On the existence of private unemployment insurance with advance information on future job losses

Denderski, P.; Stoltenberg, C.A.

DOI
10.1016/j.jpubeco.2023.104946

Publication date
2023

Document Version
Final published version

Published in
Journal of Public Economics

License
CC BY

Link to publication

Citation for published version (APA):
Denderski, P., & Stoltenberg, C. A. (2023). On the existence of private unemployment insurance with advance information on future job losses. *Journal of Public Economics*, *224*, Article 104946. https://doi.org/10.1016/j.jpubeco.2023.104946

Journal of Public Economics 224 (2023) 104946

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

On the existence of private unemployment insurance with advance information on future job losses ☆

Piotr Denderski [a,b,*], Christian A. Stoltenberg [c]

[a] Institute of Economics, Polish Academy of Sciences, Nowy Świat 72, 00-330 Warsaw, Poland
[b] University of Leicester School of Business, Department of Economics, Finance and Accounting, Brookfield, London Road, LE2 1RQ Leicester, United Kingdom
[c] University of Amsterdam and Tinbergen Institute, Postbus 15867, 1001 NJ Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 31 August 2022
Revised 30 June 2023
Accepted 3 July 2023
Available online 12 July 2023

Keywords:
Advance information
Subjective expectations
Adverse selection
Unemployment insurance
Repeated interactions
Dynamic contracts

ABSTRACT

We study the existence of a profitable unemployment insurance market in a dynamic economy with adverse selection rooted in workers' advance information on future job losses. The new feature of the model is that the insurer and workers interact repeatedly. Repeated interactions make it possible to threaten workers with exclusion from future insurance benefits after a default on insurance premia. With exclusion, the insurer can offer not only insurance against unemployment risk itself but also against bad news about a future job loss. We discipline our model with estimates of the willingness to pay for unemployment insurance and the costs of adverse selection in the US. Our quantitative results illustrate that private unemployment insurance could be profitable for an exclusion length of one year. To stimulate the emergence of a private unemployment insurance market, policymakers can facilitate the creation of a registry that archives past defaults on insurance premia.

1. Introduction

Unemployment is one of the most severe risks that many people face. Even with unemployment benefits provided by the government, it is well documented that becoming unemployed forces people to make substantial and undesirable cuts in their consumption (e.g., Gruber, 1997). Furthermore, Braxton et al. (2020) provide evidence that the unemployed resort to costly self-insurance by defaulting on their debt with the consequence of long-lasting detrimental effects on their credit score and financial liberty. In light of this, why isn't there a thriving private unemployment insurance market in the US?.

The profitability of insurance in existing markets is plagued by two types of private information frictions: moral hazard and adverse selection. To study the existence of a profitable unemployment insurance market, Hendren (2017) extends the static binary loss adverse selection model originally proposed by Rothschild and Stiglitz (1976) with moral hazard. He demonstrates that only adverse selection, and not moral hazard is relevant for market existence, echoing the classical result of Shavell (1979).[1] Eventually, Hendren (2017) argues quantitatively that the reason for the missing market is that unemployment insurance policies would be too adversely selected to be profitable.

In the Rothschild and Stiglitz model, risk-averse agents have private advance information on their idiosyncratic endowment risk, which leads to an adverse selection problem. Conditional on this information, but before the endowment shock realizes, they decide whether to accept an insurance contract offered by a risk-neutral insurer or to merely consume their endowments in autarky. This decision occurs at a single instance, which is why we refer to it as a single-interaction model. This parsimonious framework abstracts, by construction, from long-run interactions between agents and the insurer. We show that allowing for long-run interactions matters for the existence of a profitable unemployment insurance market and offers new insights on overcoming the negative effects of adverse selection for market existence.

☆ We are thankful to our editor, Nathaniel Hendren, and two anonymous referees for valuable suggestions. We would also like to thank Aris Boukouras, Kyle Herkenhoff, Kristóf Madarász, Marcelo Pedroni, and the audiences at the European Economic Association Annual Conference and the seminar at the University of Leicester for helpful comments and discussions.

* Corresponding author at: University of Leicester School of Business, Department of Economics, Finance and Accounting, Brookfield, London Road, LE2 1RQ Leicester, United Kingdom.

E-mail addresses: piotr.denderski@le.ac.uk (P. Denderski), c.a.stoltenberg@uva.nl (C.A. Stoltenberg).

[1] While moral hazard matters for the profitability of insurance in *existing* markets, Hendren (2017) shows that it does not matter at the margin between existence and non-existence.

https://doi.org/10.1016/j.jpubeco.2023.104946
0047-2727/

To study the profitability of private unemployment insurance, we develop a dynamic version of the binary loss model. Private advance information arrives every period and is modeled as a private signal that informs agents about their subsequent endowment shock. After the signal realizes, agents decide in each period whether to sustain the insurance contract on offer. With repeated interactions – and this is the crucial difference to the single-interaction case – the insurer has the possibility to threaten agents with exclusion from future insurance in case they decide to reject the contract. Exclusion fundamentally alters agents' incentives to participate in the insurance market. Not only the current but also the future benefits of insurance matter when agents decide whether to accept an insurance contract offered to them. As a consequence, their willingness to pay for insurance increases.

The benefits of insurance in the dynamic economy are larger than in the single-interaction economy because they include additional insurance possibilities. In a dynamic economy, agents can insure not only the fundamental endowment shocks but also the future private signal realizations, that is, "bad news" received in the future. Insurance against bad news cannot be provided in the single-interaction environment because today's news has already been realized. Consequently, the threat of exclusion from future insurance is a powerful one in the dynamic context. It can even prompt agents who don't face any unemployment risk in the current period to purchase insurance. The inclusion of these agents in insurance is vital for profitability because they constitute good risks today. In the single-interaction model, these agents are unwilling to pay for insurance and drop out of the insurance scheme.

As our main theoretical result, we provide a no-trade condition for the dynamic economy with the length of exclusion as the key parameter, which generalizes the earlier contributions. The possibility of exclusion is what distinguishes our dynamic framework from previous work. Without exclusion, the no-trade condition reduces to the one provided in the earlier literature. Intuitively, the longer the agents are excluded from future insurance, the higher their willingness to pay for insurance is today. This makes the provision of insurance more profitable to the insurer and is reflected in a more restrictive no-trade condition in the dynamic economy. On the other hand, the lower the likelihood of the good signal agents receiving bad news in the future, the tougher it is to offer insurance at a profit. In the extreme case of them never receiving bad news, the no-trade condition again reduces to the one derived for the single-interaction model.

We employ our model to study the issue of the missing unemployment insurance (UI) market in the US. While the missing market is a universal phenomenon, our focus on this country is motivated by the fact that the importance of private advance information on future job losses is well documented for the US; Stephens (2004) finds that individual subjective job-loss expectations carry predictive power for subsequent job losses, even when public information available to an econometrician is taken into account. Through the lens of our model, we revisit the issue of the missing private unemployment insurance market in the US and ask: do repeated interactions and market exclusion matter for its existence?.

To illustrate their quantitative importance, we inform our theoretical model with existing estimates of individuals' willingness to pay for UI and the costs of adverse selection stemming from the Panel Study of Economic Dynamics (PSID) and the Health and Retirement Survey (HRS). The willingness to pay estimates do include self-insurance via savings and the existing public unemployment insurance as the status quo. These estimates imply markups to pay for UI between 30 and 90 percent. Thus, the question is whether it is profitable to provide private insurance in addition to the already existing insurance possibilities against unemployment. Without exclusion from insurance in the future, the cost of adverse selection far exceeds agents' willingness to pay, and private unemployment insurance is not profitable. This changes when we allow for exclusion. For the UI markup of 60 percent, the model implies that if agents are excluded for approximately one year after a default, unemployment insurance can be provided at a profit.

The exclusion length for profitable unemployment insurance is well below the exclusion length in the US unsecured credit market. After filing for private bankruptcy, individuals face a loss in financial liberty of up to 10 years, according to Chapter 7. Compared to this exclusion length, the exclusion of one year to render the UI market profitable does not appear long.[2] A relatively short exclusion length suffices to render the unemployment insurance market profitable because the benefits of future insurance are quantitatively important. Thereby, the future benefits of insurance stem predominantly from insurance against bad news that was not accounted for in the previous literature.

The quantitative importance of exclusion for the profitability of private UI is a robust finding. Its importance prevails even when we allow for permanent unobserved heterogeneity. Indeed, our exclusion thresholds are virtually unchanged when we use estimates of the degree to which unemployment risk is clustered on the individual level from Morchio (2020). While the one-year ahead job-loss indicators in the HRS are also correlated with realized job losses two years and beyond, the implied signal persistency is relatively modest, hardly affecting the necessary exclusion lengths for a profitable UI provision. Further, it does not depend on whether we consider an economy with infinitely or a life-cycle economy with finitely lived agents. A life-cycle version of the model is relevant because the threat of exclusion is of no concern to agents who are retiring. We find that a similar length of exclusion is enough for a profitable insurance provision in this case. The insurer can exploit that agents' willingness to pay for insurance differs with age. The insurer provides less attractive UI to the young cohorts – who value insurance the most – and extracts more resources from them to subsidize the better insurance offered to individuals nearing retirement, which prevents the contract from unraveling.

Our quantitative results suggest that adverse selection rooting in individual foreknowledge of a future job loss, while still relevant, is unlikely to be the sole cause of the missing unemployment insurance market. Arguably, there isn't a thriving private UI market in the US but our results have clear policy implications for how to kick-start the market. To make the exclusion threat credible, it is helpful to have a registry system that collects information on defaults on UI policies, similar to the successful registry with information on credit defaults. Creating and maintaining such a registry entails fixed costs, and economic policy can contribute to overcoming them. One difference to the credit market is that we show that exclusion is not only important for the well-functioning of the UI market, but actually indispensable for its existence. Thus, the launch of the registry system must be pre-announced and in place when the first UI policies are sold.

Related literature In a related paper, Braxton et al. (2020) study optimal public UI in a dynamic economy with temporary exclusion from the credit market of defaulting agents due to search frictions. Our contribution is to highlight the importance of contractual

[2] The reason why exclusion matters is different in the two markets. In credit markets, agents that were bad risks are punished with exclusion (Bond and Krishnamurthy, 2004). In the case of unemployment insurance, however, the threat of exclusion keeps agents that are good risks on board. The incentive problem is similar to the one in the market for long-term care where Finkelstein et al. (2005) argue that individuals that learn to be good risks have an incentive to select out of the insurance contract.

exclusion for the existence of a profitable UI market hampered by adverse selection stemming from private information on future job losses. Our paper is closely related to Hendren (2017), who also studies the conditions for profitable unemployment insurance. We generalize his theoretical results in a dynamic economy to accentuate the role of market exclusion and insurance against bad news for the existence of a profitable UI market.

We also contribute to the literature on the optimal design of unemployment insurance over the life cycle. While the previous literature mainly focuses on moral hazard as the private information friction, we offer new insights into the design of optimal insurance because we investigate the role of adverse selection. Michelacci and Ruffo (2015) find that younger and not older workers should receive more generous UI because the implicit costs of moral hazard are mitigated by long-term career concerns in their case. We find that unemployment insurance for older workers should be subsidized because exclusion from future insurance is less relevant for them.

Our theoretical approach shares similarities with the literature on the welfare effects of advance information in efficient risk sharing. Hirshleifer (1971) and Schlee (2001) show that advance information can make risk-averse agents ex-ante worse off if such information leads to the evaporation of risks that otherwise could have been shared in a competitive equilibrium with full insurance. Allowing also for the insurance of news as well as fundamental risk, Denderski and Stoltenberg (2020) analyze the social value of better public information when agents also have private advance information about future income shocks. None of these papers study the role of advance information and exclusion for the existence of a profitable UI market.

In Section 2, we present the model. In Section 3, we provide and discuss our theoretical results on the existence of insurance. Section 4 contains a quantitative application of theory to unemployment insurance and the last section concludes.

2. Environment

Time is infinite and there is a unit mass of agents who, when employed, earn income y.[3] Each period, agents receive a private signal n about their probability to be unemployed and suffer an income loss $l, 0 < l < y$, with their income dropping to $y - l$. Throughout, we use *unemployment* and *income loss* interchangeably. The shock to their endowment realizes before agents consume. The signal is i.i.d. across agents and over time with two realizations, *good* or *bad news*, $n \in \{g, b\}$. When $n = b$, agents become unemployed with probability one. For $n = g$, the probability of an income loss is $0 < p < 1$. The probability to receive bad news is $0 < \mu < 1$.[4]

To facilitate comparison with earlier work, we consider contracts that prescribe consumption that solely depends on the current realizations of the signal and the endowment shock. A consumption allocation prescribed by such a contract is denoted by $c = \{c_{gy}, c_{gl}, c_{bl}\}$, with c_{gy} as consumption in case of no loss, and c_{gl} (c_{bl}) as consumption in the event of unemployment after receiving a good (bad) private signal. Thus, the insurance contracts and the corresponding consumption allocations are memoryless.

Agents discount future utilities with $0 < \beta < 1$ and have identical expected-utility preferences over consumption streams. The instantaneous utility function $u : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is strictly increasing, strictly concave and satisfies the Inada conditions. In particular, we define $\tilde{w}(c)$ to be the lifetime expected utility implied by a particular allocation before any risk has been resolved:

$$\begin{aligned}\tilde{w}(c) &\equiv (1-\beta)\mathbb{E}\sum_{t=0}^{\infty}\beta^t u(c_t)\\ &= \mu u(c_{bl}) + (1-\mu)\left[pu(c_{gl}) + (1-p)u(c_{gy})\right]. \qquad (1)\end{aligned}$$

Each period after the realization of the private signal, but before the endowment shock occurs, agents have the option to default to autarky. In case of default, agents consume their endowments and – which is the new element here – are excluded from insurance for $N \geqslant 0$ periods when they default. The individual rationality constraint of good-signal agents can be compactly written as:

$$\begin{aligned}&(1-\beta)\left[pu(c_{gl}) + (1-p)u(c_{gy})\right] + \beta\tilde{w}(c) \geqslant\\ &(1-\beta)[pu(y-l) + (1-p)u(y)]\\ &+ (1-\beta)\left[\beta(1+\beta+\ldots\beta^{N-1})U^{Aut} + \beta^{N+1}(1+\beta+\ldots)\tilde{w}(c)\right],\end{aligned}$$

or:

$$\begin{aligned}&(1-\beta)\left[pu(c_{gl}) + (1-p)u(c_{gy})\right] + \beta(1-\beta^N)\tilde{w}(c) \geqslant\\ &\quad(1-\beta)[pu(y-l) + (1-p)u(y)] + \beta(1-\beta^N)U^{Aut},\end{aligned}$$

with the value of autarky given by

$$U^{Aut} = \mu u(y-l) + (1-\mu)[pu(y-l) + (1-p)u(y)]. \qquad (2)$$

With memoryless contracts and i.i.d. private signals, the continuation value $\tilde{w}(c)$ depends only on the future, but not on the current realization of signal and endowment.[5] Autarky captures the status quo in which individuals rely on their private savings and credit limits as well as on public unemployment insurance to smooth their consumption in case of unemployment.[6]

For the following, it is convenient to express allocations in terms of utility. Let $C : R \to R^+$ be the inverse of the utility function u. With u strictly increasing and strictly concave, C is strictly increasing and strictly convex. A memoryless utility allocation is then denoted by $h = \{u(c_{gy}), u(c_{gl}), u(c_{bl})\} = \{h_{gy}, h_{gl}, h_{bl}\}$ with the corresponding memoryless consumption allocation as $c = \{C(h_{gy}), C(h_{gl}), C(h_{bl})\}$. Equivalently, we will write $\tilde{w}(h)$ instead of $\tilde{w}(c)$. The set of implementable (or constrained feasible) allocations is defined as follows.

