

Risk Aversion and Double Marginalization*

Soheil Ghili[†]
Yale

Matt Schmitt[‡]
UCLA

April 17, 2019

Abstract

In vertical channels, eliminating double marginalization with a two-part tariff may not be possible due to downstream firms' risk aversion. When demand is uncertain, contracts with large fixed fees expose the downstream firm to more risk than contracts that are more reliant on variable fees. In equilibrium, contracts may thus rely on variable fees, giving rise to double marginalization. Counterintuitively, we show that increased demand risk or risk aversion can actually *mitigate* double marginalization. We also characterize several sufficient conditions under which increased risk or risk aversion does exacerbate double marginalization. We conclude with an application to channel coordination.

*We thank Larry Samuelson for detailed and insightful comments.

[†]E-mail: soheil.ghili@yale.edu

[‡]E-mail: matthew.schmitt@anderson.ucla.edu

1 Introduction

Following the pioneering work of Spengler (1950), double marginalization has become a foundational concept in the analysis of channels and vertical markets.¹ In empirical studies of vertical integration, the degree to which double marginalization is mitigated is often a primary focus (e.g., Chipty (2001); Crawford et al. (2018)). It is also well understood that alternatives to vertical integration of a channel can also remedy double marginalization (e.g., Mathewson and Winter (1984)). One such alternative is a two-part tariff.² By charging a linear fee equal to marginal cost, the upstream firm can maximize joint surplus – eliminating double marginalization – and then capture the entirety of that surplus via the fixed fee (e.g., Oi (1971)). In Jeuland and Shugan (1983) and numerous follow up papers, two-part tariffs have been studied as one of the main means to eliminate double marginalization.

While two-part tariffs can in principle eliminate double marginalization in vertical channels, there may be obstacles to their implementation. In this paper, we focus on one such possible obstacle: risk aversion.³ If the exact level of trade is uncertain at the time of contracting and the downstream firm is averse to that uncertainty, the upstream firm might need to “insure” the downstream firm against the risk of low demand volume. With a two-part tariff, doing so involves charging a smaller fixed fee, extracting rents via a larger linear fee instead. This contract adjustment lowers the variance of the downstream firm’s profits by making the payment to the upstream firm more sensitive to the realized demand volume. While such a contract insures the downstream firm against risk, it also gives rise to double marginalization as it involves a linear fee above the upstream firm’s marginal cost. This same basic argument has been made in several prior papers examining the use of two-part tariffs (e.g., Rey and Tirole (1986); Hayes (1987); Png and Wang (2010); Joskow (2012)).

Is double marginalization exacerbated by increases in downstream demand risk or risk aversion? The argument above suggests that the answer is likely yes: e.g., the more risk-averse the downstream firm, the more that the upstream firm will need to insure that firm from low demand. We show, however, that the analysis is complicated by two potentially competing economics forces. The first force – which we call “risk insurance” – is as already

¹Consider the countless intermediate microeconomics/industrial organization courses for which double marginalization is assuredly the first example of the importance of vertical market structure. For a more extensive treatment of vertical relations (including double marginalization), see, e.g., Katz (1989).

²Hereafter, we always refer to the fixed component of a two-part tariff as the “fixed fee” and the variable component as the “linear fee.”

³To be clear, there may be other reasons why firms may not utilize two-part tariffs besides risk aversion, such as intrabrand competition (Rey and Vergé (2008)), incomplete contracts (Iyer and Villas-Boas (2003)), and asymmetric information (Maskin and Riley (1984)).

described. As the magnitude of demand risk (or the downstream firm’s aversion to that risk) increases, the downstream firm assigns higher value to the fixed fee relative to the linear fee. The consequence is upward pressure on the linear fee and, hence, in double marginalization. The second force – which we call “risk compensation” – is that the value of trade for the downstream firm decreases in demand risk or downstream risk aversion. Ceteris paribus, the set of two-part tariff contracts that are acceptable to the downstream firm shrinks as risk or risk aversion increase. This force puts downward pressure both on the fixed fee and the linear fee, possibly *mitigating* double marginalization.⁴ To our knowledge, this force has not been mentioned in the literature.

We capture these forces in a bilateral monopoly model with a single upstream firm U and a single downstream firm D , where D is risk-averse. Firm U makes a take-it-or-leave-it contract offer to D consisting of a fixed fee T and a linear fee t . The level of downstream demand for the product is uncertain and D must accept or reject U ’s offer prior to the resolution of that uncertainty. If D accepts the offer, D sets the final price to consumers and then demand is determined as a function of the price and the resolution of the demand uncertainty. The greater the equilibrium linear fee, the greater the extent of double marginalization.

Weighing the effects of risk insurance and risk compensation is a complicated task even with functional form assumptions on the downstream demand function, the upstream production function, the distribution of demand risk, and D ’s risk attitude. We begin by providing sufficient conditions under which double marginalization increases in the level of demand risk and downstream risk aversion for any demand and production function. When D ’s utility function exhibits constant absolute risk aversion (CARA), double marginalization increases in the level of risk aversion, irrespective of the distribution of demand risk (Theorem 1).⁵ When the demand risk follows a uniform distribution, double marginalization increases in the level of demand risk, irrespective of D ’s risk attitude (Theorem 2). In other words, Theorems 1 and 2 specify sufficient conditions under which risk insurance always dominates risk compensation. Next, with concrete counterexamples, we demonstrate that the effect of risk insurance does *not* always dominate the effect of risk compensation. As a result, increases in risk or risk aversion can, in some circumstances, mitigate double marginalization (Theorem 3). One feature of these counterexamples is that D ’s utility function declines rapidly as D ’s profit falls below some “catastrophic” value. To illustrate the importance of this feature, we provide a refinement that places further restrictions on D ’s risk attitude to rule out the

⁴The terminology “risk compensation” derives from the idea that the disutility faced by the downstream firm from increases in risk or risk aversion must be “compensated for” by the upstream firm.

⁵A similar result obtains in Png and Wang (2010), who restrict their attention to the CARA case.

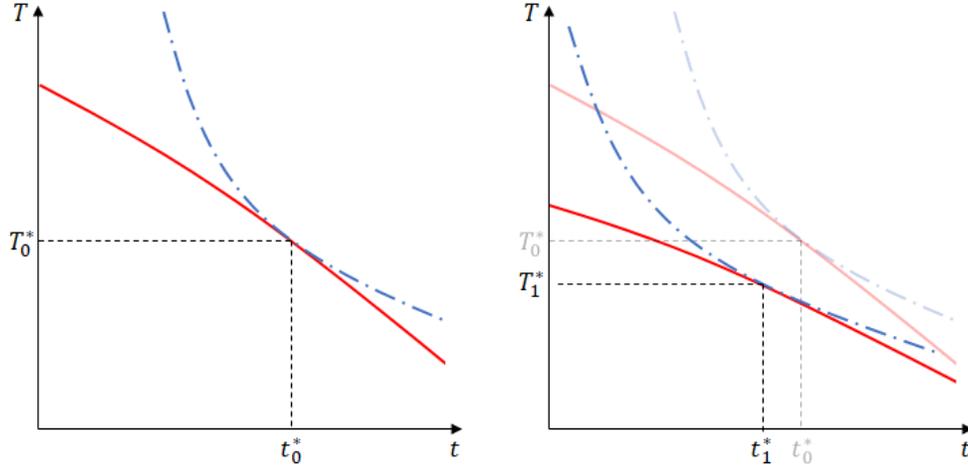


Figure 1: Equilibrium Two-Part Tariffs In both panels, the solid line denotes combinations of the linear fee t and fixed fee T such that D receives zero utility. The dashed line is U 's highest iso-profit curve that touches the solid line. The equilibrium two-part tariff is determined by the tangency point between U 's iso-profit curve and D 's zero-expected-value curve. An increase in risk or risk aversion can be thought of as generating (a) a counter-clockwise rotation (risk insurance) and (b) an inward shift (risk compensation) of D 's zero-expected-value curve. The right panel indicates that one possible consequence of an increase in risk or risk aversion is a lower equilibrium linear fee.

existence of this kind of catastrophic value (Theorem 4). Last, we show that certain demand always yields less double marginalization than uncertain demand, as does a risk-neutral D compared to a risk-averse D (Theorem 5).

To gain intuition for the effect of risk insurance and risk compensation on the equilibrium two-part tariff, consider Figure 1. In the model, the equilibrium two-part tariff is determined by the tangency point between U 's iso-profit curve and D 's zero-expected-value curve, as depicted in the left panel of the figure. Now consider the impact of an increase in risk (or risk aversion). First, D 's iso-utility curve that passes through the original equilibrium rotates counter-clockwise, reflecting increased disutility from increases in the fixed fee T relative to the linear fee t . This rotation captures the effect of risk insurance, which operates similarly to the substitution effect in demand theory. However, due to the increase in risk (or in downstream risk aversion), the original equilibrium contract may no longer be acceptable to D . The new set of contracts delivering zero utility to D will be an inward shift of the original set of contracts delivering zero utility to D . This shift captures the effect of risk compensation, which operates similarly to the income effect in demand theory. As drawn, the increase in risk has the effect of *reducing* the equilibrium linear fee, and hence reducing double marginalization as well. The aim of the paper is to explore circumstances under which this result can and cannot occur, and to demonstrate how seemingly innocuous functional

form assumptions can abstract away from important economic forces.

