Bundled Risks in Insurance Markets

Adam Solomon*

December 7, 2022

Abstract

All insurance products are bundles. For example, medical insurance bundles oncology and cardiology insurance. What determines which risks are insured together, and is this welfare enhancing? Theoretically, I introduce selection - costs correlated with willingness-to-pay - into a model of bundling under perfect competition. I show that the correlation structure between risk types, and whether the selection is adverse or advantageous, are central to firm incentives to bundle and the welfare consequences. With adverse selection, firm and social incentives are aligned, whereas with advantageous selection they are misaligned. Empirically, I focus on long-term care insurance and show the steep discounts given to couples (≈25%) can be explained by cost saving correlation. To test the theory more widely, I study the universe of risks in the HRS: long-term care, disability, annuities and life insurance. In the set of products considered, bundling occurs iff there is negative correlation between the risks, and steeper bundling discounts are given for more negatively correlated risks. I show that behavioral misperceptions of risks have a theoretically ambiguous effect on the propensity to bundle, but in the case of long-term care and mortality risk, miserpceptions increase the firm propensity to bundle and welfare when they do.

*Department of Economics, Massachusetts Institute of Technology. Email: adamsol@mit.edu. I thank Jim Poterba, Amy Finkelstein, Frank Schilbach, Abhijit Bannerjee, Stephen Morris, Johannes Spinnewijn, Daniel Gottlieb for helpful comments, the Shultz Fund for funding, and Jayson Roxas for excellent research assistance.

1 Introduction

All insurance products are bundles. Medical insurance bundles heart attack insurance with stroke insurance, inpatient coverage with outpatient coverage. General insurers will often give a discount for home and auto insurance purchased together, but not for home and renters insurance. Why are certain risks covered under a bundled policy and others not? And what are the welfare effects: Does bundling alleviate or compound the market failures due to that plague insurance markets generally?

The intuitive welfare argument in favor of bundling is that individual's should be more homogeneous with respect to a bundle of risks than any single risk. A extreme case is clear: with infinite i.i.d. risks, by the logic of the central limit theorem, all individuals are identical and there is no more selection problem. Intermediate cases are less clear: The degree to which bundling homogenizes risk depends on the correlation structure between risks: The more negatively correlated the risks the more homogeneous the bundled risk becomes. Moreover, bundling relatively independent risks reduces the probability of an individual being much higher risk than the average, but also the probability of an individual being much low risk than the average. The former is likely helpful for adverse selection, the latter unhelpful. And how these conclusions might change when the selection is advantageous and not adverse is not straightforward.

Relatedly, what are the private market incentives to introduce bundling, and how do they compare to the social value of bundling? I find that the answer to both and the wedge between the private and social incentives depends on two key factors: The correlation structure between the risks being bundled, and whether selection is adverse or advantageous. For the social planner, whether selection is adverse or advantageous, bundling does more good when types are less positively or more negatively correlated. This is because the market failure is caused by heterogeneity of risk types, and the degree to which bundling homogenizes risk type increases for more negatively correlated risks. In contrast, the private market will introduce bundling of adversely selected products when risk types are negatively correlated, but will only bundle advantageously selected products when they are positively correlated. In both cases firms chase the low cost individuals. Hence, broadly speaking, private and social incentives for bundling are aligned in the case of adverse selection but misaligned in the case of advantageous selection.

In this paper I develop the theory and empirics of product bundling in insurance markets. I study what the private and social incentives are to bundle insurance products, and how these incentives change based on the nature of selection and the correlation structure between the risks. I show how quantity discounts in insurance fit into the same theoretical framework by the same theory as bundling. I demonstrate that when contracts respond endogenously the pattern still holds: the degree to which adversely selected markets are improved by bundling increases in less positive or more negative correlation. I offer empirical evidence consistent with the theoretical model: The risks that are bundled together are more negatively correlated than those that are not. Finally, I study how behavioral frictions in the misperception of risk within each market and between markets impact the incentives to bundle.

I begin by extending the widely used Einav et al. (2010) framework, based on the celebrated Akerlof (1978) setup. There are two risks and a fixed insurance product for each. When costs increase in WTP, there is adverse selection, whereas when costs decrease with WTP there is advantageous selection.

I study the incentives for profit maximizing firms to introduce a bundling discount when previously each market was separately in equilibrium. When the markets are adversely selection firms will start bundling when the risks are negatively correlated. This is because when risks are negatively correlated the people who buy the bundle are relatively lower risk on each individual risk than those who just buy the single insurance product for that risk. On the other hand, when the markets are advantageously selected, insurers are chasing the high WTP, low cost individuals. Hence, here, it is positive correlation between WTP in each market means offering a bundle discount to attract those with high WTP for both products that is profitable.

I analyze the degree to which a mandate to bundle solves the selection problem. As correlation decreases, demand curves rotate counter-clockwise as the mass of individuals with very high or very low WTP for the bundle declines. Further, the average cost curve decreases (increases) when the markets are individually adversely (advantageously) selected. The quantity insured typically increases (decreases), mitigating the under-(over-) provision of insurance symptomatic of the market failure that results from the respective selection patterns.

I discuss how this theory of bundling one insurance product together with another also applies to quantity discounts in insurance. The latter can be thought of as bundling the first dollar of coverage with the second dollar of coverage at less than double the price. The theoretical predictions are analogous: In adversely selected markets firms should offer steeper quantity discounts when those who want to buy more insurance are relatively cheaper than those who want to buy less insurance.

I exhibit three complementary pieces of empirical evidence consistent with the theoretical predictions. First, in the context of long-term care, life insurance, disability insurance and annuities: I show that couples who buy long-term care insurance are offered a steep discount of up to 40%. For hybrid LTC and life insurance the couples discount is 5-10%, for life insurance 1-2%, zero for disability insurance, and for annuities no discount or anti-discount is offered. Examining risk data from couples in the Health and Retirement Study (HRS), I find patterns consistent with the discounts offered: Those who are married are substantially less likely to need long-term care, marginally less likely to die, and hence marginally more likely to be a costly annuitant, and no less likely (perhaps more) to make a disability claim. This confirms the theoretical prediction: discounts are offered in line with the risk reduction that those who would buy both policies (here, couples) offer over those who would only buy one policy (singles).

Second, emphasizing the finding of large couples discounts offered in LTC insurance, I complement the HRS analysis with data from a large database of policyholder claims. The data include over 60 million contract years of exposure and 600,000 claims. Controlling for all available price-relevant covariates, I find that being married strongly predicts a lower claim rate. From a base claim rate of 13%, those who are married or partnered have a 8% lower chance of claim, a reduction of almost 70%. This complements and makes robust the HRS finding in this important market.

Third, I analyze quantity discounts in different types of life insurance. The theory predicts quantity discounts should be steeper when those who buy a larger quantity of insurance are cheaper, per dollar insured, than those who buy less. I compare term life insurance - commonly held by the working age as insurance against lost income - with whole life insurance -commonly held by older people as a combination of insurance and tax-advantaged estate planning. The discounts offered in term life insurance are substantially larger than in whole life insurance: the discount offered per dollar insured when comparing a policy of face value \$ 100k with face value \$1m is 49% for term life versus 9% for whole life. I analyze a large database of insured lives in the US and find mortality patterns, per dollar of face value, consistent with this pattern. The slope of mortality with respect to policy size is very steep for term life insurance: those who buy a policy with face value 100k die twice as often as those who buy a policy with face value 1m, whereas for whole life insurance the difference is only 5%.

The analysis above assumes that individuals purchase insurance as if they perceive their risk accurately. There is lots of evidence that this is not the case. The effect that risk mispercetions may have on the public or private incentives to bundle is ambiguous. Misperceptions might be correlated within person - due to common optimism or pessimism or risk - or within a market - the high risk and the low risk might have systematically different misperceptions. In the context of individual LTC and mortality risk within the HRS, I find that misperceptions function as relatively i.i.d noise so that the bundled risk is more homogeneous when misperceived than when accurately perceived. This indicates that if risks are positively correlated misperceptions might decrease correlation and increase bundling, but is risks are negatively correlated then misperceptions might increase correlation and decrease bundling.

1.1 Literature Review

This paper contributes primarily to two literatures. The first is a well-established Industrial Organization literature on bundling products and add-on pricing. Relative to that literature, my main contribution is introducing selection (e.g. non-constant marginal costs) and studying the case of perfect competition instead of the oligopoly or monopoly.

In the IO context, with constant marginal costs, conditional on WTP being above the price, individual heterogeneity no longer matters. As such, when evaluating the profitability of introducing a bundle, for example, the relative densities of individuals in different regions of the WTP space matter, but conditional on being in a region (e.g. the 'buy bundle at price p_B') region, the distribution within that region doesn't matter.

In this paper the correlation between WTP for each product, and the correlation between WTP and cost of serving the individual will be key. In only a few IO papers is this question of correlation broached. Schmalensee (1984) and Adams and Yellen (1976) were the first to suggest, numerically and under assumptions of joint normality, that bundling was good for a monopolist's profit under independence or perfect negative dependence. Perfect negative dependance means there is no heterogeneity in consumers valuation for the bundled good, allowing a monopolist to capture the entire consumer surplus with a bundled good. Along the same lines, Bakos and Brynjolfsson (1999) note that when the number of goods approached infinity, if their valuations are independent then by the law of large numbers the heterogeneity in WTP will dissapear and a monopolist can capture all the consumer surplus.

Less parametrically, McAfee et al. (1989) give an indication of a sufficient condition for mixed bundling to dominate separate sales, which was later formalized by Chen and Riordan (2013) as a statistical property called 'right-tail increasingness'. This is a measure of correlation, although strictly stronger the version required in this paper (see definition 1). Recently, Haghpanah and Hartline (2021) show that pure bundling is optimal when the consumers with a high value for the bundle also have high values for the individual goods. This essentially says that pure bundling dominates mixed bundling when their is sufficiently high correlation in WTPs. This is because with high correlation in WTPs, there are few people attracted to only a single good over the bundle, and hence selling the single products separately is of little use.

Apart from correlation in values, the IO literature has studied other rationales for bundling. Zhou (2017) and Zhou (2021) study how bundling may soften competition between oligopolists selling differentiated products but that this conclusion depends on the number of firms. Nalebuff (2004) demonstrates that bundling can be an effective entry detterrance strategy, or a means to exploit market power in market A in a separate market B in which they have less power. Similarly, Hurkens et al. (2019) find that when two dual-product firms are ordered in terms of production efficiency in the same way for both both products, bundling can be a means for the more efficient firm to amplify its market power, or to leverage efficiency in one good into profits in the other market.

Finally, in recent work Nguyen (2018) structurally estimates a model of familial health insurance choice in Vietnam. That paper compares the current market in which individual family members can enroll separately with a forced bundling counterfactual in which the entire household must enroll, or none of the family can. Nguyen (2018) finds positive correlation of health types within a household and a positive welfare effect of forced bundling. The theory developed in this paper is complementary to such findings, and helps explain why forced bundling can increase welfare despite the market not offering it.

This paper speaks to the literature about contract design in insurance contexts. Multiple papers have documented firm incentives to change contract coverage so as to attract a better risk pool. Lavetti and Simon (2018) show that selective formulary design in Medicare Advantage plans screens out individuals with high costs in part A and B (e.g. by excluding drugs that correlate with high cost conditions). Shepard (2022) shows how some insurer networks exclude 'star hospitals' so as to remove the loyal users of these high cost hospitals from their risk pool. Cooper and Trivedi (2012) show how Medicare Advantage plans strategically included gym memberships in their plans to attract healthier and fitter enrollees.

The aforementioned papers can be thought of as examples of bundling an insurance contract with a second product or feature that is negatively correlated with original risk. The general theoretical framework introduced in this paper explains the market dynamics that the above papers have empirically documented. Moreover, it relates private incentives to bundle (as documented in the papers above) with planner incentives and indicates that forced bundling (i.e. not allowing the cream-skimming documented) is likely in the public interest. More broadly, while the above papers can be thought of the private market using a 'tag', as in Akerlof (1978), to screen individuals, correlation of insurance products can be thought of as both products simultaneously being tags for the other, and bundling as a means to take advantage of that.

2 Theory

2.1 Setup

I take a sufficient statistics approach that builds on the widely used Einav et al. (2010) (EFC) framework. A microfoundation is explored in appendix C. There are two risks and a fixed insurance contract for each. I will refer to the risks as risk 1 and 2 and the fixed insurance contracts as insurance contract 1 and 2 respectively. Each individual i is labeled by their willingness-to-pay (WTP) for each risk (w_1^i, w_2^i). I assume each WTP is bounded: $w_i \in [0, \bar{w}_i]$ for $i = 1, 2$ and write $\mathcal{W} = [0, \bar{w}_1] \times [0, \bar{w}_2]$ for the set of all permissible WTPs. Where there is no confusion, or I am referring to a generic individual, I will drop the superscripts for clarity. The insurance contracts may or may not be bundled. If they are bundled, then each individual's WTP for the bundled contract is $w_B = w_1 + w_2$. This is a strong assumption, which is microfounded in appendix C. For example, it is correct when utility has CARA functional form.

I stress that a bundled insurance product still offers the same state-contingent payoffs (except perhaps the premium) as buying the two products separately. The bundled product pays the same indemnity should risk 1 occur as does insurance contract 1, and similarly for risk 2. This is in contrast to a different policy that pays off only when some condition involving both risks is met, perhaps the sum of losses exceeds some deductible.

Each individual has heterogeneous costs. The cost of insuring an individual of WTP (w_1, w_2) is given by $\phi_1(w_1)$ if they buy contract 1, $\phi_2(w_2)$ if they buy contract 2, and the affine $\phi_B(w_B) = \phi(w_1 + w_2) = \phi(w_1) + \phi(w_2)$ if they buy the bundle contract. I assume that the functions ϕ_1, ϕ_2, ϕ_B are continuous and differentiable. Further assume that $\phi_1'(w_1) > 0, \phi_1''(w_1) > 0$ and similarly for ϕ_2 and hence for ϕ_B. This assumes adverse selection that weakly increases in type. I study the case of advantageous selection in section 2.3. But given the assumption of adverse selection and the monotonicity of WTP and cost, I will simply refer to the high risk as those with high WTP and high cost.

In the population there is some joint distribution of (w_1, w_2). For a population X we write F_X for the joint CDF function, F_1 for the CDF of the risk 1 marginal distribution and F_2 for the CDF of risk 2 marginal distribution. I will primarily be interested in populations X and Y that have the same marginal distributions F_1, F_2 but different joint distributions F_X, F_Y due to different correlation structures. This is explained in section 2.1.1.

The supply side of the market consists of infinitely many identical firms who compete on price. Each firm chooses a vector of prices to offer (p_1, p_2, p_B) where the bundle price will have to be lower than the sum of individual prices in order. In equilibrium all firms will offer the same prices at which total profits will be zero, and so we assume each firm receives a representative sample of consumers and hence makes zero profit individually.

Write the demand $D_m(p_m; p_{-m})$ in market $m = 1, 2, B$ as a function of price p_m holding fixed the other two prices $p_{-m} = \{1, 2, B\} \setminus \{m\}$. Demand for market m will be given by the mass of individuals who prefer purchasing product m at price p_m rather than either of the two other products at their prices or purchasing nothing. Formally, for product 1, under distribution of types X, demand will be

$$D_1(p_1; p_2, P_B) = \int_{\mathcal{W}} \mathbb{1}\left(WTP_1 - p_1 > \max\{WTP_2 - p_2, WTP_B - p_B, 0\}\right) dF_X(w).$$

And similarly for product 2 and B. Given prices the type-space $\mathcal{W}$ will be partitioned into the set of types who purchase only product 1, those who only purchase product 2, those that purchase neither, and those that purchase both, perhaps at a bundle discount. For clarity write these groups as, suppressing the dependence on prices, respectively

$$\mathcal{W}_1 = \{w \in \mathcal{W} : WTP_1 - p_1 > \max\{WTP_2 - p_2, WTP_B - p_B, 0\}\}$$

and similarly for $\mathcal{W}_2, \mathcal{W}_0$ and $\mathcal{W}_B$.

Write marginal costs in market 1, fixing prices in the other markets, as $MC_1(p_1; p_2, p_B) = \mathbb{E}\phi(w_{p_1;p_2,p_B})$ where $w_{p_1;p_2,p_B} = (w_1, w_2, w_B)$ solves $w_1 - p_1 = \max\{w_2 - p_2, w_B - p_B, 0\}$. This simply says that the marginal cost is the average of all the types marginal to product 1 at the given prices. The outside option to buying product 1 for these marginal types may be any of: buying product 2, buying both products or buying neither. Marginal costs in the other two markets are defined similarly.

Average cost in market 1 at prices $p_1; p_2, p_M$ is simply the average cost of all those buying product 1.

$$AC_1(p_1; p_2, p_B) = \mathbb{E}\left[\phi_1(w_1) \mid w \in \mathcal{W}_1\right]$$

and similarly for all those buying products 2 and B.