Definition 1 (*Implementable allocations*). An allocation $h = \{h_{gy}, h_{gl}, h_{bl}\}$ is *implementable* if the following statements hold for all periods $t \geqslant 0$.

1. h is resource feasible

$$\begin{aligned}&\mu C(h_{bl}) + (1-\mu)\left[pC(h_{gl}) + (1-p)C(h_{gy})\right] \leqslant \mu(y-l)\\ &+ (1-\mu)[p(y-l) + (1-p)y], \qquad (3)\end{aligned}$$

2. h is incentive compatible

$$(1-\beta)h_{bl} + \beta\tilde{w}(h) \geqslant (1-\beta)h_{gl} + \beta\tilde{w}(h), \qquad (4)$$

$$\begin{aligned}&(1-\beta)\left[ph_{gl} + (1-p)h_{gy}\right] + \beta\tilde{w}(h)\\ &\geqslant (1-\beta)[ph_b + (1-p)u(0)] + \beta\tilde{w}(h), \qquad (5)\end{aligned}$$

3. h is individually rational

[3] Opting for an infinite horizon allows us to analytically characterize the necessary and sufficient condition for the absence of trade. In Section 4, we also study a life-cycle economy.

[4] We relax the i.i.d. assumption in Section 3.2.1.

[5] With contracts that are contingent on the history of endowment and signal realizations, this is not the case. We cover the relevance of history-dependent contracts for the existence of profitable insurance in Section 3.2.2.

[6] Following the logic of Chetty (2006), average consumption levels in employment and unemployment together with their corresponding marginal utilities are a sufficient statistic for agents' willingness to pay for insurance, irrespective of the factors shaping the consumption-savings choice. We further discuss this in Section 3.2.4.

$$(1-\beta)\left[ph_{gl}+(1-p)h_{gy}\right]+\beta\left(1-\beta^{N}\right)\tilde{w}(h) \geqslant (1-\beta)[pu(y-l)+(1-p)u(y)+\beta\left(1-\beta^{N}\right)U^{Aut}, \quad (6)$$

$$(1-\beta)h_{bl}+\beta(1-\beta^{N})\tilde{w}(h) \geqslant (1-\beta)u(y-l)+\beta(1-\beta^{N})U^{Aut}. \quad (7)$$

Note that the two incentive compatibility constraints simplify to:

$$h_{bl} \geqslant h_{gl}, \quad (8)$$

$$ph_{gl}+(1-p)h_{gy} \geqslant ph_{bl}+(1-p)u(0), \quad (9)$$

which resemble the corresponding conditions in Hendren (2013). The two individual rationality constraints (6)–(7) are different than in his paper when there is exclusion, that is, $N > 0$.[7]

Per construction, the no-trade allocation, $\{u(y), u(y-l), u(y-l)\}$, is implementable. As Hendren (2013), we focus on implementable allocations to study the existence of a profitable insurance market. The question is whether there exists an alternative implementable allocation to autarky that is cost-efficient. More formally, such an allocation is defined as follows.

Definition 2 (*Cost-efficient allocation*). A cost-efficient allocation h^* is implementable and maximizes the slack on the resource constraint:

$$h^* = \arg\max_{h_{bl},h_{gl},h_{gy}} \Pi = [(1-\mu)p+\mu](y-l)+(1-\mu)(1-p)y - \mu C(h_{bl}) - (1-\mu)\left[pC\left(h_{gl}\right)+(1-p)C\left(h_{gy}\right)\right]. \quad (10)$$

A cost-efficient allocation is an allocation that a profit-maximizing monopolist insurer chooses subject to individual rationality and incentive constraints.[8]

3. Analysis

In this section, we derive our main theoretical result: a no-trade condition with repeated interactions and exclusion.

3.1. Existence of insurance

As an intermediate step leading to the main theoretical result, we characterize cost-efficient allocations with trade. We show that if such allocations exist, they feature insurance, which creates slack on the resource constraint.

Lemma 1 (*Cost-efficient allocation with trade*). Let $h^* = \left\{h^*_{gl}, h^*_{bl}, h^*_{gy}\right\}$ be a cost-efficient allocation with trade. Then, the following statements hold.

(i) Incentive constraints of bad-signal agents hold with equality so that the utility of agents who incur an income loss is equalized across signal realizations, $h^*_{bl} = h^*_{gl} = h^*_l$, and the incentive constraints of good-signal agents are slack.

(ii) Individual rationality constraints of good-signal agents hold with equality; the individual rationality constraints of bad-signal agents are slack.

(iii) h^* is characterized by $u(y-l) < h^*_l \leqslant h^*_{gy} < u(y)$.

The proof is provided in Appendix A. The logic of the proof can be summarized as follows. For part (i), agents with good signals have a better outside option than the bad-signal agents which is reflected in the cost-efficient allocation due to individual rationality constraints, and their expected utility exceeds the one of the bad-signal agents. Thus, only agents with a bad signal have the incentive to report a good signal realization but not vice versa.

For part (ii), the convexity of resource costs implies that it is optimal to equalize the consumption of all agents who suffer an income loss. Excluding the bad-signal agents from the contract completely and only transferring resources within the pool of good signal agents is prevented by private information, as captured in the incentive constraints (4)–(5). Perfect insurance, $h_{gy} = h_l$, minimizes the resource costs and satisfies incentive constraints. However, perfect insurance may violate the individual rationality constraints of good-signal agents because these agents must agree to transfer resources not only to the job losers with a good signal but also to the ones with a bad signal. The latter benefit from the insurance contracts directly by receiving transfers from the employed good-signal agents but also indirectly because the continuation value with insurance is higher than without trade. These two benefits together render the individual rationality constraints of bad-signal agents slack. Lemma 1, therefore, implies that the problem of finding a cost-efficient allocation simplifies to:

$$\max_{h_{gy},h_l}[(1-\mu)p+\mu](y-l)+(1-\mu)(1-p)y - [\mu+(1-\mu)p]C(h_l)-(1-\mu)(1-p)C\left(h_{gy}\right), \quad (11)$$

subject to individual rationality constraints with that of the good-signal agents holding with equality. As autarky is implementable and yields zero profits, the insurer has no incentive to choose an allocation that requires more resources, implying that an optimal allocation satisfies resource feasibility, though not necessarily with strict equality.

We proceed to our main theoretical result and show under which conditions autarky is the only implementable allocation and, therefore, also the cost-efficient allocation.

Theorem 1 (*No Trade*). The autarky allocation $h = \{u(y), u(y-l), u(y-l)\}$ is the only implementable allocation if and only if

$$\frac{u\prime(y-l)}{u\prime(y)} \leqslant \underbrace{T_s(p,\mu)T_r(p,\mu,\beta,N)}_{T_d(p,\mu,\beta,N)}, \quad (12)$$

with $T_s(p,\mu)$ as the single-interaction pooled price ratio:

$$T_s(p,\mu) = \frac{\mu+(1-\mu)p}{(1-\mu)p} = \frac{\mathbb{E}[P|P \geqslant p]}{1-\mathbb{E}[P|P \geqslant p]}\,\frac{1-p}{p},$$

$T_r(p,\mu,\beta,N)$ as the diminishing factor resulting from repeated interactions:

$$T_r(p,\mu,\beta,N) = \frac{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p}{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p+\beta\left(1-\beta^{N}\right)\mu}, 0 < T_r \leqslant 1, \frac{\partial T_r}{\partial N} < 0,$$

$\mathbb{E}[P|P \geqslant p]/(1-\mathbb{E}[P|P \geqslant p])$ as the relative unemployment probability:

[7] To ease the exposition, we assume that when an agent with a good signal lies about the signal, their consumption is zero in case of employment, $u(0)$. This choice minimizes the insurer's resource costs and results in non-binding incentive constraints of agents with a good signal. However, even an infinitely small garnishment of income in case of good-signal agents lying about their signal, who nevertheless do not lose a job, has identical implications.

[8] The question of insurance market existence boils down to whether a monopolist insurer can incur profits or not. Focusing on monopolistic insurance provision circumvents the issue of potentially non-existent competitive Nash equilibria when insurers can poach insurees from their competitors. Other papers studying monopolistic provision of insurance include Stiglitz (1977) and Chade and Schlee (2012).

$$\frac{\mathbb{E}[P|P \geqslant p]}{1-\mathbb{E}[P|P \geqslant p]} = \frac{(1-\mu)p+\mu}{1-(1-\mu)p-\mu}$$

and $0 < T_d(p,\mu,\beta,N) \leqslant T_s(p,\mu)$.

The proof is provided in Appendix B.[9] The no-trade condition (12) generalizes earlier findings by allowing for repeated interactions in the insurance market. For $N = 0$, the no-trade condition can be rearranged, yielding:

$$\frac{p}{1-p} \times \quad \frac{u'(y-l)}{u'(y)} \leqslant \frac{\mathbb{E}[P|P \geqslant p]}{1-\mathbb{E}[P|P \geqslant p]}, \tag{13}$$

which is the condition provided in Hendren (2017, 2013) for a single interaction. The ratio of marginal utilities $u'(y-l)/u'(y)$ is the marginal rate of substitution of consumption absent any private unemployment insurance transfers, which captures the good-signal agents' willingness to pay for insurance. The ratio $p/(1-p)$ is the insurer's actuarially fair cost of transferring resources between the good signal agents without and with an income loss. Finally, on the right side of (13), we have the true cost of such a transfer, one that accounts for the additional resources going to bad signal agents. With adverse selection due to private information, the true costs exceed the actuarial fair costs such that:

$$\frac{\mathbb{E}[P|P \geqslant p]}{1-\mathbb{E}[P|P \geqslant p]} \frac{1-p}{p} > 1.$$

The main message from Theorem 1 is that it is easier to sustain insurance for $N > 0$. With $N > 0$, opting for autarky becomes costlier because agents also sacrifice a part of their future insurance benefits. This can be seen by re-arranging the no-trade condition as follows:

$$\underbrace{\left[\frac{1}{T_r(p,\mu,\beta,N)}\right]}_{\geqslant 1} WTP \leqslant T_s(p,\mu).$$

On the left-hand side, the willingness to pay increases due to repeated interactions because $T_r(p,\mu,\beta,N) \leqslant 1$ is decreasing in N. The tightness of the no-trade condition depends on the appeal of future insurance benefits to good-signal agents. Thereby, the increased attractiveness of the insurance contract with repeated interactions does not simply stem from extending the benefits in case of a single interaction to multiple periods but also from additional insurance possibilities.

First, insurance is more relevant for good-signal agents because it becomes more likely they will need it. For one period, the probability of a good-signal agent suffering an income loss l is p, which constitutes the risk she likes to insure away. For N future periods, the probability for an agent who always receives a good signal to suffer an income loss at least once is $1-(1-p)^{N+1}$. Thus, in the infinite limit, the probability of incurring a loss at least once is one, which increases the attractiveness of insurance for good-signal agents.

Second, with repeated interactions, there can be insurance even in the case of $p = 0$, that is, when the good signal agents are certain of remaining employed in the current period. The case of $p = 0$ is a natural one to consider in the context of unemployment insurance because there are many formal and informal institutions (e.g., employment protection legislation, no-compete clauses, and seniority of position) that temporarily eliminate the risk of unemployment for some agents. Such *safe* agents are not willing to pay for insurance in the static model because the signals have already been realized and can, therefore, no longer be insured, which results in an infinitely large single-interaction pooled price ratio T_s:

$$\lim_{p \mapsto 0} \frac{\mathbb{E}[P|P \geqslant p]}{1-\mathbb{E}[P|P \geqslant p]} \frac{1-p}{p} = \infty. \tag{14}$$

However, the dynamic pooled price ratio T_d is finite for an arbitrarily small value of p and positive exclusion length, and is given by:

$$\lim_{p \mapsto 0} T_d(p,\mu,\beta,N>0) = \frac{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]}{(1-\mu)\beta\left(1-\beta^N\right)} > 1. \tag{15}$$

Hence, for some parameter values, safe agents might be willing to accept the insurance contract in the current period. The reason for them to do so is that rejecting the contract now foregoes the value of insurance against receiving *bad news* (receiving a bad signal) in the future.

3.2. Discussion of Theorem 1

Theorem 1 summarizes how i.i.d. private signals on income losses shape the profitability of memoryless insurance contracts with a deterministic return from autarky after N periods. In this section, we discuss each of these assumptions. We turn first to the i.i.d. assumption and study persistent private signals.[10] Then, we analyze the no-trade condition when the insurance contracts are not memoryless but also depend on the past realizations of the endowments and signals instead. Next, we study how the no-trade condition is affected when return from autarky is not deterministic but random, as in the literature on sovereign default (e.g., Arellano, 2008) or search for credit (e.g., Braxton et al., 2020). Eventually, we argue that Theorem 1 also applies when households self-insure against job losses via savings.

3.2.1. Persistent signals

The effect of the exclusion threat on market existence depends on how likely the good-signal agents are to receive bad news in the future. Because of that, the analytically convenient assumption of i.i.d. signals is not innocuous. Here, we relax it and study how the persistence of private signals affects the no-trade condition. We show that increases in signal persistence reduce the ability of exclusion to render insurance provision profitable.[11]

Consider a Markov chain over private signals with transition probabilities $\rho_g = \text{Prob}(n' = g|n = g)$ and $\rho^b = \text{Prob}(n' = b|n = b)$, for $\rho^g, \rho^b \in [0,1]$. This specification nests i.i.d. signals as a special case: $\rho^b = 1-\rho^g = \mu$. Unless both signals are perfectly persistent (in which case μ has to be specified independently), the invariant distribution of this Markov chain is unique. In that case, the measure of agents who receive a bad signal depends on signal transition probabilities:

$$\mu(\rho^b,\rho^g) = (1-\rho^g)/(2-\rho^b-\rho^g). \tag{16}$$

With memoryless insurance contracts, signal persistency does not alter the incentive constraints but it affects the individual rationality constraints.[12] For example, the individual rationality constraint of agents with a good signal now reads:

$$\begin{aligned}&(1-\beta)\left[ph_{gl}+(1-p)h_{gy}\right]+\beta\left(1-\beta^N\right)\left[(1-\rho^g)\,\tilde{w}(b)+\rho^g\,\tilde{w}(g)\right]\\ &\geqslant (1-\beta)[pu(y-l)+(1-p)u(y)]+\beta\left(1-\beta^N\right)\left[(1-\rho^g)U^{\text{Aut}}(b)+\rho^g U^{\text{Aut}}(g)\right],\end{aligned} \tag{17}$$

[9] This is also the threshold for an implementable allocation chosen by a benevolent social planner to improve welfare relative to autarky.