The results of the paper have applications to a variety of topics in marketing, economics, and operations. To illustrate one such application, we examine the impact of demand risk on channel coordination (i.e., the profitability of the vertical channel as a whole) in a simple benchmark model. In the model, an upstream manufacturer sells to a downstream retailer via a two-part tariff.⁶ Downstream demand is uncertain. In the case where the retailer is risk-neutral, channel coordination is unaffected by demand risk. The intuition is straightforward: since two-part tariffs are available, the upstream manufacturer charges the linear fee that maximizes the profitability of the channel, and then extracts those profits via the fixed fee. If the retailer is risk-averse, on the other hand, the analysis is considerably more nuanced. Depending on the parameters of the model, increases in demand risk and risk aversion can either worsen or improve channel coordination: i.e., the effects are not necessarily monotonic. These results suggest that incorporating risk aversion into models of channel coordination can fundamentally alter the analysis. The example emphasizes that complexities in the behavior of the channel in the presence of demand uncertainty can arise even when other parts of the environment are relatively simple. The tension between risk insurance and risk compensation can generate nuanced effects on channel behavior even without introducing other sources of channel complexity (e.g., bilateral oligopoly instead of bilateral monopoly, bargaining over contract terms instead of take-it-or-leave-it offers, etc.).

Another possible application of our results is in justifying some of the common assumptions made in empirical work on vertical markets. In the rapidly growing literature on bargaining in vertical markets,⁷ for instance, researchers typically assume that contracts take the form of linear fees. Our results suggest that one possible barrier to more complex contracts – like two-part tariffs – is demand risk. An assumption of linear fees can therefore potentially be motivated not only by an examination of actual contracts, but also in a more microfounded way by appealing to the nature of demand risk in the industry (and downstream aversion to that risk). Depending on the specific nature of the demand risk, our results show that increased risk may either strengthen or weaken the justification for assuming linear fees, and similarly for increased risk aversion.

Finally, it is worth noting at the outset that the standard assumption in the literature is to model firms as risk-neutral. However, we believe that studying the implications of

⁶The example can also be interpreted within the context of a vertically integrated firm with separately managed manufacturing and retailing divisions.

⁷Recent examples include Gowrisankaran et al. (2015), Ho and Lee (2017), Crawford et al. (2018), Ghili (2018), Ho and Lee (2018), and Liebman (2018), among many others.

firm risk aversion for market outcomes is highly pertinent (see, e.g., Asplund (2002) and Banal-Estañol and Ottaviani (2006) for other examples). Firms may behave as if they were risk-averse for many reasons, from tax consequences to liquidity constraints to delegating decision-making power to risk-averse managers (e.g., Amihud and Lev (1981); Nance et al. (1993)). In recent litigation, the US Third Circuit Court of Appeals ruled that risk aversion on the behalf of a branded pharmaceutical manufacturer plausibly explained the firm conduct under examination.⁸ Beyond business-to-business transactions, our results are also relevant for firms’ pricing strategies when the buyer is the final consumer. An important question in this literature is the optimal use of two-part tariffs – or, more generally, volume discounts – for the seller. Empirical evidence suggests that risk aversion and other behavioral forces can lead to inefficient contracts (e.g., Lambrecht and Skiera (2006); Narayanan et al. (2007)). Our analysis provides additional theoretical backing for these empirical findings.

The rest of the paper is organized as follows. Section 2 sets up the model and develops preliminary results. Section 3 presents the main results. Section 4 discusses extensions of the model. Section 5 illustrates an application of the results to channel coordination. Section 6 concludes.

2 Model

There is a single upstream firm U producing a single product and a single downstream firm D . The two firms play a game according to the following timeline:

1. Firm U offers a two-part tariff contract (T, t) to firm D , where T denotes the fixed fee and t denotes the linear fee.
2. Firm D decides whether to accept or reject the offer. If it rejects, the game ends and each firm receives zero profits.
3. If D accepts the offer, it decides the per-unit downstream price p at which it sells the product to final consumers.
4. The demand volume q for the product is realized according to $q = \alpha \cdot Q(p)$, where the continuous, non-negative, and weakly decreasing function $Q(p)$ is the expected level of

⁸In re: Wellbutrin XL Antitrust Litigation, No. 15-2875 (3d Cir. 2017). Risk aversion has been a major focus of recent debates concerning the effects of “reverse payment” patent settlements in the industry (e.g., Edlin et al. (2013); Harris et al. (2014)).

demand for the product at price p .⁹

Multiplier α is a realized draw from probability distribution f over non-negative real values. As a normalization without loss of generality, we assume that the expected value of α is always one. Let \mathcal{F} denote the set of all f that satisfy this normalization.

At the end of stage four, the profits to U and D are given by:

$$\pi_U(T, t|\alpha) = T + t \cdot \alpha \cdot Q(p(t)) - C(\alpha \cdot Q(p(t))) \quad (1)$$

$$\pi_D(T, t|\alpha) = -T + (p(t) - t) \cdot \alpha \cdot Q(p(t)), \quad (2)$$

where $C(\cdot)$ is U 's cost function and $p(t)$ is the optimal price charged by D (a function of the linear fee t).¹⁰

We assume that U is risk-neutral, whereas D evaluates profit realizations according to an increasing and concave function $v(\cdot)$. Without loss of generality, we assume that $v(0) = 0$ and $v'(0) = 1$.¹¹ Let \mathcal{V} denote the set of all such functions. The expected payoffs to U and D from the contract are thus given by:

$$\mathbf{E}\pi_U(T, t) = \int_0^\infty \pi_U(T, t|\alpha) f(\alpha) d\alpha \quad (3)$$

$$\mathbf{E}v_D(T, t) = \int_0^\infty v(\pi_D(T, t|\alpha)) f(\alpha) d\alpha. \quad (4)$$

2.1 Preliminary results

In this section, we develop several illustrative results to demonstrate the presence of (i) risk insurance and (ii) risk compensation, the forces that govern the response of the equilibrium contract to increases in risk or risk aversion. Throughout, we assume that the equilibrium contract (T^*, t^*) is unique.¹² This assumption is not crucial for the results, but substantially simplifies the presentation as the comparative statics thus involve comparing real numbers instead of set order comparisons. In the appendix, we discuss mild assumptions that would deliver uniqueness.

⁹This formulation assumes that D is uncertain only about the volume of demand, not the price elasticity. The main benefit of this assumption is that D 's optimal price does not depend on the value of α .

¹⁰That is, $p(t) = \operatorname{argmax}_p (p - t) \cdot Q(p)$.

¹¹If the derivative at zero does not exist, we assume that the *right* derivative at zero – which is guaranteed to exist by concavity – is equal to one.

¹²Note that this automatically rules out equilibria in which D rejects U 's offer (T^*, t^*) . In such cases, U could offer (T, t^*) for any $T > T^*$ and it would still get rejected, bringing about the same payoffs to both U and D . Thus any such (T, t^*) would also be part of an equilibrium, contradicting uniqueness.

Unlike the task of weighing the effects of risk insurance and risk compensation against one another, establishing the fact that each of these two forces exists is more straightforward. We begin with a definition characterizing the slope of D 's indifference curves.

Definition 1. Given (f, v) , denote D 's marginal rate of substitution between the linear fee and the fixed fee at contract (T, t) by $\Delta(T, t|f, v)$. $\Delta(T, t|f, v)$ is defined as:

$$\Delta(T, t|f, v) = \frac{\frac{\partial \mathbf{E}v_D(T, t)}{\partial t}}{\frac{\partial \mathbf{E}v_D(T, t)}{\partial T}}.$$

We now examine how D 's marginal rate of substitution changes with demand risk and risk aversion. For the following results, we assume that f is a uniform distribution over the interval $[1 - \sigma, 1 + \sigma]$ for some $\sigma \in [0, 1]$. The purpose of this assumption is to simplify the meaning of increased demand risk: given uniformity, increases in risk are captured by increases in σ . Increases in risk aversion are captured by increases in the concavity of v .

Proposition 1. (*risk insurance*)

a. For all $v \in \mathcal{V}$ and $\sigma_1, \sigma_2 \in [0, 1]$ such that $\sigma_2 \geq \sigma_1$:

$$\Delta(T, t|\sigma_2, v) \leq \Delta(T, t|\sigma_1, v) \quad \forall T, t.$$

b. For all $\sigma \in [0, 1]$ and $v_1, v_2 \in \mathcal{V}$ such that v_2 is more concave than v_1 :

$$\Delta(T, t|\sigma, v_2) \leq \Delta(T, t|\sigma, v_1) \quad \forall T, t.$$

Proof. In the appendix.