I assume that $D_m(p_m; p_{-m}) - AC_m(p_m; p_{-m})$ monotonically decreases in p_m for any p_{-m}. It follows that there is a unique equilibrium in each market m taking the other prices p_{-m} as given. This equilibrium price p_m^* solves $D_m(p_m^*; p_{-m}) - AC_m(p_m^*; p_{-m})$ I further assume, and empirically verify, that the general equilibrium of these markets, (p_1^*, p_2^*, p_M^*) is unique for each distribution on $\mathcal{W}$.

2.1.1 Correlation Orders

The comparative static of interest is the correlation of risk type 1 with risk type 2. To fix everything else means in this context to fix the marginal distributions in the population of risk type 1 and 2 and vary only the part of the joint distribution that encodes the correlation structure. (1)

Define $\Gamma(F_1, F_2)$ to be the set of joint distribution functions with marginals F_1 and F_2. Following Shaked and Shanthikumar (2007)(2) and Denuit et al. (2006), the correlation order that will be of

(1)This is coherent via Sklar's theorem.In particular, the density f_X of every joint distribution F_X can be factored as

$$f_X(w_1, w_2) = c\left(F_1(w_1), F_2(w_2) \cdot f_1(w_1) \cdot f_2(w_2)\right)$$

where F_1, F_2 are the CDFs of the marginals and f_1, f_2 the densities of the marginals.

(2)Here this ordering is called Positive Quadrant Dependance (PQD) primary interest is defined

Definition 1. *Suppose $X, Y \in \Gamma(F_1, F_2)$ and have CDFs F_X, F_Y respectively. I say that X is be less correlated than Y or that X precedes Y in the correlation order, written as $X \precsim Y$ when*

$$X \precsim Y \iff F_X(w_1, w_2) \leq F_Y(w_1, w_2), \quad \textit{for all } (w_1, w_2) \in \mathcal{W}.$$

Intuitively, this says that w_1, w_2 are more likely to both be small under the more correlated Y than under X. This notion of ordering of correlation is often called 'Positive Quadrant Dependance (PQD)' in the literature.[3]. This is equivalent to writing

$$X \precsim Y \iff (1 - F_X(w_1, w_2)) \leq (1 - F_Y(w_1, w_2)), \quad \text{for all } (w_1, w_2) \in \mathcal{W}.$$

This says that w_1, w_2 are also more likely to both be large under Y than X. This comports with the usual intuitive notion of correlation: knowing w_1 is small/large means w_2 is more likely small/large the higher the correlation is, and conversely for negative correlation it becomes more likely that when w_1 is large w_2 is small and vice versa.

This correlation order generalizes the familiar linear correlation coefficient. In particular, if $X \precsim Y$ then $\rho(X) \leq \rho(Y)$. If the random variables were discrete a similar implication holds for both of Spearman and Kendall's rank correlation coefficients.

Example 1. *As an example, if both X and Y are jointly normally distributed with identical marginals, then this ordering aligns with the intuitive ordering based on the correlation coefficient. That is, $X \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_X)$ and $Y \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_Y)$ then we have*

$$X \precsim Y \iff \rho_X \leq \rho_Y.$$

Note that this definition can (and will) be applied to other distributions related to X, Y. For example, we could define the order $\succsim^{\boldsymbol{w}}$ to mean that Definition 1 holds with the truncated distributions $F_{X>\underline{\boldsymbol{w}}}, F_{Y>\underline{\boldsymbol{w}}}$ instead of the unconditional distributions F_X, F_Y.

Definition 2. *Suppose $X, Y \in \Gamma(F_1, F_2)$ and have CDFs F_X, F_Y respectively. I say that X is be less correlated than Y or that X precedes Y in the truncated correlation order at $\boldsymbol{w} = (\underline{w}_1, \underline{w}_2)$, written as $Y \succsim^{\boldsymbol{w}} X$ when*

$$Y \succsim^{\boldsymbol{w}} X \iff F_{X_1>w_1} \leq F_{Y_1>w_1} \textit{ and } \iff F_{X_2>w_2} \leq F_{Y_2>w_2}, \quad \textit{for all } (w_1, w_2) \in \mathcal{W}.$$

There isn't always an exact relationship between the correlation of the non-truncated distributions X, Y and that of the truncated distributions. But typically the correlation order of the underlying distrubtions is retained under truncation. For example, (e.g. see Kotz (2000)):

[3] See for example, Shaked and Shanthikumar (2007) and Denuit et al. (2006)

Example 2. *If both X and Y are jointly normally distributed with identical marginals, $X \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_X)$ and $Y \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho_Y)$ and we write X_w, Y_w for the (singly) truncated distributions at w then we have*

$$\rho_X \leq \rho_Y \iff \rho_{X_w} \leq \rho_{Y_w}.$$

Finally, to obtain sharper results I will sometimes make assumptions about how the marginal distributions relate to the joint distribution. One easy interpretable assumption is below, but in appendix **??** I present many alternative copula assumptions that suffice as well.

Definition 3. *Define $X \in \Gamma(F_1, F_2)$ to have Farlie-Gumbel-Morgenstern (FGM) form when the CDF of the joint density function has the form*

$$F_X(w_1, w_2) = F_1(w_1) \cdot F_2(w_2) \cdot [1 + \theta\,(1 - F_1(w_1))\,(1 - F_2(w_2))]$$

with $\theta \in [-1, 1]$ governing the level of dependence between the marginals.

Assumption 1. *Assume that X and Y either:*

- *Satisfy definition 3, or*
- *Are jointly normal.*

2.2 Bundling Incentives and Impacts

2.2.1 Incentives to offer a small bundled discount

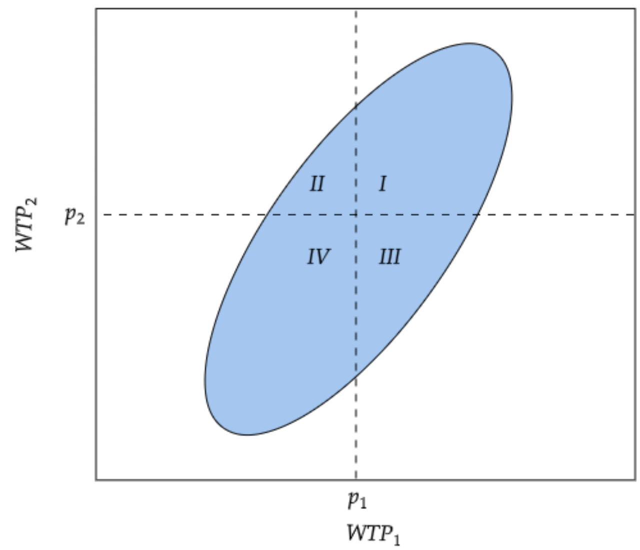
Begin with both risk markets individually in equilibrium. Because the marginals are assumed the same, when there is no bundling the prices and quantities in the seperate product equilibria are independent of the correlation structure.

Fact 1. *For any two distributions with the same marginals, $X, Y \in \Gamma(F_1, F_2)$, if there is no bundling, the equilibrium prices and quantities in each market are the same: $q_1^X = q_1^Y \equiv \bar{q_1}, \quad q_2^X = q_2^Y \equiv \bar{q_2}, \quad p_1^X = p_1^Y \equiv \bar{p_1}, \quad p_2^X = p_2^Y \equiv \bar{p_2}$.*

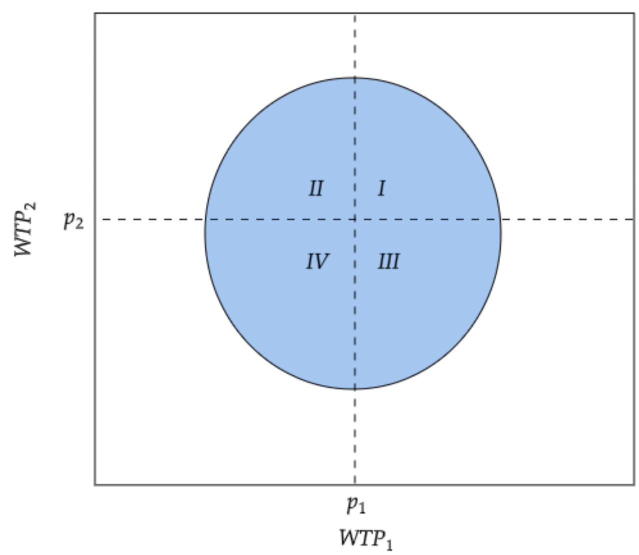
At these prices both contracts separately break even.

A graphical intuition is helpful. With adverse selection, I will interchangably refer to risk, WTP and cost. Consider figures 1a, 1b and 1c. Recall that the marginal distribution of each risk is fixed, and only the correlation structure between the risks is varied. Figure 1a illustrates strong positive correlation between the two WTP/risk/cost types, figure 1b exhibits independence between the two markets, whereas 1c shows strong negative correlation.

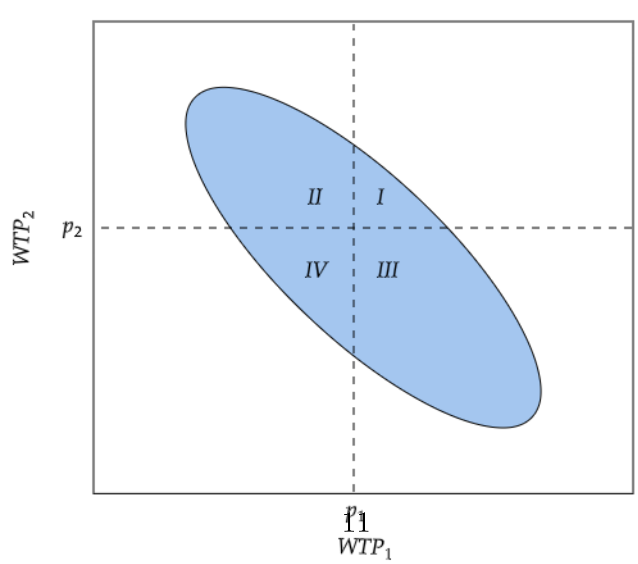
Fact 1 means that the seperate insurance markets break even at the same prices p_1, p_2 no matter which correlation structure we are in. Graphically it means the average cost of insuring groups I and III exactly equals the price p_1, and the average cost of insurance groups I and II exactly equals the price p_2. This is true for all three correlation structures.



(a) **Positive Correlation Between WTP/Costs**



(b) **Independence Between WTP/Costs**



(c) **Negative Correlation Between WTP/Costs**

The key question is: When would it be profitable for an insurer to offer the bundled contract at a price of $p_B = \bar{p}_1 + \bar{p} - \epsilon$ for some small $\epsilon > 0$?

First, when an insurer offers an ϵ bundle discount, they attract to their bundled contract exactly the set of people who previously chose to buy both products separately. Recalling that those separate markets previously broke even, the profitability of the small discount will hinge on whether the people who chose to buy both products 1 and 2 are cheaper in terms of risk 1 (risk 2) than those who only bought contract 1 (contract 2) before.

Graphically this means: Does group I have lower cost than group III (in terms of risk 1) and lower cost than group II (in terms of risk 2)? If group I is lower cost, then the fact that groups I and III together broke even in the risk 1 market means that now attracting just group I will be profitable in market 1, and similarly in market 2.

Under positive correlation, as in figure 1a, group I is more expensive than group III (II) in terms of risk 1 (2). Bundling is therefore a bad idea, as the price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ will not cover the costs of group I. Hence bundling will result in a loss, and there is no incentive to do it.

Under independence, as in figure 1b, groups I and III (II) are the same cost in terms of risk 1 (2) and hence offering a bundling discount will result in (arbitrarily close to) zero profit.

Conversely, under negative correlation, as in figure 1c, group I is less expensive than group III (II) in terms of risk 1 (2). Bundling is therefore a good idea, as the price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ will more than cover the costs of group I. Hence bundling will result in a profit, and firms will have an incentive to do it.

This is formalized in the following characterization:

Proposition 1. *Suppose $X, Y, Z^{\perp} \in \Gamma(F_1, F_2)$ with $Y \succsim^{\bar{\boldsymbol{p}}} Y$. Denote the profit earned per person on a bundle contract offered at price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ by π^{ϵ}. We have the following:*

1. $\pi^{\epsilon}_{Z^{\perp}} = 0$
2. *If $Y \succsim^{\bar{\boldsymbol{p}}} X \succsim^{\bar{\boldsymbol{p}}} Z^{\perp}$ then $\pi^{\epsilon}_Y \leq \pi^{\epsilon}_X \leq \pi^{\epsilon}_{Z^{\perp}} = 0$, and conversely if $Z^{\perp} \succsim^{\bar{\boldsymbol{p}}} X \succsim^{\bar{\boldsymbol{p}}} Y$ then $\pi^{\epsilon}_Y \geq \pi^{\epsilon}_X \geq \pi^{\epsilon}_{Z^{\perp}} = 0$*

In line with Figures 1a and 1c, this says that the profitability of introducing a small bundling discount increases the less correlated (or more negatively correlated) the risks are. When the risks are independent, as in figure 1b, which I label $Z^{\perp}$, the profitability of introducing a small bundling discount is zero. This is intuitive, as when a bundling discount is offered those attracted to the bundle are those with high w_1 and high w_2 whereas those who don't buy have only high w_1 or high w_2 or neither. In the case of independence, of those who previously bought product 1, those that now buy the bundle (i.e. those who have w_2 large) have the same expected cost as those who stick with only product 1 (those who have w_2 small) as conditioning on w_2 being large or small doesn't change the expectation of risk 1 type, by independence. So those that buy the bundle look the same on risk 1 as those who only buy product 1, and the same on risk 2 as those who only buy product 2. Since those markets were at equilibrium previously, it follows that the bundle also just breaks even, hence $\pi^{\epsilon}_{Z^{\perp}} = 0$.

Proposition 1 demonstrates when there is an incentive for a firm to deviate from seperate equilibria and begin offering a bundle. It suggests that the more negative the correlation, the lower the bundle price, and the higher the price in the individual markets. The following proposition formalizes this.

Proposition 2. *Suppose $X, Y \in \Gamma(F_1, F_2)$ with $Y \succsim^{\overline{p}} X$. Suppose economy Y is at a mixed bundling equilibrium (zero profits earned) with prices p_1, p_2, p_B. At those same prices in economy X:*

1. *The bundle would be profitable*

2. *The seperate products would make a loss.*

This strongly suggests that in equilibrium, as correlation decreases, the bundle price declines and the single product price increases. Indeed in all empirical simulations this is the case.

2.2.2 Managed Competition - pure bundling vs no bundling

When the regulator or market designer does not allow single insurance products to be sold, each firm can only choose a price p_B at which they sell the bundled product, and each consumer chooses between the bundle or nothing. I refer to this situation as 'pure bundling'. The following proposition characterizes how the pure bundling equilibrium changes with the correlation structure.