[10] For completeness, we consider public instead of private signals in Appendix C.

[11] In Appendix D, we provide a detailed description of the model with persistent signals.

[12] In Appendix D, we argue that the results in Lemma 1 also apply to persistent signals. In particular, the same pattern of binding constraints applies, resulting in $h_{gl} = h_{bl} = h_l$.

where $\tilde{w}(g), \tilde{w}(b), U^{\text{Aut}}(g), U^{\text{Aut}}(b)$ are the lifetime utilities of agents with good and bad signals with insurance and in autarky, respectively.

To see how persistency of signals affects the tightness of the no-trade condition, consider the extreme case of perfectly persistent signals, $\rho^g = \rho^b = 1$. Then, in particular, the individual rationality constraint of good signal agents (17) does not depend on the exclusion length and simplifies to:

$$ph_{gl} + (1-p)h_{gy} \geqslant pu(y-l) + (1-p)u(y).$$

For perfectly persistent signals, the no-trade condition is therefore – ceteris paribus – identical to the one in a single-interaction model.

Arbitrary changes in ρ^b and ρ^g do not merely capture changes in persistency. According to (16), they also affect the fraction of bad-signal agents in the invariant distribution, and with it, the total amount of resources in the economy, $[(1-\mu)p+\mu](y-l)+(1-\mu)(1-p)y$. Thus, to isolate the effect of signal persistence, we will vary signal transition probabilities in a way that keeps the resources constant, as captured in the following definition.

Definition 3 (*Resource-neutral combinations of signal persistency*). Given p and income endowments y and $y-l$, pairs of (ρ^g, ρ^b) are resource-neutral if they yield the same μ in the invariant distribution of private signals.

Consider varying bad-signal persistence ρ^b. Solving (16) for ρ^g yields that adjusting ρ^g as follows:

$$\rho^g(\rho^b) = \frac{1-\mu(2-\rho^b)}{1-\mu} \tag{18}$$

is *resource neutral* for a given μ. In the following proposition, we generalize Theorem 1 to persistent signals and show that increases in signal persistency result in a less restrictive no-trade condition.

Proposition 1 (*No trade: persistent private signals*). Consider persistent signals.

(i) The autarky allocation $h = \{u(y), u(y-l), u(y-l)\}$ is the only implementable allocation if and only if

$$\frac{u\prime(y-l)}{u\prime(y)} \leqslant T_s(p,\mu)T_r(p,\beta,N,\rho^g,\rho^b), \tag{19}$$

with $T_s(p,\mu)$ as defined in Theorem 1, $0 < T_r(p,\beta,N,\rho^b,\rho^g) \leqslant 1$. Furthermore, it is

$$\frac{\partial T_r[p,\beta,N,\rho^g,\rho^b]}{\partial N} < 0, \text{if} \quad 0 \leqslant \rho^g, \rho^b < 1,$$

and

$$\lim_{\rho^b\to\mu, \rho^g\to 1-\mu} T_r(p,\beta,N,\rho^g,\rho^b) = T_r(p,\mu,\beta,N).$$

(ii)Consider a resource-neutral increase in ρ^b such that $T_r[p,\beta,N,\rho^g,\rho^b] = T_r[p,\beta,N,\rho^b]$. Then

$$\partial T_r[p,\beta,N,\rho^b]/\partial\rho^b > 0,$$

that is, the diminishing factor of repeated interactions is increasing in signal persistence.

The exact definition of the diminishing factor $T_r(p,\beta,N,\rho^g,\rho^b)$ and the proof are provided in Appendix D. The main message is that increases in signal persistency are detrimental to market existence. For agents with a good signal, the more persistent the good signal is, the less likely is that these agents become unemployed in the future. As a consequence, these agents value future insurance less, which reduces the profitability of the insurance contract. For less-than-perfectly-persistent signals, increasing exclusion length tightens the no-trade condition.

As an immediate corollary of Proposition 1, suppose that the no-trade condition (12) is satisfied for $N = 0$, but violated for some N with i.i.d. signals. From part (i), the limiting value of $T_r(p,\beta,N,\rho^g,\rho^b)$ for $\rho^b \to \mu$ and $\rho^g \to 1-\mu$ is $T_r(p,\mu,\beta,N)$. According to part (ii), given N there exists a resource-neutral lower bound for persistency $\underline{\rho}^b > \mu$ from which onward no trade is the only implementable allocation. In Section 4.3.2, we will compute such a lower bound for the US and compare it to estimates of signal persistency.

3.2.2. History-dependent insurance contracts

To ensure comparability with earlier work, we studied insurance contracts that prescribe consumption solely on the basis of the current signal and endowment realization. However, a risk-neutral insurer can increase profits by offering insurance contracts that track the history of signals and endowments for two reasons. First, these contracts allow smoothing consumption of agents not only across different states in the current period but also over time, which increases agents' willingness to pay for insurance and relaxes the individual rationality constraints. Second, the insurer can use both contemporaneous utilities h and continuation values $\tilde{w}$ to provide agents the incentives to truthfully report their private signal. Hence, agents can be screened to alleviate the costs of adverse selection. Compared to memoryless insurance contracts, history-dependent contracts therefore expand the parameter region in which unemployment insurance is profitable.[13]

3.2.3. Random return after default

After default, agents in our environment are allowed back into insurance for sure after being excluded for N periods. Assume alternatively that leaving autarky is possible every period, but random, and let $\theta \in [0,1]$ be the probability of this happening.[14] We elaborate more on this version of the model in Appendix E where we show that, ceteris paribus, the higher the probability to return is, the more likely it is that autarky is the only implementable allocation. We also derive the following explicit relationship between N and θ that yields an identical no-trade condition:

$$\theta = \frac{\beta^N(1-\beta)}{1-\beta^{N+1}}.$$

3.2.4. Savings

Self-insurance via savings is an important mechanism through which households shield their consumption from income fluctuations. For the purpose of assessing agents' willingness to pay for insurance, though, knowing the consumption choices in employment and unemployment is enough. Due to an envelope-condition logic similar to the one applied in Chetty (2006), the circumstances in the light of which these choices were made are irrelevant. In particular, this logic applies when the consumption choices of individuals are the result of precautionary savings. In our parsimonious model, as there are no savings, *all* agents consume y in employment and $y-l$ in unemployment. It can be shown that the no-trade condition that we have derived is robust

[13] A more extensive description of the existence of insurance with history-dependent contracts is presented in Section 1 of the Online Appendix together with accompanying numerical results.

[14] The expected time until re-entry (including the current period) is, therefore, $1/\theta$.

to savings as long as y and $y - l$ correspond to *average* consumption choices of the employed and the unemployed.[15]

4. Quantitative results

In this section, we study the quantitative implications of our theoretical results for the private unemployment insurance market in the US. Our calibrated model illustrates that relatively short exclusion lengths, shorter than those for personal bankruptcy, could be sufficient to render the private UI market profitable.

4.1. Estimates and calibrated parameters

We begin by reviewing the estimates for the willingness to pay for unemployment insurance and the pooled price ratio in the US. Afterward, we describe how we calibrate the structural parameters of the model presented in the previous section.

We target annual estimates of the willingness to pay (WTP) for UI based on observed consumption drops after a job loss in the Panel Study of Income Dynamics (PSID). Further, we inform our model with estimates of the pooled price ratio (PPR) and the mean of the subjective job loss probability distribution based on the Health and Retirement Survey (HRS), Prob(U), which are summarized in Table 1.[16]

The willingness to pay estimates – for relative risk aversion $\sigma \in \{1, 2, 3\}$ – one-for-one correspond to the ratio of marginal utilities in the model. The pooled price ratio in the model T_s is parametrized by p and μ. To identify these parameters, on top of the estimate of the pooled price ratio, we use the mean of the subjective job loss probability distribution. In the model, there are two subjective unemployment probabilities, Prob($U|n$), conditional on the signal n. Thus, the mean of the job loss probability in the model is

$$\begin{aligned} &\mathrm{E}[\mathrm{Prob}(U|n)] = \mathrm{Prob}(U|n=g)\mathrm{Prob}(n=g) \\ &+\mathrm{Prob}(U|n=b)\mathrm{Prob}(n=b) = p(1-\mu) + \mu, \end{aligned} \tag{20}$$

and we choose $\{\mu, p\}$ to solve

$$\mathrm{E}[\mathrm{Prob}(U|n)] = 0.031 \tag{21}$$

$$T_s(p, \mu) = 4.36, \tag{22}$$

which results in $p = 0.0073$ and $\mu = 0.0239$.[17] For the discount factor, we choose a standard annual value of $\beta = 0.96$, implying an annual real interest rate of four percent. The estimate of the pooled price ratio by far exceeds the willingness to pay in all three cases, clearly satisfying the no-trade condition for a single interaction (13). Correspondingly, Hendren (2017) concludes that unemployment insurance contracts in the US would be too adversely selected to be profitable.

4.2. Exclusion and the existence of profitable unemployment insurance

To assess the quantitative importance of repeated interactions for unemployment insurance, we compute the number of exclusion periods necessary for the existence of unemployment insurance in the US. In the next step, we compare this number to alternative exclusion periods observed in reality.

Table 1
Estimates and calibrated parameters: baseline.

	Estimate/Parameter	Value
WTP	Willingness to pay	$\{1.29, 1.58, 1.87\}$
PPR	Pooled price ratio, $\inf[T_s]$	4.36
Prob(U)	Mean, subjective job loss distribution	0.031
μ	Measure of bad-signal agents	0.0239
p	Job loss probability, good-signal	0.0073
β	Annual discount factor	0.960

Notes: WTP for unemployment insurance with relative risk aversion $\sigma = \{1, 2, 3\}$. Pooled price ratio as point estimate for the minimum pooled price ratio (semi-parametric), $\inf[T_s]$, evaluated at the mass point Prob(U) = 0.031.

Given p, μ, β, we compute the shortest exclusion needed for the existence of unemployment insurance, N_{min}, as follows:

$$N_{min} \equiv N : \frac{u\prime(y-l)}{u\prime(y)} - T_d(p, \mu, \beta, N) = 0. \tag{23}$$

The resulting values are reported in the first row of Table 2. Depending on risk aversion, the minimum number of exclusion periods for the existence of unemployment insurance varies between less than one year for $WTP = 1.87$ to approximately two and a half years for $WTP = 1.29$. In the second row of Table 2, we display the expected number of exclusion periods when return is random.[18]

Why does a relatively short exclusion length suffice? The probability to receive a good signal and become unemployed in the current period is relatively small with $(1-\mu)p = 0.0071$, but the probability to receive bad news and become unemployed in the future, $\mu \times 1 = 0.0239$, is over three times higher.

To see this more clearly, consider alternatively the case of *safe* good signal agents, so that $p = 0$ instead. We then set $\mu = 0.031$ to match the mean of the subjective unemployment beliefs distribution Prob(U). With $p = 0$, the single-interaction pooled price ratio becomes arbitrarily large, in line with Eq. (14), and profitable insurance cannot be provided for any finite willingness to pay estimates. In the dynamic model, however, there can be insurance, which stems – by construction — exclusively from insurance against bad news, as encapsulated by (15). To illustrate this, we compute the necessary exclusion length for $\sigma = 2$. The exclusion length mildly increases from $N_{min} = 1.18$ in the baseline to $N_{min} = 1.89$.

Suppose alternatively that p and μ are as in the baseline model, but $N = 0$, such that insurance against bad news is irrelevant. Consider the increase in p necessary to render insurance profitable, again adjusting μ to match Prob(U) = 0.031. We find that p must be almost three times larger than in the baseline. Therefore, the profitability of insurance provision in the dynamic model stems indeed mainly from the possibility to provide insurance against bad news.

Are the minimum exclusion lengths long or short? Arguably, there is no thriving private unemployment insurance market in the US, which makes it difficult to say whether the number of exclusion periods in Table 2 are realistic. To put the numbers into perspective, we compare them to two real-world analogues: the loss in financial liberty resulting from private bankruptcy and the exclusion time after sovereign default.

[15] The corresponding derivations are available in Section 2 of the Online Appendix.

[16] For the WTP, we use Hendren (2017)'s baseline estimates reported in Table 6 on Page 1808. The resulting numbers imply markups to pay for additional UI ranging from approximately 30 to about 90 percent. The baseline estimates for the PPR and Prob(U) can be found in Table 8 on Page 1814.

[17] All the results in this section are robust to approximating the distribution of subjective beliefs with more than just two points. We consider a version of the model with a distribution of beliefs that features safe, uninformed and bad signal agents in Section 3 of the Online Appendix

[18] Comparing the first and the second row, we find that the expected length of exclusion is slightly larger than in the case of a fixed-length exclusion $N_{\min}$. The reason for this is as follows. When return from autarky is random, there is a non-zero probability to return from autarky earlier than with a deterministic exclusion, making autarky more attractive to the agents. To compensate for this and to render insurance more attractive than autarky, the expected length of exclusion must be larger than in the deterministic case.

Table 2
Existence of insurance: minimum number of exclusion periods.

	WTP, $u'(c_u)/u\prime(c_e)$, for various σ		
	1.29 ($\sigma = 1$)	1.58 ($\sigma = 2$)	1.87 ($\sigma = 3$)
Fixed exclusion, N_{min}	2.6803	1.1774	0.6962
Random exclusion, $\mathbb{E}[N_{min}]$	2.8906	1.2309	0.7207

Notes: Baseline specification. Minimum number of exclusion periods for the existence of unemployment insurance, N_{min}, and expected exclusion length $\mathbb{E}[N_{min}] = 1/\theta - 1$ as functions of the willingness to pay, WTP.

In the US, individuals can file for private bankruptcy according to Chapter 7 (about 71% of filings) or Chapter 13 (29% of filings). In both cases, bankruptcy appears on the individual's credit history for a given time period with the consequence that it is either impossible to receive unsecured credit or only possible at a high interest rate premium. The loss of access to financial markets closely resembles the idea of exclusion from insurance. The private bankruptcy entry appears in the individual credit history for either ten (Chapter 7) or seven years (Chapter 13). In light of these numbers, a necessary exclusion length of up to three years does not appear to be unrealistically large.

Another possibility is to compare the exclusion numbers to the time until countries gain re-access to international financial markets after sovereign default. For example, Schmitt-Grohé and Uribe (2017) in Chapter 13 provide estimates on how long exclusion after a sovereign default lasted on average for a sample of countries in the years 1974–2014. Excluding the period of default itself, it takes on average about 9 years until countries can borrow again some amount (partial re-access) and about 15 years until they can borrow an amount exceeding 1% of their GDP (full re-access). Both estimates are well above the length of the exclusion period necessary for the existence of profitable unemployment insurance that we compute.

