s_k \) and bears close similarity to a covariance between (relative) marginal utility and income levels again [although the last factor is not expressed as deviations from expected values, it still says that the benefit of SI is large if high (relative) marginal utilities of income coincide with low (relative) levels of wealth]. Here, the relative moral hazard effect of PI relative to SI given by \( e_k/e_{\bar{W},1-r} \), the rationing due to adverse selection \( (t^*_k-t_k) \), and the leakage due to the crowding-out effect \( (1-r_k) \) all enter as qualifying factors.

Clearly, the averaged marginal utility of income depends on their values in the loss and the no-loss state, which in turn are determined by the curvature of the risk utility function. The relationship with the coefficient of (relative) risk aversion is established by Chetty (2006) who also shows that in spite of later modifications, an influential finding by Bailey (1978) continues to be valid. It states that the optimal amount of UI benefits is determined by (1) the mitigation of the drop in consumption achieved by UI, (2) the coefficient of relative risk aversion, and (3) the elasticity of unemployment duration with respect to UI benefits. The analysis by Chetty and Saez (2009) shows that these parameters are indeed crucial. In (37.2), the mitigation of the drop in consumption is reflected by \( \bar{W}_k/\bar{W} \), the role of (differences in) risk aversion indicated by \( \bar{u}_k/\bar{u} \), while the elasticity of duration has become the elasticity of earnings \( e_{\bar{W},1-r} \).

Finally, Chetty and Saez (2009) explicitly relate the SI contribution rate \( \tau \) to a PI contribution rate \( \hat{\tau} \), which is weighted by population shares, relative incomes, and relative marginal utilities,

\[
\hat{\tau} = -\frac{\tau}{1-\tau} + \frac{1-\hat{\tau}}{e_{\bar{W},1-r}} \cdot \frac{\text{Cov}(u', W)}{\bar{W}}.
\]  

Clearly, this defines an optimal crowding-out relation in that a higher (optimal) value of \( \tau \) implies a lower value \( \hat{\tau} \), ceteris paribus (first term on the right-hand side). However, the PI component is also to be credited for its contribution to consumption smoothing (indicated by the average covariance between marginal utility \( u' \) and income \( W \), which is usually negative), qualified by the crowding-out factor, with \( \hat{\tau} \) symbolizing an average value of \( r_k \) weighted in the same way as \( t_k \). Equations (37.3) and (37.4) together indeed define an optimal division of labor between PI and SI in the case of unemployment and similar risks such as workplace accidents.

### 37.4 The Crowding-Out Phenomenon

#### 37.4.1 The Public Choice View of Social Insurance

Contrary to the basic assumption adopted in Sect. 37.3, government in a democracy cannot act as a benevolent dictator but must act on behalf of citizens who know their risk type and position in the income and wealth distribution. Referring back to Fig. 37.1 above, it is evident that the high risks would always gain from a continued expansion of SI to the detriment of PI. At first sight, such a shift appears even more likely if individuals not only differ in terms of risk but also in terms of income. As voters they would have an incentive to see SI expanded because it also redistributes income from the rich to the poor. However, it is the preferences of the median voter who is decisive in determining whether or not a proposal finds a majority in the political process. Indeed, the middle class may favor a mixed system because on the one hand it can benefit from the better deal offered by PI; on the
other, SI enables a general expansion of coverage but would in the limit entail an excessive burden in terms of income redistribution (Gouveia 1997).

In the following the model by de Donder and Hindricks (2003) is presented because it seems best suited to explain the shifting division of labor between PI and SI once voters are offered a choice between them. In this way, it provides a strong theoretical basis for the crowding-out findings to be reported in Sect. 37.4.2 below. Indeed, the model predicts support for SI against PI even by voters who are slightly richer and are slightly better risks than average regardless of the distribution of income and risk in the country. This sweeping prediction goes beyond Casamatta et al. (2000), who merely derive conditions that shift the division of labor between PI and SI in favor of SI. It comes at a price, however. Rather than the conventional expected utility (EU) framework, Yaaris (1987) dual formulation is adopted by the authors. In this formulation, risk aversion is expressed not by the concavity of the risk utility function but by the concavity of a weighting function \( \phi(p) \) defined over probabilities \( p \). This permits to express the utility \( V^P \) derived from a PI contract directly in terms of predetermined wealth or income \( W \),

\[
V^P = \varphi(p)(W - P - k \cdot 1) + (1 - \varphi(p))(W - P)) = W - P - \varphi(p)(1 - k), \tag{37.5}
\]

with \( \varphi(p) = (1 + A)p \) (\( A \) is explained below).