Proposition 1 states that an increase either in the magnitude of the risk faced by D or in D 's risk aversion changes D 's marginal rate of substitution in favor of the linear fee. That is, D 's disutility from being charged an additional dollar in the fixed fee relative to an additional dollar in the linear fee increases with risk or risk aversion. This result is intuitive because – unlike fixed fee payments – linear fee payments scale with the demand volume, and hence “insure” D against the risk in α .

If the risk insurance effect established by Proposition 1 was the only force present, one would expect the equilibrium linear fee to always increase in risk or risk aversion. However, higher risk or risk aversion also make non-trade (i.e., D rejecting U 's offer in stage two of the game) a more attractive option to D . To battle this effect, U has to offer a more attractive

contract, which puts downward pressure on both the fixed fee and the linear fee. Proposition 2 below formalizes the idea of risk compensation.

Proposition 2. (*risk compensation*)

a. For all $\sigma \in [0, 1]$ and $v \in \mathcal{V}$, the equilibrium contract (T^*, t^*) satisfies:

$$\mathbf{E}v_D(T^*, t^* | \sigma, v) = 0.$$

b. For all $v \in \mathcal{V}$ and $\sigma_1, \sigma_2 \in [0, 1]$ such that $\sigma_2 \geq \sigma_1$:

$$\mathbf{E}v_D(T, t | \sigma_1, v) \geq \mathbf{E}v_D(T, t | \sigma_2, v) \quad \forall T, t.$$

c. For all $\sigma \in [0, 1]$ and $v_1, v_2 \in \mathcal{V}$ such that v_2 is more concave than v_1 :

$$v_1^{-1}(\mathbf{E}v_D(T, t | \sigma, v_1)) \geq v_2^{-1}(\mathbf{E}v_D(T, t | \sigma, v_2)) \quad \forall T, t.$$

(To clarify, part c of the proposition states that the certainty equivalent for T, t is larger under v_1 than under v_2 .)

Proof. In the appendix.

Proposition 2 shows that if risk or risk aversion increases from the current configuration (f, v) , then the equilibrium contract under (f, v) may no longer be acceptable to D . D receives zero utility in equilibrium (part a), so an increase in risk (part b) or risk aversion (part c) may make the current equilibrium deliver D less utility than the outside option. Hence, U needs to adjust by lowering either the fixed fee, the linear fee, or both. Thus, the negative utility to D from trade due to extra risk or risk aversion must be “compensated for” by U .

The fact that risk insurance and risk compensation have opposing effects on the equilibrium linear fee suggests that it is essential to compare the sizes of these two forces to understand the overall direction of the response of the equilibrium linear fee – and thus double marginalization. We turn to this task next.

Example: risk and double marginalization

To motivate the relevance of demand risk in generating double marginalization, we first prove – under specific functional form assumptions on the demand and cost functions $Q(p)$

and $C(q)$ – that the equilibrium contract can give rise to either zero or complete double marginalization depending on D 's level of risk aversion.

Proposition 3. *Let $C(q) = cq$ for some $c \geq 0$ and $Q(p) = a - p$ for some $a > c$. Then:*

a. If f has no mass at any point other than 1 or $v(x) = x$:

$$t^*|f, v = c.$$

b. If f has mass at 0 and v is “extremely concave”¹³:

$$t^*|f, v = \frac{a + c}{2}.$$

Proof. In the appendix.

Note that $\frac{a+c}{2}$ is the optimal linear fee if no fixed fee is allowed (i.e., $T = 0$). Thus, Proposition 3 demonstrates that both zero and complete double marginalization are possible in equilibrium. In the main results that follow, our core goal is essentially to understand whether the degree of double marginalization is monotone in the level of risk and D 's risk aversion, without relying on the illustrative assumptions of Proposition 3.

3 Main Results

This section is organized as follows. In section 3.1, we show that if D has a constant absolute risk aversion (CARA) utility function, then increases in risk aversion exacerbate double marginalization (irrespective of the demand risk distribution f). In section 3.2, we show that if demand risk is uniform, then increases in demand risk exacerbate double marginalization (irrespective of D 's utility function v). In section 3.3, we provide counterexamples showing that these results are not fully general: increases in risk or risk aversion can *mitigate* double marginalization. Sections 3.4 and 3.5 provide further discussion and results.

¹³By “extremely concave,” we mean a v that makes D only care about the worst possible demand state (i.e., $\alpha = 0$). Such a function can be constructed as the limiting case of iteratively applying a strictly concave transformation $v_0 \in \mathcal{V}$ on the identity function.

3.1 CARA utility

We first consider constant absolute risk aversion (CARA) utility functions defined as follows:

$$v(x) = \frac{1 - \exp(-rx)}{r}, \quad (5)$$

where r is the coefficient of risk aversion. For any $r > 0$, the CARA utility function is a member of \mathcal{V} . In addition, if $r_2 \geq r_1$, a CARA utility function with risk coefficient r_2 is more concave than a CARA utility function with risk coefficient r_1 . Theorem 1 shows that if v is CARA, then the equilibrium linear fee increases in the risk coefficient r , irrespective of the demand risk distribution f .

Theorem 1. *Let v_1 and v_2 denote CARA utility functions with risk coefficients r_1 and r_2 (respectively). For all $r_2 \geq r_1 > 0$ and all $f \in \mathcal{F}$:*

$$t^*|f, v_1 \leq t^*|f, v_2.$$

That is, with CARA utility, the equilibrium linear fee – and thus the degree of double marginalization – increases in the magnitude of downstream risk aversion. Png and Wang (2010) obtain a similar result.

Proof. The full proof is in the appendix, but we discuss the proof strategy here to build intuition. Denote the equilibrium contract for f and v_1 by (T', t') . Following Definition 1, denote D 's marginal rate of substitution in this equilibrium by $\Delta(T', t'|f, v_1)$. To prove the theorem, we evaluate the overall direction of the change in D 's marginal rate of substitution as a result of two changes: (i) concavifying v_1 to v_2 , and (ii) lowering the fixed fee from T' to the new level that satisfies D 's zero-expected-value condition (part a of Proposition 2) given the more risk-averse utility function v_2 .¹⁴ Denote this adjusted fixed fee by $T(v_2)$.¹⁵ We should expect that the equilibrium linear fee will increase if D becomes relatively more sensitive to changes in the fixed fee even after being fully compensated for the increase in risk aversion. That is, if:

$$\Delta(T(v_2), t'|f, v_2) \leq \Delta(T', t'|f, v_1). \quad (6)$$

The equilibrium contract is determined by the tangency point between U 's highest iso-profit

¹⁴As stated in the main text, Proposition 2 assumes uniform demand risk. However, part a (and part c) of the proposition holds for all $f \in \mathcal{F}$. See the appendix for details.

¹⁵That is, $T(v_2)$ is the fixed fee such that $\mathbf{E}v_D(T(v_2), t'|f, v_2) = 0$.

curve and the zero-expected-value curve for D , where both U and D have the same marginal rate of substitution between the linear fee and the fixed fee. Neither concavifying v_1 to v_2 nor lowering the fixed fee from T' to $T(v_2)$ changes U 's marginal rate of substitution. Thus, if the two changes overall increase D 's sensitivity to the fixed fee relative to the linear fee – i.e., if the inequality in (6) holds – U will charge a higher linear fee in the new equilibrium.

To see that the inequality in (6) holds when v_1 and v_2 are CARA, first note by part b of Proposition 1 that:

$$\Delta(T', t'|f, v_2) \leq \Delta(T', t'|f, v_1), \quad (7)$$

because v_2 is more concave than v_1 .¹⁶ In addition, it follows directly from the CARA utility function (equation (5)) that:

$$\Delta(T(v_2), t'|f, v_2) = \Delta(T', t'|f, v_2). \quad (8)$$

Together, (7) and (8) imply that the inequality in (6) holds. The particular feature of CARA utility in the proof is that constant *absolute* risk aversion implies that changes in the fixed fee alone do not affect D 's marginal rate of substitution. Therefore, risk compensation does not impact the linear fee with CARA utility, whereas risk insurance puts upward pressure on the linear fee, thus resulting in increased double marginalization.

3.2 Uniform demand risk

Theorem 1 shows that when D 's utility function is CARA, the degree of double marginalization increases as D becomes more risk-averse – irrespective of the demand risk distribution f . We now provide a similar theorem for the magnitude of demand risk – namely, when demand risk is uniform – without imposing any additional assumptions on D 's risk attitude.

Theorem 2. *Let f_1 and f_2 denote uniform distributions on the intervals $[1 - \sigma_1, 1 + \sigma_1]$ and $[1 - \sigma_2, 1 + \sigma_2]$ (respectively). For all $\sigma_1, \sigma_2 \in [0, 1]$ such that $\sigma_2 \geq \sigma_1$ and all $v \in \mathcal{V}$:*

$$t^*|f_1, v \leq t^*|f_2, v.$$

That is, with uniform demand risk, the equilibrium linear fee – and thus the degree of double marginalization – increases in the magnitude of the demand risk.

¹⁶As stated in the main text, Proposition 1 assumes uniform demand risk. However, part b of the proposition holds for all $f \in \mathcal{F}$. See the appendix for details.