4.3. Robustness exercises

In this section, we explore three deviations from the baseline that have the potential to overturn the result that unemployment insurance can be profitable for a relatively short exclusion. Firstly, we allow for permanent unobserved heterogeneity in unemployment risk. Secondly, we study exclusion with persistent instead of i.i.d. signals. Thirdly, we consider an economy with a finite instead of an infinite life span of agents.

4.3.1. Permanent unobserved heterogeneity

In reality, workers differ not only with respect to their advance information on a future job loss but also in their willingness to pay for unemployment insurance due to their permanent characteristics. In particular, we introduce a new type of agent in the model, one that does not face any unemployment risk. This extension is motivated by recent findings of Morchio (2020) who estimates that approximately 60% of workers are never unemployed in the US.[19]

To account for these findings, we assume that a fraction ν of agents is *never* at risk of unemployment. Hence, their type is fixed, and their layoff risk is $\tilde{p} = 0$ forever.[20] Such agents are, therefore, *permanently* safe from unemployment. We further assume that their type is unobservable by the insurer. The remaining fraction $1 - \nu$ of agents are specified as in the baseline model, that is, μ of them have a layoff risk of one, and $1 - \mu$ a job-loss probability of p. We then investigate quantitatively whether such permanent differences in layoff risk matter for the length of exclusion necessary to support the existence of a private UI market.

The permanently safe agents do not value insurance at all, and they consume y at all times. For the same reason, they also do not have any incentive to lie about their type. However, good signal and bad signal agents might have the incentive to misreport their type as permanently safe. To prevent this, the optimal allocation prescribes consumption in case of unemployment to the permanently safe agents that does not exceed consumption in autarky in case of unemployment. While the consumption in that state does not matter for the permanently safe agents because it has a probability of zero, any consumption equal to or below $y - l$ eliminates all incentives of agents with unemployment risk to pretend to be of the permanently safe type.[21] Eventually, Lemma 1 applies to the remaining $1 - \nu$ population of workers, and the optimal allocation is constrained by the same pattern of individual rationality and incentive constraints as in the baseline case.

Accounting for the permanently safe agents, the profits of the insurer in Eq. (10) are diminished by $1 - \nu$. The resources available for trade and the share of agents in the population that seek insurance shrink by the same factor. This implies that the existence of the permanently safe agents does not affect the insurer's optimality conditions, and the no-trade condition is again given by Theorem 1.

However, the presence of the permanently safe agents has to be taken into account when calibrating the parameters μ and p. To see this, note that the left hand side of condition (21) that ensures we match the mean of the job loss probability now becomes:

$$\mathrm{E}[\mathrm{Prob}(U|n)] = (1 - \nu)[p(1 - \mu) + \mu]. \tag{24}$$

Hence, given ν, we re-calibrate p and μ to solve the moment conditions (21)–(22), with the mean job loss probability given by (24). The more agents are permanently safe, the higher p and μ must be to match it. That is, the remaining pool of agents becomes riskier.

We display the exclusion length necessary for the existence of profitable insurance as a function of ν for the three different values of the willingness to pay in Fig. 1. The main takeaway is that the exclusion length necessary for market existence hardly responds to the fraction of permanently safe agents. For $\nu = 58\%$, the number stemming from Morchio (2020), the necessary exclusion length for $\sigma = 2$ increases from 1.1774 in the baseline to 1.2195.[22] This value of ν is the largest fraction of permanently safe agents consistent with Morchio (2020)'s findings because never being unemployed in the past is not equivalent to not facing any unemployment risk.

Formally, the reason for this result is that the product on the right-hand side of Eq. (12), $T_s(p, \mu)T_r(p, \mu, \beta, N)$, hardly increases in response to the increases in p, μ. The logic is as follows. T_s is not affected by the increases in the parameters because the calibration requires the increases to yield $T_s = 4.36$. Thus, the necessary increase in N_{min} exclusively stems from the parameters' effect on T_r, which summarizes how good signal agents value future insurance benefits; the lower T_r is, the higher the value of future insurance is. The two parameters affect T_r in an opposite way: while T_r increases in p, it decreases in μ. When p increases, the good and the bad signal agents face more similar unemployment risk, which reduces the value of future insurance. However, when μ increases, receiving a bad signal and suffering a certain income loss becomes

[19] Morchio (2020) uses the data on the National Longitudinal Survey of Youth 1979 cohort. Gregory et al. (2021) conduct a similar exercise using Longitudinal Employer-Household Dynamics data.

[20] By choosing $\tilde{p} = 0$ we not only maintain the tractability of the model, but we also pose the strongest case in favor of permanent unobserved heterogeneity in unemployment risk as the root of the missing private UI market.

[21] To see this, note that in case the consumption of the permanently safe agents were equal to $y - l$ in case of unemployment, the two incentive constraints replicate the individual rationality constraints of good and bad signal agents and are therefore obsolete.

[22] This is a conservative estimate because we are assuming that the realized unemployment risk in Morchio (2020) equals the ex-ante risk in the model.

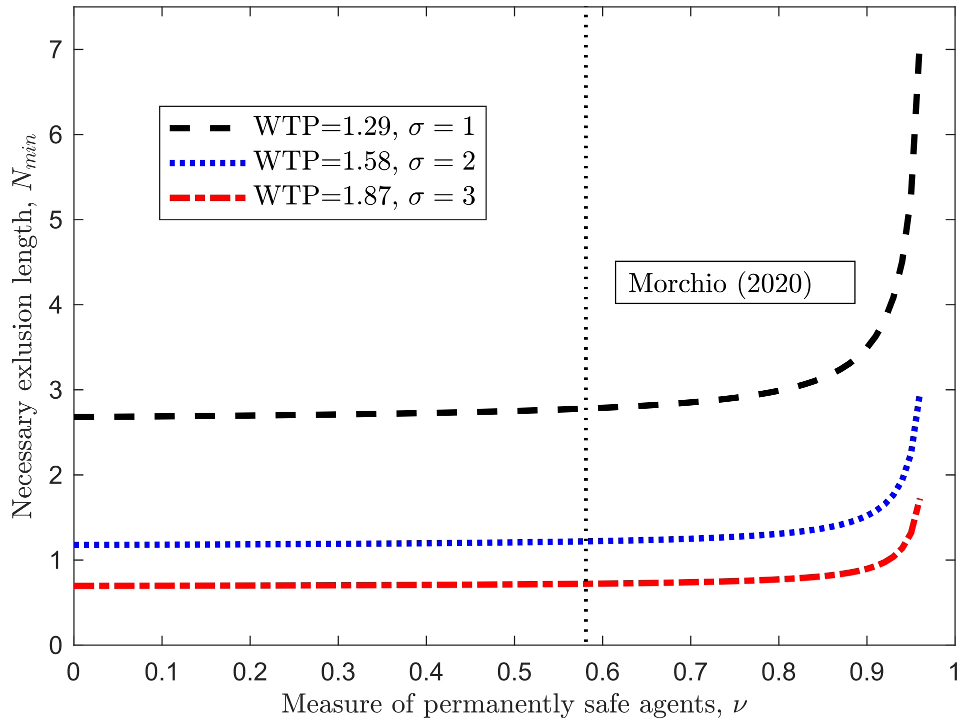

Fig. 1. Minimum number of exclusion periods for the existence of unemployment insurance, N_{min}, as function of the measure of the permanently safe agents ν for different willingnesses to pay, WTP.

more likely, which increases future insurance benefits. The former effect is only weakly stronger than the latter and the necessary exclusion length increases only mildly compared to the baseline.

Quantitatively, all three functions are essentially flat until $\nu = 0.9$. At this point, p and μ must increase strongly to match the unemployment probability of 0.031. As a consequence, the necessary exclusion length spikes upwards. For $\nu \geqslant 0.969 = 1 - 0.031$, the mean job loss probability can neither be matched for any $p < 1$ nor $\mu < 1$. Overall, the existence of a permanent type of agents without any unemployment risk does not qualify as an explanation for the missing market.[23]

4.3.2. Persistent signals

In the baseline specification, we assumed signals to be i.i.d. In Proposition 1, we show that unemployment insurance with exclusion is profitable if and only if signals are not too persistent. In this section, we quantitatively investigate this persistence threshold for the US.

As a first step, we compute the persistency threshold as follows. Given an exclusion length for which unemployment insurance is profitable with i.i.d. signals, we compute the minimum transition probability ρ^b that makes insurance unprofitable. To be more precise, given some $N = \overline{N}$, we compute the minimum threshold $\underline{\rho}^b$ that satisfies

$$\underline{\rho}^b = \min\left\{\rho^b : \frac{u\prime(y-l)}{u\prime(y)} - T_s(p,\mu)T_r\left[p,\beta,\overline{N},\rho^g(\rho^b),\rho^b\right] = 0\right\},$$

with ρ^g chosen in a resource-neutral way given $\mu = 0.0239$.[24] In doing so, we keep the unemployment probability and the pooled price ratio as listed in Table 1.

Table 3
Existence of insurance: persistency threshold with finite exclusion.

	WTP, $u\prime(c_u)/u\prime(c_e)$, for various σ		
	1.29 ($\sigma = 1$)	1.58 ($\sigma = 2$)	1.87 ($\sigma = 3$)
Persistency threshold, $\underline{\rho}^b$	0.7877	0.9327	0.9635
Prob($n\prime = b \mid n = g$)	0.0052	0.0016	0.0009
$\rho^b = 0.3192, N_{min}$	2.8602	1.2227	0.7171

Notes: Persistent signals. First and second row: minimum bad-signal transition probability $\underline{\rho}^b$ that renders unemployment insurance unprofitable given an exclusion length of $N = 5$, and the corresponding probability of good-signal agents to receive a bad signal, $1 - \rho^g$, as functions of the willingness to pay, WTP. Third row: minimum exclusion length for the estimated signal persistency $\rho^b = 0.3192$ as a function of WTP.

Table 3 displays the minimum transition probability $\underline{\rho}^b$ that renders unemployment insurance unprofitable for a given exclusion length of $N = 5$ – half the exclusion length as in case of bankruptcy according to Chapter 10 – for different degrees of relative risk aversion. The persistency thresholds vary between 0.79 and 0.96 (see Row 1). Overall, these figures suggest that unless signals are very persistent, unemployment insurance is profitable for an exclusion length of five years.[25]

A natural question is what the actual persistence in the elicited subjective job-loss probabilities in the HRS data is. In the Online Appendix, Hendren (2017) provides evidence of the persistence of the elicitations. He regresses the subjective job-loss loss probabilities in period t on the unemployment indicator in years $t + j, j = 1, \ldots 8$ (separately for each year), and finds the coefficients

[23] Another relevant dimension of permanent heterogeneity can be the differences in how forward-looking the agents are. To this end, we assume that the agents who are exposed to job loss risk have a lower discount factor. We report the results of the robustness check of decreasing β in Table 3 in the Online Appendix, Section 5.

[24] We plot the set of resource-neutral signal persistencies consistent with our calibration on Fig. 1 in Section 5 of the Online Appendix.

[25] One can use the persistency thresholds to compute transition probabilities from employment-to-employment, Prob($E\prime|E$), and unemployment-to-unemployment, Prob($U\prime|U$). Consider the lowest persistency threshold, for WTP = 1.29, the transition probabilities amount to 0.99 and 0.79, respectively. For the US, Hobijn and Şahin (2009) estimate annual probabilities to remain in employment and unemployment, of 0.90 and 0.07, respectively.

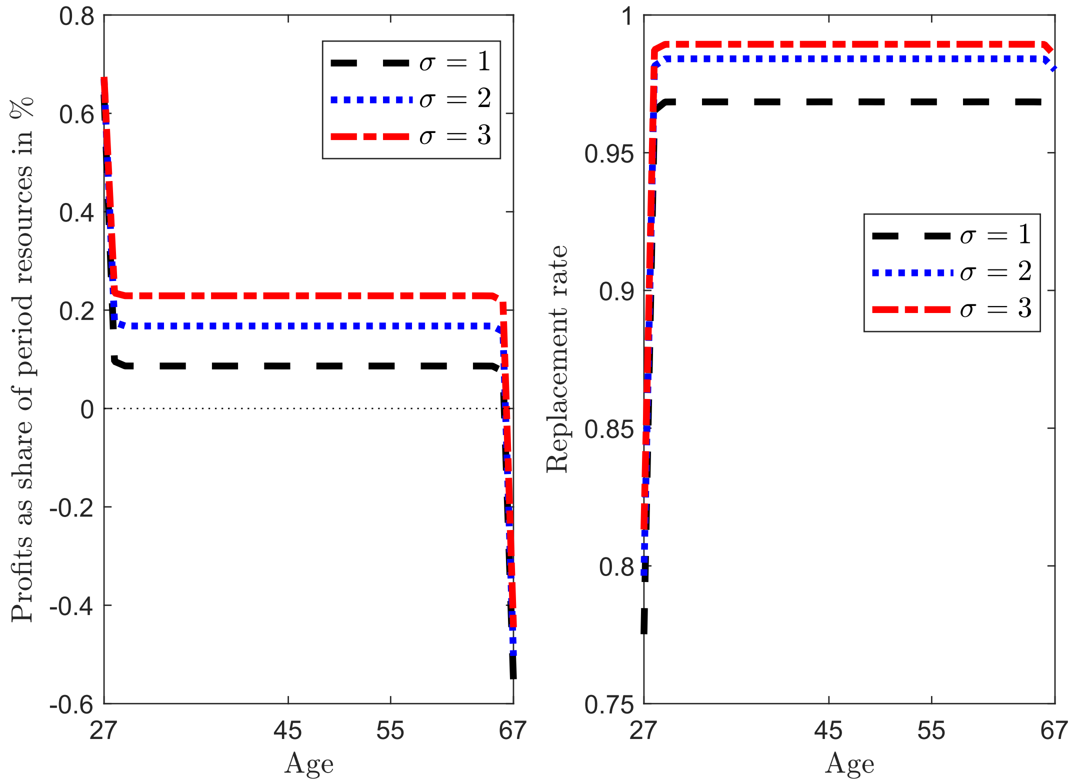

Fig. 2. Life-cycle economy. Insurer profits(left panel) and replacement rate (right panel), defined as the ratio of consumption of unemployed agents over that of employed agents at each tenure in the labor market, c_l^k/c_{gy}^k, per cohort for different risk aversion, lifetime exclusion, $N_k = K - k$.

to be significantly positive, ranging from 0.20 for $j = 1$ to 0.05 for $j = 8$. To get an estimate of the implied signal persistence ρ^b in our model, we choose ρ^b to minimize the sum of equally-weighted squared differences between the regression coefficients in the data and the ones estimated on data generated by the Markov chain for signals. We estimate $\rho^b = 0.3192$. Given this persistency, we compute the necessary exclusion length N_{min} for profitable UI. The resulting exclusion length is displayed in the third row of Table 3. The exclusion lengths are very similar to the ones obtained in the benchmark in Table 2.[26]

4.3.3. Finite planning horizon

With an infinite horizon, an exclusion length of up to three years is found to be enough to render UI profitable. A natural question is whether this conclusion changes when agents' planning horizon is shorter than three years, for example, because of retirement. This is a relevant question because the employed individuals in Hendren (2017)'s HRS sample are up to 64 years old. The (full) retirement age in the US is 67, which makes considering a finite planning horizon relevant, at least for some of the individuals in the HRS sample. For this reason, we study a life-cycle economy.[27]

Agents work for K periods and then they retire. Each period a new generation is born.[28] All generations are of equal measure set to $\frac{1}{K}$ so that the total measure of all agents in the economy is equal to 1. In period t, the working age of a generation who entered the labor market in $t - k$ is k, and we will use k as generation index. An important feature of the life-cycle economy, absent in the infinite horizon baseline, is that the insurer can condition insurance premia and benefits on publicly observable age, which allows for intergenerational transfers. As in the infinite horizon economy, the incentive compatibility constraints of the bad signal agents are binding, and for each k, the utility allocation comprises two elements, $h^k = \left\{h_l^k, h_{gy}^k\right\}, k \in \{1, \dots, K\}$. Taking this into account, we define the working age k expected utility as:

$$\tilde{w}^k = (1-\beta)\left[(\mu + (1-\mu)p)h_l^k + (1-\mu)(1-p)h_{gy}^k\right],$$

while the expected utility in autarky in every period is independent of k and given by:

$$\widetilde{U}^{Aut} = (1-\beta)[(\mu + (1-\mu)p)u(y-l) + (1-\mu)(1-p)u(y)].$$

With a finite working life, a cohort can not be excluded for longer than the remaining time until retirement. For example, a total exclusion length of two periods is only relevant for all agents that have at least two years until they retire. With exclusion of $N \geqslant 1$ future periods and $k < K$, the individual rationality constraints read:[29]

$$(1-\beta)h_l^k + \sum_{t=k+1}^{\min\{k+N,K\}} \beta^{t-k}\tilde{w}^t \geqslant (1-\beta)u(y-l) + \sum_{t=k+1}^{\min\{k+N,K\}} \beta^{t-k}\widetilde{U}^{Aut},$$

and:

[26] The results are robust to alternatively weighing the squared deviation of each regression coefficient by the inverse of its standard error.