Therefore, utility amounts to a weighted average of wealth levels in the loss and the no-loss state, respectively. In the loss state, income is net of the premium \( P \) and the rate of coinsurance \( k \) on a loss that is normalized to 1. This normalization makes the loss a fixed quantity, thus precluding any moral hazard effect, in contradistinction to Sect. 37.2.3. In the no-loss state, the premium still needs to be paid. The two income levels are weighted by a weighting function \( \varphi(\cdot) \) whose parameter \( A \) reflects risk aversion; for simplicity, \( A \) is assumed independent of \( W \). Note that (37.5) implies choice of full coverage \( (k = 0) \) even if the premium exceeds its fair value \( p \) (which is commonly observed behavior, however). This differs from the prediction derived from EU theory (see Fig. 37.1 again). It is the consequence of the fact that risk aversion has a first-order influence here but merely a second-order influence in EU theory (stemming from the concavity of the risk utility function).

The SI contract can offer full coverage \( [(1 - k) = 1] \) since it suffers neither from moral hazard nor risk selection. On the other hand, contributions are scaled according to the individual’s wealth compared to the average \( \bar{W} \), reflecting the redistribution characteristic of SI. Therefore, the utility value of SI is given by (37.6),

\[
V^S = W - (W/\bar{W})\bar{p} \tag{37.6}
\]

Before comparing (37.5) and (37.6) de Donder and Hindricks (2003) proceed to impose the self-selection constraint that renders separating PI contracts viable, in analogy to Fig. 37.1. They apply the criterion introduced by Mailath (1987),

\[
dV^P/d\bar{p} = 0 \quad \text{for} \quad \bar{p} \to p, \tag{37.7}
\]

stating that an individual cannot gain by announcing a risk type \( \hat{p} \) arbitrarily close to the true one. This permits them to derive an optimal coverage function, relating \( k^* \) to \( p \) and \( A \). One can now define a wealth level \( W^{\circ} \) where the two utilities are equal. After substitution of \( k \) in (37.5), the pertinent condition reads

\[
W^{\circ} - (1 + A)p + Ap(\bar{p}^{1/A})^{1/A} = W^{\circ} - (W^{\circ}/\bar{W})\bar{p} \tag{37.7}
\]

The new symbol \( \bar{p} \) is the maximum loss probability in the population; it is part of the optimal benefit function. Therefore, the last term on the left-hand side of (37.7) expresses the benefit in financial terms that an individual characterized by a loss probability \( p \) relative to \( \bar{p} \) can derive from the separating PI contract. Taking the log of this term and differentiating it w.r.t. \( A \) results in a positive value, indicating that this benefit increases with risk aversion, as one would expect. However,
the weighting function $\phi(\cdot)$ causes the premium paid (the term $-\left(1 + A\right)p$ on the left-hand side) to increase as well. Solving condition (37.7) for the wealth level $W^o$ indicating indifference between PI and SI yields

$$\frac{W^o}{W} = \left[1 + A - A \left(\frac{p}{\bar{p}}\right)^{1/A}\right] \frac{p}{\bar{p}} = \left[1 + \varphi(A, p)\right] \frac{p}{\bar{p}},$$

(37.8) with $\varphi(A, p) = A \left[1 - (p/\bar{p})^{1/A}\right]$. This condition can be interpreted as follows.

Consider an individual with the mean loss probability $\bar{p}$; since $\phi(\cdot) > 0$, $W^o/\bar{W}$ must exceed one implying that such an individual votes in favor of SI although his or her income is above average. Conversely, assume that the pivotal voter has below-average wealth. In that event, he or she continues to support SI rather than PI if his or her loss probability $p$ corresponds to the average $\bar{p}$; indeed, it could even be somewhat below average. Ultimately, the reason for this dominance of SI lies with the term $(p/\bar{p})$, which reflects the necessity of sustaining separating contracts. This detracts from the benefit of PI, whereas SI does not need to observe this restriction.

The final step is forming the derivative of $V^s$ with respect to $1 - k^s$, which the authors show to be positive; in addition, its value does not depend on $1 - k^s$, which implies the corner solution with $1 - k^s = 1$, i.e., full coverage by SI. Moreover, individuals with the crucial income $W^o$ turn out to be indifferent to an expansion of SI, while those with less income (the majority in a country with a skewed wealth distribution) prefer to have full SI coverage. For this reason, a mixed system cannot be sustained politically.

In this way, the analysis provided by de Donder and Hindricks (2003) comes close to explaining the expansion of SI since its inception. It could be interpreted as the transition to a political equilibrium in which a loss of a given type is covered by SI exclusively. The puzzle then becomes why this expansion seems to have slowed down recently (see Table 37.1 again). A possible explanation is that several exogenous changes (eg., increased migration, to be discussed in Sect. 37.6) cause increasing costs of enforcement for SI, causing a loading in the SI contribution $\bar{P}$ of (37.5) that may push the threshold value of $W^o$ sufficiently below $\bar{W}$ to establish a stalemate between supporters and opponents of SI.