Proposition 3. *Suppose $X, Y \in \Gamma(F_1, F_2)$ with $Y \succsim X$. The following comparative statics hold:*

- *The average cost curve is everywhere lower under less correlated distributions: $AC_X(p) \leq AC_Y(p)$ for all p.*

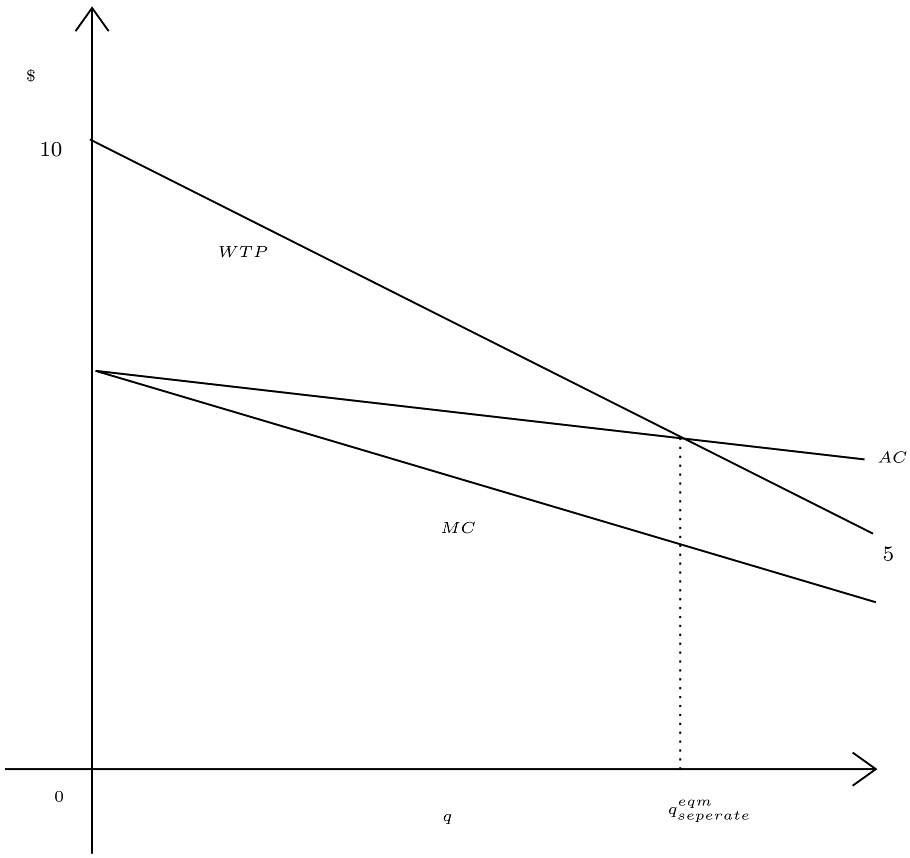
- *The equilibrium bundle price increases in correlation: $p_B^X \leq p_B^Y$.*

- *Under assumption 1, for 'large' markets, i.e. if $q_B^Y \geq \underline{q}$ for some $\underline{q}$ then the equilibrium quantity insured increases under the less correlated X: $q_B^X \geq q_B^Y$.*

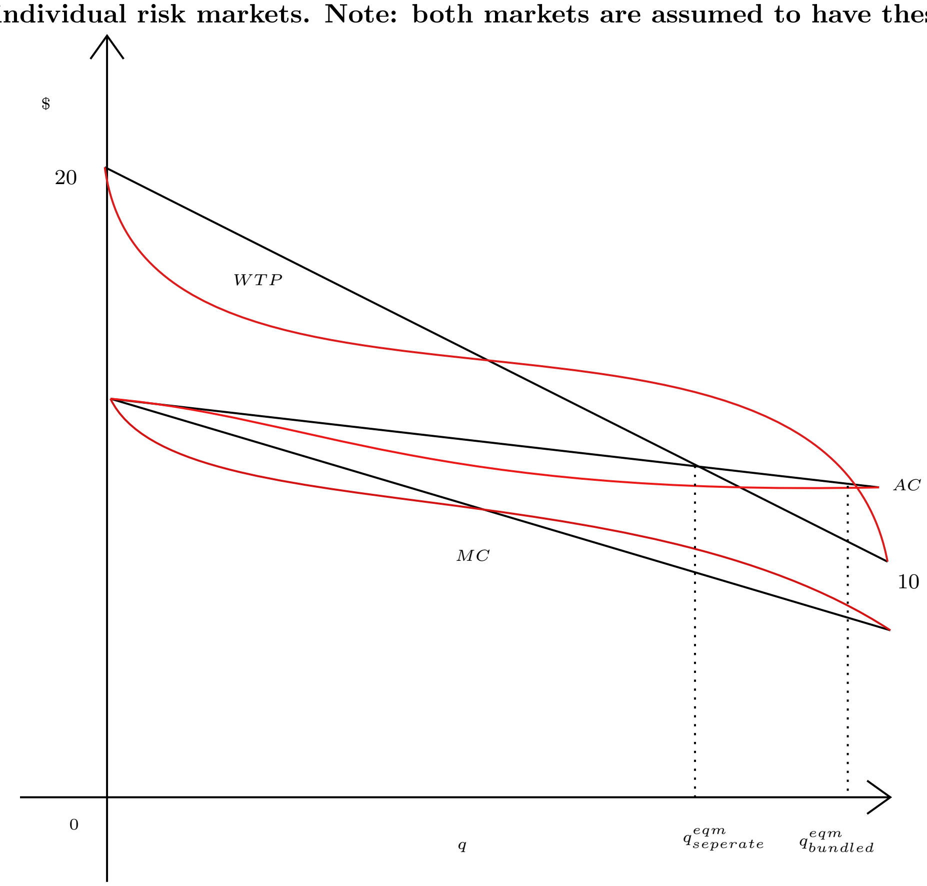
This proposition says that the average cost and equilibrium price always declines under less (or more negatively) correlated risk distributions. The equilibrium quantity usually increases, a sufficient condition for which is that under the more correlated Y, more than $\underline{q}$ of the market is insured. $\underline{q}$ is distribution specific, but is defined in the proof and typically not that large. For instance, if X, Y are jointly normal, $\underline{q} = \frac{1}{2}$.

The proposition is illustrated in an example in figures 2a and 2b. Suppose *both* individual risk markets have WTP uniformly distributed on $[5, 10]$, and standard EFC declining cost curves, as illustrated in figure 2a.

In the case that the individual risks are perfectly correlated the WTP in the bundled market will be uniformly distributed on $[10, 20]$ and the cost curves will also double. The equilibrium quantity will be the same as in the individual markets, and forced bundling will do nothing. These are the black curves in 2b.



(a) **The individual risk markets. Note: both markets are assumed to have these curves.**



(b) **The bundled market under two different correlation structures: Perfect correlation (black) vs less than perfect correlation (red).**

Figure 2: EFC style graphs for individual and bundled risk markets.

However, when the individual risks are less than perfectly correlated, there will be fewer people with very high WTP or very low WTP for the bundle, and more near the middle. This is shown by the rotation of the *red* WTP curve in figure 2b. Similarly the marginal cost (which is 1-1 with WTP) curve will rotate. The proposition says that the average cost will everywhere decline, which is why the red AC curve is below the black AC curve in figure 2b. Hence, when we are in the part of the WTP curve where less correlation means higher WTP, the equilibrium quantity insured will increase under the red curves relative to the individual markets or the perfectly correlated black curves.

Intuitively this proposition says that when correlation declines, or becomes more negative, the adverse selection problem becomes less severe. This is because the WTP for the bundle is a mixture of WTP for risk 1 and risk 2, and as those become less correlated, everyone's costs become flatter and more homogenous, lessening the adverse selection problem.

The planner wants quantity insured to increase. Under less correlated risks, if only bundling is allowed the quantity will in fact increase. Similarly, proposition 1 says that for sufficiently negatively correlated risks, the private market will move in the direction of bundling anyway. In this sense the planner's objectives are aligned with and achieved by the private market: when it is socially useful to bundle (negative correlation) the private market does so. This will not be the case under advantageous selection, as studied in the next section.

2.3 Advantageous selection

The only change needed to study advantageous selection is that instead of the cost functions increasing in WTP, they decrease: $\phi_1', \phi_2', \phi_B' < 0$ as well as assuming, in a symmetrical sense, assuming submodularity of the cost functions instead of supermodularity. Below are analogues to propositions 1 and 3 that demonstrate the similarities and differences between adverse and advantageous selection. However, the market failiure that occurs under advantageous selection is distinct from under adverse selection. Under advantageous selection the equilibrium quantity is *greater* than the socially optimal quantity. This is because individuals with $MC > WTP$ end up insured because they are subsidized by the cheaper people that have higher WTP.

The planner wants the equilibrium quantity to move downwards. When the planner mandates pure bundling: we will see that this movement of equilibrium quantity toward socially optimal quantity occurs, as in the case of adverse selection, under less positive or more negative correlation. This is for the same reason: as types become less correlated, their sum become more homogenous, viciating the advantageous (or adverse) selection problem that is caused by heterogeneity in WTP and cost types.

However, the private market will not have the same incentives. Under adverse selection, the private market wanted to introduce bundling under negative correlation. Under advantageous selection the private market will introduce bundling only when the type correlation is positive. This is because when types are positively correlated and costs are advantageously selected, those that buy the bundle are likely high on both types, which means low on all costs, which is very attractive to firms. These facts are formalized below.

Proposition 4. *Suppose $X, Y, Z^{\perp} \in \Gamma(F_1, F_2)$ with $X \precsim Y$. Denote the profit earned per person on a bundle contract offered at price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ by π^{ϵ}. We have the following:*

1. $\pi^{\epsilon}_{Z^{\perp}} = 0$
2. *If $Y \succsim^{\bar{\boldsymbol{p}}} X \succsim^{\bar{\boldsymbol{p}}} Z^{\perp}$ then $\pi^{\epsilon}_{Y} \geq \pi^{\epsilon}_{X} \geq \pi^{\epsilon}_{Z^{\perp}} = 0$, and conversely if $Z^{\perp} \succsim^{\bar{\boldsymbol{p}}} X \succsim^{\bar{\boldsymbol{p}}} Y$ then $\pi^{\epsilon}_{Y} \leq \pi^{\epsilon}_{X} \leq \pi^{\epsilon}_{Z^{\perp}} = 0$*

And in the case of mandated pure bundling, the analogue is:

Proposition 5. *Suppose $X, Y \in \Gamma(F_1, F_2)$ with $X \precsim Y$. The following comparative statics hold:*

- *The average cost curve is everywhere higher under less correlated distributions: $AC_X(p) \geq AC_Y(p)$ for all p.*
- *The equilibrium bundle price decreases in correlation: $p_B^X \geq p_B^Y$.*
- *Under assumption 1, for 'small' markets, i.e. if $q_B^Y \leq \underline{q}$ for some $\underline{q}$ then the equilibrium quantity insured decreases under the less correlated X: $q_B^X \leq q_B^Y$.*

Bundling can still solve the problem of advantageous selection when types are more and more negatively correlated by raising the equilibrium price and quantity. However, the private market has no incentive ti introduce bundling when types are negatively correlated. In contrast, the more positively correlated types are the greater are private incentives for bundling while the lesser are social incentives for bundling. This disjunction

2.4 Discussion and relation to 'standard' bundling

In this section I discuss how the results above relate to the typical patterns observed in the standard IO bundling settings.

In that literature more negative correlation is good for mixed bundling over separate sales in that the price discrimination it allows for has more bite. With enough positive correlation, there is less and less to be gained by bundling, as there are very few people who would buy the bundle without buying both products . With negative correlation, separating off those that have moderate values for both goods through a bundle leaves a larger group of consumers who have high value for one good and low for the other, and whose high valuations can then be more profitably exploited by raising the price than if the moderate WTP in both were mixed in, in which case raising the price would have a larger marginal (quantity) effect.[(4)]

Similarly, when considering pure bundling, negative correlation means a flatter demand curve, so a lower grand bundle price will have a larger marginal quantity effect. In the extreme case pointed

[(4)]The ratio monotonicity results of recent papers such as Haghpanah and Hartline (2021), Yang (2021) and Ghili (2022) are of the same flavor: those that have high valuations for the grand bundle must find the products in the grand bundle substitutable, so that conditional on valueing one product highly they do not value the other.

out by Adams and Yellen (1976), perfect negative correlation allows a grand bundle to extract all of the surplus.

The results here for the planner's choice to mandate bundling have a similar flavor. Negative correlation makes mandated pure bundling better for overcoming the selection problem as it homogenizes costs so that the wedge between average and marginal cost (the root of the market failiure) is made smaller. Whereas homogeneity is attractive to a monopolist in an IO setting as it reduces the tradeoff between the marginal and infra-marginal price incentives (i.e. it 'flattens' the demand curve, cost is already assumed constant) here the value to the planner is from a flattened cost curve.

Contrastingly, firm incentives in this competitive model are distinct from the monopolist. A monopolist internalizes the externality that their bundling choices has on single markets, competitive firms do not: There is an incentive cream skim through the bundle and shift losses onto other firms who offer the single products. The desirability of bundling is increased whenever those who have high WTP for both products are lower cost than those who have low WTP for one product and high for the other. Under adverse selection this means negative correlation, whereas under advantageous selection this means positive correlation.

What matters here is the slope of the cost curve, not the slope of the difference between demand and cost. The latter is always decreasing es no matter the selection structure (advantageous or adverse or neither) and indeed for the monopolist as well.

2.5 Quantity discounts as bundling

Closely related bundling different products together is the common practice of quantity discounts. Instead of bundling two seperate products together, a quantity discount can be thought of as bundling two units of the same product together. This fits easily into the notation and modeling introduced thus far and it will provide the motivation for some of the empirical tests.

Suppose a company is selling insurance with a variable amount of coverage. A leading example would be life insurance with different face values. Define the seperate products as 'life insurance with a \$100,000 face value' (the WTP for which are w_1 and w_2) and then the bundled product as 'life insurance with a \$200,000 face value' (with WTP $w_B = w_1 + w_2$). Typically w_1 and w_2 are different, as you can think of these as WTP for the first \$100,000 of life insurance and the second \$100,000 life insurance. In that case we would expect $w_1 \geq w_2$. But things will be easier if we enforce symmetry, and so I will randomly assign half of the consumers to have their WTP for the first \$100,000 denoted by w_1 and WTP for the second \$100,000 by w_2 (hence $w_1 \geq w_2$) and for the other half of the consumers to have their WTP for the first \$100,000 denoted by w_2 and the second \$100,000 by w_2 (hence $w_2 \geq w_1$). This means that the marginal distributions of W_1 and W_2 are the same as they are both an equal mix of people buying their first \$100,000 of insurance and people buying their second \$100,000. Moreover, the seperate market equilibrium prices and quantities will be equal: $\bar{q}_1 = \bar{q}_2$ and $\bar{p}_1 = \bar{p}_2$.

Throughout I will assume adverse selection for ease of exposition and because the empirical examples conform to this. Positive correlation would mean that those who most want the first

\$100,000 of life insurance most want the second \$100,000, and conversely for negative correlation. One might expect that if demand for life insurance is mostly driven by risk then we would see positive correlation. On the other hand, negative correlation might be explained by those who want \$200,000 of insurance being high wealth individuals whose mortality risk is not that high. Hence those that buy both might be lower cost, per \$100,000, than those that buy just one. This would generate negative correlation.

With this labelling, the earlier propositions still apply. The incentives to bundle will be positive for negative correlation, zero for indepdence and negative for positive correlation. And the degree to which forced bundling will expand the market and fix the adverse selection increases the less positive or more negative the correlation is. The empirical tests in section 3.3 give evidence for these propositions in the context of quantity discounts for different types of life insurance.

2.6 Endogenous contracts

Until this point I have assumed that the insurance contracts under consideration are fixed, in particular that they do not change as the correlation structure changes, beyond being bundled or not. In this section I show that the qualitative conclusions of the prior sections hold in a simplified model where the contracts offered endogenously respond to the correlation structure. The model closely follows Crocker and Snow (2011). Here, however, I focus on the (locally) competitive equilibrium which will not allow cross-subsidization between types.

There is a unit mass of individuals who begin with wealth w. There are two states of the world, L(oss) and NL and two types, $t = H, L$ who have different probabilities of experiencing a loss: $p^H > p^L$. Conditional on a loss occuring, there are two possible (exhaustive and exclusive) versions of this loss, which I refer to as *perils*. For simplicity assume both perils lead to a loss of the same size l. Write θ^H and θ^L for the likelihood that peril 1 occurs, conditional on a loss, for the H and L types respectively. Peril 2 occurs with the complementary probability. Without loss of generality assume $\theta^H \geq \theta^L$.

An insurance contract specifies a premium to be paid α to be paid in all states of the world in exchange for indemnities ι_1, ι_2 to be paid in case a loss occurs due to peril 1 or 2 respectively. On net, an insurance contract promises consumption of $c_0 = W - \alpha$ in the no-loss state, and consumption of $c_1 = w - \alpha - l + \iota_1$ if the loss occurs due to peril 1, with c_2 defined similarly.

If an individual of type $t = H, L$ purchases an insurance contract that promises consumption vector $\boldsymbol{c} = (c_0, c_1, c_2)$ their expected utility is given by

$$V^i(\boldsymbol{c}) = (1 - p^t)u(c_0) + p^t\theta^t u(c_1) + p^t(1 - \theta^t)u(c_2).$$

The flow utility function $u(\cdot)$ is assumed twice continuously differentiable and weakly concave. When evaluating welfare, I do so according to, where α is the pareto weight:

$$W = \alpha V^H \left(\boldsymbol{c}^H\right) + (1 - \alpha)V^L \left(\boldsymbol{c}^L\right).$$

As in section 2.2 I am interested in two questions: when will bundling occur and when will it be welfare improving. To that end I define:

Definition 4. *If a contract promises* $c_1 = c_2$ *then say that contract is unbundled. If* $c_1 \neq c_2$ *then that contract is bundled.*

A bundled contract means there is some cross-subsidization across perils so that the contract only breaks even when both perils are insured. Conversely an unbundled contract doesn't differientiate between perils and hence breaks even no matter which of the perils occurs. In this sense it could be unbundled to be peril specific while remaining weakly profitable. This conceptually mirrors the choice to be made by an insurer in section 2.2.1.

Definition 5. *Define the degree of correlation between types as* $\rho = \theta_L/\theta_H$*.*

When $\rho = 1$ types are 'perfectly correlated' in the sense that conditional on a loss, it is equally likely to be from the same peril for each type. As ρ decreases to zero the peril causing the loss is more likely to be different for each type. Nevertheless, as ρ changes the expected loss doesn't change, nor does the overall probability of some loss occurring, simply which of the two perils.

The first proposition speaks to when policies will be bundled in equilibrium. This is essentially Theorem 1 of Crocker and Snow (2011).