[27] In the main text, we focus on the main features of this economy. We provide further details, including the insurer's profit maximization problem that defines the cost-efficient allocation in Section 4 of the Online Appendix.

[28] Alternatively, consider an insurer that can transfer resources between periods for a single cohort of agents.

[29] For deterministic exclusion, we assume here N to be an integer. The formulae can be adapted to have $N \in \mathbb{R}_{\geqslant 0}$, for example, by taking an appropriately weighted average of the expected utility of the contract and the outside option in the last non-integer part of the exclusion period.

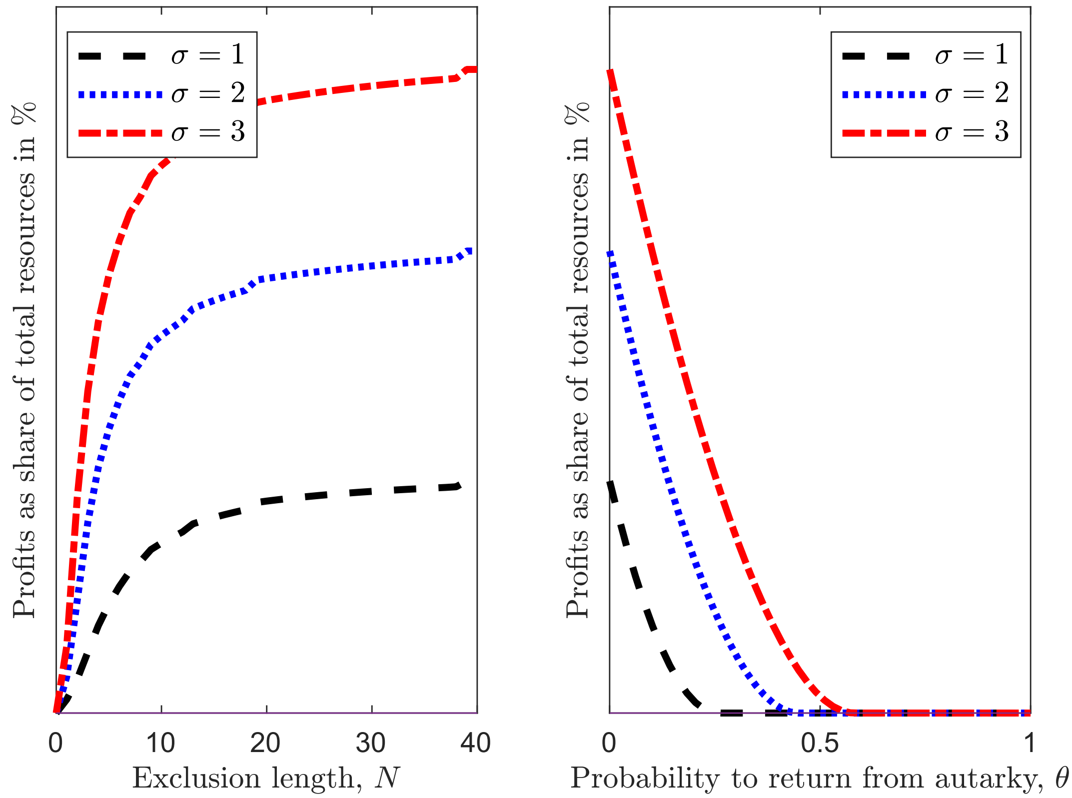

Fig. 3. Life-cycle economy. Insurer profits as share of total resources in % for different exclusion length (left panel) and probability to return from autarky (right panel) and risk aversion.

$$(1-\beta)\left[ph_l^k + (1-p)h_{gy}^k\right] + \sum_{t=k+1}^{\min\{k+N,K\}} \beta^{t-k}\widetilde{w}^t \geqslant$$
$$(1-\beta)[pu(y-l) + (1-p)u(y)] + \sum_{t=k+1}^{\min\{k+N,K\}} \beta^{t-k}\widetilde{U}^{Aut}.$$

When either $N = 0$ or $k = K$, these constraints involve instantaneous utilities only:

$$(1-\beta)h_l^k \geqslant (1-\beta)u(y-l), \tag{25}$$
$$(1-\beta)\left[ph_l^k + (1-p)h_{gy}^k\right] \geqslant (1-\beta)[pu(y-l) + (1-p)u(y)], \tag{26}$$

and the individuality rationality constraints are like the ones in the single-interaction economy.

To numerically compute cost-efficient allocations, we set $y - l = 1$ and then compute l to yield the ratio of marginal utilities reported as the willingness to pay in Table 1. The values of β, p and μ are taken from Table 1, the total working life is set to $K = 40$.

Cost-efficient allocations in the life-cycle economy are characterized by two key properties. As illustrated in Fig. 2, the insurer extracts relatively more profits from the younger generations. The reason is that the younger generations have a higher willingness to pay for insurance because they can become unemployed multiple times in the future. Correspondingly, agents that are close to retirement value unemployment insurance less. Hence, it is relatively easier to extract resources from the younger generations. Second, the insurer provides better unemployment insurance to the older generations as indicated by a higher replacement rate or smaller relative difference between consumption of employed and unemployed agents. Agents that are close to retirement have a lower future value of insurance. To keep these agents in the insurance scheme, the insurer has to offer them relatively better unemployment insurance than the younger generations.

As illustrated in Fig. 3, we find very similar total exclusion lengths for the finite horizon economy as the ones computed in Table 2. If exclusion is not possible, that is, when either $N = 0$ or $\theta = 1$, the insurer can not improve upon the no-trade allocation and makes zero profits. However, we find that already for $N = 1$ the insurer incurs positive profits in case of deterministic exclusion for all three risk-aversion values (left panel). For the random return specification, the threshold values for θ are 0.24, 0.43, and 0.57, for $\sigma = \{1, 2, 3\}$ (right panel). These values imply expected necessary exclusion lengths of 3.08, 1.33, and 0.75 years, respectively.

5. Conclusions

The main takeaway from our analysis is that the future benefits of insurance matter for the existence of a profitable UI market. These benefits naturally emerge when insurer and workers meet repeatedly. We find these benefits to be sizeable, such that our quantitative exercises imply that relatively short exclusion suffices to render the UI market profitable in the US.

Notably, our quantitative results do not derive from an exogenous variation on the effect of exclusion on market existence in reality. Therefore, the results have to be interpreted with caution. Nevertheless, our analytical results demonstrate that the threat of exclusion has the potential to be a powerful tool for facilitating profitable private unemployment insurance provision. Therefore, the hypothesis that adverse selection alone can explain the non-existing unemployment insurance market needs revisiting in future work.

Our theoretical and quantitative results have clear policy implications for how to stimulate the existence of the private UI market.

The threat of exclusion is pivotal and must be credible. A useful mechanism to ensure credibility is a registry system that collects information on UI premia defaults, similar to the successful system collecting information on credit defaults. However, creating such a registry system entails fixed costs, and economic policy can facilitate creating and maintaining the registry. The main difference to the credit market is that we show that exclusion is not only important for the well-functioning of the private UI market but already for its existence. Thus, the registry system must be pre-announced and up and running when the first UI policies are sold.

If other fixed costs of insurance provision existed that were not overcome by economic policy, then moral hazard could not be discarded as a force preventing the existence of a UI market anymore. Indeed, the results on the irrelevance of moral hazard by Shavell (1979) and Hendren (2017) hinge on the assumption of insurance provision subject to no fixed costs. Furthermore, if there are fixed costs of insurance provision, then the permanent differences in layoff risk, e.g., the presence of permanently safe agents, are relevant for the existence of private UI. This is because the larger the population of agents who are never willing to purchase insurance, the smaller the profits of the insurer.

The main premise of our analysis is not limited to unemployment insurance. Indeed, we expect the ability of the insurer to exclude agents from insurance following their default to also improve the profitability of insurance in existing markets with adverse selection due to advance information. We leave the exploration of the effects of exclusion in these markets for future research.

Data availability

No data was used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Proof of Lemma 1

We prove this lemma in several steps. To begin with, we show all implementable allocations must at least deliver the expected lifetime utility of autarky.

Step 1 Let h be an implementable allocation. The following statements hold.

(i) Allocation h delivers a lifetime expected utility not worse than that of autarky, $\tilde{w}(h) \geqslant U^{Aut}$.

(ii) It is $\tilde{w}(h) = U^{Aut}$ if, and only if, h is the no-trade (autarky) allocation.

(iii) Let h be an implementable allocation with trade. Then, $h_{bl} \geqslant u(y-l)$ and the individual rationality constraint of the bad-signal agents is slack.

Proof. These statements are standard results. For this reason, we just sketch the proofs and omit the details. If (i) was not true, adding the two individual rationality constraints would result in h being not implementable. The if part of statement (ii) is by definition. For the only-if part, setting $\tilde{w}(h) = U^{Aut}$ results in the following individual rationality constraints for agents with a good and bad signal

$$(1-p)h_{gy} + ph_{gl} \geqslant (1-p)u(y-l) + pu(y)$$
$$h_{bl} \geqslant u(y-l).$$

The definitions of $\tilde{w}(h)$ and U^{Aut} yield

$$\tilde{w}(h) = U^{Aut} \iff (1-\mu)\left[(1-p)h_{gy} + ph_{gl}\right] + \mu h_{bl}$$
$$= (1-\mu)[(1-p)u(y) + pu(y-l)] + \mu u(y-l).$$

Thus, both individual rationality constraints can only be satisfied if $h_{gl} = h_{bl} = u(y-l)$and $h_{gy} = u(y)$ to also meet $\tilde{w}(h) = U^{Aut}$. Hence, with trade it must be that $\tilde{w}(h) > U^{Aut}$. Statement (iii) is proven by contradiction, assuming $h_{bl} < u(y-l)$ together with (4) and (6) implies $h_{gl} < u(y-l)$ and $h_{gy} > u(y)$ which either violates resource feasibility or yields $\tilde{w}(h) < U^{Aut}$. □

In the following step, we prove Part (i) of Lemma 1.

Step 2 Let h be a cost-efficient allocation with trade. Incentive constraints of bad-signal agents hold with equality, so that utility for agents who incur an income loss is equalized across signal realizations, $h_{bl} = h_{gl} = h_l$, and the incentive constraints of good-signal agents are slack.

Proof. Given the incentive compatibility constraint of bad-signal agents, we only need to consider implementable allocations with trade which have either $h_{bl} > h_{gl}$ or $h_{bl} = h_{gl} = h_l$. For the first case, consider a cost-efficient allocation with trade, $h_0 = \{h_{gy}, h_{gl}, h_{bl}\}$, $h_{bl} > h_{gl}$. Thus, by assumption, h_0 is also implementable (satisfies all constraints). The resulting expected utility is:

$$\tilde{w}(h_0) = \underbrace{(1-\mu)(1-p)h_{gy}}_{\tilde{w}_y(h_0)} + \underbrace{(1-\mu)ph_{gl} + \mu h_{bl}}_{\tilde{w}_l(h_0)}$$

with $\tilde{w}_y(h_0)$ as the utility of agents without and $\tilde{w}_l(h_0)$ as the expected utility of agents with income loss. The resource costs $\widetilde{C}(h_0)$ of this allocation are:

$$\widetilde{C}(h_0) = \underbrace{(1-\mu)(1-p)C(h_{gy})}_{\widetilde{C}_y(h_0)} + \underbrace{(1-\mu)pC(h_{gl}) + \mu C(h_{bl})}_{\widetilde{C}_l(h_0)}$$

Next, let ε_b^{ir} be the utility slack in the bad-signal agents individual rationality constraint (7) and $\varepsilon_b^{ic} = h_{bl} - h_{gl}$ the utility slack in the incentive constraint of bad-signal agents (4). By assumption, $\varepsilon_b^{ir} > 0$ and $\varepsilon_b^{ic} > 0$. Consider a perturbation $h_1 = \{h_{gy}, h_{gl} + \delta_g, h_{bl} - \delta_b\}$ such that $\delta_g = \frac{\mu\delta_b}{(1-\mu)p}$ and the expected utility of the alternative allocation is given by

$$\tilde{w}(h_1) = \underbrace{(1-\mu)(1-p)h_{gy}}_{\tilde{w}_y(h_1)} + \underbrace{(1-\mu)p(h_{gl} + \delta_g) + \mu(h_{bl} - \delta_b)}_{\tilde{w}_l(h_1)} = \tilde{w}(h_0).$$

This perturbation keeps the incentive compatibility of good-signal agents satisfied (as it was already met by h_0, implementable by assumption, and now we are decreasing h_{bl} and increasing h_{gl}). It also trivially satisfies the good-signal agents individual rationality constraint (the continuation value $\tilde{w}(h_1) = \tilde{w}(h_0)$ is unchanged and we have increased h_{gl}). We need to ensure that the remaining two constraints are also satisfied. The individual rationality constraint of bad-signal agents is satisfied for $\delta_b \leqslant \varepsilon_b^{ir}/(1-\beta)$. The incentive compatibility constraint of bad-signal agents requires:

$$h_{bl} - \delta_b \geqslant h_{gl} + \delta_g \iff \delta_b \leqslant \varepsilon_l^{ic} \frac{(1-\mu)p}{(1-\mu)p + \mu}$$