## 37.4.2 Evidence on the Crowding-Out Phenomenon: Private Saving

The work by de Donder and Hindricks (2003) presented in the preceding section predicts that SI replaces (“crowds out”) PI in the political process. The underlying reason is that PI is beset with the problem of adverse selection, which makes it more costly than SI, causing an expansion of SI to be efficiency enhancing. But what if SI crowds out private saving, which cannot be claimed to be subject to adverse selection effects? After all, banks are hardly concerned about “risky depositors” who have to be staved off by paying them low interest rates. Using macroeconomic data and viewing so-called SSW as exogenous (in contradistinction to Sect. 37.3), Feldstein (1974) seemed to present conclusive evidence of such an effect (see the Introduction again). However, Leimer and Lesnou (1982) found his findings to be mainly caused by an error in the calculation of the SSW variable.

The decisive support for the “crowding-out” hypothesis with regard to private saving came from Hubbard et al. (1995), who provide both a theoretical basis and empirical evidence based on individual data. They note that many SI programs (eg., food stamps in the USA) are means-tested with regard to assets rather than with regard to earnings. These programs have two effects, both of which may depress private saving, thus causing what can be called a welfare-reducing “fiscal externality”. First, they reduce uncertainty...
concerning future consumption levels, making precautionary saving less important. Second, they increase (often sharply) the opportunity cost of accumulating wealth especially for low-income households who would otherwise qualify for the SI program. The data come from the US panel study of income dynamics (PSID). Using education of the family head as an indicator of lifetime income, the authors indeed find that wealth in 1984 (measured as the number of years it could replace estimated permanent income) does not have the same age profile across educational levels. It exhibits the hump shape predicted by the life-cycle consumption model only among households with a college degree; for low-income households, the age profile is rather flat. The authors consider three explanations for this difference: the bequest motive for high-income families, a higher share of consumption during retirement covered by SI among low-income households, and a lower rate of time preference among the poor. They show that these explanations fail to provide a convincing explanation of observed patterns, leaving a fourth: asset means-tested SI.

Hubbard et al. (1995) then build a model designed to determine optimal consumption over an individual’s lifetime that incorporates uncertainty with regard to lifespan, earnings (taking into account unemployment insurance, however), and out-of-pocket medical expenses. Estimates of uncertainty are derived from the error variances of regressions of these variables on (powers of) age. The consumption floor is derived from all SI programs that are asset means-tested (such as food stamps and housing assistance), but not, eg., unemployment insurance which is earnings-tested. They then solve the stochastic dynamic optimization problem and use its solution to simulate age profiles for consumption and wealth. When either uncertainty is neglected or the consumption floor provided by SI is set to the counterfactual value of $1,000 (1984 prices), their simulations fail to replicate the age-wealth profiles observed in the PSID. However, when uncertainty is combined with the actual $7,000 consumption floor, the replication is quite close. This permits the authors to conclude that asset-tested SI crowds out savings of low-income households.

Using PSID once more, Gruber (1997) analyzes the consumption smoothing benefits of US unemployment insurance. For the period 1968–1987 (but without the recession year 1973 for lack of data, likely causing an underestimate), he finds that an increase in the income replacement rate of 10 percentage points reduces the fall in consumption by 2.65% during the unemployment spell. If this fall were zero, then the comitant reduction in income would be entirely to the detriment of private saving. However, the smoothing of consumption is only partial, implying that the crowding-out effect of unemployment insurance is partial as well.

Gruber and Yelowitz (1999) single out the US Medicaid program to test for the crowding-out effect of means- and asset-tested SI on private saving by low-income households. Between 1984 and 1993, expenses of Medicaid increased by 500%; however, this expansion occurred at a markedly different pace between US member states, creating a source of exogenous variation. Criteria for eligibility were relaxed; in particular, the income threshold was lifted from 1987. The authors create a variable they call “Medicaid eligible dollars” (MED) by multiplying, for each member of a family, the Medicaid benefit with an imputed likelihood of being eligible. Their dependent variable is the household’s net wealth; it is indeed found to decrease by an estimated 29% for every $1,000 of MED, pointing to a substantial crowding-out effect, according to which the expansion of Medicaid induced a reduction of 8.2% over time. The estimate of 2.9% drops to 2% if there had been no asset test but increases to 5.3% given an asset test, underscoring the importance of this feature prevalent in many SI programs.