Proof. In the appendix. The first step in the proof for Theorem 2 is similar to Theorem 1 in that it examines the effects of risk insurance and risk compensation on D 's marginal rate of substitution. However, the implementation of the proof is substantially more complicated because equation (8) need not hold once the CARA assumption is relaxed. In short, the effect of risk compensation on D 's marginal rate of substitution can be positive, putting downward pressure on the linear fee. Therefore, unlike the CARA case, the case with generic v presents a real tension between risk insurance and risk compensation. Our proof employs multiple geometric and algebraic ideas that make the comparison between the two forces possible. We demonstrate that, under the assumptions of Theorem 2, risk insurance dominates risk compensation, leading to a monotone reaction of the equilibrium linear fee to the amount of demand risk.

3.3 Counterexamples

The two theorems presented so far both describe conditions under which the overall impact of increased risk or risk aversion is to exacerbate double marginalization. A natural follow-up question is whether these results also hold in general: Theorem 3 below establishes that the answer is no.

Theorem 3.

a. There exist $f \in \mathcal{F}$ and $v_1, v_2 \in \mathcal{V}$ such that v_2 is more concave than v_1 and:

$$t^*|f, v_1 > t^*|f, v_2.$$

b. There exist $v \in \mathcal{V}$ and $f_1, f_2 \in \mathcal{F}$ such that f_2 is a mean preserving spread of f_1 and:

$$t^*|f_1, v > t^*|f_2, v.$$

That is, increases in risk or risk aversion can result in a *lower* equilibrium linear fee (and hence less double marginalization).

Proof. We prove Theorem 3 with a concrete example. Let $C(q) = 0$ and $Q(p) = 4 - p$. α is drawn from a probability distribution with three points: 0.4, 1, and 1.6. The probability that $\alpha = 0.4$ is $r \in [0, 0.5]$, as is the probability that $\alpha = 1.6$. Let f_1 denote this distribution with $r = 0.03$ and f_2 denote this distribution with $r = 0.10$. f_2 is a mean preserving spread

of f_1 . The utility function v_1 is given by:

$$v_1(x) = \begin{cases} x, & x \geq 0 \\ 2x, & -2 \leq x < 0 \\ 12 + 8x, & x < -2. \end{cases} \quad (9)$$

That is, v_1 is continuous piecewise linear with a kink at $x = 0$ and another kink at $x = -2$. The slope of v_1 is 2 between $x = 0$ and $x = -2$ and 8 for $x < -2$. Let v_2 take the same form as v_1 but with steeper slopes for $x < 0$: the slope of v_2 is 3 between $x = 0$ and $x = -2$ and 12 for $x < -2$. The utility function v_2 is more concave than v_1 .

To demonstrate part a of Theorem 3, we solve for the equilibrium linear fee for v_1 compared to v_2 , holding the distribution of risk at f_1 : $t^*|f_1, v_1 = 0.294$ and $t^*|f_1, v_2 = 0.241$. That is, the increase in risk aversion reduced the equilibrium linear fee by 18 percent. To demonstrate part b of Theorem 3, we solve for the equilibrium linear fee for f_1 compared to f_2 , holding D 's utility function at v_1 : $t^*|f_1, v_1 = 0.294$ and $t^*|f_2, v_1 = 0.207$. That is, the increase in risk reduced the equilibrium linear fee by 30 percent. These reductions in linear fees then pass through to downstream prices. The example for part a of the theorem implies 1.2 percent lower downstream prices and the example for part b implies 2.0 percent lower downstream prices.¹⁷

3.4 Discussion and a refinement

The economics of the result that increases in risk or risk aversion can mitigate double marginalization is best understood by considering the possible effects of risk compensation on D 's marginal rate of substitution. Upon an increase in risk or risk aversion, U must compensate D by lowering either the fixed fee or the linear fee (or both). This compensation shifts the distribution of D 's possible realized profits to the right. This rightward shift of D 's profit distribution can substantially alter D 's marginal rate of substitution if D 's utility function shows strong aversion to profits below some ‘‘catastrophic’’ level (-2 in the example that proves Theorem 3). Due to the stark difference between D 's risk attitude above and below this catastrophic level, a rightward shift of D 's profit distribution can lead to D 's utility function over the new range of profits exhibiting *effectively less* risk aversion. As a result, there may be less need for U to provide risk insurance, and hence U may benefit from

¹⁷For part a of Theorem 3, it is also possible to construct an example with a uniform f in which double marginalization is mitigated by an increase in risk aversion. Such an example is available upon request.

offering a linear fee that results in less double marginalization.¹⁸

The possible presence of such a catastrophic profit level is arguably more than a mere theoretical curiosity. For instance, consider the possibility of bankruptcy. If large fixed fees are capable of exposing a downstream firm to possible bankruptcy, it is plausible that the firm would become substantially more risk-averse in the profit range that makes the firm prone to bankruptcy. Another possibility is risk aversion on the part of a manager in charge of procurement. Suppose that D is a large firm that delegates the procurement of a particular product to a manager. In this case, v can be thought of as the manager's utility function defined over D 's profit from the specific product that U sells. This utility function might rapidly steepen once the profit goes below a level that the manager begins to worry about his or her job security.

To confirm this intuition about the counterexamples in section 3.3, we prove a theorem that places additional restrictions on v to rule out the existence of a catastrophic profit level at which D 's risk attitude changes substantially. To state the theorem, it is helpful to first establish some notation. For every $v \in \mathcal{V}$ and $L > 0$, let $a_{v,L}$ be the number that satisfies $\int_{a_{v,L}}^{L+a_{v,L}} v(x)dx = 0$. That is, $a_{v,L}$ is the beginning of the interval of length L that has an expected value of zero (according to v) if the distribution of α is uniform.

Definition 2. *Function $v_2 \in \mathcal{V}$ is **globally more risk-averse** than $v_1 \in \mathcal{V}$ if for every $L > 0$, v_2 on the interval $[a_{v_2,L}, L + a_{v_2,L}]$ is more concave than v_1 on the interval $[a_{v_1,L}, L + a_{v_1,L}]$ in the following sense:*

$$\frac{-v_2(a_{v_2,L})}{v_2(L + a_{v_2,L})} \geq \frac{-v_1(a_{v_1,L})}{v_1(L + a_{v_1,L})}. \quad (10)$$

To gain intuition about how the expressions in (10) measure concavity, note that the mean of v_i ($i \in \{1, 2\}$) over $[a_{v_i,L}, L + a_{v_i,L}]$ is zero. Since v_i is increasing, the more concave v_i is, the closer this mean is to $v_i(L + a_{v_i,L})$. The less concave v_i is, the closer the mean is to $v_i(a_{v_i,L})$. Thus, a measure of concavity of the function over this interval is $\frac{-v_i(a_{v_i,L})}{v_i(L + a_{v_i,L})}$, with larger values indicating more concavity.

Theorem 4. *Let f denote a uniform distribution on the interval $[1 - \sigma, 1 + \sigma]$. For all $\sigma \in [0, 1]$ and $v_1, v_2 \in \mathcal{V}$ such that v_2 is globally more risk-averse than v_1 :*

$$t^*|f, v_1 \leq t^*|f, v_2.$$

¹⁸See the last paragraph of this subsection for a discussion of this intuition as it relates to Theorem 2, in which increases in demand risk always exacerbate double marginalization even if v exhibits the catastrophe-aversion explained here.

Proof. In the appendix.

Compared to Theorem 1 for CARA utility, Theorem 4 substantially relaxes the assumptions on D 's risk attitude, but requires a stronger assumption on the demand risk (uniform). One consequence of Theorem 4 is that different classes of utility functions can be classified according to whether they become globally more risk-averse as they are concavified. If so, those utility functions will imply higher equilibrium linear fees – and hence double marginalization – as they become more concave (i.e., for the case of uniform demand risk).¹⁹

One additional point worth further discussion is the comparison between Theorem 2 and Theorem 3. Theorem 2 shows that when f is uniform, increased risk exacerbates double marginalization regardless of v : i.e., even if v features the catastrophe-aversion that motivated the development of Theorem 4. A natural question is thus how the nature of the demand risk in the counterexample for Theorem 3 differs from the uniform case. With uniform demand risk, demand becomes riskier by *stretching out*. In the counterexample for Theorem 3, on the other hand, demand becomes riskier by having higher probabilities on the *same extreme points*. In the latter case, increases in demand risk leave the range of possible profits for D unchanged. Thus, after U provides risk compensation by lowering the either the fixed fee or the linear fee (or both), the overall effect on D 's minimum possible profit realization is positive. This change implies that the resulting range of possible profits for D covers a smaller range beyond the catastrophic level (if such a level exists), which reduces the need for U to provide risk insurance. With uniform demand risk, however, an increase in risk decreases the minimum possible profit for D , which amplifies the need for U to provide risk insurance.