Thus, $\delta_b \leqslant \min\left\{\frac{\varepsilon_l^{ir}}{1-\beta}, \varepsilon_l^{ic}\left[\frac{(1-\mu)p}{(1-\mu)p+\mu}\right]\right\}$ ensures that the perturbed allocation is implementable. The perturbed allocation requires resource costs $\widetilde{C}(h_1)$:

$$\widetilde{C}(h_1) = \underbrace{(1-\mu)(1-p)C(h_{gy})}_{\widetilde{C}_y(h_1)} + \underbrace{(1-\mu)pC(h_{gl}+\delta_g) + \mu C(h_{bl}-\delta_b)}_{\widetilde{C}_l(h_1)}$$

The difference in required resources, for an arbitrarily small δ_b is

$$\widetilde{C}(h_1) - \widetilde{C}(h_0) = (1-\mu)pC'(h_{gl})\delta_g - \mu C'(h_{bl})\delta_b = \mu\delta_b[C'(h_{gl}) - C'(h_{bl})] < 0,$$

where the last inequality follows from strict convexity of resource costs with $h_{gl} < h_{bl}$, implying that h_0 cannot be a cost-efficient allocation. Thus, a cost-efficient allocation is characterized by binding incentive constraints of bad-signal signals and $h_{bl} = h_{gl} = h_l$. Using this, Step 1 and resource feasibility imply $h_l > u(y-l)$ and $h_{gy} < u(y)$. Finally, the incentive constraint of the good-signal agents can only be met with equality if $c_{gy} = 0$ now that we have established $h_{gl} = h_{bl}$. However, this can not be optimal because $u(c)$ satisfies Inada conditions, $\lim_{c\to 0} u\prime(c) = \infty$ and hence an infinitely small redistribution back from c_l to c_{gy} yields infinite improvements in lifetime utility, increasing profits. Thus, setting $c_{gy} = 0$ can not be optimal and the good-signal agent incentive compatibility constraint must be slack. □

So far we established that only incentive constraints of good-signal agents are slack as well as individual rationality constraints of bad-signal agents. Further, we have shown that $h_l > u(y-l)$ and $h_{gy} < u(y)$. In the remaining two steps, we demonstrate first that individual rationality constraints of good-signal agents are binding, completing the proof of Lemma 1, Part (ii), and afterwards that $h_{gy} \geqslant h_l$ at the cost-efficient allocation, completing the proof of Lemma 1, Part (iii).

Step 3 Let h be a cost-efficient allocation with trade. Then the individual rationality constraint of the good-signal agents are binding.

Proof. Suppose not, such that there is not only slack $\varepsilon_b^{ir} > 0$ but also a slack $\varepsilon_g^{ir} > 0$ in the good-signal agents individual rationality constraint (6). Let $h_0 = \{h_l, h_{gy}\}$ be the cost-efficient allocation with trade in that case. This cannot be the cost-efficient allocation because further resources can be saved by choosing an allocation $h_1 = \{h_l, h_{gy} - \delta\}$ as long as

$$\delta \leqslant \min\left\{\frac{\varepsilon_b}{\beta(1-\beta^N)(1-\mu)(1-p)}, \frac{\varepsilon_g}{1-\beta+\beta(1-\beta^N)(1-\mu)(1-p)}\right\},$$

which ensures that individual rationality is satisfied for good and bad-signal agents. Choosing such allocation decreases the resource costs by $(1-\mu)(1-p)[C(h_{gy}) - C(h_{gy}-\delta)]$, contradicting that the good-signal individual rationality constraints are not binding at the cost-efficient allocation. □

Step 4 Let h be a cost-efficient allocation with trade. Then, $h_{gy} \geqslant h_l$.

Proof. Suppose not, and without loss of generality, let $h_l = h_{gy} + \varepsilon$ for $h_0 = \{h_l, h_{gy}\}$ with $\varepsilon > 0$ arbitrarily small. Then, an insurer can reduce resource costs by choosing the following allocation: $h_1 = \{h_{bl} = h_l, h_{gl} = h_l - \delta_{gl}, h_{gy} + \delta_{gy}\}$ with $\delta_{gl} = \frac{1-p}{p}\delta_{gy}$, and $\delta_{gy} \leqslant \frac{\varepsilon}{2}$. This perturbation does not affect individual rationality constraints and satisfies incentive constraints. While h_1 is not cost-efficient, it is implementable. By convexity of resource costs it marginally increases the insurer's profits by $(1-\mu)(1-p)[C'(h_{gy}+\varepsilon) - C'(h_{gy})]\delta_{gy}$. □

Appendix B. Proof of Theorem 1

The proof proceeds in two parts. First, we show that the no-trade condition is sufficient. In the second part, we show the no-trade condition is also necessary.

Part I: No-trade condition is sufficient We show that if the no-trade condition holds, autarky is the cost-efficient allocation. From Lemma 1, we consider the problem of choosing $\{h_{gy}, h_l\}$ to maximize (11) subject to the individual rationality constraints of good-signal agents (6):

$$[(1-\beta)p + \beta(1-\beta^N)[\mu + (1-\mu)p]]h_l + [(1-\beta)(1-p) + \beta(1-\beta^N)(1-\mu)(1-p)]h_{gy} \geqslant U^{Aut}(g)^N,$$

with $U^{Aut}(g)^N = (1-\beta)[pu(y-l) + (1-p)u(y)] + \beta(1-\beta^N)U^{Aut}$. The first-order conditions for h_{gy}, h_l are

$$(1-\mu)(1-p)C\prime(h_{gy}) = \lambda_g^{IR}[(1-\beta)(1-p) + \beta(1-\beta^N)(1-\mu)(1-p)]$$
$$[\mu + (1-\mu)p]C\prime(h_l) = \lambda_g^{IR}[(1-\beta)p + \beta(1-\beta^N)[\mu + (1-\mu)p]],$$

with $\lambda_g^{IR} \geqslant 0$ as the multiplier on the constraint (6). The two first order conditions can be combined into a single first order condition

$$\frac{\mu + (1-\mu)p}{(1-\mu)p}\left\{\frac{[1-\beta\mu-\beta^{N+1}(1-\mu)]p}{[1-\beta\mu-\beta^{N+1}(1-\mu)]p + \beta(1-\beta^N)\mu}\right\} = \frac{C'(h_{gy})}{C'(h_l)} = \frac{u'(c_l)}{u'(c_{gy})}.$$

Next, if the no-trade condition holds, and using (12), it follows

$$\frac{u'(y-l)}{u'(y)} \leqslant \frac{\mu + (1-\mu)p}{(1-\mu)p}\left\{\frac{[1-\beta\mu-\beta^{N+1}(1-\mu)]p}{[1-\beta\mu-\beta^{N+1}(1-\mu)]p + \beta(1-\beta^N)\mu}\right\} = \frac{u'(c_l)}{u'(c_{gy})}.$$

From Lemma 1, we have $c_l \geqslant y-l$ and $c_{gy} \leqslant y$. With decreasing marginal utility, this is only possible if $c_l = y-l$ and $c_{gy} = y$. Hence, when the no-trade condition holds, the cost-efficient allocation is the no-trade allocation.

Part II: No-trade condition is necessary We show that if the no-trade condition is not satisfied, there is an implementable allocation with trade which generates a slack in the resource constraint. Consider a perturbed allocation $h = \{h_{gy} = u(y-\delta), h_l = u(y-l+\gamma)\}$ with $\delta > 0, \gamma > 0$, with δ, γ arbitrarily small. By definition, this is an allocation with trade which also satisfies all other results in Lemma 1, but we don't know if this perturbed allocation is resource feasible. Net resources saved by this perturbation are $(1-\mu)(1-p)\delta - [\mu + (1-\mu)p]\gamma$. Observe that when this is non-negative, this perturbed allocation is implementable. By Lemma 1, the allocation implies the following pattern of binding individual rationality constraints

$$\tilde{w}(g)^N = U^{Aut}(g)^N, \tag{27}$$
$$\tilde{w}(b)^N > U^{Aut}(b)^N, \tag{28}$$

where $\tilde{w}(g)^N, \tilde{w}(b)^N$ denote the left-hand side and $U^{Aut}(g)^N, U^{Aut}(b)^N$ the right-hand side of Eqs. (6) and (7), respectively. Before the signal realizes, the relevant expected utilities for N future periods are

$$\tilde{w}^N = (1-\mu)\tilde{w}(g)^N + \mu\tilde{w}(b)^N$$
$$U^{Aut,N} = (1-\mu)U^{Aut}(g)^N + \mu U^{Aut}(b)^N$$

Let $\tilde{w}(g), \tilde{w}(b)$ and $U^{Aut}(g), U^{Aut}(b)$ be the lifetime utilities of good- and bad-signal agents as the corresponding limiting expressions

$N \to \infty$ of both sides of the individual rationality constraints (27) and (28). The following relationships between lifetime utilities and the period-N utilities apply

$$\tilde{w}(g)^N = \tilde{w}(g) - \beta^{N+1}\tilde{w} \tag{29}$$

$$\tilde{w}(b)^N = \tilde{w}(b) - \beta^{N+1}\tilde{w} \tag{30}$$

$$U^{Aut}(g)^N = U^{Aut}(g) - \beta^{N+1}U^{Aut} \tag{31}$$

$$U^{Aut}(b)^N = U^{Aut}(b) - \beta^{N+1}U^{Aut}. \tag{32}$$

Let ϵ be the slack on the individual rationality constraint of bad-signal agents:

$$\tilde{w}(b)^N - U^{Aut}(b)^N = \epsilon > 0.$$

Individual rationality constraints of good-signal agents are binding, $\tilde{w}(g)^N - U^{Aut}(g)^N = 0$. Using (30), (32) as well as (29), (31), gives the following restrictions

$$\tilde{w}(b)^N - U^{Aut}(b)^N = \epsilon = \tilde{w}(b) - U^{Aut}(b) + \beta^{N+1}\left(U^{Aut} - \tilde{w}\right)$$
$$\tilde{w}(g)^N - U^{Aut}(g)^N = 0 = \tilde{w}(g) - U^{Aut}(g) + \beta^{N+1}\left(U^{Aut} - \tilde{w}\right),$$

Using the definition of lifetime utilities of good and bad-signal agents in the insurance contract and in autarky, one eventually gets

$$\tilde{w} - U^{Aut} = \frac{\mu\epsilon}{1-\beta^{N+1}}.$$

For the individual rationality constraint of the bad-signal agents (28) combined with (30) and (32), it follows that:

$$\tilde{w}(b)^N - U^{Aut}(b)^N = \epsilon = (1-\beta)u'(y-l)\gamma + \beta\left(1-\beta^N\right)\frac{\mu\epsilon}{1-\beta^{N+1}}$$

such that ϵ can be written as

$$\epsilon = \frac{(1-\beta)\left(1-\beta^{N+1}\right)\gamma}{1-\beta^{N+1}-\beta\left(1-\beta^N\right)\mu}u'(y-l). \tag{33}$$

For the binding individual rationality constraint of good-signal agents (27) we have:

$$0 = \tilde{w}(g)^N - U^{Aut}(g)^N$$
$$= (1-\beta)[-(1-p)u'(y)\delta + pu'(y-l)\gamma] + \beta\left(1-\beta^N\right)\frac{\mu\epsilon}{1-\beta^{N+1}}$$

This implies:

$$\frac{(1-\mu)(1-p)\delta}{[\mu+(1-\mu)p]\gamma} = \frac{u'(y-l)}{u'(y)}\frac{(1-\mu)p}{\mu+(1-\mu)p} \times \frac{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p+\beta\left(1-\beta^N\right)\mu}{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p}. \tag{34}$$

Suppose that the no-trade condition is violated:

$$\frac{u'(y-l)}{u'(y)} > \frac{\mu+(1-\mu)p}{(1-\mu)p}\left\{\frac{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p}{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p+\beta\left(1-\beta^N\right)\mu}\right\}.$$

Using this, (34) boils down to:

$$\frac{(1-\mu)(1-p)\delta}{[\mu+(1-\mu)p]\gamma} > 1 \Rightarrow (1-\mu)(1-p)\delta - [\mu+(1-\mu)p]\gamma > 0,$$

thus, net resources are positive, and the no-trade condition is not only sufficient but also necessary.

Monotonicity of $T_r(p,\mu,\beta,N)$ in N. Let's rewrite the definition of $T_r(p,\mu,\beta,N)$ as:

$$T_r(p,\mu,\beta,N) = \frac{f(p,\mu,\beta,N)}{f(p,\mu,\beta,N)+g(\mu,\beta,N)}$$

with: $f(p,\mu,\beta,N) = \left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p$ and $g(\mu,\beta,N) = \beta\left(1-\beta^N\right)\mu$. Then, $\frac{\partial T_r}{\partial N} = \frac{\frac{\partial f}{\partial N}(f+g)-\left(\frac{\partial f}{\partial N}+\frac{\partial g}{\partial N}\right)f}{[f+g]^2} = \frac{\frac{\partial f}{\partial N}g-\frac{\partial g}{\partial N}f}{[f+g]^2}$. We have:

$$\frac{\partial f(p,\mu,\beta,N)}{\partial N} = -\beta^{N+1}(1-\mu)p\ln(\beta),$$
$$\frac{\partial g(\mu,\beta,N)}{\partial N} = -\beta^{N+1}\mu\ln(\beta).$$

Thus:

$$\frac{\partial f}{\partial N}g - \frac{\partial g}{\partial N}f = -\beta^{N+1}\ln(\beta)\mu p(1-\beta) < 0,$$

as we have assumed $0 < \beta < 1, \mu > 0$ and $p > 0$.

Appendix C. Public signals

Individual rationality and private information together restrict insurance possibilities. To clarify the importance of both frictions, we show that individual rationality alone does not prevent the existence of insurance. Suppose alternatively that agents receive *public* signals on their income-loss probability that have otherwise the same properties as the private ones. Thus, the insurer is only constrained by resource feasibility and individual rationality. In this environment, we find the following result.

Proposition 2 (*Public information*). Autarky is not the cost-efficient allocation when signals are public.

We begin with the intuition behind this result and provide the proof below. Any insurance transfer between good-signal agents unambiguously increases their lifetime utility. The increase in lifetime utility also increases the continuation value of bad-signal agents and thus indirectly their lifetime utility. Hence, individual rationality constraints of bad-signal agents are satisfied even when they receive no transfers from good-signal agents. The key difference is that with public signals, incentive constraints are absent and such a transfer scheme is implementable. With private information, such transfer scheme would violate the incentive constraints of bad-signal agents, increasing its costs beyond what is actuarially fair. Therefore, individual rationality alone cannot explain the absence of a profitable private insurance market.