37.4.3 Evidence on the Crowding Out Phenomenon: Private Health Insurance

The study by Gruber and Yelowitz (1999) does not provide evidence bearing on the crowding out of private insurance by public insurance [unless one is willing to acknowledge (precautionary) saving as a type of private self-insurance, as in...
Kimball (1990)]. This gap is filled by Cutler and Gruber (1996), whose theoretical background can be illustrated by Fig. 37.2. The straight line \( ABC \) depicts the budget constraint of an individual whose income is low enough to qualify for an SI program. Since “all other goods” corresponds closely to income, the slope of the budget constraint (inversely) reflects the amount of income that has to be sacrificed for additional insurance coverage, i.e., the premium per unit coverage provided by PI. If risk averse, the individual is characterized by an indifference curve making \( H^* \) the optimum; for a less risk-averse type, the optimum is \( L^* \). Now let the government offer an SI program with a take-it-or-leave-it benefit falling short of \( L^* \) at a subsidized contribution rate. This causes the budget constraint to become \( ABSC \). If the program does not require a contribution, thus effectively amounting to public welfare, the budget constraint becomes \( ABMC \). The H-type individual is unaffected; however, the L-type is predicted to typically opt for the SI program at point \( S^0 \) or \( M^0 \), respectively. These corner points dominate because both constitute accumulation points induced by many types of indifference curves with differing slopes. In either case, the amount of PI demanded decreases.

Now consider a poorer individual with budget constraint \( A'B'C'' \), with the two preference types unchanged (homothetic preferences are assumed). Absent the SI program, the H-type opts for \( H^{**} \). When the program becomes available, the L-type shifts to it for sure (to point \( S' \) or \( M' \), respectively). However, even the H-type is now predicted to opt for the SI alternative, with the result that SI crowds out PI regardless of type. The reason is that for a low-income individual, an SI option of a given amount has a much higher relative importance than for a higher-income individual, a difference which may swamp the influence of a preference structure that in principle favors the PI alternative (see \( H^{**} \) again). This effect is reminiscent of Hubbard et al. (1995), who also emphasized the impact of SI on low-income households. It may be counteracted to some extent when the amount of PI rather than the likelihood of having it is considered, since the transition from \( H^{**} \) to \( S' \) (or \( M' \)) entails a larger reduction of PI coverage by the richer individual than the transition from \( H^{**} \) to \( S \) (or \( M \)) by the poorer individual.

Cutler and Gruber (1996) again take the expansion of US Medicaid (in particular benefits for mothers with children) as the exogenous impulse to study its impact on the demand for private health insurance. They ignore considerations of adverse selection in PI that might contribute to its crowding out by the public alternative. To some extent, this can be justified by noting that much of US health insurance is contracted through employers, who by imposing open enrolment on health insurers prevent them from cream skimming. The authors start by noting that about two-thirds of mothers and children who became eligible had PI coverage initially. As in Gruber and Yelowitz (1999), they estimate the probability of eligibility for children and women of childbearing age. For children, eligibility has a positive...
effect on the take-up of Medicaid but a negative one on the likelihood of having private health insurance, amounting to a crowding out of 31% per percentage point of Medicaid expansion. For women, both the effect on take-up of Medicaid and (negative) on private health insurance are insignificant, but if taken at face value, they even suggest a crowding-out rate in excess of 100%. For the elderly, Medicaid coverage for long-term care (LTC). Notably, in 1998 the State Children’s Health Insurance Program (SCHIP) was initiated; however, beneficiaries have to prove that they have been without private health insurance (usually due to a job loss) for between 6 and 12 months, depending on the state. Indeed, Lo Sasso and Buchmueller (2004) concluded that these waiting periods serve to significantly reduce crowding out. In their re-estimation using a different data set, Gruber and Simon (2008) are unable to replicate this finding, estimating overall crowding-out rates for SCHIP to be similar to those for Medicaid. As to Medicaid, they obtain rates ranging from 24 to 81%, bracketing their original estimate of 49%. The large variation is caused by differing assumptions regarding the changes in composition of the subgroup that has both private and public coverage.

Another instance of Medicaid displacing PI is coverage for long-term care (LTC). Brown and Finkelstein (2008) emphasize that contrary to the Medicare program (for the elderly), Medicaid (for the poor) does not permit PI to top up benefits; rather, PI (which provides partial coverage) must be used up until wealth falls to the low Medicaid threshold. Using a dynamic optimization algorithm, the authors derive willingness-to-pay (WTP) values for LTC coverage as currently offered in the presence of Medicaid, to find a value of $20,700 (women) and -$18,200 (men) at the 30th percentile of the US wealth distribution. The WTP values turn positive no sooner than at the 60th percentile. They then assume fair premiums for PI combined with full LTC coverage—only to still find negative WTP values up to at least the 50th percentile. The cause is the implicit tax on PI which reaches values close to 100% for wealth levels below the 30th percentile. As a consequence, full loadings on PI approach 100% in this domain, making coverage very expensive indeed. While the authors do not explicitly calculate crowding-out rates, their simulations clearly suggest substantial crowding-out effects of Medicaid in spite of the fact that its coverage is partial only.