3.5 Can increased risk or risk aversion *always* mitigate double marginalization?

The analysis thus far asks, under different conditions, whether it is true that increases in risk or risk aversion *always exacerbate* double marginalization. What these results leave unanswered is whether there are general conditions under which increases in risk or risk aversion *always mitigate* double marginalization. The answer – provided that those conditions do not rule out (i) certain demand and (ii) risk-neutrality – is no.

¹⁹CARA utility functions become globally more risk-averse as they are concavified. It is harder to analytically verify the conditions of Definition 2 for constant relative risk aversion (CRRA) utility functions, but our simulations over a wide range of values suggest that CRRA utility functions also become globally more risk-averse as they are concavified. Thus, Theorem 4 can likely be invoked for CRRA as well.

Theorem 5.

a. If $v \in \mathcal{V}$ is strictly concave and $f_1, f_2 \in \mathcal{F}$ are such that f_1 is degenerate (i.e., $F(x) = 1_{x \geq 1}$) and f_2 is non-degenerate, then:

$$t^*|_{f_1, v} < t^*|_{f_2, v}$$

b. If $f \in \mathcal{F}$ is non-degenerate and $v_1, v_2 \in \mathcal{V}$ are such that $v_1(x) = x$ and v_2 is strictly concave, then:

$$t^*|_{f, v_1} < t^*|_{f, v_2}$$

Proof. In the appendix.

Part a of Theorem 5 states that, independent of features of v , an uncertain f yields a higher linear fee – and hence more double marginalization – than a certain f . Part b of Theorem 5 states that, independent of features of f , a risk-averse v yields a higher linear fee – and hence more double marginalization – than a risk neutral v . Therefore, there are no general conditions on f or v under which increases in risk or risk aversion *always* mitigate double marginalization. The intuition for these results is best understood when considering the specific case of a linear cost function. In that case, as part a of Proposition 3 illustrates, zero risk or zero risk aversion will result in a linear fee equal to marginal cost, thus eliminating double marginalization. The proofs in the appendix address the more general case where the cost function is not required to be linear.

4 Extensions

The model developed in this paper was chosen to illustrate the economic forces of risk insurance and risk compensation in as parsimonious a model as possible. Given that a crucial part of the analysis is to study how the shape of the downstream utility function can lead to counterintuitive results, we chose to be as flexible as possible with respect to parametric assumptions. Consequently, we needed to be more parsimonious with other aspects of the model in order to retain the feasibility of attaining analytical results. In this section, we provide an informal discussion of some of the most natural possible extensions of the model. With each extension, we argue that risk insurance and risk compensation likely remain economically relevant forces in determining the shape of contracts.

4.1 Bargaining over the contract

As modeled, U makes a take-it-or-leave-it contract offer to D . Suppose instead that the contract is determined by Nash bargaining between U and D (e.g., Binmore et al. (1986)). That is, where the equilibrium contract is the contract that maximizes $[\mathbf{E}\pi_U(T, t)]^\zeta [\mathbf{E}v_D(T, t)]^{1-\zeta}$, subject to the constraints that U and D receive non-negative payoffs. $\zeta \in [0, 1]$ captures U 's "bargaining power." The model we examine is equivalent to the case where U has all of the bargaining power ($\zeta = 1$). For other ζ , the solution still involves equating U and D 's marginal rates of substitution between the linear fee and the fixed fee. Risk compensation is also still present because an increase in risk or risk aversion may make non-trade a relatively more attractive option for D , which acts similarly to an increase in the value of D 's outside option. In short, when contracts are determined via Nash bargaining, risk insurance and risk compensation continue to govern the responsiveness of the equilibrium contract to changes in risk or risk aversion. Crucially, this is not to say that bargaining does not matter for the double marginalization problem. Papers such as Iyer and Villas-Boas (2003) clearly highlight the role of bargaining. Our conjecture, however, is that bargaining does not change the fact that there is a tension between risk insurance and risk compensation. Hence, bargaining does not change the fact that increased risk or risk aversion might have non-monotonic effects on the extent of double marginalization.

4.2 Upstream risk aversion

As modeled, U is risk-neutral while D is risk-averse. Suppose instead that U is risk-averse while D is risk-neutral. Risk insurance and risk compensation are still present in that case, but work in the opposite direction. In contrast to the analysis above, increases in risk or risk aversion put downward pressure on the linear fee (and upward pressure on the fixed fee) via risk insurance and upward pressure on both the linear fee and the fixed fee via risk compensation. If both U and D are risk averse, outcomes will be determined in part by the relative strength of upstream and downstream risk aversion. Our results are thus likely to be most directly relevant in cases where D is more risk-averse than U . Empirically, to the extent that risk aversion is correlated with firm size, our results may be most applicable to contracting between large upstream suppliers and comparatively smaller downstream buyers.

4.3 Non-iso-elastic risk

Our model assumes that demand risk is iso-elastic. That is, the only uncertainty faced by D is over the volume of demand. In reality, D may also face uncertainty over the price elasticity of final customers. The chief difficulty in allowing non-iso-elastic risk is that D 's optimal downstream price would become a function not only of the linear fee, but also of the realized price elasticity. Nevertheless, we expect that risk insurance and risk compensation remain relevant. When there is considerable risk of highly price-sensitive consumers and D has sufficient aversion to that risk, U may optimally provide risk insurance by offering a contract that is more heavily reliant on the linear fee than the fixed fee. Moreover, when the risk over the price elasticity increases, U will need to offer more attractive contracts to D – i.e., provide risk compensation – to prevent D from deviating to the outside option of non-trade.

4.4 Downstream oligopoly

Instead of a single downstream firm D , one could think of K downstream firms D_1 through D_K . These downstream firms could face iso-elastic risks (e.g., uncertainty about the total size of the market) and/or non-iso-elastic risks (e.g., uncertainty about each firm's competitive position within the market). In both cases, increased risk or risk aversion would likely (i) make downstream firms less willing to accept high fixed fees compared to linear fees (risk insurance), and (ii) require U to offer better terms to sustain trade (risk compensation). Of course, downstream competition may also introduce other reasons for U to charge linear fees above cost, such as softening price competition downstream (e.g., Rey and Vergé (2008)).

Similar to the issue of bargaining, our discussion here does not argue against the general importance of retailer competition for channel coordination and double marginalization. Numerous papers (e.g., Marx and Shaffer (2007); Bryan and Hovenkamp (2016); Inderst and Shaffer (2018); Staahl Gabrielsen et al. (2018)) study the role of such competition. Our point is that we expect risk insurance and risk compensation to be relevant forces in the presence of retailer competition, and that understanding the tension between the two may be important in understanding the behavior of the market in the presence of demand risk.

4.5 Upstream oligopoly

Instead of a single upstream firm U , one could think of K upstream firms U_1 through U_K . In this case, both risk insurance and risk compensation will still be present, though upstream

competition may weaken the effects of risk compensation. If D makes a large profit in equilibrium due to upstream competition, it is less likely for D to lose interest in trade upon increased risk or risk aversion. Thus, we conjecture that the relative strength of risk insurance compared to risk compensation increases in the degree of upstream competition.

5 Application: Channel Coordination

One application of the preceding results is to channel coordination. Consider a situation in which an upstream manufacturer sells to a downstream retailer who then sells the products to consumers, where demand is uncertain.²⁰ The upstream manufacturer is risk-neutral and the downstream retailer is risk-averse. How do the risk preferences of the downstream retailer affect the expected profits of the channel as a whole? We show that the impact of demand risk and risk aversion can have nuanced implications for the answer to this question. In particular, we show that increases in demand risk or risk aversion do not necessarily have monotonic effects on channel coordination. In some cases, increases in demand risk or risk aversion worsen channel coordination, while in other cases such increases improve channel coordination.

Model

We adopt extremely simple functional forms to demonstrate the relevance of downstream risk aversion in even the simplest benchmark model. We assume that the upstream manufacturer produces at a constant marginal cost (c) and that the retailer incurs no additional cost beyond what it pays to the upstream manufacturer.²¹ Downstream demand (q) is linear in the price charged by the retailer (p) but is uncertain: $q = \alpha \cdot (a - bp)$, where a and b are parameters and α is drawn from a probability distribution with three points: $1 - \beta$, 1 , and $1 + \beta$, with $0 < \beta < 1$. The probability that $\alpha = 1 - \beta$ is $r \in [0, 0.5]$, as is the probability that $\alpha = 1 + \beta$. r can therefore be interpreted as the probability of “extreme” demand. In short, demand is uncertain and increases in r make it more so (in a mean-preserving spread sense).

The upstream manufacturer is risk-neutral but the downstream retailer is risk-averse. In particular, the retailer evaluates profit realizations x according to the continuous piecewise

²⁰Or, analogously, a situation in which a vertically integrated firm’s manufacturing and retailing divisions are separately managed by decision-makers with the specified (division-specific) preferences.

²¹In the vertically integrated formulation, this payment is the transfer price chosen by the manufacturing division.

linear utility function:

$$v(x) = \begin{cases} x, & x \geq 0 \\ \gamma x, & \theta \leq x < 0 \\ \gamma\theta + \lambda \cdot (x - \theta), & x < \theta. \end{cases} \quad (11)$$

In words, the retailer's utility function has two kinks: one at 0 and the other at $\theta < 0$. The slope of the retailer's utility function between 0 and θ is $\gamma > 0$, and the slope beyond θ is $\lambda > 0$. The risk-neutral case corresponds to $\gamma = \lambda = 1$.