Proof. To see why autarky cannot be a cost-efficient allocation, consider an insurance scheme that involves only transfers between good-signal agents, so that $h_{gy} = u(y-\delta)$ and $h_{gl} = u\left(y-l+\gamma_g\right), \delta, \gamma_g > 0$, in line with resource feasibility:

$$(1-p)\delta \geqslant p\gamma_g.$$

The scheme provides insurance, leads to higher utility by concavity, to lower resource costs by convexity, and therefore at least weakly dominates autarky in terms of resources saved. What remains to be shown is that the scheme is also implementable. First, such a scheme affects the individual rationality constraints of good-signal agents as follows

$$(1-\beta)\left[pu'(y-l)\gamma_g - (1-p)u'(y)\delta\right]$$
$$+\beta\left(1-\beta^N\right)(1-\mu)\left[pu'(y-l)\gamma_g - (1-p)u'(y)\delta\right]$$
$$= [u'(y-l)-u'(y)]\left[(1-\beta)+\beta\left(1-\beta^N\right)(1-\mu)\right]p\gamma_g \geqslant 0.$$

The second line uses resource feasibility, the third one that $\left[(1-\beta)+\beta\left(1-\beta^N\right)(1-\mu)\right], p, \gamma_g$ are strictly positive; The positive sign then follows for $l > 0$, and strict concavity of utility. Per construction, such a scheme is consistent with resource feasibility and also satisfies individual rationality constraints of bad-signal agents who remain at autarky. Thus, autarky cannot be a cost-efficient allocation. □

Appendix D. Persistent signals

In this section, we first describe the environment with persistent signals. Afterwards, we argue that the results of Lemma 2 apply and provide the proof of Proposition 1.

Constraints The incentive-compatibility constraints read:

$$(1-\beta)h_{bl} + \beta\left(\rho^b\tilde{w}(b) + \left(1-\rho^b\right)\tilde{w}(g)\right) \\ \geqslant (1-\beta)h_{gl} + \beta\left(\rho^b\tilde{w}(b) + \left(1-\rho^b\right)\tilde{w}(g)\right),$$

and:

$$(1-\beta)\left[ph_{gl} + (1-p)h_{gy}\right] + \beta((1-\rho^g)\tilde{w}(b) + \rho^g\tilde{w}(g)) \geqslant \\ (1-\beta)[ph_{bl} + (1-p)u(0)] + \beta((1-\rho^g)\tilde{w}(b) + \rho^g\tilde{w}(g)).$$

The currently reported realisation of the signal only matters for instantaneous utilities. Thus, the continuation values can be scrapped again, leading to identical IC constraints as before:

$$h_{bl} \geqslant h_{gl}, \tag{35}$$
$$\left[ph_{gl} + (1-p)h_{gy}\right] \geqslant ph_{bl} + (1-p)u(0). \tag{36}$$

The individual rationality constraints can be written as:

$$(1-\beta)\left[ph_{gl} + (1-p)h_{gy}\right] + \beta\left(1-\beta^N\right)\left(\left(1-\rho^g\right)\tilde{w}(b) + \rho^g\,\tilde{w}(g)\right) \geqslant \\ (1-\beta)[pu(y-l) + (1-p)u(y)] + \beta\left(1-\beta^N\right)\left[(1-\rho^g)U^{\mathrm{Aut}}(b) \\ +\rho^g U^{\mathrm{Aut}}(g)\right], \tag{37}$$

and:

$$(1-\beta)h_{bl} + \beta\left(1-\beta^N\right)\left(\rho^b\tilde{w}(b) + \left(1-\rho^b\right)\tilde{w}(g)\right) \\ \geqslant (1-\beta)[pu(y-l) + (1-p)u(y)] + \beta\left(1-\beta^N\right) \\ \left[(1-\rho^g)U^{\mathrm{Aut}}(b) + \rho^g U^{\mathrm{Aut}}(g)\right]. \tag{38}$$

The lifetime utilities with insurance, $\tilde{w}(g), \tilde{w}(b)$, and in autarky, $U^{\mathrm{Aut}}(g), U^{\mathrm{Aut}}(b)$, are recursively defined by

$$\tilde{w}(g) = (1-\beta)\left[ph_l + (1-p)h_{gy}\right] + \beta[\rho^g\tilde{w}(g) + (1-\rho^g)\tilde{w}(b)], \tag{39}$$
$$\tilde{w}(b) = (1-\beta)h_l + \beta\left[\rho^b\tilde{w}(b) + \left(1-\rho^b\right)\tilde{w}(g)\right], \tag{40}$$

and

$$U^{\mathrm{Aut}}(g) = (1-\beta)\left[ph_{gl} + (1-p)h_{gy}\right] + \beta\left[\rho^g U^{\mathrm{Aut}}(g) \\ +(1-\rho^g)U^{\mathrm{Aut}}(b)\right], \tag{41}$$
$$U^{\mathrm{Aut}}(b) = (1-\beta)h_{bl} + \beta\left[\rho^b U^{\mathrm{Aut}}(b) + \left(1-\rho^b\right)U^{\mathrm{Aut}}(g)\right]. \tag{42}$$

Properties of cost-efficient allocations Lemma 1 is unaffected by relaxing the assumptions of i.i.d. signals. Note that Step 1 of the proof of this Lemma rely solely on the IC constraints and the profit function of the insurer. Hence, it goes through unchanged when signals are arbitrarily persistent. Similar reasoning as in the case of the i.i.d. signals yields that the insurer optimally sets $h_{gl} = h_{bl}$ in a cost-efficient allocation with trade as it improves profits relative to offering $h_{gl} > h_{bl}$ (Step 2). Next, providing expected utility to agents that yields the IR constraint of the good signal agents slack cannot be profit-maximizing either (Step 3). Together with $h_{gl} = h_{bl}$, if an allocation features trade, then the IR constraint of the bad agents must be slack. Finally, the insurer can also incur additional profits relative to an allocation which has $h_l > h_{gy}$ and hence, in any cost-efficient allocation with trade $h_{gy} \geqslant h_l$ (Step 4).

Re-writing the individual rationality constraints of agents with a good signal Before we prove Proposition 1, it is useful to write the individual rationality constraint of agents with a good signal in a compact form as summarized in the following lemma.

Lemma 2. The individual rationality constraint of agents with a good signal can be written as follows

$$\left[(1-\beta)p + \beta\left(1-\beta^N\right)\left(\rho^g a_l^g + (1-\rho^g)a_l^b\right)\right]h_l \\ + \left[(1-\beta)(1-p) + \beta\left(1-\beta^N\right)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]h_y \\ \geqslant \left[(1-\beta)p + \beta\left(1-\beta^N\right)\left(\rho^g a_l^g + (1-\rho^g)a_l^b\right)\right]u(y-l) \\ + \left[(1-\beta)(1-p) + \beta\left(1-\beta^N\right)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]u(y) \tag{43}$$

with coefficients:

$$a_l^g = \frac{(1-\beta)\left(p + \frac{\beta(1-\rho^g)}{1-\beta\rho^b}\right)}{1-\beta\left(\rho^g + \frac{\beta(1-\rho^g)\left(1-\rho^b\right)}{1-\beta\rho^b}\right)} > 0, \quad a_y^g \\ = \frac{(1-\beta)(1-p)}{1-\beta\left(\rho^g + \frac{\beta(1-\rho^g)\left(1-\rho^b\right)}{1-\beta\rho^b}\right)} > 0, \tag{44}$$

and:

$$a_l^b = \frac{(1-\beta) + \beta\left(1-\rho^b\right)a_l^g}{1-\beta\rho^b} > 0, \quad a_y^b = \frac{\beta\left(1-\rho^b\right)a_y^g}{1-\beta\rho^b} > 0. \tag{45}$$

Proof. Using $h_{gl} = h_{bl} = h_l$ in Eqs. 39,40, we can solve for the lifetime utilities to get

$$\tilde{w}(b) = \frac{(1-\beta)h_l + \beta\left(1-\rho^b\right)\tilde{w}(g)}{1-\beta\rho^b} \\ \tilde{w}(g) = (1-\beta)\left[ph_l + (1-p)h_y\right] \\ + \beta\left(\rho^g\tilde{w}(g) + (1-\rho^g)\frac{(1-\beta)h_l + \beta\left(1-\rho^b\right)\tilde{w}(g)}{1-\beta\rho^b}\right).$$

Therefore:

$$\tilde{w}(g) = (1-\beta)\left(p + \frac{\beta(1-\rho^g)}{1-\beta\rho^b}\right)h_l + (1-\beta)(1-p)h_y \\ +\beta\left(\frac{\rho^g\left(1-\beta\rho^b\right)+\beta(1-\rho^g)\left(1-\rho^b\right)}{1-\beta\rho^b}\right)\tilde{w}(g),$$

and:

$$\left[1-\beta\left(\rho^g + \frac{\beta(1-\rho^g)\left(1-\rho^b\right)}{1-\beta\rho^b}\right)\right]\tilde{w}(g) \\ = (1-\beta)\left(p + \frac{\beta(1-\rho^g)}{1-\beta\rho^b}\right)h_l + (1-\beta)(1-p)h_y.$$

Hence, $\tilde{w}(g)$ satisfies:

$$\tilde{w}(g) = a_l^g h_l + a_y^g h_y,$$

with coefficients a_l^g and a_y^g as in Eq. (44). Given the parameter restrictions, the numerator of both coefficients is positive; the denominator is positive because the expression in brackets is smaller than unity. These also imply

$$\tilde{w}(b) = a_l^b h_l + a_y^b h_y,$$

with coefficients a_l^b and a_y^b as in Eq. (45) and the positive signs follow from $a_l^g, a_>^g 0$. Plugging these representations of $\tilde{w}(g)$ and $\tilde{w}(b)$ into Eq. (37) and redoing the same steps for the outside options (41)–(42) results in the IR constraint as written in Eq. (43). □

In the next paragraph, we provide the proof of Proposition 1.

Proof of Proposition 1

(i)The proof follows the same steps as in Theorem 1. In particular, the profit function is (10). Profits are maximized subject to the individual rationality constraints of agents with a good signal (43).

Sufficiency The first order conditions for a profit-maximising monopolist insurer are:

$$(1-\mu)(1-p)C'(h_y) = \lambda_g^{IR}[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)] \quad (46)$$

$$(\mu + (1-\mu)p)C'(h_l) = \lambda_g^{IR}[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]. \quad (47)$$

where λ_g^{IR} is the Lagrange multiplier on the good signal IR constraint. Then, the sufficient condition can be obtained from combining the two into:

$$\frac{(1-\mu)(1-p)}{\left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]}C'(h_y) = \frac{\mu + (1-\mu)p}{[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]}C'(h_l)$$

and rearranging and imposing the inequality on the marginal utility in the allocation with trade and autarky:

$$\frac{u'(y-l)}{u'(y)} \leqslant \frac{\mu + (1-\mu)p}{(1-\mu)(1-p)} \frac{\left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]}{[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]} = \frac{C'(h_y)}{C'(h_l)} = \frac{u'(c_l)}{u'(c_y)}$$

Multiplying the right hand side of this condition by $p/(1-p)$ and its reciprocal reveals that indeed it can be written as $T_s(p,\mu) \times T_r(p,\beta,N,\rho^g,\rho^b)$ where:

$$T_r(p,\beta,N,\rho^g,\rho^b) = \frac{p}{1-p} \frac{\left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]}{[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]}.$$

Necessity We consider a perturbed allocation $h = \{h_{gy} = u(y-\delta), h_l = u(y-l+\gamma)\}$ with $\delta > 0, \gamma > 0$ with δ, γ arbitrarily small. As before, we don't know if this perturbed allocation is resource feasible. Net resources generated by this perturbation are $(1-\mu)(1-p)\delta - [\mu + (1-\mu)p]\gamma$.

We know that the IR constraint of the good-signal agents holds with equality. Then, using the derivative approximation to differences in utility levels for this perturbed allocation we obtain:

$$[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]u'(y-l)\gamma - \left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]u'(y)\delta = 0,$$

and this is equivalent to:

$$\frac{u'(y-l)}{u'(y)} \frac{(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)}{(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)} = \frac{\delta}{\gamma}. \quad (48)$$

Now, suppose the no-trade condition is violated:

$$\frac{u'(y-l)}{u'(y)} > \frac{\mu + (1-\mu)p}{(1-\mu)(1-p)} \times \frac{\left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]}{[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)]}.$$

Plugging this into Eq. (48), we obtain:

$$\frac{\mu + (1-\mu)p}{(1-\mu)(1-p)} < \frac{\delta}{\gamma} \Longleftrightarrow (1-\mu)(1-p)\delta - (\mu + (1-\mu)p)\gamma > 0$$

Thus, assuming that the no-trade condition is violated implies that the perturbed allocation generates positive net resources and hence improves profits of the insurer relative to autarky.

Next, regarding the monotonicity of $T_r(p,\beta,N,\rho^g,\rho^b)$ with respect to N, taking logs:

$$\ln\left[T_r(p,\beta,N,\rho^g,\rho^b)\right] = \ln(p) + \ln\left[(1-\beta)(1-p) + \beta(1-\beta^N)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right] - \ln\left[(1-\beta)p + \beta(1-\beta^N)(\rho^g a_l^g + (1-\rho^g)a_l^b)\right] - \ln(1-p)$$

and hence:

$$\frac{\partial}{\partial N}\ln\left[T_r(p,\beta,N,\rho^g,\rho^b)\right] = \frac{-\beta^{N+1}\ln(\beta)\left[\rho^g a_y^g + (1-\rho^g)a_y^b\right]}{\left[(1-\beta)(1-p) + \beta\left(1-\beta^N\right)\left(\rho^g a_y^g + (1-\rho^g)a_y^b\right)\right]} + \frac{\beta^{N+1}\ln(\beta)\left[\rho^g a_l^g + (1-\rho^g)a_l^b\right]}{\left[(1-\beta)p + \beta\left(1-\beta^N\right)\left(\rho^g a_l^g + (1-\rho^g)a_l^b\right)\right]}.$$

The sign of the derivative is negative whenever the following expression is positive

$$(1-p)\left[\rho^g a_l^g + (1-\rho^g)a_l^b\right] - p\left[\rho^g a_y^g + (1-\rho^g)a_y^b\right] = (1-p)\left[a_l^b + \rho^g(a_l^g - a_l^b)\right] - p\left[a_y^b + \rho^g\left(a_y^g - a_y^b\right)\right].$$

It is

$$a_y^g - a_y^b = a_y^g\left[1 - \frac{\beta(1-\rho^b)}{1-\beta\rho^b}\right] = a_y^g \frac{1-\beta}{1-\beta\rho^b} \text{and}$$

$$a_l^g - 1 = -\frac{(1-\beta)(1-p)}{1 - \beta\left[\rho^g + \frac{\beta(1-\rho^g)(1-\rho^b)}{1-\beta\rho^b}\right]} = -a_y^g,$$

which implies

$$a_l^g - a_l^b = \frac{(1-\beta)(a_l^g - 1)}{1-\beta\rho^b} = -\frac{1-\beta}{1-\beta\rho^b}a_y^g.$$

Hence:

$$(1-p)\left[a_l^b + \rho^g(a_l^g - a_l^b)\right] - p\left[a_y^b + \rho^g\left(a_y^g - a_y^b\right)\right] = (1-p)a_l^b - pa_y^b - \rho^g\frac{1-\beta}{1-\beta\rho^b}a_y^g$$

After quite some algebra, the last equality is

$$(1-p)a_l^b - pa_y^b - \rho^g\frac{1-\beta}{1-\beta\rho^b}a_y^g = \frac{(1-p)(1-\rho^g)}{1-\beta(\rho^b + \rho^g - 1)} > 0.$$

Thus, $\frac{\partial}{\partial N}\ln\left[T_r(p,\beta,N,\rho^g,\rho^b)\right] < 0 \Rightarrow \frac{\partial}{\partial N}T_r(p,\beta,N,\rho^g,\rho^b) < 0.$

Regarding the limiting value of $T_r(p,\beta,N,\rho^g,\rho^b)$, one gets

$$\lim_{\rho^b \to \mu, \rho^g \to 1-\mu} T_r(p,\beta,N,\rho^g,\rho^b) = \frac{p(1-p)\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]}{(1-p)\left[p(1-\beta\mu-\beta^{N+1}(1-\mu) + \beta(1-\beta^N)\mu\right]} = \frac{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p}{\left[1-\beta\mu-\beta^{N+1}(1-\mu)\right]p + \beta\left(1-\beta^N\right)\mu} = T_r(p,\mu,\beta,N).$$

(ii) To sign the partial derivative $\partial T_r(p,\beta,N,\rho^g,\rho^b)/\partial\rho^b$, we need to sign the responses of numerator and denominator of $T_r(p,\beta,N,\rho^g,\rho^b)$ with respect to ρ^b. We consider resource-neutral changes in ρ^b as formalized in Definition 3. This implies that ρ^g is a function of μ and ρ^b such that $T_r(p,\beta,N,\rho^g,\rho^b) = T_r(p,\mu,\beta,N,\rho^b)$.