A few other countries have experienced a crowding out of private LTC coverage by public programs, notably Japan, South Korea, and the UK, where the National Health Service covers LTC expenditure as well. However, the evidence is descriptive rather than quantitative (Le Corre 2012).

37.5 Could the Division of Labor Between Private and Social Insurance Be Improved?

For all their merits, the contributions discussed up to this point have one weakness in that they all consider one risk impinging on the individual at a time. However, PI and SI both consist of a multitude of lines of insurance, in response to the fact that there are a great many impulses affecting individual assets, which comprise wealth but also health and skills [Zweifel and Eisen (2012), Ch 1.6]. For instance, an illness episode lowers not only the level of health but indirectly wealth (through reduced labor income) and skills (through depreciation of human capital). The same is true of an accident, of disability, and unemployment, resulting in a
The Division of Labor Between Private and Social Insurance

Table 37.2  Comparison of nominal rates of return on private and social insurance

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private $r_G$</td>
<td>Social $\dot{W}$</td>
</tr>
<tr>
<td>USA</td>
<td>6.92</td>
<td>6.89</td>
</tr>
<tr>
<td>Sweden</td>
<td>8.22</td>
<td>7.69</td>
</tr>
<tr>
<td>UK</td>
<td>8.58</td>
<td>8.55</td>
</tr>
<tr>
<td>Japan</td>
<td>5.25</td>
<td>8.33</td>
</tr>
</tbody>
</table>

Source: OECD economic outlook (2008)

Table 37.3  Correlations of trend deviations in US private insurance, 1965–2004

<table>
<thead>
<tr>
<th></th>
<th>PLID</th>
<th>PLIDI</th>
<th>PLAI</th>
<th>PHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLID</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLIDI</td>
<td>-0.0534</td>
<td>0.7435</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>PLAI</td>
<td>-0.0227</td>
<td>0.8896</td>
<td>-0.3417$^a$ (0.0282)</td>
<td>1.000</td>
</tr>
<tr>
<td>PHI$^b$</td>
<td>0.4830$^a$ (0.0033)</td>
<td>0.1358 (0.4366)</td>
<td>0.4584$^a$ (0.0056)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

PLID: life insurers’ death payments; PLID: life insurers’ disability payments; PLAI: life insurers’ annuity payments; PHI: health insurance payments

$^a$ Coefficient significant at the 5% and 1% level, respectively

$^b$ 1970–2004; another source of data leads to different results

positive correlation between these three assets. Ettner (1996) finds a positive correlation between health and wealth, Ashenfelter and Rouse (1998) one for health and skills, and Kenkel (1991) one for health and skills. The basic intuition is that in this situation, a risk-averse individual benefits from PI and SI coverage jointly neutralizing these correlations. As will be shown below, this requires deviations of paid benefits from expected value to be negatively correlated across lines of insurance. Failure to satisfy this condition would indicate the lines where the division of labor between PI and SI could be improved.

Consider two out of the three assets, $X$ and $Y$, say. Let $X$ be health transformed into monetary units. Since both changes in the probability of survival and in the quality of life can be valued using WTP and time trade-off techniques, the asset “health” can be made commensurable to the other asset, $Y$ [Zweifel et al. (2009), Ch 2.4]. Let this asset $Y$ be skills; their value can be ascertained from wage differentials on the labor market. Adopting a $(\mu, \sigma)$-framework, expected returns of PI must be close to a (quasi-) risk-free rate of return because of regulators’ solvency concerns. As is well known (Samuelson 1958), SI also has a rate of return, given by the growth of the labor income from which benefits for the retired are paid. Benefits of social disability insurance tend to change in step with labor income as well. To the extent that social health insurance is financed through a payroll tax, its benefits develop very much in line with wage income (assuming budget balance for SI).

In all, the expected rate of return of private insurance can be approximated by the return on long-term government bonds (given that they constitute the most important asset of most insurers), $(r_G)$, while that of social insurance, to the growth of wage income ($\dot{W}$). This constitutes an acceptable indicator even in countries where SI is integrated in the public budget, causing its benefits to depend on tax revenue; however, tax revenue importantly depends on labor incomes. Over a longer-run horizon, these two rates have not differed markedly according to internationally comparable OECD estimates (see Tables 37.2 and 37.3). For example, between 1960 and 2007, the average nominal interest rate on long-term government bonds was 6.92% in the USA, while nominal earnings grew by 698% (US domestic sources even arrive at a somewhat lower rate). However, during the turbulent later 1970s $\dot{W}$ exceeded $r_G$. In Sweden $(r_G)$ has exceeded $(\dot{W})$ on the long run but again with a temporary reversal in 1974–1979. Thus the expected rates of return on private and social insurance may be set equal as a first approximation.
This allows the analysis of comparative performance to be limited to the variance component of \((\mu, \sigma)\). Finally, since premiums and contributions to private and social insurance are predictable in most situations, they are treated as nonstochastic quantities in what follows, which permits focusing on the properties of the payment side.