Timing

In the first stage, the upstream manufacturer makes a two-part tariff contract offer (T, t) to the downstream retailer. In the second stage, the downstream retailer chooses whether or not to accept the contract, accepting all offers delivering expected utility of at least 0. In the third stage, the downstream retailer chooses a price p . In the fourth stage, α is revealed and demand q is realized. The payoff to the upstream manufacturer is $T + q \cdot (t - c)$ and the payoff to the downstream retailer is $v(-T + q \cdot (p(t) - t))$.²²

Parameterization

For the subsequent numerical results, we set $a = 8$, $b = 1$, $c = 3$, $\beta = 0.6$, and $\theta = -2$ (similar to the examples in the proof of Theorem 3 for consistency). The remaining parameters – r , γ , and λ – capture the level of demand risk (r) and the degree of risk aversion (γ and λ). The key point of the exercise is to illustrate how the impact of demand risk and risk aversion on channel coordination depends on these parameters. For demand risk r , we scan over a fine grid of values between 0 and 0.5. For risk aversion, we begin with the risk neutral case before adding increasing levels of risk aversion. The four different utility functions that we examine are depicted in Figure 2 and described in the notes below the figure.

Results

In this framework, maximizing expected channel profit requires a linear fee equal to upstream marginal cost ($= 3$) in order to eliminate double marginalization. Given this linear fee, consumers face a retail price of $p(t^*) = \frac{a+bt^*}{2b} = \frac{8+3}{2} = 5.5$, yielding the channel its maximal expected profit ($= 6.25$). The impact of demand risk and risk aversion on channel

²²As with the preceding theoretical results, given the form of the demand function the retailer's optimal price, $p(t) = \frac{a+bt}{2b}$, does not depend on the value of α .

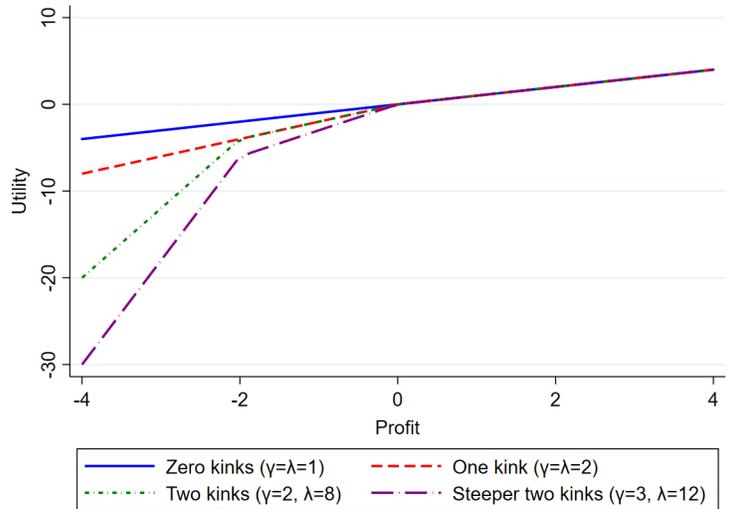


Figure 2: Utility Functions $v(x)$ “Zero kinks” refers to the risk-neutral utility function with $\gamma = \lambda = 1$. “One kink” refers to the utility function with $\gamma = \lambda = 2$. “Two kinks” refers to the utility function with $\gamma = 2$ and $\lambda = 8$. “Steeper two kinks” refers to the utility function with $\gamma = 3$ and $\lambda = 12$.

coordination is therefore determined by whether they result in linear fees closer to or further away from 3.

Figure 3 summarizes the impact of demand risk and risk aversion on channel coordination for the parameter values outlined above. Panel (a) of the figure plots the equilibrium linear fee, while panel (b) plots expected channel profit. In the case where the downstream retailer is risk-neutral, the level of demand risk has no impact on channel coordination. The intuition is straightforward: since two-part tariffs are available, the upstream manufacturer charges a linear fee to maximize channel profit and then captures the entirety of the surplus via the fixed fee.

To see how risk aversion can alter the analysis, first consider the “one kink” utility function, in which the downstream retailer’s utility function exhibits a kink at zero where the slope doubles. In that case, while the upstream manufacturer would prefer to offer an efficient contract with $t = 3$ and $T = 6.25$ (as in the risk-neutral case), such a contract is no longer tolerable to the downstream retailer. When the retailer has to pay a large fixed fee and faces demand risk, the retailer becomes exposed to negative profit realizations that greatly harm its utility. To make the contract tolerable to the retailer, the manufacturer needs to reduce the fixed fee and instead capture surplus via the linear fee. The greater the probability of extreme demand r , the greater the upward pressure on the linear fee as opposed to the fixed fee. Thus, as demand risk increases, channel coordination worsens.

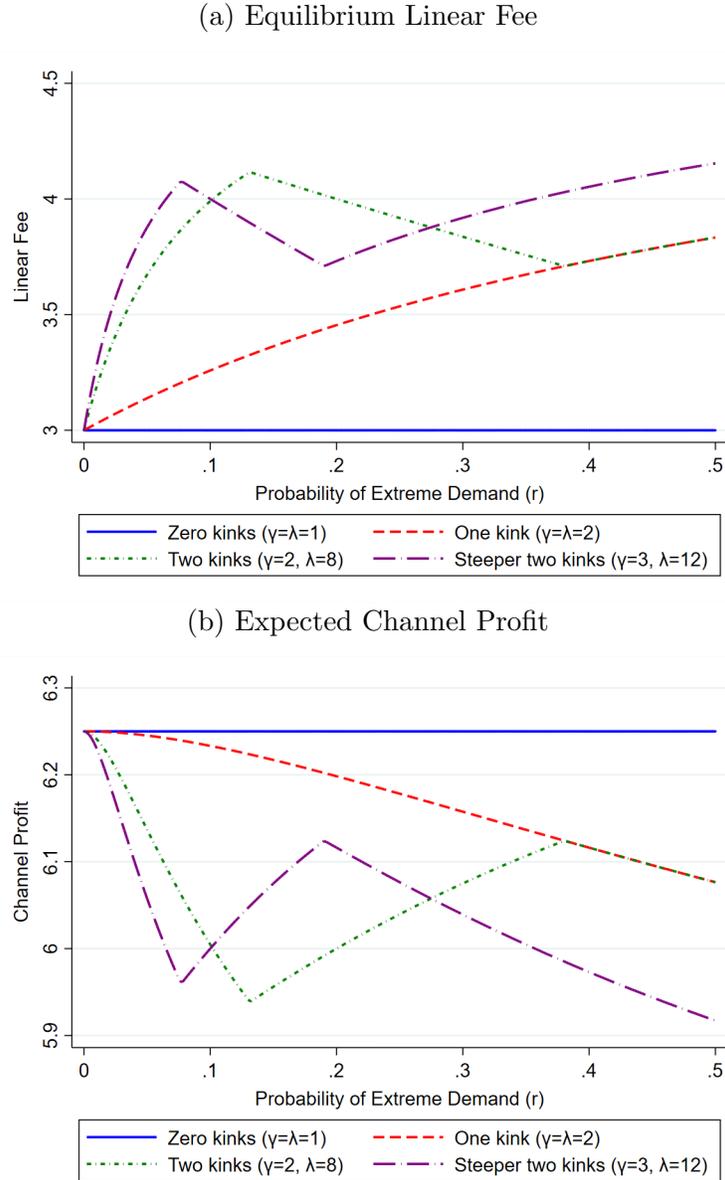


Figure 3: Demand Risk/Risk Aversion and Channel Coordination

The remaining utility functions demonstrate that the way in which risk aversion affects the analysis is not always so simple. Consider the “two kink” utility function, in which the downstream retailer’s utility function exhibits a second kink: for profit realizations less than -2, such losses are eight times as harmful as equivalent magnitude gains are beneficial. While the upstream manufacturer must still rely on the linear fee to capture surplus in this case, the effect is non-monotonic in the level of demand risk.

The intuition for this result is analogous to the intuition for the counterexamples in the proof of Theorem 3. The equilibrium fixed fee when $r = 0$ results in downstream profit realizations that are less than -2 for the low-demand state ($\alpha = 1 - \beta$), but this outcome never occurs because $r = 0$. As r increases, the manufacturer needs to compensate the retailer with a smaller fixed fee. Eventually, the equilibrium fixed fee becomes small enough that the retailer’s minimum profit realization is -2, at which point the retailer becomes effectively *less* risk-averse since profit realizations beyond the second kink of the utility function no longer occur. There is therefore an intermediate range of r such that the equilibrium fixed fee *increases* as r increases, while the linear fee *decreases*. This non-monotonic relationship gives rise to the pattern shown in the figure, where increases in demand risk can improve channel coordination.