We will show that:

$$\frac{\partial}{\partial\rho}\left[\rho^g a_y^g + (1-\rho^g)a_y^b\right] > 0 \text{and} \frac{\partial}{\partial\rho}\left[\rho^g a_l^g + (1-\rho^g)a_l^b\right] < 0,$$

which implies that $\partial T_r(p,\mu,\beta,N,\rho^b)/\partial\rho^b > 0$ because the numerator and the denominator of $T_r(p,\mu,\beta,N,\rho^b)$ are strictly positive. The proof proceeds in several steps.

Step 1 The denominator of a_l^g and a_y^g is strictly decreasing in ρ^b.

Proof. Using that with resource-neutral changes it is $\rho_g = [1-\mu(2-\rho^b)]/(1-\mu)$, the derivative of the denominator of a_l^g and a_y^g with respect to ρ^b is

$$\frac{d}{d\rho^b}\left[\frac{1}{\beta}-\rho^g-\frac{\beta(1-\rho^g)(1-\rho^b)}{1-\beta\rho^b}\right]=\frac{d}{d\rho^b}\frac{(\beta\mu-2\mu-\beta\rho^b+\mu\rho^b+1)}{(\mu-1)(1-\beta\rho^b)}$$
$$=\frac{\mu(1-\beta)^2}{(1-\beta\rho^b)^2(\mu-1)}=-\frac{\mu(1-\beta)^2}{(1-\beta\rho^b)^2(1-\mu)}<0.$$

Thus, the denominator of a_l^g and a_y^g is strictly decreasing in ρ^b. □

Step 2 The numerator of $T_r(p,\mu,\beta,N,\rho^b)$is strictly increasing in ρ^b.

Proof. The derivative of the numerator of $T_r(p,\mu,\beta,N,\rho^b)$ with respect to ρ^b is

$$\frac{d}{d\rho^b}\left[(1-\beta)(1-p)+\beta(1-\beta^N)\left(\rho^g a_y^g+(1-\rho^g)a_y^b\right)\right]$$
$$=\beta(1-\beta^N)\frac{d}{d\rho^b}\left[\rho^g a_y^g+(1-\rho^g)a_y^b\right]$$
$$=\beta(1-\beta^N)\frac{d}{d\rho^b}\left[1+\frac{\mu(1-p)(1-\rho^b)}{\mu+\beta\rho^b-\beta\mu-1}-p\right]$$
$$=\beta(1-\beta^N)\frac{\mu(1-\beta)(1-\mu)(1-p)}{(\mu-\beta\mu+\beta\rho^b-1)^2}>0.$$

□

Step 3 The denominator of $T_r(p,\mu,\beta,N,\rho^b)$is strictly decreasing in ρ^b.

Proof. Differentiating the denominator of $T_r(p,\mu,\beta,N,\rho^b)$we get:

$$\frac{d}{d\rho^b}\left[(1-\beta)p+\beta(1-\beta^N)\left(\rho^g a_l^g+(1-\rho^g)a_l^b\right)\right]$$
$$=\beta(1-\beta^N)\frac{d}{d\rho^b}\left[\rho^g a_l^g+(1-\rho^g)a_l^b\right]$$
$$=\beta(1-\beta^N)\frac{d}{d\rho^b}\left[p-\frac{\mu(1-p)(1-\rho^b)}{\mu-\beta\mu+\beta\rho^b-1}\right]$$
$$=-\beta(1-\beta^N)\frac{\mu(1-\beta)(1-\mu)(1-p)}{(\mu-\beta\mu+\beta\rho^b-1)^2}<0.$$

□

Appendix E. Random return from autarky

An alternative possibility to model exclusion from insurance comes from the literature on sovereign default. With probability $0\leqslant\theta\le 1$, agents are offered an insurance contract again after defaulting as in Eaton and Gersovitz (1981) and Arellano (2008). With a constant probability to return each period, the average number of total exclusion periods (including teh current period) is $E(N)=1/\theta$. Incentive constraints and resource feasibility are unaffected by this change, and so is the definitions of lifetime utility $\tilde{w}(h)$. The value of autarky U^{Aut} now satisfies the following recursive equation:

$$U^{Aut}=(1-\beta)\{\mu u(y-l)+(1-\mu)[pu(y-l)+(1-p)u(y)]\}$$
$$+\beta\theta\tilde{w}+\beta(1-\theta)U^{Aut} \tag{49}$$

Individual rationality constraints of good and bad-signal agents now read

$$\underbrace{(1-\beta)\left[pu\left(y-l+\gamma_g\right)+(1-p)u(y-\delta)\right]+\beta(1-\theta)\tilde{w}}_{\tilde{w}(g)}\geqslant$$
$$\underbrace{(1-\beta)[pu(y-l)+(1-p)u(y)]+\beta(1-\theta)U^{Aut}}_{U^{Aut}(g)}, \tag{50}$$

and

$$\underbrace{(1-\beta)u(y-l+\gamma)+\beta(1-\theta)\tilde{w}}_{\tilde{w}(b)}\geqslant\underbrace{(1-\beta)u(y-l)+\beta(1-\theta)U^{Aut}}_{U^{Aut}(b)}. \tag{51}$$

In this economy, the no-trade condition is as follows.

Theorem 2 (*No Trade*). The autarky allocation $\{y, y-l, y-l\}$ is the only implementable allocation if and only if

$$\frac{u'(y-l)}{u'(y)}\leqslant\underbrace{T_s(p,\mu)T_r(p,\mu,\beta,\theta)}_{T_d(p,\mu,\beta,\theta)}$$

with $T_s(p,\mu)$ as the single-interaction pooled price ratio

$$T_s(p,\mu)=\frac{\mu+(1-\mu)p}{(1-\mu)p}=\frac{\mathbb{E}[P|P\geqslant p]}{1-\mathbb{E}[P|P\geqslant p]}\frac{1-p}{p}$$

and with $T_r(p,\mu,\beta,\theta)$ as a diminishing factor resulting from repeated interactions

$$T_r(p,\mu,\beta,\theta)=\frac{p[1-(1-\theta)\beta\mu]}{p+(1-p)\beta(1-\theta)\mu},\quad 0<T_r\leqslant 1$$

and $0<T_d(p,\mu,\beta,\theta)\leqslant T_s(p,\mu)$ as the repeated-interactions pooled price ratio for $0\leqslant\theta\le 1$.

Proof. The proof follows analogous steps as the one of Theorem 1.

Sufficiency We arrive at the following constraint on the FOC of the profit maximisation problem of an insurer (the right-hand side equality):

$$\frac{u'(y-l)}{u'(y)}\leqslant\frac{\mu+(1-\mu)p}{(1-\mu)p}\frac{p[1-\beta(1-\theta)\mu]}{p[1-\beta(1-\theta)\mu]+\beta(1-\theta)\mu}=\frac{u'(c_l)}{u'(c_{gy})},$$

Imposing the no-trade condition (the left hand side inequality) implies this can only be satisfied by $c_l=y-l$ and $c_{gy}=y$.

Necessity As in the proof of Theorem 1, we consider an allocation with trade, $h=\{h_{gy}=u(y-\delta), h_l=u(y-l+\gamma)\}$, which, if implementable, yields the following for the individual rationality constraints:

$$\tilde{w}(g)=U^{Aut}(g) \tag{52}$$
$$\tilde{w}(b)-U^{Aut}(b)=\epsilon>0. \tag{53}$$

Following similar steps as in the proof of Theorem 1, the slack ϵ in the bad-signal agents individual rationality constraint of bad-signal agents is

$$\epsilon=\frac{(1-\beta)}{1-\beta(1-\theta)\mu}u'(y-l)\gamma. \tag{54}$$

We get the counterpart to (34) from the binding individual rationality constraints of the good-signal agents:

$$0=\tilde{w}(g)-U^{Aut}(g)$$
$$=(1-\beta)[-(1-p)u'(y)\delta+pu'(y-l)\gamma]+\beta(1-\theta)\mu\epsilon.$$

Using (54), we get:

$$\frac{(1-\mu)(1-p)\delta}{[\mu+(1-\mu)p]\gamma} = \frac{u'(y-l)}{u'(y)}\frac{(1-\mu)p}{\mu+(1-\mu)p} \times \frac{p[1-\beta(1-\theta)\mu]}{p[1-\beta(1-\theta)\mu]+\beta(1-\theta)\mu}. \tag{55}$$

Assuming that the no-trade condition is violated yields

$$\frac{u'(y-l)}{u'(y)} > \frac{\mu+(1-\mu)p}{(1-\mu)p}\frac{p[1-\beta(1-\theta)\mu]}{p[1-\beta(1-\theta)\mu]+\beta(1-\theta)\mu},$$

combining with (55) implies that the allocation h generates positive resource savings. Hence, autarky cannot be the cost-efficient allocation, concluding the proof. □

Equivalence between deterministic and random exclusion Comparing Theorem 1 and Theorem 2, we can establish equivalence between the two formulations if the diminishing factors $T_r(p,\mu,\beta,\theta)$ and $T_r(p,\mu,\beta,N)$ are equal. The two factors for fixed-length and random-length exclusion are

$$T_r(p,\mu,\beta,N) = \left\{\frac{[1-\beta\mu-\beta^{N+1}(1-\mu)]p}{[1-\beta\mu-\beta^{N+1}(1-\mu)]p+\beta(1-\beta^N)\mu}\right\}$$

and

$$T_r(p,\mu,\beta,\theta) = \frac{p[1-\beta(1-\theta)\mu]}{p[1-\beta(1-\theta)\mu]+\beta(1-\theta)\mu}.$$

The equivalence between the two is that $T_r(p,\mu,\beta,N=0) = T_r(p,\mu,\beta,\theta=1) = 1$ and $\lim_{N\mapsto\infty} T_r(p,\mu,\beta,N) = T_r(p,\mu,\beta,\theta=0) = \frac{p(1-\beta\mu)}{p(1-\beta\mu)+\beta\mu}$. The explicit relationship between N and θ is:

$$\theta = \frac{\beta^N(1-\beta)}{1-\beta^{N+1}},$$

which is confirmed by direct evaluation of the two no-trade conditions.

Appendix F. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jpubeco.2023.104946.

References

Arellano, C., 2008. Default risk and income fluctuations in emerging economies. Am. Econ. Rev. 98, 690–712.
Bond, P., Krishnamurthy, A., 2004. Regulating exclusion from financial markets. Rev. Econ. Stud. 71, 681–707.
Braxton, J.C., Herkenhoff, K.F., Phillips, G.M., 2020. Can the unemployed borrow? Implications for public insurance. Working Paper 27026, National Bureau of Economic Research.
Chade, H., Schlee, E., 2012. Optimal insurance with adverse selection. Theoret. Econ. 7, 571–607.
Chetty, R., 2006. A general formula for the optimal level of social insurance. J. Public Econ. 90, 1879–1901.
Denderski, P., Stoltenberg, C.A., 2020. Risk sharing with private and public information. J. Econ. Theory 186, 104988.
Eaton, J., Gersovitz, M., 1981. Debt with potential repudiation: theoretical and empirical analysis. Rev. Econ. Stud. 48, 289–309.
Finkelstein, A., McGarry, K., Sufi, A., 2005. Dynamic inefficiencies in insurance markets: evidence from long-term care insurance. Am. Econ. Rev. 95, 224–228.
Gregory, V., Menzio, G., Wiczer, D.G., 2021. The alpha beta gamma of the labor market. Tech. rep., National Bureau of Economic Research.
Gruber, J., 1997. The consumption smoothing benefits of unemployment insurance. Am. Econ. Rev. 87, 192–205.
Hendren, N., 2013. Private information and insurance rejections. Econometrica 81, 1713–1762.
Hendren, N., 2017. Knowledge of future job loss and implications for unemployment insurance. Am. Econ. Rev. 107, 1778–1823.
Hirshleifer, J., 1971. The private and social value of information and the reward to inventive activity. Am. Econ. Rev. 61, 561–574.
Hobijn, B., Şahin, A., 2009. Job-finding and separation rates in the OECD. Econ. Lett. 104, 107–111.
Michelacci, C., Ruffo, H., 2015. Optimal life cycle unemployment insurance. Am. Econ. Rev. 105, 816–859.
Morchio, I., 2020. Work histories and lifetime unemployment. Int. Econ. Rev. 61, 321–350.
Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Q. J. Econ. 90, 629–649.
Schlee, E.E., 2001. The value of information in efficient risk-sharing arrangements. Am. Econ. Rev. 91, 509–524.
Schmitt-Grohé, S., Uribe, M., 2017. Open Economy Macroeconomics. Princeton University Press.
Shavell, S., 1979. On moral hazard and insurance. Q. J. Econ. 93, 541–562.
Stephens, M., 2004. Job loss expectations, realizations, and household consumption behavior. Rev. Econ. Stat. 86, 253–269.
Stiglitz, J.E., 1977. Monopoly, non-linear pricing and imperfect information: the insurance market. Rev. Econ. Stud. 44, 407–430.