Proposition 6. *The equilibrium will feature only unbundled contracts when* $\rho = 1$*. Whereas for* $\rho < 1$ *the low types will receive a bundled contract.*

When $\rho = 1$, that is, when there is 'perfect correlation' between the two types, there is no benefit to bundling the perils and cross-subsidizing between them. On the other hand, when $\rho < 1$, and one peril is more likely than the other to affect the low type, then a bundled contract can be more attractive to the low types. Specifically, a contract that slightly overpays against the peril comparatively more likely to afflict the low types, and underpays against the peril comparatively more likely to afflict the high types will impose a second order utility cost on the low types (as marginal utility will be slightly unequal in each of the peril states) but a first order gain as the IC constraint that is holding the low types to partial insurance will be loosened.

This logic makes the next proposition clear:

Proposition 7. *Welfare is minimized at* $\rho = 1$ *and monotonically increases as* ρ *falls to zero.*

As ρ decreases companies can offer a contract to the low types that redistribute from H's more likely peril to L's. Low types will only buy the new contract if their welfare improves, and H types are getting full insurance throughout. Hence as soon as the equilibrium contracts change due to a ρ change, welfare must increase, analogously to the logic explained by Crocker and Snow (2011). As ρ declines, the 'comparative advantage' enjoyed by L at one peril vs the other increases, and therefore does the screening benefits possible by over-paying in that peril and underpaying in the other.

Overall, the qualitative message from this section coheres with the prior sections. In a setting with adverse selection, when the less correlation there is between types in what type of loss will occur, fixing the overall probability of *some* loss occurring, the more easily can the types be screened apart and the incentive constraint holding the low types to partial insurance be relaxed.

In section 2.2, bundling was introduced when the types with WTP for one contract had relatively low WTP for the other. Similarly here, the degree to which the WTPs for the different types of losses differ encourages bundling to the benefit of the firms and the planner.

3 Empirical Tests of the Theory

In this section I test the predictions of the theory in two different settings. The prediction being tested is that, in adversely selected markets, bundling discounts should be offered if the risk types are negatively correlated and not offered if positively correlated.

To speak to this prediction I need two ingredients: data on discounts offered in actual insurance markets and a measure of the correlation between risk types for each of products in the bundle.

First, I will study discounts given to spouses or partners when they buy long-term care, life insurance or annuities together versus seperately. In this case, the individual risks are the risk for spouse 1 needing insurance (e.g. long-term care insurance) and the risk for spouse 2 needing insurance. I will use representative population data from the Health and Retirement Study (HRS) and study how the risk types/WTP of each member of a couple compare with single individuals (for whom the WTP for 'spouse's insurance' is zero). I find that the correlation between spouses risk type for long-term care are strongly negatively correlated and this is consistent with a large couples discount offered in the private market of up to 40%. I find that there is a small negative correlation in mortality risk type between spouses, consistent with a small couples discount when purchasing life insurance, and a zero or small anti-discount when purchasing annuities.

Second, in view of the equivalence established in section 2.5, I will compare the quantity discounts offered in term life insurance with whole/universal life insurance. The quantity discounts will be measured in terms of how much it costs per 100k face value of life insurance when different total face values are purchased. For concreteness, suppose that the bundled product is a life insurance product with face value of 200k. The products being bundled can be thought of as 'the first 100k face value of life insurance' and the 'the second 100k of life insurance'. The difference between the price of a 200k face value policy and twice the price of a 100k face value policy is the quantity discount. I show these discounts are substantially higher for term life insurance than for whole/universal life. Correspondingly, I use a large life insurance policyholder mortality experience study to show that indeed the negative correlation is much higher for term life insurance holders than for whole/universal life insurance holders. Negative correlation means that the risk *per 100k exposed*, is lower for those that buy 200k face value than those who only buy 100k.

The two settings are complementary. In both cases the premia and discounts will be measured in the actual US market. But in the first setting risk types will be measured in a population survey, while in the second setting risk types will be measured in a study that only includes those that have insurance. The drawback of using population risk data is that it will leave open the possibility that risk propensity doesn't inform insurance purchase, and so even if the former is negatively correlated the latter might not be. The second setting addresses this by using data on actual insurance holders and demonstrating the predicted correlation is present there.

3.1 Discounts for Spouses or Partners

When spouses or partners purchase insurance from the same company they receive a discount for some products but not for others. In this section I study four products and the spousal[(5)] discount offered in each: Long-term care insurance, life insurance, hybrid LTC and life insurance, and life annuities. In all cases only individual (i.e. non-group) products are included.

First I measure the average discount offered in each product when spouses apply and are accepted together. The data for life insurance comes from ?? (COM), which collects historical quotes for over 100 life insurance issuers. The data for LTC policies come from LTC Quote Plus (StrateCision (2022b)), which quotes from 9 different issuers. The data for hybrid life-LTC policies come from ComboCompare (StrateCision (2022a)) which quotes from 6 different hybrid products. Annuity prices come from the Annuity Shopper (ImmediateAnnuities.com (ImmediateAnnuities.com)). Disability (non-group) discounts come directly from underwriting guides available publicly.

A full specification of which combination of policy parameters were used to generate the quotes is in appendix. The general idea is straightforward for LTC, life and hybrid products. I simply compare the price of two identical policies when bought as a couple versus when bought individually. For example, two 10 year term life insurance policies from Ameritas Life Insurance Corp, each with a face value of $1 million for a 40 year old man and a 40 year old women in the highest rating category cost $370 and $330 per month when purchased separately, but $660 when purchased together, a discount of $40 or 6%. In the case of disability insurance, I was unable to obtain direct prices, but did find the underwriting guides for 12 different products. None offered couples discounts.

The only exception are annuity prices. To measure a discount (or anti-discount) offered to couples I construct a payoff equivalent version of a joint life annuity out of two single life annuities, and compare the prices. In particular, for each company at each time period, I take the payout offered to a joint and survivor annuity bought at a cost of $100,000 that has payments that reduce to 50% after one spouse dies. For example, for a joint annuity for a couple consisting of a 65 year old man and 60 year old woman, Nationwide promised payouts of $428 per month reducing to $214 after one spouse days. This could alternatively be constructed by spending $44,958 on a 65 year old male single life annuity from Nationwide (the quoted payout rate is $476 per month per $100,000) and $53,768 on a 60 year old female single life annuity (the quoted rate is $398 per month per $100,000). This will generate $428 in income per month while both annuitants are alive, and $214 after one dies. But this synthetic version of the joint annuity costs only $98,727. Hence the joint product is sold at an *anti-discount* of $1,273, or 1.275%. That final number, -1.275%, is what I record for each company x year x joint policy combination.

The resulting average and range of discounts for each product are in table 1.

The theory predicts that a larger bundling discount should be offered to married couples when the WTP/risk of spouses are negatively correlated, when we consider single individuals to have a

[(5)] For brevity I will omit the qualifier 'and partnership'.

Product	LTC	Hybrid LTC/LI	Life Insurance	Disability	Annuities
Mean Discount	25.4%	4.20%	1.01%	0%	-3.50%
95% CI	(24.3%, 26.4%)	(3.57%,4.84%)	(0.004%, 1.7%)	-	(-5.05%, -1.96%)
Range	(19.7%,35.0%)	(0.09%, 10.7%)	(0%, 33.3%)	-	(-11.3% 5.5%)
N	114	60	257	12	567

Table 1: For LTC/Hybrid/LI/Disability, these are literal couples discounts. For annuities, it is the discount/cost of arbitraging a 50% joint and survivor annuity with two single life annuities. There were no couples discounts in any disability products surveyed.

WTP for spouses insurance of zero. We will test the following precise prediction, which follows from the proof of proposition 1. Recall, given the construction, that the risks are symmetric, in that who are labelled spouse 1 or 2 is random.

In terms of the regions in figures 1a - 1c, the theory predicts that the lower the average risk in region I relative to II or III, the greater the incentive to bundle. In particular, for price p, define the percentage difference between the average risk in regions I and II as follows:

$$\Delta\% = \frac{\mathbb{E}\left[p_1|p_2 > p, p_1 > p\right] - \mathbb{E}\left[p_1|p_2 \leq p, p_1 > p\right]}{\mathbb{E}\left[p_1|p_2 \leq p, p_1 > p\right]} = \frac{\mathbb{E}\left[\text{risk} \mid \text{region II}\right] - \mathbb{E}\left[\text{risk} \mid \text{region I}\right]}{\mathbb{E}\left[\text{risk} \mid \text{region II}\right]}.$$

In this notation, the prrice patterns yields the following hypothesis:

Hypothesis 1. *For the different risks:* $\Delta\%^{LTC} > \Delta\%^{Hybrid} > \Delta\%^{Life} > \Delta\%^{Annuity}$.

The hypothesis is in terms of risk (i.e. probabilities) which I can measure, not in terms of insurance WTP or costs.

The data I use come from the HRS. It is a biennial survey that has been conducted since 1990, so that long-term outcomes as as long-term care or mortality by particular elderly ages has been observed for much of the sample. To test the hypothesis I first predict the probabilities of needing long-term care or of death for each respondent and their spouse in the HRS. I measure mortality risk by the probability of death by age 75, and long-term care risk by the probability of needing care in the next 10 years.

I then assign a probability of zero to any non-existent spouse of a single respondent. I then randomly label each person in the couple as either spouse 1 or 2 with equal probability so that the two risks are symmetrical. I then compute $\mathbb{E}\left[p_1|p_2 > p, p_1 > p\right]$, $\mathbb{E}\left[p_1|p_2 \leq p, p_1 > p\right]$ and their difference for various p. To compare across risks with different ranges, write p_X for the Xth quantile. I will consider the six top-most deciles, since only about 60% of my sample have life insurance and fewer than 10% of the sample has long-term care.

The results are in table 2 below.

These results are consistent with the discounts observed. A large married couple LTC discount is due to the fact that, at any equilibrium price between $p_{30} \leq p \leq p_{90}$, those who are attracted by a bundling discounts are between 10 and 25% less likely to need long-term care than those who would buy one policy but not the bundle. Further supplementary policyholder evidence in appendix 3.2 shows that, among those who hold LTC insurance, being married is associated with an 9% lower risk of making a claim, relative to a base rate of 13%.

p_X	Long-term Care Risk		Mortality Risk		Disability Risk	
	$\mathbb{E}$ [risk|region II]	$\Delta\%$	$\mathbb{E}$ [risk|region II]	$\Delta\%$	$\mathbb{E}$ [risk|region II]	$\Delta\%$
p_{30}	0.14^{***}	-0.22^{***}	0.35^{***}	-0.17^{***}	0.28^{***}	0.2^{***}
	(0)	(0.01)	(0)	(0.01)	(0)	(0.06)
p_{40}	0.16^{***}	-0.14^{***}	0.39^{***}	-0.06^{***}	0.31^{***}	0.11^{**}
	(0)	(0.01)	(0)	(0.01)	(0)	(0.04)
p_{50}	0.18^{***}	-0.11^{***}	0.45^{***}	-0.02	0.35^{***}	0.04
	(0)	(0.01)	(0)	(0.01)	(0)	(0.04)
p_{60}	0.2^{***}	-0.08^{***}	0.51^{***}	0.01	0.39^{***}	0.08^{**}
	(0)	(0.02)	(0)	(0.01)	(0)	(0.04)
p_{70}	0.24^{***}	-0.07^{***}	0.59^{***}	0	0.46^{***}	0.05
	(0)	(0.02)	(0)	(0.01)	(0)	(0.03)
p_{80}	0.3^{***}	-0.06^{***}	0.7^{***}	0	0.54^{***}	-0.01
	(0)	(0.02)	(0)	(0.01)	(0)	(0.03)
p_{90}	0.44^{***}	-0.08^{**}	0.83^{***}	0.01	0.67^{***}	-0.05
	(0)	(0.03)	(0)	(0.01)	(0)	(0.04)
n	110294	110294	85575	85575	1938	1938

*** = significant at the 1 % level.

Table 2: For the three risks under study, the difference in risk between the bundle buyers (region II) and the non-bundle buyers (region I) at various prices.

On the other hand, there is a smaller, often indistinguishable from zero, difference in mortality risk between those who would buy the bundle at a small discount and those who would buy a single product. This is consistent with the evidence: small, close to zero, bundling discounts in the life insurance market and small anti-discounts in the annuity market.

3.2 Further Evidence from LTC Policyholders

A possible drawback of the HRS data is that I can measure only underlying risk, not actual insurance purchases. In this section I provide evidence from a large experience study of LTC insurance policyholders consistent with the HRS evidence. A downside of policyholder data is that this may be a sample already induced to buy by couples discounts. Given my data I cannot rule this out. Nevertheless, the HRS and policyholder evidence are complementary and together more convincing.

The data come from the Society of Actuaries (SOA) SOA (2016) who periodically run 'experience studies'. This experience study collected data from LTC insurance policyholders from 18 different insurers from 2000-2016. The outcomes of interest were claim rates for different types of long-term care benefits. The data contain approximately 60 million contract years of exposure and 600,000 claims. The SOA amalgamated contract level data such that each observation contains multiple years of exposure for the individuals with the same covaraites. This is controlled for in my analysis,

	Coefficient on Single (β)		
Outcome	Without Controls	With Controls	Overall Claim Rate
All claims	0.17*** (0.004)	0.09*** (0.004)	0.13*** (0.002)
NH claims	0.06*** (0.001)	0.03*** (0.001)	0.05*** (0.001)
ALF claims	0.04*** (0.001)	0.02*** (0.001)	0.03*** (0.001)
HHC claims	0.07*** (0.003)	0.04*** (0.003)	0.04*** (0.001)
N	1,507,275	1,507,275	1,507,275

Table 3: Results from estimating equations (3.1) with different outcomes, with and without controls. The rightmost column is the sample average claims per contract-year. *** = significant at the 1 % level.

but explains how 60 million contract years of exposure becomes an N of just over 1 million.

I run 8 different regressions. All combinations of four different outcome variables (total claims, nursing home claims (NH), assisted living facility claims (ALF), home health care (HHC) and two different specifications, with and without controls, are included in the table. Note that I don't observe price, although I do observe the above factors that enter into pricing.

The equation estimated is

$$outcome_i = \alpha_0 + \alpha_1 LivesExposed_i + \beta Single_i + \gamma Controls_i + \epsilon_i. \tag{3.1}$$

The outcome is either all claims, NH claims, ALF claims or HHC claims. The controls, when included, contain the following exhaustive set of variables that are priced upon: , Coverage Type , Issue Age, Current Age, Issue Year , Premium Class, Underwriting Type, Coverage Type, Inflation Rider, Rate Increase Flag, Restoration of Benefits Flag, NH Daily Benefit, ALF Daily Benefit, HHC Daily Benefit, NH Benefit Period, ALF Benefit Period , HHC Benefit Period, NH Elimination Period, ALF Elimination Period, HHC Elimination Period.

Table 3 below illustrates the coefficient on being single (as opposed to married) without controls (column 1), with controls (column 2) and column 3 is the overall claim rate per contract-year exposed to allow for interpretation of the effect of being single.

We see that those who are single have substantially higher claim rates than those who are married. Even with controls,and relative to the overall claim rate, the probability of making a claim is approximately 69%, 60%, 66% and 100% higher for all claims, NH claims, ALF claims and HHC claims respectively.

This is evidence that being married strongly predicts the probability of utilizing LTC insurance. This is consistent with the substantial discounts offered to married couples.

Lapsation as an alternative explanation? One alternative explanation is that married policyholders might lapse less. In the context of long-term care, in which many years of premia payments typically precede any possible claims, lapsation is in the interests of the insurer. I give evidence in appendix D.2 that this is not the case. If anything, those receicing marital discounts lapse less, making them less profitable along that dimension. This is consistent with many years of experience studies conducted by the SOA.

3.3 Quantity Discounts in Life Insurance

Two of the most popular forms of life insurance are term and whole life insurance. Term products are straight morality insurance: if the policyholder dies while the policy is in force their beneficiaries will receive a payout. There is no surrender or cash out value if the policyholder no longer wishes to continue paying premia. The premia are level for a fixed term, often 10 or 20 years, after which they increase substantially and almost all policyholders lapse.

Whole life insurance also features defined premia that may rise over time or be fixed, but are committed to at purchase. Whole life is designed to be held for longer periods of time, has a cash value that can be borrowed against or redeemed should the policyholder decide to end coverage while alive.

Term insurance is more purely an insurance product aimed at replacing lost income if the policyholder dies in their (typically) prime age working years. As such, it is plausible that mortality among term life insurance policyholders displays a income gradient. On the other hand, whole life insurance funcitons more like a savings account that often never leads to a death payout (see, e.g. Gottlieb and Smetters (2021)). Thought of as a savings account, there is less reason to beeleive that one $1 million policy looks that different from 10 $100k policies, which explains some of that lack of mortality gradient the data will show.