Specifically, denote by \(X^a + x\) the final value of health capital, with \(X^a\) the asset value after expected payment by insurance.\(^2\) Under ideal circumstances, \(X^a\) would be a fixed (optimal) quantity. However, medical care may fail to restore the health stock to its optimal value; more importantly in the present context, any copayment gives rise to an unexpected variation (denoted by \(x\)).

Likewise, the total value of skills can be split into \(Y^a\) (after expected insurance benefits) and \(y\) (unexpected variation in benefits). Unexpected benefit variation even occurs in SI. For example, a beneficiary may learn that a degree of negligence was found on his or her part in a workplace accident, causing workers’ compensation to be curtailed. Conversely, benefits may be higher than expected because, e.g., the presence of a child triggers a supplementary benefit.

Under these circumstances, total asset variance is given by

\[
\text{var}(X^a + x + Y^a + y) = \text{Var}(X^a) + \text{Var}(x) + \text{Var}(Y^a) + \text{Var}(y) + 2\text{Cov}(X^a, x) + 2\text{Cov}(X^a, y) + 2\text{Cov}(Y^a, y) + 2\text{Cov}(Y^a, x) + 2\text{Cov}(x, y). 
\]

(37.9)

Here, \(\text{Var}(X^a)\) and \(\text{Var}(Y^a)\) remain positive to the extent that insurance stipulates a degree of copayment in response to moral hazard or as a property of separating contracts (see Sects. 37.2.1 and 37.2.3). Under the usual assumption that expected and unexpected components of insurance benefits are uncorrelated, \(\text{Cov}(X^a, x)\) and \(\text{Cov}(Y^a, y)\) can be set to zero. Moreover, expected benefits under one title (\(X\), say) presumably are not related to unexpected benefits under another title (\(Y\)). Thus, \(\text{Cov}(X^a, y) = \text{Cov}(Y^a, x) = 0\). Also, while expected benefits are negatively correlated with variations in the covered asset, some of the post-insurance variation usually remains. Thus, the positive covariance between the two assets will still show after payment of expected benefits, making \(\text{Cov}(X^a, Y^a) > 0\). The one term related to insurance that can be used for minimizing total asset variance therefore is \(\text{Cov}(x, y)\). The more strongly negative the covariance between unexpected deviations form expected benefits, the smaller the total asset variance. Thus, the performance of the insurance system as a whole from the consumer’s point of view can be gauged by the direction and amount of correlation in its unex-

pected benefit components.

By fitting the time series of insurance benefits to quadratic time trends to (roughly) account for both their expansion and inflation in the economy, residuals can be calculated that reflect unexpected deviations of benefits from their expected value. These residuals are then used to determine correlation coefficients \(\rho_{x,y}\) between two lines of insurance. For a given pair of standard errors \(\{\sigma_x, \sigma_y\}\), this coefficient indicates the size of \(\text{Cov}(x, y)\) (see Eugster et al. 2012 for details).

In the case of the USA (Table 37.3), four lines of personal insurance can be distinguished, reflecting data availability. There are only two signifi-

antly positive correlations out of six, between unexpected benefit variations in life insurers’ disability payments (PLID) and health payments (PHI) [cell (1,3) of Table. 37.3] and between PHI and life insurers’ annuity payments (PLAI) [cell (3,3)]. One correlation is significantly negative.

In an earlier study, using data for 1972–1992,
### Table 37.4 Correlations of trend deviations in US social insurance, 1980–2004

<table>
<thead>
<tr>
<th></th>
<th>SDCB</th>
<th>SWCB</th>
<th>SOACB</th>
<th>SPSB</th>
<th>SSB</th>
<th>SFCB</th>
<th>SUB</th>
<th>SHB</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDCB</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWCB</td>
<td>0.9336(^a)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOACB</td>
<td>0.3033</td>
<td>0.0989</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPSB</td>
<td>0.3115</td>
<td>0.3846</td>
<td>0.3477 (0.1128)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSB</td>
<td>0.7867(^a)</td>
<td>0.6613(^a)</td>
<td>0.4395(^a) (0.0407)</td>
<td>−0.1444</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFCB</td>
<td>0.7769(^a)</td>
<td>0.7101(^a)</td>
<td>0.2728 (0.2193)</td>
<td>−0.0626</td>
<td>0.8888(^a)</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUB</td>
<td>0.6085(^a)</td>
<td>0.6293(^a)</td>
<td>0.4030 (0.0629)</td>
<td>0.5833(^a)</td>
<td>0.3508</td>
<td>0.2444</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>SHB</td>
<td>0.9443(^a)</td>
<td>0.8588(^a)</td>
<td>0.3194 (0.1473)</td>
<td>0.1141</td>
<td>0.8884(^a)</td>
<td>0.8863(^a)</td>
<td>0.4690(^a)</td>
<td>1.000 (\times)</td>
</tr>
</tbody>
</table>