The final utility function (“steeper two kinks”) demonstrates that analogous results can be obtained for an increase in risk aversion rather than demand risk. In particular, there is an intermediate range of r such that making the downstream retailer more risk averse – increasing γ from 2 to 3 and λ from 8 to 12 – improves channel coordination.

Discussion

As the above example illustrates, incorporating risk aversion into the analysis of channel coordination can have nuanced and economically meaningful effects, even in extremely simplified settings. The impact of demand uncertainty on channel coordination has been studied in the past. However, it has usually been analyzed in the context of other effects, such as quality and inventory (Jerath et al. (2017)), revenue sharing (Cachon and Lariviere (2005)), and information sharing (Gal-Or et al. (2008); Dukes and Liu (2010); Dukes et al. (2011)). The direct impact of risk and risk aversion on double marginalization is generally considered monotone and simple in the literature. The example above – and the analysis in the paper more generally – demonstrates otherwise. A more complete consideration of the downstream firm’s risk attitude reveals that the impact of risk and risk aversion on double marginalization need not be monotone.

We would also like to point out an interesting connection between our results and Gan et al. (2005). Gan et al. (2005) argue that channel coordination can be achieved when the retailer is “downside risk-averse.” They define downside risk aversion by having a constraint on the probability that the retailer’s ex-post profit is below a certain cutoff. For instance, the retailer is downside risk averse if she does not accept any contract that would imply a probability higher than α that her ex-post profit is less than some cutoff β . This specific

kind of risk aversion is very similar to our notion of catastrophe-aversion. Hence, the result of Gan et al. (2005) that the channel can be coordinated under this kind of risk aversion mirrors our result that increases in catastrophe-aversion can improve channel coordination.

6 Conclusion

This paper explores how double marginalization in vertical markets is affected by downstream demand risk and aversion to that risk. In a bilateral monopoly model with two-part tariffs, we show that two economic forces with potentially opposing effects must be considered. (i) *Risk insurance*: if risk or risk aversion increases, the upstream firm needs to insure the downstream firm from outcomes in which the realized demand is low. Doing so involves a shift from the fixed fee to the linear fee, which smooths profits over demand realizations. (ii) *Risk compensation*: if risk or risk aversion increases, the upstream firm needs to compensate the downstream firm for the lessened attractiveness of trade. Doing so involves a decrease in the fixed fee, linear fee, or both. Therefore, the overall effect of increased risk or risk aversion on the equilibrium linear fee – and thus double marginalization – is unclear.

We use the model to demonstrate that the interaction between these two forces has a counterintuitive outcome: sometimes, increased risk or risk aversion can mitigate double marginalization. This result illustrates that the appealing intuition that increased risk or risk aversion will exacerbate double marginalization is not complete. We also characterize circumstances under which double marginalization is indeed always exacerbated when risk or risk aversion increases: (1) when a CARA utility function concavifies; (2) when uniform demand risk spreads out; and (3) when any utility function becomes “globally more risk-averse” (see Definition 2) and the demand risk is uniform. We also show that, although increased risk or risk aversion *sometimes* mitigates double marginalization, certain demand always generates less double marginalization than uncertain demand, as does downstream risk-neutrality compared to risk-aversion.

To illustrate an application of these results, we apply the model to the context of channel coordination. How do demand risk and risk aversion affect the profitability of a vertical channel? We show that increased demand risk and risk aversion do not always have monotonic effects on channel coordination. Increases in demand risk or risk aversion may worsen channel coordination, or improve it. Beyond the specific example, the main takeaway is that the impact of demand risk on channel coordination can be complex simply because of the nature of the retailer’s risk attitude, which affects the relative strengths of risk insurance

and risk compensation.

Research in this area could be further pursued in several directions. On the theory side, one direction is to allow for a more general contract set than two-part tariffs. “Share-of-wallet” (SOW) contracts are especially interesting. A SOW contract gives a larger discount to a buyer when a greater share of the buyer’s total purchases is accounted for by the seller. In contrast to regular volume discount contracts, which give discounts as the *absolute* purchase volume increases, SOW contracts give discounts as *relative* purchase volume increases. SOW contracts are commonly utilized in multiple business-to-business settings, such as transactions between medical device producers and hospitals (Mojir and Sudhir (2017)). SOW contracts can potentially combine the advantages of linear fees and volume discounts. Like volume discounts, SOW contracts incentivize buyers to increase the volume of their purchases. But similar to linear fees, SOW contracts do not punish buyers if downstream demand is weak. In our view, an analysis of the performance of SOW contracts under uncertainty is an interesting topic for future work.

Empirically, an interesting avenue of research is to develop tests of theoretical mechanisms that are thought to affect the shape of contracts. Consider the mechanism studied in this paper: risk and risk aversion. Conditional on data availability, one approach to testing the theory is to examine whether industries characterized by demand risk tend to rely more heavily on linear fees.²³ Note, however, that our analysis suggests that reliance on linear fees is not necessarily monotonically increasing in the magnitude of demand risk. Despite the potential challenges along these lines, we believe that a better empirical understanding of the factors influencing the shape of contracts would be a major advance.

References

- Amihud, Y. and B. Lev (1981). Risk reduction as a managerial motive for conglomerate mergers. *Bell Journal of Economics* 12(2), 605–617.
- Asplund, M. (2002). Risk-averse firms in oligopoly. *International Journal of Industrial Organization* 20(7), 995–1012.
- Banal-Estañol, A. and M. Ottaviani (2006). Mergers with product market risk. *Journal of Economics & Management Strategy* 15(3), 577–608.

²³For a simple example of measuring demand risk by industry, see, e.g., Dyer, J., N. Furr, and C. Lefrandt (2014, September 11). “The Industries Plagued by the Most Uncertainty.” *Harvard Business Review*.

- Binmore, K., A. Rubinstein, and A. Wolinsky (1986). The nash bargaining solution in economic modelling. *RAND Journal of Economics* 17(2), 176–188.
- Bryan, K. and E. Hovenkamp (2016). Profitable double marginalization. *Available at SSRN 2734513*.
- Cachon, G. P. and M. A. Lariviere (2005). Supply chain coordination with revenue-sharing contracts: strengths and limitations. *Management science* 51(1), 30–44.
- Chipty, T. (2001). Vertical integration, market foreclosure, and consumer welfare in the cable television industry. *American Economic Review* 91(3), 428–453.
- Crawford, G. S., R. S. Lee, M. D. Whinston, and A. Yurukoglu (2018). The welfare effects of vertical integration in multichannel television markets. *Econometrica* 86(3), 891–954.
- Dukes, A., E. Gal-Or, and T. Geylani (2011). Who benefits from bilateral information exchange in a retail channel? *Economics Letters* 112(2), 210–212.
- Dukes, A. and Y. Liu (2010). In-store media and distribution channel coordination. *Marketing Science* 29(1), 94–107.
- Edlin, A., S. Hemphill, H. Hovenkamp, and C. Shapiro (2013). Activating actavis. *Antitrust* 28(1), 16–23.
- Gal-Or, E., T. Geylani, and A. J. Dukes (2008). Information sharing in a channel with partially informed retailers. *Marketing Science* 27(4), 642–658.
- Gan, X., S. P. Sethi, and H. Yan (2005). Channel coordination with a risk-neutral supplier and a downside-risk-averse retailer. *Production and Operations Management* 14(1), 80–89.
- Ghili, S. (2018). Network formation and bargaining in vertical markets: The case of narrow networks in health insurance. *Working Paper*.
- Gowrisankaran, G., A. Nevo, and R. Town (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review* 105(1), 172–203.
- Harris, B. C., K. M. Murphy, R. D. Willig, and M. B. Wright (2014). Activating actavis: A more complete story. *Antitrust* 28(2), 83–89.
- Hayes, B. (1987). Competition and two-part tariffs. *Journal of Business* 60(1), 41–54.

- Ho, K. and R. S. Lee (2017). Insurer competition in health care markets. *Econometrica* 85(2), 379–417.
- Ho, K. and R. S. Lee (2018). Equilibrium provider networks: bargaining and exclusion in health care markets. *forthcoming, American Economic Review*.
- Inderst, R. and G. Shaffer (2018). Managing channel profits when retailers have profitable outside options. *Management Science* 65(2), 642–659.
- Iyer, G. and J. M. Villas-Boas (2003). A bargaining theory of distribution channels. *Journal of Marketing Research* 40(1), 80–100.
- Jerath, K., S.-H. Kim, and R. Swinney (2017). Product quality in a distribution channel with inventory risk. *Marketing Science* 36(5), 747–761.
- Jeuland, A. P. and S. M. Shugan (1983). Managing channel profits. *Marketing Science* 2(3), 239–272.
- Joskow, P. L. (2012). Vertical integration. *Antitrust Bulletin* 57(3), 545–586.
- Katz, M. L. (1989). Vertical contractual relations. *Handbook of Industrial Organization* 1, 655–721.
- Lambrecht, A. and B. Skiera (2006). Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases. *Journal of Marketing Research* 43(2), 212–223.
- Liebman, E. (2018). Bargaining in markets with exclusion: An analysis of health insurance networks. *Working Paper*.
- Marx, L. M. and G. Shaffer (2007). Upfront payments and exclusion in downstream markets. *The RAND Journal of Economics* 38(3), 823–843.
- Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *RAND Journal of Economics* 15(2), 171–196.
- Mathewson, G. F. and R. A. Winter (1984). An economic theory of vertical restraints. *RAND Journal of Economics* 15(1), 27–38.
- Mojir, N. and K. Sudhir (2017). A structural model of organizational buying: Innovation adoption under share of wallet price contracts.