In this section I explore the quantity discounts offered in term life insurance and whole life insurance.[(6)] I find that term life insurance attracts a substantially larger quantity discount than whole life. The theory predicts that this is due to those who have high WTP for higher quantity of term life insurance being lower risk, relative to the difference in whole life insurance, per dollar of insurance, than those who only want to buy a lower quantity of term insurance. This is what I find. I begin by illustrating the differential quantity discounts offered for the two products.

Table 4 shows the average prices per 100k of face value for term and whole life policies of different sizes. The prices are the universe of policies offered in ?? (COM) for a 40 year old male in January 2022. THere are stark differnces in the quantity discounts offered. Naturally whole life premia, which are guranteed to never rise, are higher than 10 year term prices for a 40 year old male or

[(6)] I include in whole life insurance any policies that feature premia and a death benefit guranteed for life. This includes some policies labelled named universal life. I define the latter as featuring variable payout and death benefits that can be higher or lower as the policyholder prefers, subject to some minima.

Product	Term Life			Whole Life		
	Price per 100k FV			Price per 100k FV		
Face Value (FV)	Mean)	Range	N	Mean	Range	N
\$50k	289.04	(153, 680)	120	1048.92	(637, 2034)	60
\$100k	175.36	(95, 417)	475	859.74	(583, 1621)	115
\$250k	149.42	(54, 380)	585	798.06	(559, 1585)	131
\$500k	99.64	(42, 380)	568	787.38	(559, 1573)	131
\$1m	90.32	(35, 380)	578	783.84	(559, 1567)	131
\$10m	84.26	(8.6, 380)	549	766.78	(68, 1561.6)	131

Table 4: Average prices per 100k face value for policies of different face value. All policies offered in ?? (COM) to a 40 year old male in January 2022 are included.

female. But the discount given to those who buy more term life is much steeper than to those that buy more whole life insurance. The price of \$50,000 of term life is 3.2 times as much per dollar as for \$1,000,000, while the price of \$50,000 of whole life is only 1.3 times as muchper dollar as \$1,000,000.

Figure **??** shows the same pattern, except now the discount is computed relative to the 100k policy, and the set of policies included is slightly different. When computing the discount per 100k that a 500k policy receives versus a 100k policy, only companies that offered both a 100k and 500k policy were included. This explains the differences between the figure and the table. The broad pattern is identical: the discount is much higher for term than whole life insurance.

Per the theory, this suggests two tests. First, that there are non-trivial quantity discounts suggests that those who buy more life insurance, term or whole life, are lower risk per dollar than those who buy less. Second, this gradient of risk per dollar with respect to face value should be more sharply negative for term life insurance and for whole. This will bear out in the data.

The risk data I use are from the Society of Actuaries Life Insurance Mortality Experience Study SOA (2016). The study collected policyholder data from 91 life insurers for the period 2009 - 2016. It features 410 million life-years exposed, \$85 trillion in exposed face value, and 4 million actual deaths and \$217 billion in payouts.

To test the theory, I need to compare mortality risk of those who buy term versus whole life insurance and at various different face amounts of the policy.

To be concrete I want to test the following:

Hypothesis 2. *Holding all other pricing factors constant, those that purchase higher face value policies should have lower mortality risk.*

The gradient of mortality risk with respect to higher face values should be higher for term life than for whole life.

A breakdown of the lives and dollars exposed, and deaths and payouts, for each of these combinations is in the following table 5.

Product		Term Life		Whole Life	
Face Value		Lives	Dollars	Lives	Dollars
\$25k- \$49	Exposed Mortality Rate	3,132,211 0.61%	\$90,108,302,691 0.58%	36,628,908 0.85%	\$1,106,103,451,522 0.89%
\$50k- \$99	Exposed Mortality Rate	8,106,706 0.39%	\$457,498,237,859 0.37%	31,964,996 0.66%	\$1,910,823,841,349 0.69%
\$100k- \$249	Exposed Mortality Rate	37,650,270 0.22%	\$4,996,866,077,735 0.20%	26,013,086 0.53%	\$3,365,376,478,388 0.54%
\$250k- \$499k	Exposed Mortality Rate	34,846,346 0.11%	\$10,014,688,285,766 0.11%	5,911,227 0.46%	\$1,787,931,040,035 0.49%
\$500k- \$990k	Exposed Mortality Rate	23,429,274 0.09%	\$13,039,792,965,080 0.09%	2,413,536 0.49%	\$1,425,705,040,952 0.50%
\$1m- \$9.9m	Exposed Mortality Rate	14,165,447 0.08%	\$19,947,131,154,479 0.08%	1404744. 0.60%	\$2,366,995,318,122 0.70%
\$10m+	Exposed Mortality Rate	83,846 0.09%	\$1,119,422,717,399 0.10%	26,383 1.14%	\$371,509,238,870 1.00%

Table 5: Lives and dollars exposed and mortality rates by term vs whole life and for different face value bands. *** = significant at the 1 % level.

This offers suggestive evidence for both parts of hypothesis 2. To rigorously establish this, we need to control for all other demographic factors that go into pricing. A standard way in the insurance industry to compare mortality risks across groups is to compute an actual versus expected ratio for deaths or for dollars payed out. The expected basis is typically taken from an industry life table used for pricing.

In this case, so as to control for the demographic factors that go into pricing (e.g. gender, age, risk class, smoking status) but not for face value or term vs whole, I use the 2015 Valuation Basic Table (VBT) as the expected basis. The 2015 VBT gives differentiated mortality rates by age and sex and smoking status, and supplemental relative risk tables to account for risk classes. In the analysis that follows I compute the following ratio for each product type p (term vs whole) $\times$ face value fv band combination:

$$A/E^{lives}_{p,fv} = \frac{\text{Actual Deaths}}{\text{2015 VBT Expected Deaths}}, \quad A/E^{\$}_{p,fv} = \frac{\text{Actual Payouts}}{\text{2015 VBT Expected Payouts}}.$$

I take the expected basis as ground truth, and compute standard errors on the actual deaths by using the normal approximation to the binomial distribution (e.g. each death is an independent Bernoulli trial). The A/E ratios with standard errors are reported in table **??** and figure **??**.

The mortality patterns are consistent with the theory. The mortality decline, per dollar insured, is steep and negative for term life. This coheres with the large quantity discounts offered. For whole life: there is a decline in the actual over expected mortality at low face values, consistent with the quantity discount offered between $50k and other face values in figure **??**. But from $100k onwards, the mortality slope for whole life policies is essentially flat or rising. This is consistent with essentially no further discount being offered beyond the $100/$250k region for whole life policies. Note that an anti-discount (prices per $100k rising with face value) is not feasible owing to the non-exclusivity of these contracts: An individual could always buy two $500k policies were they cheaper than a $1 million policy.

4 Behavioral Analysis

There is a plethora of evidence that individuals err when purchasing insurance.[7] In this section I focus misperception of risks and the resulting over or under-insurance. The primary question is, assuming some sort of misperception occurs in the individual risk markets, does the bundled market compound or mitigate that misperception? The answer will turn on what type of misperception occurs in the individual risk markets, and the correlation of the errors across risk markets.

I adapt the notation and setup of Spinnewijn (2017) to my setting. As before, index each person i's true value of insurance (if they had no behavioural frictions) for risk 1 and 2 by (w_1^i, w_2^i). When I speak of a generic individual I will drop the superscripts. Individual i may make some behavioural mistake in their valuation of insurance, such as risk mis-perception, so that their perceived value is instead $(\hat{w}_1^i, \hat{w}_2^i)$. Denote the difference between true and perceived value by $(\epsilon_1^i, \epsilon_2^i) = (\hat{w}_1^i, \hat{w}_2^i) - (w_1^i, w_2^i)$. Individuals purchase insurance contract 1, for example, if $\hat{w}_1 \geq p_1$ but would maximize their utility (and welfare) if they purchased only when $w_i \geq p_1$.

The costs of insuring individual i in contract 1 i is given by $\phi_1(w_1)$ and similarly for product 2 or the bundle. Most importantly, the cost of insuring the individual depends on their true risk type, not their (mis)perceived $\hat{w}$. Misperceptions will affect purchase decisions, but conditional on that decision, they will not affect costs. Throughout the section I assume that selection is adverse, $\phi' > 0$, although it is not hard to adapt these definitions and propositions to the advantageous selection case.

There are two broad patterns I will analyze here. This makes clear the main conceptual considerations. First, the perceived value (i.e. the demand) curve in a particular contract market might be a clockwise rotation of the true value curve. Second, the demand curve might be an anti-clockwise rotation of the true value curve. I define these now.

Definition 6. *For product $j = 1, 2$ if $\frac{\partial}{\partial p_j} E\left[\epsilon_j \mid \hat{w}_j \geq p_j\right] \geq 0$ with equality at p'_j then I say that the errors in market j are risk-independent centered at p'_j.*

Misperceptions that are uncorrelated with risk type $E(\epsilon_j \mid w_j) = 0$ will generate a demand curve that obeys definition 6[8]. This is because, even if errors have mean zero, individuals that draw a high error will end up with higher demand, on average, than those that draw a low error. Hence, if we condition on $w_j \geq p_j$ the error of the marginal type will be higher the higher p_j is.

Inversely:

Definition 7. *For product $j = 1, 2$ if $\frac{\partial}{\partial p_j} E\left[\epsilon_j \mid \hat{w}_j \geq p_j\right] \leq 0$ with equality at p'_j then I say that the errors in market j are risk dependent centered at p'_j.*

Misperceptions that are correlated with risk in the standard 'S' shape as in Solomon (2021) will lead to demand curves that satisfy definition 7. The 'S' shape is generated by individuals thinking

[7] For example, see CITE Odea? Some Gruber? Myself?)

[8] See Proposition 2 in Spinnewijn (2017) for details

their risk is closer to some notion of the mean. High risks think they are lower than in truth and low risks think they are higher. This naturally generates demand curves that obey definition 7.

How does the misperception in the bundled market relate to the misperceptions in the individual markets? The crucial ingredient is how the errors in the individual markets correlate. Again I will focus on two broad patterns: independence or positive correlation of errors between risk markets.

Definition 8. *Suppose $E[\epsilon_1 \mid \epsilon_2] = E[\epsilon_1]$ for all ϵ_2 and vice versa for ϵ_2 conditioning on ϵ_1, then I say that errors are indepedent between risks, and write this distribution of errors as $E^{\perp}$.*

Definition 9. *For product $j = 1, 2$ if the distribution of errors E exceeds the independence between risks case (definition 8) in the correlation order, $E \succsim E^{\perp}$, then I say that errors are correlated between risks.*

All four combinations of errors within (definitions 6 and 7) and between (definitions 8 and 9) risk markets are plausible. If all errors are simply random draws then independence within a risk market and between risk markets will ensue. If individuals draw a level of optimism or pessimism independently of their risk type, but this is constant across both risks then within a market errors may be uncorrelated with risk type but between markets errors may be strongly correlated. On the other hand, if within each market risk type correlates with misperception, then the degree of correlation between risk types between markets will determine the degree of correlation of misperception between markets.

The effects of misperpcetions between markets can cancel out or compound the effects of misperceptions between markets. To isolate just the effect of varying the between market correlation in errors, I assume that in the bundled market the cost and true WTP curves do not change, only the perceived WTP curve $\hat{W} = \hat{W}_1 + \hat{W}_2$ will change as the errors become more or less correlated. I am agnostic about what causes the between market correlation. It might be fixed effect of individual optimism or pessimism, or it might be correlations between risk and misperception within a market and correlation between risks in the two different markets.

This is encapsulated by the following table:

	Perceived value is a ____________ rotation of actual value	
	Clockwise	Counter-clockwise
Within each risk market i	$Corr(w_i, \epsilon_i)$ not too negative	$Corr(w_i, \epsilon_i)$ negative
Between risk markets i and j	$Corr(\epsilon_i, \epsilon_j)$ large	$Corr(\epsilon_i, \epsilon_j)$ not too positive

Table 6: How correlation between error and risk within each risk market and between errors across markets influences the wedge between true and perceived WTP curves.

This means there are four types of correlation structure that may obtain: Errors within the markets can be strongly negatively correlated with risk, or not. Errors between markets can be positively correlated, or not.

I focus on whether the wedge between actual and perceived value caused by subjective misperceptions of risk are accentuated or mitigated by bundling the two risks together. If misperceptions are correlated within person, perhaps due to a person fixed effect of optimism or pessimism, then perceived value will (more of a) clockwise rotation of actual value than if errors were independent or negatively correlated within person.

If the market failiure is under provision of insurance due to adverse selection, and the market is 'large' (as in proposition 3), then the planner desires an anti-clockwise-rotation of demand. This could occur, for example, if errors are indepedent or negatively correlated across people and errors are indepedent of risk.

On the other hand if the market failure is over-provision of insurance due to advantageous selection, and the market is large, then the planner prefers perceived WTP to be a clockwise rotation of true WTP. This might occur with more positive correlation between errors and/or between error and risk within each market.

Whether errors in risk perception add to or compound the effects of bundling, and hence whether correcting the misperceptions is helpful to market operations depends on what type of correlation pattern in the errors obtains within and between markets, as well as the nature of the market failiure that the planner wishes to correct.

4.1 Empirical Example of Misperceptions Interacting with Bundling

As an empirical example, I use estimates of mortality and long-term care risk for the same individual in the HRS. For all 65 to 70 year olds, I predict the probability of death by 80, and I predict the probability of entering a nursing home in the next 5 years. Label this by p_r^i where the person is labelled by i and the risk by r =long-term care or mortality. Individuals are also surveyed about their subjective perceptions of these probabilities, analogously labeled $\hat{p}_r^i$. I define the error of person i in risk r =long-term care, mortality, as the subjective elicitation minus my objective prediction:

$$\epsilon_r^i = \hat{p}_r^i - p_r^i.$$

In the table below are coeffieicnts from the regression of $\epsilon^i_{\text{mortality}}$ on ϵ^i_{LTC} either controlling for mortlaity risk or not. Specifically I estimate:

$$\epsilon^i_{\text{mortality}} = \alpha + \beta\epsilon^i_{\text{LTC}} + \varepsilon \tag{4.1}$$

$$\epsilon^i_{\text{mortality}} = \alpha + \beta\epsilon^i_{\text{LTC}} + \gamma p^i_{\text{mortality}} + \varepsilon \tag{4.2}$$

Table 7 shows, in the first column, that there is a correlation between errors an individual makes in one risk perception and in the second risk perception. This might be due to person fixed optimism or pessimism, or it might be due to correlation in risk and a tendency for those with high risk to under-perceive their risk, and vice versa. The second column clarifies this: The strong negative coefficient on mortality risk is consistent with the 'S' shaped pattern that risk misperceptiosn typically display: the high risk underestimate their risk and the low risk overestimate it.

But holding mortality risk fixed, there is still a smaller but positive association between making

Coefficient Estimates	Model 4.1	Model 4.2
ϵ^i_{LTC}	0.21^{***} (0.01)	0.16^{***} (0.01)
$p^i_{\text{mortality}}$	-	-0.63^{***} (0.01)
Constant	-0.005^{**} (0.002)	0.24^{***} (0.004)
N	21,697	21,697

*** = 1% significance, ** =5% significance

Table 7: OLS estimates of models 4.1 and 4.2

an error in the LTC risk and in the mortality risk. This is consistent with there being an individual fixed component of misperception: those that think themselves closer to the average do so across risks, that isn't due to their risk level.

This suggests that misperceptions may counteract the underlying correlation between risk types. The positive correlation in risk types makes bundling relatively undesirable for a planner and for a firm. But to the extent that errors are person specific and risk independent, this adds weakens the correlation between the *perceived* risks relative to the true risks.

p_X	Objective Risk		Subjective Risk	
	$\mathbb{E}$ [risk|region II]	$\Delta\%$	$\mathbb{E}$ [risk|region II]	$\Delta\%$
p_{10}	0.42^{***}	0.24^{***}	0.45^{***}	0.01
	(0)	(0.01)	(0)	(0.01)
p_{20}	0.45^{***}	0.25^{***}	0.49^{***}	-0.03^{***}
	(0)	(0.01)	(0)	(0.01)
p_{30}	0.49^{***}	0.24^{***}	0.53^{***}	-0.03^{***}
	(0)	(0.01)	(0)	(0.01)
p_{40}	0.53^{***}	0.21^{***}	0.58^{***}	-0.05^{***}
	(0)	(0.01)	(0)	(0)
p_{50}	0.58^{***}	0.18^{***}	0.63^{***}	-0.03^{***}
	(0)	(0.01)	(0)	(0)
p_{60}	0.63^{***}	0.14^{***}	0.67^{***}	-0.03^{***}
	(0)	(0)	(0)	(0)
p_{70}	0.7^{***}	0.1^{***}	0.73^{***}	-0.02^{***}
	(0)	(0)	(0)	(0)
p_{80}	0.78^{***}	0.06^{***}	0.84^{***}	-0.01^{***}
	(0)	(0)	(0)	(0)
p_{90}	0.88^{***}	0.03^{***}	0.95^{***}	0.01^{***}
	(0)	(0)	(0)	(0)

*** = 1% significance

Table 8: Savings when bundling mortality and long-term care risk together according to objective (predicted) versus subjective (elicited) probabilities.