SDCB: disability cash benefits; SWCB: worker’s compensation cash benefits; SOACB: old age cash benefits; SPSB: paid sick leave benefits; SSB: survivors’ benefits total; SFCB: family cash benefits; SUB: unemployment benefits; SHB: health benefits

\(^a\)Coefficient significant at the 5 and 1% level, respectively

---

one positive and one negative correlation (out of six) were found [Zweifel (2000) where the admissibility of a microeconomic interpretation of the aggregate data is also discussed]. Still, one might argue that risk diversification by PI could be better, calling for a shift in favor of SI. According to the philosophy of social insurance, a beneficiary should always be able to count on a minimum level of consumption (which corresponds to a minimum of total assets). On these grounds, a better record in terms of variance reduction might be expected for SI than for PI. Somewhat surprisingly, however, the evidence does not point to a superior performance of SI (see Table 37.4). Unexpected variations in benefits correlate positively in no less than 15 out of 28 cases, and there is not a single negative correlation of statistical significance. An almost perfect positive correlation is noted for workers’ compensation (SWCB) and disability benefits (SDCB) [cell (2,1)]. One may be tempted to argue that these two types of benefit are triggered by a common impulse. However, this is an argument explaining positive correlation between expected benefits, not unexpected deviations. Again, the earlier study based on 1972–1992 data found six positive correlations (but also three negative ones) out of 21 possible cases. Therefore, SI in the USA seems to expose rather vulnerable individuals to excessive asset variance.

A final consideration is that risk diversification could be also achieved by private insurance filling unexpected gaps left by social insurance and vice versa. However, when the four private insurance lines distinguished are juxtaposed with the seven lines of social insurance (not shown here), this expectation fails to be confirmed. Out of 28 correlation coefficients, none is significantly negative. On the whole, then, PI and SI fail to complement each other in a way that would contribute to a maximum reduction of consumers’ total asset variance.

### 37.6 Conclusions and Outlook

This contribution takes as its starting point the well-known potential failures of private insurance (PI) markets, moral hazard, and adverse selection, which might justify government intervention, possibly in the guise of social insurance (SI). Since moral hazard effects are found to beset SI at least as much as PI, adverse selection constitutes the crucial market failure. Indeed, some degree of mandatory SI has the potential to improve the welfare of both high and low risks; the benefit for...
low risks is that private insurers can provide them with more complementary coverage without having high risks infiltrate the low-premium contract designed for them.

However, this analysis was found to be insufficient on three counts. First, the policy issue invariably is the inability of high risks rather than low risks to obtain sufficient PI coverage. This problem can be addressed by arguing that SI homogenizes the conditional loss distribution confronting private insurers offering complementary coverage, serving to reduce the especially large loading charged to high risks. Second, the optimal amount of SI and hence the optimal division of labor between PI and SI is left undetermined. Since the first layer (provided by SI) and the second layer (provided by PI) together usually do not cover the loss fully (as in health insurance and unemployment insurance), there is a net loss remaining. However, this net loss depends on the gross loss which in turn varies with the combined degree of coverage. Therefore, there is an overall moral hazard effect which must be balanced against the benefit of risk pooling afforded by additional SI coverage. The third criticism is that while this optimization (presumably by a benevolent government) may determine the division of labor between PI and SI at a given point of time it fails to explain the historical expansion of SI to the detriment of PI. For a theoretical explanation of this development, one has to take recourse to the interests of the voting public. Indeed, if one is willing to depart from the standard expected-utility framework, it becomes possible to predict that an expansion of SI will always find a majority.

In the literature, this change in the division of labor has become known as the crowding-out phenomenon. The empirical evidence almost exclusively comes from the USA. It focuses on the rapid expansion of Medicaid (for the poor), which is found to depress private saving (this is of interest because in the context of savings deposits, the problems of moral hazard and adverse selection are much less prevalent than in PI). More recently, crowding-out effects have been identified as a cause for the sluggish development of private LTC insurance.