To confirm this, I simulate the increase or reduction in risk acheived by a tiny bundling discount in two cases: when individuals know and purchase according to their true risk, and when individuals purchase according to their subjectively (mis-)perceived risk. The table below shows the relative mortality risk reductions achieved by attracting those with high mortality and high LTC risk versus those with high mortality but low LTC.

The second column shows that, according to objective risks, offering a bundling discount is a bad idea for firms: those attracted are of higher risk than those who are not. But the fourth column shows that, according to subjective risks, bundling is a more profitable strategy, with those attracted to teh bundle of a similar lower mortality risk than those who are not.

This shows that the risk independent noise introduced by the misperceptions allow for the bundle buyers to be less expensive with the noise than without it. The misperceptions weaken the link between WTP and cost, and mean that those most willing to pay for the bundle are more like those who are not. Since we began with positive correlation between risks, this noise moves us toward independence, i.e. toward less correlation, which is good for bundling and expanding the market.

As a consequence, bundling two misperceived can help the adverse selection problem over and above the 'S' shape misperception that prevails in each risk market.

This is specific to the particular pattern of risk and errors in this setting. Because mortality and LTC risk were correlated within person, whereas errors were closer to independent, subjective risk was less correlated than objective risk. Were mortality and LTC risk negatively correlated within person, then adding iid errors would increase correlation toward zero. The interaction of misperceptions on a bundled market is a subtle and context specific matter which this section has broadly introduced.

5 Conclusion

This paper offers a theoretical foundation and empirical evidence for the pervasive practice of bundling insurance products. The theory predicts that bundling is a cream-skimming device: Firms will bundle if and only if those that buy both products are cheaper than those who buy one. For the planner, bundling is useful when it homogenizes the risk pool, which occurs when the risks are more negatively (or less positively) correlated. Empirically, I offer evidence consistent with firm behavior that the model predicts: I show that larger couples discounts are offered to products in which being in a couple has the greatest negative correlation with risk. And I show that quantity discounts (which can be thought of as bundling the first dollar of coverage with the second) in term versus whole life insurance follow the same pattern: term life insurance features steep quantity discounts and high risk reductions amongst high quantity policyholders, whereas whole life has minimal discounts and flat or increasing risk as a function of the quantity of insurance purchased.

Broadly this paper speaks to a literature about the optimal scope of insurance products: which risks should be insured together and which separately. Future research might apply these theoretical results to different risk markets to clarify the limits and robustness of the empirical findings.

Bibliography

COMPULIFE | Life Insurance Quote Software.

Adams, W. J. and J. L. Yellen (1976). Commodity Bundling and the Burden of Monopoly. *The Quarterly Journal of Economics 90*(3), 475–498.

Akerlof, G. A. (1978). The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *The American Economic Review 68*(1), 8–19.

Bakos, Y. and E. Brynjolfsson (1999, December). Bundling Information Goods: Pricing, Profits, and Efficiency. *Management Science 45*(12), 1613–1630.

Cambanis, S., G. Simons, and W. Stout (1976, December). Inequalities for Ek(X, Y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 36*(4), 285–294.

Chen, Y. and M. H. Riordan (2013). Profitability of Product Bundling*. *International Economic Review 54*(1), 35–57.

Cooper, A. L. and A. N. Trivedi (2012). Fitness memberships and favorable selection in medicare advantage plans. *New England Journal of Medicine 366*(2), 150–157. PMID: 22236225.

Crocker, K. J. and A. Snow (2011). Multidimensional Screening in Insurance Markets with Adverse Selection. *The Journal of Risk and Insurance 78*(2), 287–307.

Denuit, M., J. Dhaene, M. Goovaerts, and R. Kaas (2006, May). *Actuarial Theory for Dependent Risks: Measures, Orders and Models.* John Wiley and Sons.

Einav, L., A. Finkelstein, and M. R. Cullen (2010, August). Estimating Welfare in Insurance Markets Using Variation in Prices*. *The Quarterly Journal of Economics 125*(3), 877–921.

Ghili, S. (2022, April). A Characterization for Optimal Bundling of Products with Non-Additive Values. Technical report. arXiv:2101.11532 [econ, q-fin] type: article.

Gottlieb, D. and K. Smetters (2021, August). Lapse-Based Insurance. *American Economic Review 111*(8), 2377–2416.

Haghpanah, N. and J. Hartline (2021, May). When Is Pure Bundling Optimal? *The Review of Economic Studies 88*(3), 1127–1156.

Hurkens, S., D.-S. Jeon, and D. Menicucci (2019, August). Dominance and Competitive Bundling. *American Economic Journal: Microeconomics 11*(3), 1–33.

ImmediateAnnuities.com. Annuity shopper buyer's guide.

Kotz, S. (2000). *Continuous multivariate distributions. Volume 1, Models and applications* (2nd ed. ed.). Wiley series in probability and statistics. New York: John Wiley and Sons, Inc.

Lavetti, K. and K. Simon (2018, August). Strategic Formulary Design in Medicare Part D Plans. *American Economic Journal: Economic Policy 10*(3), 154–192.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989, May). Multiproduct Monopoly, Commodity Bundling, and Correlation of Values*. *The Quarterly Journal of Economics 104*(2), 371–383.

Nalebuff, B. (2004, February). Bundling as an Entry Barrier*. *The Quarterly Journal of Economics 119*(1), 159–187.

Nguyen, A. (2018, May). Household Bundling to Reduce Adverse Selection: Application to Social Health Insurance. Technical Report ID 3173424, Rochester, NY.

Schmalensee, R. (1984). Gaussian Demand and Commodity Bundling. *The Journal of Business 57*(1), S211–S230.

Shaked, M. and J. G. Shanthikumar (2007, April). *Stochastic Orders*. Springer Science & Business Media. Google-Books-ID: rPiToBK2rwwC.

Shepard, M. (2022, February). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. *American Economic Review 112*(2), 578–615.

SOA (2016). 2009-2016 individual life insurance mortality experience report. Technical report, Society of Actuaries.

Solomon, A. (2021, February). Imperfect Private Information in Insurance Markets. Technical Report 3787556, Rochester, NY.

Spinnewijn, J. (2017, February). Heterogeneity, Demand for Insurance, and Adverse Selection. *American Economic Journal: Economic Policy 9*(1), 308–343.

StrateCision (2022a, June). Combocompare. https://www.ltca.com/Combo.htm. https://www.ltca.com/Combo.htm.

StrateCision (2022b, June). Ltc quote plus. https://www.ltca.com/qp.htm. https://www.ltca.com/qp.htm.

Yang, F. (2021, November). Costly Multidimensional Screening. Technical Report 3915700, Rochester, NY.

Zhou, J. (2017). Competitive Bundling. *Econometrica 85*(1), 145–172.

Zhou, J. (2021, June). Mixed bundling in oligopoly markets. *Journal of Economic Theory 194*, 105257.

A Extended Theory Section

The assumptions in the main theory section were typically about truncated distributions. These are attractive as they correspond to the exactly the individuals that would show up in a dataset of the insured. But in the case when one has data on populations, perhaps because the insurance market doesn't exist, or because counter-factual insurance policies are being considered, assumptions on the global distribution of types might be preferable. In this section I present some results that are close equivalents of results in the main paper except with different assumptions used. Moreover, results that extend the theory in the main paper that require alternative or additional assumptions are studied.

An alternatively sufficient condition for many of the results that follow is that the change in total cost when correlation increases has the same sign as the change in average cost, conditional on being in some set S. Note that, fixing prices, if the joint distribution changes there are two effects:

1) those that buy products $1, 2$ or B have higher or lower costs and 2) the mass of people who buy a product changes. The second effect is of sole importance in case of no adverse or advantageous selection, such as the IO literature. I focus on the first effect. The assumption is essentially an assumption that the second effect is small. I refer to the first effect as the cost effect and the second as the quantity effect.

Note that the average cost of all those who buy a product $i = 1, 2, B$ can be written as

$$AC = \frac{\mathbb{E}\left[\mathbb{1}\left(w \in W_i\right) \times \phi_i(w)\right]}{\mathbb{E}\left[\mathbb{1}\left(w \in W_i\right)\right]}.$$

When the distribution changes the change in the value of the numerator is effect (1) and the denominator is effect (2). The following assumption therefore assumes (2) is small and will be an alternatively sufficient condition for the propositions of the main theory section to go through.

Assumption 2. *Given two distributions $X, Y \in \Gamma(F_1, F_2)$, we say that the quantity effect is small if*

$$\mathbb{E}_X\left[\mathbb{1}\left(w \in W_i\right)\right] \approx \mathbb{E}_Y\left[\mathbb{1}\left(w \in W_i\right)\right].$$

Given these alternate definitions we can restate the main results of section 2 in terms of correlation structures of the entire distributions with the additional assumptions. In particular, all assumptions about $Y \succsim^{\bar{\boldsymbol{p}}} X$ or $Y \succsim^{\backslash p_B} X$ are changed to simpler assumptions on $Y \succsim X$ and the additional assumption 2.

First the results about adverse selection, in which $\phi' > 0$

Proposition 8. *(Analogue of Proposition 1). Suppose $\phi' > 0$. Suppose $X, Y, Z^{\perp} \in \Gamma(F_1, F_2)$ with $X \precsim Y$. Denote the profit earned per person on a bundle contract offered at price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ by π^{ϵ}. We have the following:*

1. $\pi^{\epsilon}_{Z^{\perp}} = 0$
2. *Suppose Y, X, Z satisfy assumption 2. If $Y \succsim X \succsim Z^{\perp}$ then $\pi^{\epsilon}_Y \leq \pi^{\epsilon}_X \leq \pi^{\epsilon}_{Z^{\perp}} = 0$, and conversely if $Z^{\perp} \succsim X \succsim Y$ then $\pi^{\epsilon}_Y \geq \pi^{\epsilon}_X \geq \pi^{\epsilon}_{Z^{\perp}} = 0$*

Proposition 9. *(Analogue of Proposition 3). Suppose $\phi' > 0$. Suppose $X, Y \in \Gamma(F_1, F_2)$ and satisfy assumption 2 with $Y \succsim X$. The following comparative statics hold:*

- *The average cost curve is everywhere lower under less correlated distributions: $AC_X(p) \leq AC_Y(p)$ for all p.*
- *The equilibrium bundle price increases in correlation: $p^X_B \leq p^Y_B$.*
- *Under assumption 1, for 'large' markets, i.e. if $q^Y_B \geq \underline{q}$ for some $\underline{q}$ then the equilibrium quantity insured increases under the less correlated X: $q^X_B \geq q^Y_B$.*

And the analogues for the case of advantageous selection, $\phi' < 0$.

Proposition 10. *(Analogue of Proposition 4). Suppose $\phi' < 0$. Suppose $X, Y, Z^{\perp} \in \Gamma(F_1, F_2)$ with $X \precsim Y$. Denote the profit earned per person on a bundle contract offered at price $p_B = \bar{p}_1 + \bar{p} - \epsilon$ by π^{ϵ}. We have the following:*

1. $\pi^{\epsilon}_{Z^{\perp}} = 0$
2. *Suppose Y, X, Z satisfy assumption 2. If $Y \succsim X \succsim Z^{\perp}$ then $\pi^{\epsilon}_Y \geq \pi^{\epsilon}_X \geq \pi^{\epsilon}_{Z^{\perp}} = 0$, and conversely if $Z^{\perp} \succsim X \succsim Y$ then $\pi^{\epsilon}_Y \leq \pi^{\epsilon}_X \leq \pi^{\epsilon}_{Z^{\perp}} = 0$*

Proposition 11. *(Analogue of Proposition 5). Suppose $\phi' < 0$. Suppose $X, Y \in \Gamma(F_1, F_2)$ and satisfy assumption 2 with $Y \succsim X$. The following comparative statics hold:*

- *The average cost curve is everywhere higher under less correlated distributions: $AC_X(p) \geq AC_Y(p)$ for all p.*
- *The equilibrium bundle price decreases in correlation: $p_B^X \leq p_B^Y$.*
- *Under assumption 1, for 'small' markets, i.e. if $q_B^Y \leq \underline{q}$ for some $\underline{q}$ then the equilibrium quantity insured decreases under the less correlated X: $q_B^X \leq q_B^Y$.*

A.1 Equilibrium Dynamics for Mixed Bundling

In section 2 the equilibrium dynamics for pure bundling and purely seperate sales were analyzed. In this section, under stronger assumptions, I study the comparative statics of the mixed bundlign equilibrium with respect to changing correlation.

Throughout I assume adverse selection. The dynamics for advantageous selection are similar but inverted.

Again the useful benchmark case is when w_1 and w_2 are independent. As before denote the independent joint distribution by $Z^{\perp}$. Recall from section 2 that under $Z^{\perp}$ there is an equilibrium in which no bundling occurs and the seperate markets clear at their prices. By single the assumption of single crossing, this is the unique equilibrium. Similarly this is the case for all $X \in \Gamma(F_1, F_2)$ with $X \succsim Z^{\perp}$. When there is no or positive correlation, there is no incentive to introduce any bundling discount.

Hence the dynamics of interest are for various degrees of negative correlation. I will compare the equilibrium outcomes under distributions $X, Y \in \Gamma(F_1, F_2)$ with X more negatively correlated than Y which is more negatively correlated than independent $Z^{\perp}$. The exact correlation ordering needed will vary for different propositions.

First, I show where profits and losses would be mad if the equilibrium prices under Y were offered but the true distribution was X.

Proposition 12. *Assume $Y, X, Z^{\perp} \in \Gamma(F_1, F_2)$ with $X \precsim Y \precsim X^{\perp}$, as well as assumption 2, and label the equilibrium prices under Y as $\boldsymbol{p}^Y = (p_1^Y, p_2^Y, p_B^Y)$. If all firms offer prices $\boldsymbol{p}^Y$ when the joint distribution is actually X, then the profits earned in each product market will be: $\pi_1^X(\boldsymbol{p}^Y) \leq 0, \pi_1^Y(\boldsymbol{p}^Y) \leq 0$ and $\pi_1^B(\boldsymbol{p}^Y) \geq 0$.*

This suggests but does not definitely prove that under X the equilibrium prices in the single markets must rise relative to Y and the price of the bundle must fall. While all numerical simulations suggest this to be the case, the analytical assumptions needed to ensure this are so strong as to be basically just assume the result.

Nevertheless, empirically and concetpually it is safe to follow the suggestion: When the types become more negatively correlated under mixed bundling, as under pure bundling, the types that buy the bundle become cheaper and those that buy the single products become more expensive.

B Proofs

B.1 Preliminary Results

Definition 10. *We say that X precedes Y in the stop-loss order, written $X \precsim_{SL} Y$, iff*

$$\mathbb{E}_X\left[\max\{X - d, 0\}\right] \leq \mathbb{E}_Y\left[\max\{Y - d, 0\}\right] \quad \textit{for all } d \in \mathbb{R}.$$

The quantity $E(\max\{X-d,0\})$ is known as the stop-loss premium. For example, it is the expected loss an insurer faces when they insure risk X with a deductible d.

The main proposition to be used comes from Denuit et al. (2006) or a variant from Cambanis et al. (1976).

Proposition 13. *(Denuit et al. (2006)) Suppose* $X, Y \in \Gamma(F_1, F_2)$. *If* $X \precsim Y$ *then* $\psi(X) \precsim_{SL} \psi(Y)$ *for any non-decreasing super-modular function* ψ.

Alternatively, per Cambanis et al. (1976), if ψ *is supermodular and right continous, the stop-loss ordering holds, and if* ψ *is submodular and right continuous the stop-loss ordering is reversed.*

This says that when X is less correlated than Y, the stop-loss of any supermodular function of the margins is higher under Y than X for any stop-loss premium. In particular, $w_1 + w_2 = w_B$ is super-modular, and by assumption so is $\phi_B(w_1 + w_2)$ and so the expectation of the excess of these functions relative to a deductible d is always higher under Y than X.

Next, I show that for any joint distribution satisfying assumption 1, the WTP for the bundle W_B 'rotates' as the correlation changes.

Proposition 14. *Suppose* X, Y *satisfy assumption 1. Then if* $X \succsim Y$ *the demand curve for the bundle under* X*, relative to the demand curve under* Y *satisfies:*

- $WTP_X(q) \leq WTP_Y(q)$ *for* $q \in [0, \underline{q}]$ *and then*
- $WTP_X(q) \geq WTP_Y(q)$ *for* $q \in [\underline{q}, 1]$.

Proof. **Joint Normality**

In the case of joint normality with identical margins, the CDF of the convolution $W_1 + W_2$ is normally distributed with mean $\mu_X + \mu_Y$ and standard deviation $\sigma_{W_B} = \sqrt{\sigma_{W_1}^2 + \sigma_{W_2}^2 + \rho\sigma_{W_1}\sigma_{W_2}}$. That the CDF's of W_B (and hence the WTP for the bundle) are rotated as described is a standard fact about two normal distributions with the same mean and differing variances.

FGM form

Writing $u(w_1) = F_1(w_1), v = F_2(w_2)$ as the uniformly distributed CDFs of marginals, assuming that X and satisfied definition 3 means that

$$F_X(w_1, w_2) = u \cdot v \cdot [1 + \rho\,(1-u)\,(1-v)]\,.$$

Differentiating we get that the density is:

$$f_X(w_1, w_2) = 1 + \rho(1 - 2w_1)(1 - 2w_2).$$

Hence the CDF of the convolution $W_B = W_1 + W_2$ can be written as

$$\begin{aligned} Prob(W_B < s) &= \int_0^s \int_0^{s-w_2} f_X(w_1, w_2)\,dw_1\,dw_2 \\ &= \frac{1}{6}s^2(\rho(s-3)(s-1)+3) \end{aligned}$$

We are interested in how this CDF changes with the correlation parameter ρ. Hence differentiating we have

$$\frac{\partial Prob(W_B < s)}{\partial \rho} = \frac{1}{6}(s-3)(s-1)s^2.$$

The derivative is weakly positive on $[0,1]$, and becomes weakly negative on $[1,2]$. Hence for all $s \geq 1$ the CDF of the sum gets lower under more correlation, while for $s \leq 1$ the CDF of the sum gets higher under more correlation. This is what we needed to show.

■

B.2 Proof of Proposition 1

Proof. Suppose that the equilibrium prices in markets 1 and when there is no bundling are given by $\bar{p}_1$ and $\bar{p}_2$ respectively. Then a bundle is introduced at price $p_B = \bar{p}_1 + \bar{p}_2 - \epsilon$. For small enough *epsilon* > 0 individuals will buy the bundle at price p_B are those who separately bought products 1 and 2 before. Hence the set that purchase the bundle, which I denote $\mathcal{W}_B$ is given by $\mathcal{B} = \{w \in \mathcal{W} : w_1 \geq \bar{p}_1 \wedge w_2 \geq \bar{p}_2\}$. As before, the sets fo types that only buy 1 or 2 after the bundle is offered are denoted by $\mathcal{W}_1$ and $\mathcal{W}_2$.

To the first point, note simply that by the definition of independence, and that the market for product 1 was initially in equilibrium with $E[\phi(w_1) \mid \mathcal{W}_1] = \bar{p}_1$, we have

$$\begin{aligned}\mathbb{E}[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \geq \bar{p}_2] &= \mathbb{E}[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2] \\ &= \mathbb{E}[\phi_1(w_1) \mid w_1 \geq \bar{p}_1] \\ &= \bar{p}_1\end{aligned}$$

Similarly

$$\mathbb{E}[\phi_2(w_2) \mid w_1 \geq \bar{p}_1 \wedge w_2 \geq \bar{p}_2] = \bar{p}_2.$$

Hence the total cost of selling the bundle is $\bar{p}_1 + \bar{p}_2$, and so this breaks even for small ϵ, i.e. $\pi^{\epsilon}_{Z^{\perp}} \to 0$ as $\epsilon \to 0$.

To prove the rest, consider two correlation structures X and Y with $Y \succsim^{\bar{\boldsymbol{p}}} X$. This assumption means that, in particular,

$$P_X(W_1 > w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1) \leq P_Y(W_1 > w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1) \tag{B.1}$$

We then have:

$$\begin{aligned}\mathbb{E}_X[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \geq \bar{p}_2] &= \int_{\bar{p}_1}^{\overline{w}_1} \phi(w_1) P_X(W_1 = w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1)\, dw_1 \\ &= [\phi(w_1)(P_X(W_1 \leq w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1) - 1)]_{w_1=\bar{p}_1}^{w_1=\overline{w}_1} \\ &+ \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1)(1 - P_X(W_1 \leq w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1))\, dw_1 \quad \text{(integrating by par} \\ &= \phi(\bar{p}_1) + \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1) P_X(W_1 > w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1)\, dw_1 \\ &\leq \phi(\bar{p}_1) + \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1) P_Y(W_1 > w_1 \mid W_2 \geq \bar{p}_2 \wedge W_1 \geq \bar{p}_1)\, dw_1 \\ &= \mathbb{E}_Y[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \geq \bar{p}_2].\end{aligned}$$

where the inequality follows from $\phi_1'(w_1) > 0$ (adverse selection) and equation (B.3).

Similarly we conclude that

$$\mathbb{E}_X\left[\phi_2(w_2) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right] \leq \mathbb{E}_Y\left[\phi_2(w_2) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right].$$

Hence we have that

$$\mathbb{E}_X\left[\phi_B(w_B) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right] \leq \mathbb{E}_Y\left[\phi_B(w_B) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right]$$

as required.

■

B.3 Proof of Proposition 2

Proof. The second part of the proposition is almost identical to the proof of proposition 1. Simply replace $w_2 \geq p_2$ with $w_2] \leq p_2$ and inversely use

$$P_X\left(W_1 > w_1 \mid W_2 \leq \bar{p}_2 \wedge W_1 \geq \bar{p}_1\right) \geq P_Y\left(W_1 > w_1 \mid W_2 \leq \bar{p}_2 \wedge W_1 \geq \bar{p}_1\right). \tag{B.2}$$

To prove the first part of the proposition we proceed as follows. The set of bundle buyers are those satisfying $w_B > p_B \wedge w_1 \geq p_B - p_2 \wedge w_2 \geq p_B - p_1$. Writing $p_B = p_1 + p_2 - \epsilon$ this set is equivalently: $w_B > p_B \wedge w_1 \geq p_1 - \epsilon \wedge w_2 \geq p_2 - \epsilon$. We compute their average cost under distribution X, which by assumption was equal to the price, and show that the average cost under Y is higher. First for risk 1:

$$\begin{aligned}
\mathbb{E}_X\left[\phi_1(w_1) \mid \text{buy bundle}\right] = &\int_{p_1-\epsilon}^{p_1} \phi(w_1) P_X\left(W_1 = w_1 \mid W_2 \geq p_B - w_1 \wedge W_1 \geq p_1\right) dw_1 \\
&+ \int_{p_1}^{\overline{w}_1} \phi(w_1) P_X\left(W_1 = w_1 \mid W_2 \geq p_2 \wedge W_1 \geq p_1\right) dw_1 \\
&= \left[\phi(w_1)\left(P_X\left(W_1 \leq w_1 \mid W_2 \geq p_B - w_1 \wedge W_1 \geq p_1\right) - 1\right)\right]_{w_1=p_1-\epsilon}^{w_1=p_1} \\
&+ \left[\phi(w_1)\left(P_X\left(W_1 \leq w_1 \mid W_2 \geq p_2 \wedge W_1 \geq p_1\right) - 1\right)\right]_{w_1=p_1}^{w_1=\overline{w}_1} \\
&+ \int_{p_1-\epsilon}^{p_1} \phi'(w_1)\left(1 - P_X\left(W_1 \leq w_1 \mid W_2 \geq p_B - w_1 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&+ \int_{p_1}^{\overline{w}_1} \phi'(w_1)\left(1 - P_X\left(W_1 \leq w_1 \mid W_2 \geq p_2 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&= \phi(p_1 - \epsilon) + \int_{p_1-\epsilon}^{p_1} \phi'(w_1)\left(P_X\left(W_1 > w_1 \mid W_2 \geq p_B - w_1 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&+ \int_{p_1}^{\overline{w}_1} \phi'(w_1)\left(P_X\left(W_1 > w_1 \mid W_2 \geq p_2 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&\leq \phi(p_1 - \epsilon) + \int_{p_1-\epsilon}^{p_1} \phi'(w_1)\left(P_Y\left(W_1 > w_1 \mid W_2 \geq p_B - w_1 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&+ \int_{p_1}^{\overline{w}_1} \phi'(w_1)\left(P_Y\left(W_1 > w_1 \mid W_2 \geq p_2 \wedge W_1 \geq p_1\right)\right) dw_1 \\
&= \mathbb{E}_Y\left[\phi_1(w_1) \mid \text{buy bundle}\right].
\end{aligned}$$

The same works to show that $\mathbb{E}_Y\left[\phi_2(w_2) \mid \text{buy bundle}\right] \leq \mathbb{E}_Y\left[\phi_2(w_2) \mid \text{buy bundle}\right]$ and hence that the cost of selling the bundle is lower under X than under Y. Since by assumption of a competitive equilibrium $\mathbb{E}_X[\phi_B(w_1,w_2) \mid \text{buy bundle}]-p_B=0$, we then have that $\mathbb{E}_Y[\phi_B(w_1,w_2) \mid \text{buy bundle}]-p_B \leq 0$ makes a loss, as required.

■

B.4 Proof of Proposition 3

Proof. Consider two correlation structures X and Y with $Y \succsim X$. This assumption means, in words, that knowing W_2 is large (larger than $p_B - w$ in particular) says more about W_1 being large under Y than under X. Formally,

$$P_X(W_1 > w_1 \mid W_2 \geq p_B - w_1) \leq P_Y(W_1 > w_1 \mid W_2 \geq p_B - w_1) \tag{B.3}$$

(B.4)

For brevity I write this set as

$$\mathcal{W}_{\backslash p_B} = \{w \in \mathcal{W} : W_B \geq p_B\} \tag{B.5}$$

We then have

$$\begin{aligned}
\mathbb{E}_X\left[\phi_1(w_1) \mid \mathcal{W}_{\backslash p_B}\right] &= \int_0^{\overline{w}_1} \phi(w_1) P_X(W_1 = w_1 \mid W_2 \geq p_B - w_1)\, dw_1 \\
&= \left[\phi(w_1)\left(P_X(W_1 \leq w_1 \mid W_2 \geq p_B - w_1) - 1\right)\right]_{w_1=\bar{p}_1}^{w_1=\overline{w}_1} \\
&+ \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1)\left(1 - P_X(W_1 \leq w_1 \mid W_2 \geq p_B - w_1)\right) dw_1 \qquad \text{(integrating by parts)} \\
&= \phi(\bar{p}_1) + \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1) P_X(W_1 > w_1 \mid W_2 \geq p_B - w_1)\, dw_1 \\
&\leq \phi(\bar{p}_1) + \int_{\bar{p}_1}^{\overline{w}_1} \phi'(w_1) P_Y(W_1 > w_1 \mid W_2 \geq p_B - w_1)\, dw_1 \\
&= \mathbb{E}_Y\left[\phi_1(w_1) \mid W_2 \geq p_B - w_1\right].
\end{aligned}$$

Similarly we conclude for $\phi_2(w_2)$ and hence for $\phi_B(w_B)$.

This establishes that for a fixed price, the average cost of those who buy increases under Y relative to X when Y is more correlated than X in this sense.

This immediately implies that the equilibrium price is lower under X than under Y, since

$$p_B^Y = \mathbb{E}_Y\left[\phi_1(w_1) \mid \mathcal{W}_{\backslash p_B^Y}\right] \geq \mathbb{E}_X\left[\phi_1(w_1) \mid \mathcal{W}_{\backslash p_B^Y}\right]$$

and hence if firms sold the bundle to a population with true distribution X at price p_B^Y they would make a profit, and by the assumption of single crossing the true equilbirium price under X must be lower.

Finally, for the third part, by proposition 14, the demand curve for the bundle under X is above that for Y for all $q \geq \bar{q}$ where $\bar{q}$ is defined in the proof of proposition 14. That means, for any quantity insured beyond $\bar{q}$, the gap between demand and AC is higher under X than Y. In particular, at the equilibrium quantity q_B^Y the gap is zero under Y and positive under X, hence $q_B^X > q_B^Y$ as required.

■

B.5 Proof of Proposition 4

Proof. Essentially identical to the proof of proposition 1 except that since $\phi' < 0$ the inequality in costs is reversed under more correlated Y than X. In particular, if $X \precsim Y$ then

$$\mathbb{E}_X \left[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right] \geq \mathbb{E}_Y \left[\phi_1(w_1) \mid w_1 \geq \bar{p}_1 \wedge w_2 \leq \bar{p}_2\right]$$

and similarly for ϕ_2 and hence for phi_B.

■

B.6 Proof of Proposition 5

Proof. Similar to 3, except since ϕ_B is decreasing we apply proposition 13 to $-\phi_B$ which is increasing and supermodular by assumption. It

■

B.7 Proof of Proposition 7

Before the proof, I the key equilibrium characterization result from Crocker and Snow (2011).

Proposition 15. *(Crocker and Snow (2011)) In any constrained efficient allocation (in particular in the equilibrium with no cross-subsidization between types):*

- *The high types receive full insurance*
- *The low types receive partial insurance, limited by the high type's IC constraint*
- *The low types IC constraint is slack.*

Now we can prove proposition 7.

Proof. From proposition 15 the high type will get full insurance ($\boldsymbol{c}^*$)) and the low type will get the contract that solves:

$$\max V^L(\boldsymbol{c}^L) \tag{B.6}$$
$$\text{subject to} \tag{B.7}$$
$$IC_H : V^H(\boldsymbol{c}^*) \geq V^L(\boldsymbol{c}^L) \tag{B.8}$$
$$\text{Zero profit: } \pi_L = 0 \tag{B.9}$$

Substituting in the zero profit constraint gives the Lagrangian:

$$\mathcal{L} = V^L + \mu IC_H = u(c^*)(\mu - \mu p^H) - (\mu - \mu p^H + p^L - 1)u\left(\frac{p^L(c^1\theta^L + c^2(-\theta^L) + c^2 + l - 2w) + w}{1 - p^L}\right)$$
$$+ u(c^1)(\theta^L p^L - \theta^H \mu p^H) + \theta^H \mu p^H u(c^*) + \theta^H \mu p^H u(c^2) - \mu p^H u(c^2) - \theta^L p^L u(c^2)$$
$$+ p^L u(c^2) - \theta^H \mu p^H u(c^*) + \mu p^H u(c^*)$$

The change in welfare due to a decrease in θ^L is, since the high types get full insurance, proportional simply to the change in the maximized $V^L(\boldsymbol{c}^L)$. Hence, by the envelope theorem, after substituting in the zero profit constraint, the derivative of V^L and hence welfare with respect to θ^L is $\frac{\partial \mathcal{L}}{\partial \theta^L}$.

To compute this, first we calculate the value of μ, the multiplier, at the optimum. The first order condition with respect to c_1 is:

$$\frac{\theta^L p^L(\mu - \mu p^H + p^L - 1)u'\left(\frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}\right)}{p^L - 1} + u'(c_1)(\theta^L p^L - \theta^H \mu p^H) = 0$$

Solving this we get:

$$\mu = \frac{\theta^L(p^L - 1)p^L\left(u'\left(\frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}\right) + u'(c_1)\right)}{\theta^L(p^H - 1)p^L u'\left(\frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}\right) + \theta^H p^H(p^L - 1)u'(c_1)}$$

at the optimum.

Hence $\frac{\partial \mathcal{L}}{\partial \theta^L}$ after substituting in for μ is given by

$$\frac{\partial \mathcal{L}}{\partial \theta^L} = p^L\left(\frac{(c_1 - c_2)u'(c_1)(\theta^H p^H(p^L - 1) - \theta^L p^H p^L + \theta^L p^L)u'\left(\frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}\right)}{\theta^L(p^H - 1)p^L u'\left(\frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}\right) + \theta^H p^H(p^L - 1)u'(c_1)} + u(c_1) - u(c_2)\right). \tag{B.10}$$

Since $c_1 \leq c_0 = \frac{p^L(c_1\theta^L + c_2(-\theta^L) + c_2 + l - 2w) + w}{1 - p^L}$ the denominator of the fraction is negative. The numerator of the fraction is positive after noting that:

$$(\theta^H p^H(p^L - 1) - \theta^L p^H p^L + \theta^L p^L = -p_H\theta_H(1 - p_L) + p_L\theta_L(1 - p_H) < 0.$$

And since $\theta_L \leq \theta_H \iff \rho \leq 1$ this means (see Crocker and Snow (2011) Theorem 1) that $c_1 \leq c_2$ hence we have $\frac{\partial \mathcal{L}}{\partial \theta^L} < 0$ and so welfare increases as θ_L falls $\iff \rho$ goes towards zero. ■

B.8 Proof of Proposition 8

Proof. First, I show that, for a fixed price p, the total cost of those that buy the bundle $E(\phi_B(w_B) \mid w_B > p_B)$ is higher under Y than X when $X \precsim Y$. Note that $w_B > p_B$ iff $\phi(w_B) > \phi(p_B) \equiv \underline{\phi}$. Since $\phi_B(w_1 + w_2)$ is supermodular, by proposition 13, we have that

$$\phi_B(W_X) \precsim_{SL} \phi_B(W_Y).$$

In particular, at a stop-loss of $\underline{w}_B$ we have that

$$\mathbb{E}_X\left[\max\{\phi(w_B) - \underline{\phi}, 0\}\right] \leq \mathbb{E}_Y\left[\max\{\phi(w_B) - \underline{\phi}, 0\}\right].$$

This is equivalent to:

$$\begin{aligned} \mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right] \times \left(\phi(w_B) - \underline{\phi}\right)\right] &\leq \mathbb{E}_Y \left[\mathbb{1}\left[w_B > p_B\right] \times \left(\phi(w_B) - \underline{\phi}\right)\right] \\ \iff \mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B) - \mathbb{1}\left[w_B > p_B\right] \underline{\phi}\right] &\leq \mathbb{E}_Y \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B) - \mathbb{1}\left[w_B > p_B\right] \underline{\phi}\right] \\ \iff \mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B)\right] - \underline{\phi}\mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right]\right] &\leq \mathbb{E}_Y \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B)\right] - \underline{\phi}\mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right]\right] \\ \iff \mathbb{E}_X \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B)\right] &\leq \mathbb{E}_Y \left[\mathbb{1}\left[w_B > p_B\right] \phi(w_B)\right] \end{aligned}$$

where the final line follows from assumption 2.