In view of the evidence on crowding out, there has been a renewed interest in normative prescriptions that take into account moral hazard (on the part of both PI and SI), adverse selection (on the part of PI), and crowding out caused by PI. Indeed, it turns out that moral hazard can work in favor or against SI, adverse selection remains as the one effect motivating an expansion of SI, while crowding out suggests a reduced amount of SI. Finally, another approach is to view the benefits of SI and PI as assets with stochastic returns in the portfolio of an individual. Deviations from expected value should be negatively correlated both within the lines of PI and of SI—and between PI and SI. Aggregate data for the USA point to positive correlations in deviations in SI and hence a potential for an improved hedging of risks confronting citizens, possibly by PI.

Even with these insights in hand, it is difficult to answer the two crucial questions for the future, how will the division of labor between PI and SI evolve, and how should this division be changed if at all? As to the first question, the expansion of SI seems to have come to a halt in several industrial countries (see Table 37.1 again). This may be the consequence of several exogenous developments that may challenge SI more than PI. One such change is the opening of economies not only to the international flows of goods but of labor and capital as well. Private insurers have the freedom to pursue an investment policy that can benefit from the hedging provided by international capital markets, while institutions of SI are tied to their domestic capital market (provided they dispose of reserve capital at all). In view of their complexity, the theoretical models presented in this chapter invariably neglect PI benefits that can be financed by investment income. When it comes to international movements of labor, SI schemes of rich countries cannot easily follow highly skilled emigrants, whereas PI usually grants full portability of benefits. Conversely, lower-skilled immigrants may even be attracted by the generosity of SI benefits in rich countries while PI benefits reflect risk-based premiums. Emigration and immigration thus threaten to undermine the financial equilibrium of SI but not PI.
Another challenge is demographic change. It affects SI with their pay-as-you-go finance much more directly than PI, which frequently is capital-based. In addition, both theory and (rather spotty) empirical evidence suggest that the individual decisions causing demographic change in the aggregate (length of education, marriage, number of children, age at retirement, and even longevity) are in fact influenced by SI and predominantly in ways exacerbating its financing problems (Zweifel and Eugster 2008). It is conceivable that these changes explain the reversal in the trend towards expanding SI coverage noted above.

Turning to the normative issue of what the division of labor between PI and SI should be, the theory discussed certainly provides a measure of guidance. However, the underlying hypothesis is that governments implement and adjust SI reflecting reasonably informed self-interest of the voting public. One undisputed characteristic of the implementation is that it should be long term since volatility in SI disturbs the life-cycle decisions of individuals. However, in a disquieting piece of research, van Dalen and Swank (1996) find clear empirical evidence suggesting that a “solid” government such as the Dutch boosted SI programs around (re)election times, presumably in an attempt to win pivotal votes. Therefore, even if one were to know whether (and in which lines of insurance) the division of labor between PI and SI ought to be changed, the issue of whether the proposed adjustment will be truly efficiency-enhancing remains. The balance between market failures that may beset PI and political failures that may beset SI is far from evident and it may well change over time!

Acknowledgments

Suggestions and criticisms by three anonymous referees (especially by one who was particularly helpful) are gratefully acknowledged.

References


Gouveia M (1997) Majority rule and the public provision of a private good. Publ Choice 93:221-244


Samuelson PA (1958) An exact consumption-loan model of interest with or without the contrivance of money. J Polit Econ 66:321-338

Samuelson PA (1958) An exact consumption-loan model of interest with or without the contrivance of money. J Polit Econ 66:321-338


AUTHOR QUERIES

AQ1. Citation "Einav et al. (2011)" is not given in list. Please provide.
AQ2. Please check if edit to sentence starting "For simplicity, assume..." is okay
AQ3. Please provide closing parenthesis in sentence starting "It does not...”.
AQ4. Citation "OECD (2007)" is not given in list. Please provide.
AQ5. Both "NetLoss" and "Netloss" are used. Please check.
AQ6. Citation "de Donder and Hindriks" has been changed to "de Donder and Hindricks" to match with the list. Please check.
AQ7. Citation "Gruber (1997)" is not given in list. Please provide.
AQ8. Please check if the inserted citation for Table 37.2 here is appropriate.
AQ9. Text given in title has been moved as acknowledgement. Please check.
AQ10. Please check and confirm if the added year for reference "Cutler and Reber (1998)" is appropriate.
AQ11. Please check and confirm if the added year for reference "Dahlby (1981)" is appropriate.
AQ12. References "Einav et al. (2010b), Gruber (1987)" is not cited in text. Please provide citation or delete from list.
AQ13. Please update reference "Eugster et al. (2012)".
AQ14. Please check if the edit made to reference "Feldstein (1995)" is appropriate.
AQ15. Please check and confirm if the added year for reference "Le Corre (2012) is appropriate".