But since average cost can be written as

$$AC_B = \mathbb{E}_X \left[\phi(w_B) \mid w_B > p_B\right] = \frac{\mathbb{E}\left[\mathbb{1}\left(w \in W_B\right) \times \phi_B(w)\right]}{\mathbb{E}\left[\mathbb{1}\left(w \in W_B\right)\right]}$$

it follows again by assumption 2 that we have

$$\mathbb{E}_X \left[\phi(w_B) \mid w_B > p_B\right] \leq \mathbb{E}_Y \left[\phi(w_B) \mid w_B > p_B\right]$$

This shows that at every price, the average cost decreases. In particular, at the equilibrium price under Y, P_B^Y, if before we had $p_B^Y = \mathbb{E}_Y \left[\phi(w_B) - \underline{\phi} \mid w_B > p_B^Y\right]$, under X we have $p_B^Y > \mathbb{E}_X \left[\phi(w_B) - \underline{\phi} \mid w_B > p_B^Y\right]$ implying, by the assumption of single crossing, that $p_B^X < p_B^Y$ the second part of the proposition states.

Finally, for the third part, by proposition 14, the demand curve for the bundle under X is above that for Y for all $q \geq \bar{q}$ where $\bar{q}$ is defined in the proof of proposition 14. That means, for any quantity insured beyond $\bar{q}$, the gap between demand and AC is higher under X than Y. In particular, at the equilibrium quantity q_B^Y the gap is zero under Y and positive under X, hence $q_B^X > q_B^Y$ as required.

■

B.9 Proof of Proposition 9

Proof. The total cost of those that buy the bundle at a small discount is

$$\mathbb{E}\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right] \times \left(\phi_1(w_1) + \phi_2(w_2)\right)\right].$$

Both $\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right]\right.$ and $\phi_1(w_1) + \phi_2(w_2)$ are supermodular and increasing functions and hence so is their product. Then by proposition 13 with a stop-loss of zero we have

$$\mathbb{E}_X\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right] \times \left(\phi_1(w_1) + \phi_2(w_2)\right)\right] \leq \mathbb{E}_Y\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right] \times \left(\phi_1(w_1) + \phi_2(w_2)\right)\right].$$

Then by assumption 2 this implies that the average costs of those that buy the bundle are ordered in the same way:

$$\frac{\mathbb{E}_X\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right] \times \left(\phi_1(w_1) + \phi_2(w_2)\right)\right]}{\mathbb{E}_X\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right]\right]\right.} \leq \frac{\mathbb{E}_Y\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right] \times \left(\phi_1(w_1) + \phi_2(w_2)\right)\right]}{\mathbb{E}_Y\left[\mathbb{1}\left[\left[w_1 \geq p_1 \wedge w_2 \geq p_2\right]\right]\right.}.$$

Since the bundled price is the same under each distribution, due to fact 1, it follows that the profits from offering the bundled product are ordered $\pi_X^\epsilon \geq \pi_Y^\epsilon$ as required.

■

B.10 Proof of Proposition 10

Proof. Similarly to the proof of proposition 8 except we apply the logic to the increasing $-\phi_B(w_B)$ instead of $\phi_B(w_B)$ with a stop loss of $-\underline{\phi}$ not $\underline{\phi}$. In particular, the bundle is bought if $w_B \geq p_B \iff \phi(w_B) \leq \underline{\phi} \iff -\phi(w_B) \geq -\underline{\phi}$. We then conclude that

$$\mathbb{E}_X \left[-\phi(w_B) \mid w_B > p_B\right] \leq \mathbb{E}_Y \left[-\phi(w_B) \mid w_B > p_B\right].$$

Multiplying by minus one and reversing the inequality shows the first part. The rest are analagous to the proof of proposition 8. ■

B.11 Proof of Proposition 12

Proof. By definition, the equilibrium prices $\boldsymbol{p}^Y$ will lead to zero profits being earned in each product market: $\pi_1^Y(\boldsymbol{p}^Y) = \pi_2^Y(\boldsymbol{p}^Y) = \pi_B^Y(\boldsymbol{p}^Y) = 0$. I will use $\mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_B$ to denote those who buy products 1, 2 and the bundle at prices $\boldsymbol{p}^Y$. As the prices under consideration will not change in this proof, I will supress that dependance. Naturally $\mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_B$ still depend on whether the correlation structure is X or Y.

Now, if the distribution is actually X with $X \precsim Y$, I first show that the average cost of those in $\mathcal{W}_1$ and $\mathcal{W}_2$ is increasing as correlation gets more negative. To see this, note that $\mathbb{1}\left[[w_1 < p_1^B \wedge w_2 \geq p_2^B\right]$ is sub-modular and non-increasing. $\phi_1(w_1)$ is submodular and non-decreasing. Their product is then submodular. Their product times minus one is therefore supermodular.

Then by proposition 13 and the same logic as proposition 8, we have

$$\mathbb{E}_X \left[\mathbb{1}\left[[w_1 < p_1 \wedge w_2 \geq p_2\right] \times \phi_1(w_1)\right] \geq \mathbb{E}_Y \left[\mathbb{1}\left[[w_1 < p_1 \wedge w_2 \geq p_2\right] \times \phi_1(w_1)\right].$$

Then by assumption 2 this implies that the average costs of those that buy only product 1 are ordered in the same way. This shows that profits in market 1 are negative, since they were zero under Y. Identical logic works for market 2.

For the bundled market, first, define $\underline{\phi} = \phi_B(p_B)$ and note that $\psi(w_B) = (\phi_B(w_B) - \underline{\phi}) \times \mathbb{1}\left[w_B > p_B\right]$ is supermodular. This can be seen case-wise. Note by assumption $\phi_B(w_B)$ and hence $\phi_B(w_B) - \underline{\phi}$ is supermodular. Take two points $w = (w_1, w_2), u = (u_1, u_2) \in \mathcal{W}$. Assume $w_1 > u_1$ and $w_2 < u_2$ else supermodularity is trivial. If $u, w \notin \mathcal{W}_B$ then $\psi(w) = \psi(u) = \psi(w \vee u) = 0 \leq \psi(w \wedge u)$ and supermodularity follows. If all four points $u, w, u \wedge w, u \vee w \in \mathcal{B}$ then supermodularity follows from the supermodularity of $\phi_B(w_B)$. If both of $u, w \in \mathcal{W}$ but $u \vee w \notin \mathcal{W}$, hence

$$(\phi_B(u \wedge w) - \underline{\phi}) < (\phi_B(u \wedge w) - \underline{\phi}) \times \mathbb{1}\left[u \wedge w > p_B\right] = \mathbb{1}\left[u \wedge w > p_B\right] = 0,$$

then by the supermodularity of $\phi_B(w_B) - \underline{\phi}$ we have

$$\begin{aligned}
\psi(w) + \psi(u) &= \phi_B(u) - \underline{\phi} + \phi_B(w) - \underline{\phi} \\
&< \phi_B(u \wedge w) - \underline{\phi} + \phi_B(u \vee w) - \underline{\phi} && \text{as } \phi_B \text{ is supermodular} \\
&< \phi_B(u \vee w) - \underline{\phi} && \text{as } \phi_B(u \vee w) - \underline{\phi} < 0 \\
&= \psi(u \vee w) \\
&= \psi(u \vee w) + \psi(u \wedge w) && \text{as } \psi(u \wedge w) = 0.
\end{aligned}$$

Finally, if $u \in \mathcal{W}_B$ but $w \notin \mathcal{W}_B$ then by the monotonicity of ϕ_B with respect to both arguments we have $\psi(u) \leq \psi(w \vee u)$ and $\psi(w) = \psi(w \wedge u) = 0$ from which supermodularity follows.

Given that $\psi(w_B)$ is non-decreasing and supermodular, as is $\mathbb{1}[w_2 > p_B - p_1 \wedge w_1 > p_B - p_2]$, their product is also supermodular. That is,

$$\mathbb{1}[w_2 > p_B - p_1 \wedge w_1 > p_B - p_2] \times (\phi_B(w_B) - \underline{\phi}) \times \mathbb{1}[w_B > p_B] = (\phi_B(w_B) - \underline{\phi}) \times \mathbb{1}[w_B \in \mathcal{W}_B]$$

is supermodular. Then following the logic in the proof of proposition 8 and using assumption 2 again, we get that

$$\mathbb{E}_X[\phi_B(w_B) \times \mathbb{1}[w_B \in \mathcal{W}_B]] \leq \mathbb{E}_Y[\phi_B(w_B) \times \mathbb{1}[w_B \in \mathcal{W}_B]].$$

Divide by $\mathbb{E}_X[\mathbb{1}[w_B \in \mathcal{W}_B]] \approx \mathbb{E}_Y[\mathbb{1}[w_B \in \mathcal{W}_B]]$ (assumption 2 again) to get that the average cost is lower under X than under Y. Hence if zero profits were made at Y positive profits are made under X.

■

C Microfoundation

In this section I present a microfoundation for the individual WTP and cost functions that were taken as given in the main body of the paper.

Each agent begins with wealth w. There are two risks that might occur, causing a loss of l_1 and l_2 dollars respectively. Each agent i privately knows their probability of each risk occurring. The risks occur independently.[(9)] Label these p_1^i, p_2^i. Consumption utility is evaluated according to utility function $u(\cdot)$ which is twice continuous differentiable. The expected utility that an individual would derive without any insurance is

$$V^i = (1 - p_1^i)(1 - p_2^i)u(w) + p_1^i(1 - p_2^i)u(w - l_1) + (1 - p_1^i)p_2^i u(w - l_2) + p_1^i p_2^i u(w - l_1 - l_2).$$

Suppose two fixed insurance contracts are offered to this agent. An insurance contract for risk 1 specifies a premium $\rho_i$to be paid in all states of the world and an indemnity ι_1 to be paid from the insurer to the insuree in case risk 1 occurs. I summarize the random variable that the insurance contract offers by the consumption vector in the no loss and loss state of the world: $I_1 = (-p_1, -p_1 + \iota_1)$, in place of the the raw loss random variable $L_1 = (0, -l_1)$. Similarly for risk 2.

Individual i's WTP for insurance contract 1 generally depends on whether they are insured against risk 2 or not. That is, their WTP for insurance contract 1 when they are insured against risk 2, $WTP_{1,B2}$ will differ from their WTP when they are not insured against risk 2 $WTP_{1,NB2}$.

These will respectively solve:

$$\mathbb{E}[u(w - WTP_{1,B2} - I_2)] = \mathbb{E}[u(w - L_1 - I_2)]$$
$$\mathbb{E}[u(w - WTP_{1,NB2} - L_2)] = \mathbb{E}[u(w - L_1 - L_2)].$$

Below I present different assumptions that are sufficient for $WTP_{1,B2} = WTP_{1,NB2}$ and hence for the assumption in section 2.2 that there is just one WTP_1 and WTP_2 regardless of whether the other contract is purchased.

(9) The distinction drawn throughout this paper is between correlation in *probabilities* (which I focus on) and correlation in *outcomes* which in this paper I assume away.

Definition 11. *The utility function $u(\cdot)$ is of CARA form when the coefficient of absolute risk aversion, $-u''(c)/u'(c) = \alpha$ is constant.*

Proposition 16. *Suppose $u(\cdot)$ is of CARA form. Then $WTP_{1,B2} = WTP_{1,NB2}$.*

Proof. First note that for CARA u and independent risks X_1 and X_2 it is the case that $\mathbb{E}(u(X_1 + X_2)) = \mathbb{E}(u(X_1)) \cdot \mathbb{E}(u(X_2))$. Hence, we have that

$$\begin{aligned} & \mathbb{E}\left[u\left(w - WTP_{1,NB2} - L_2\right)\right] = \mathbb{E}\left[u\left(w - L_1 - L_2\right)\right] \\ \iff & \mathbb{E}\left[u\left(w - WTP_{1,NB2}\right)\right] \mathbb{E}\left[u\left(-L_2\right)\right] = \mathbb{E}\left[u\left(w - L_1\right)\right] \mathbb{E}\left[u\left(-L_2\right)\right] \\ \iff & \mathbb{E}\left[u\left(w - WTP_{1,NB2}\right)\right] = \mathbb{E}\left[u\left(w - L_1\right)\right] \end{aligned}$$

Similarly,

$$\begin{aligned} & \mathbb{E}\left[u\left(w - WTP_{1,B2} - I_2\right)\right] = \mathbb{E}\left[u\left(w - L_1 - I_2\right)\right] \\ \iff & \mathbb{E}\left[u\left(w - WTP_{1,B2}\right)\right] \mathbb{E}\left[u\left(-I_2\right)\right] = \mathbb{E}\left[u\left(w - L_1\right)\right] \mathbb{E}\left[u\left(-I_2\right)\right] \\ \iff & \mathbb{E}\left[u\left(w - WTP_{1,B2}\right)\right] = \mathbb{E}\left[u\left(w - L_1\right)\right] \end{aligned}$$

Together this implies that $WTP_{1,B2} = WTP_{1,NB2}$ as required.

■

An alternate set of assumptions that are sufficient for the WTP for insurance ocntract oen to not depend on whether contract 2 is bought or not is that both risks are small in the following sense.

D Empirical Appendix

D.1 Behavioral Analysis with LTC as the base

D.2 Lapse Rates

The data come from an SOA experience study. The dependent variable is lapsation of the contract due to any factor other than mortality. The results are similar even if mortality is included.

The equation estimated is

$$lapse_i = \alpha_0 + \alpha_1 LivesExposed_i + \beta Married_i + \gamma Controls_i + \epsilon_i. \tag{D.1}$$

The controls included are as exhaustive as possible. Specifically: Observation year, gender, policy year, attained age (grouped), issued age (group), premium payment frequency, rate increase flag, inflation coverage (grouped), elimination period length (grouped). The results are in 10 below.

We see strong evidence that voluntary lapses are lower for married couples. Voluntary lapsation is profitable for the insurer, since premia are front-loaded relative to claims. This demonstrates that it is *not* a favorable lapsation pattern that leads to a spousal discount. This gives more weight to the evidence presented in the main body of the paper that the discount is in fact due to favorable cost correlations.

p_X	Objective Risk		Subjective Risk	
	$\mathbb{E}[p_{LTC} \mid p_{LTC} > p, p_{Mort} > p]$	$\Delta\%$	$\mathbb{E}[p_{LTC} \mid p_{LTC} > p, p_{Mort} > p]$	$\Delta\%$
p_{10}	0.04***	1.05***	0.13***	1.49***
	(0)	(0.17)	(0)	(0.14)
p_{20}	0.04***	1.31***	0.14***	0.95***
	(0)	(0.16)	(0)	(0.06)
p_{30}	0.05***	1.26***	0.16***	0.61***
	(0)	(0.13)	(0)	(0.03)
p_{40}	0.05***	1.24***	0.19***	0.39***
	(0)	(0.11)	(0)	(0.02)
p_{50}	0.06***	1.23***	0.23***	0.3***
	(0)	(0.11)	(0)	(0.02)
p_{60}	0.07***	1.05***	0.27***	0.17***
	(0)	(0.09)	(0)	(0.01)
p_{70}	0.09***	0.96***	0.33***	0.13***
	(0)	(0.08)	(0)	(0.01)
p_{80}	0.12***	0.68***	0.42***	0.07***
	(0)	(0.07)	(0)	(0.01)
p_{90}	0.19***	0.3***	0.57***	0.05***
	(0)	(0.05)	(0)	(0.01)

*** = 1% significance

Table 9: Savings when bundling mortality and long-term care risk together according to objective (predicted) versus subjective (elicited) probabilities.

Table 10: Single vs Married Voluntary Lapse Rates

	Without Controls	With Controls
Single	0.038***	0.021***
	(0.003)	(0.003)
Num.Obs.	682 104	682 104
R2	0.183	0.445

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001