- Nance, D. R., C. W. Smith, and C. W. Smithson (1993). On the determinants of corporate hedging. *Journal of Finance* 48(1), 267–284.
- Narayanan, S., P. K. Chintagunta, and E. J. Miravete (2007). The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and Economics* 5(1), 1–34.
- Oi, W. Y. (1971). A Disneyland dilemma: two-part tariffs for a Mickey Mouse monopoly. *Quarterly Journal of Economics* 85(1), 77–96.
- Png, I. and H. Wang (2010). Buyer uncertainty and two-part pricing: theory and applications. *Management Science* 56(2), 334–342.
- Rey, P. and J. Tirole (1986). The logic of vertical restraints. *American Economic Review* 76(5), 921–939.
- Rey, P. and T. Vergé (2008). The economics of vertical restraints. *Handbook of Antitrust Economics*.
- Spengler, J. J. (1950). Vertical integration and antitrust policy. *Journal of Political Economy* 58(4), 347–352.
- Stahl Gabrielsen, T., B. O. Johansen, G. Shaffer, et al. (2018). When is double marginalization a problem? Technical report.

7 Appendix: Proofs

7.1 Preliminaries

Before proceeding to the proofs, we first slightly change the presentation of the model so the proofs are more easily presented and understood. We begin by the following definition:

Definition 3. Let $w(t) = (p(t) - t) \cdot Q(p(t))$, where $p(t) = \operatorname{argmax}_p (p - t) \cdot Q(p)$.

In words, $w(t)$ is D 's expected profit gross of the fixed fee T . Note that $w(t)$ does not depend on either f (demand risk) or v (downstream risk aversion). Lemma 1 below demonstrates that $w(t)$ is strictly decreasing if it is always positive. If instead $w(t)$ hits zero at some t , $w(t)$ will remain at zero thereafter.

Lemma 1. $w(t)$ is decreasing in t , and strictly decreasing in t for all t such that $w(t) > 0$.

Proof. By definition, $w(t)$ is the highest profit that D can make if it faces a linear fee of t , a fixed fee of zero, and the demand function Q . D attains this profit by optimally setting the downstream price equal to $p(t)$. Now take two linear fees t_1 and t_2 such that $t_2 > t_1$. If D charges $p(t_2)$ when facing the linear fee t_1 , it earns $\pi = (p(t_2) - t_1) \cdot Q(p(t_2))$, which is (weakly) less than D 's profit when pricing optimally at $p(t_1)$. D 's profit when facing the linear fee t_2 and pricing optimally is $w(t_2) = (p(t_2) - t_2) \cdot Q(p(t_2))$. Subtracting $w(t_2)$ from π , we have:

$$\pi - w(t_2) = Q(p(t_2)) \cdot (t_2 - t_1).$$

Since demand is non-negative and $t_2 > t_1$, we have that $\pi \geq w(t_2)$. Combined with the optimality condition $w(t_1) \geq \pi$, we have that $w(t_1) \geq w(t_2)$. Moreover, since demand is non-negative, if $w(t_2) > 0$ then $Q(p(t_2)) > 0$. In turn, this implies that $\pi > w(t_2)$ and thus $w(t_1) > w(t_2)$ as well. **Q.E.D.**

Instead of specifying a contract using the pair (T, t) as in the main text, here we instead specify a contract using $(T, w(t))$, where $w(t)$ is as specified in Definition 3. For notational convenience, we simplify this further to (T, w) . In words, w captures the component of D 's profit that is subject to demand risk while the fixed fee T captures the component of D 's profit that is certain.

By Lemma 1, there is a one-to-one mapping between t and w .²⁴ This mapping depends only on the demand function Q and *not* on the demand risk distribution f or D 's utility

²⁴Note that the mapping is not one-to-one once t is sufficiently large that $w(t) = 0$ (i.e., where D is unable to charge a profitable price that would generate positive demand). Such values of t cannot be part of an equilibrium contract outside of cases where there are no gains from trade.

function v . Since the demand function Q is fixed in all of the comparative statics outlined in the paper, presenting the contract in terms of (T, w) is essentially equivalent to presenting it in terms of (T, t) . Therefore, in what follows, we use the notation that was introduced in the main text for (T, t) instead on (T, w) . For instance, we denote U 's profit by $\pi_U(T, w)$ instead of $\pi_U(T, t)$. Given this alternate presentation of the model, it is helpful to define the slope of D 's indifference curves in the (T, w) space.

Definition 4. *Given (f, v) , denote D 's marginal rate of substitution at contract (T, w) by $\Delta(T, w|f, v)$. $\Delta(T, w|f, v)$ is defined as:*

$$\Delta(T, w|f, v) = -\frac{\frac{\partial \mathbf{E}v_D(T, w)}{\partial w}}{\frac{\partial \mathbf{E}v_D(T, w)}{\partial T}}.$$

Definition 4 reflects the same notion of D 's marginal rate of substitution as in Definition 1, but instead of the (T, t) space it is now defined in the (T, w) space. Differentiating $\mathbf{E}v_D(T, w)$ with respect to w and T , we also have the following expression for $\Delta(T, w|f, v)$:

$$\Delta(T, w|f, v) = \frac{\int_0^\infty \alpha v'(\alpha w - T) f(\alpha) d\alpha}{\int_0^\infty v'(\alpha w - T) f(\alpha) d\alpha}. \quad (12)$$

Last, many proofs examine D 's marginal rate of substitution after compensating D for the utility loss from an increase in risk or risk aversion by adjusting the fixed fee. Definition 5 formalizes this idea of adjusting the fixed fee to compensate D for a change in risk aversion.

Definition 5. *Consider the equilibrium $(T^*, w^*|f, v)$. Fix f and w^* . Let $T(v')$ denote the fixed fee that satisfies the zero-expected-value condition for D . That is, $T(v')$ satisfies:*

$$\mathbf{E}v_D(T(v'), w^*|f, v') = 0.$$

In other words, when D 's utility function changes from v to another utility function v' , $T(v')$ indicates the new fixed fee that is required to return D to zero utility, holding the linear fee at the original equilibrium value. Definition 5 is stated for changes in D 's utility function. To economize on notation, we use the same notation for changes in the demand risk distribution. That is, if the demand risk distribution changes from f to another distribution f' , $T(f')$ indicates the new fixed fee that is required to return D to zero utility, holding the linear fee at the original equilibrium value. To further reduce notation, in cases where risk or risk aversion is governed by a single parameter, we use that parameter as the argument in

$T(\cdot)$. For example, if f is a uniform distribution over $[1 - \sigma, 1 + \sigma]$, then we use the notation $T(\sigma)$. For another example, if v is a CARA utility function with risk coefficient r , then we use the notation $T(r)$.

7.2 Proofs of preliminary results (section 2.1)

Proof of Proposition 1 part a. We first prove a lemma that is equivalent to part a of Proposition 1 but in the (T, w) space. Examination of the relationship between the (T, t) space and the (T, w) space then completes the proof.

Lemma 2. *For all $v \in \mathcal{V}$ and $\sigma_1, \sigma_2 \in [0, 1]$ such that $\sigma_2 \geq \sigma_1$:*

$$\Delta(T, w|\sigma_2, v) \leq \Delta(T, w|\sigma_1, v) \quad \forall w > 0, T \in \mathbb{R}.$$

Proof of Lemma 2. Writing out equation (12) for the case of uniform demand risk, for any (σ, v) we have that:

$$\Delta(T, w|\sigma, v) = \frac{\int_{1-\sigma}^{1+\sigma} \alpha v'(\alpha w - T) d\alpha}{\int_{1-\sigma}^{1+\sigma} v'(\alpha w - T) d\alpha}. \quad (13)$$

The right hand side of this equation can be thought of as a weighted average of the α values between $1 - \sigma$ and $1 + \sigma$, where the weights are given by the function $v'(\alpha w - T)$, which is (i) positive and (ii) decreasing in α . Moving from σ_1 to $\sigma_2 \geq \sigma_1$ means that a set of new numbers is symmetrically added to the set over which the weighted average is taken. The added set is symmetric: it has numbers in $[1 - \sigma_2, 1 - \sigma_1)$ and $(1 + \sigma_1, 1 + \sigma_2]$. However, the weights on these numbers are not symmetric. Because v' is decreasing, the weights on $\alpha \in [1 - \sigma_2, 1 - \sigma_1)$ are weakly larger than the weights on $\alpha \in [1 - \sigma_1, 1 + \sigma_1]$ which are weakly larger than the weights on $\alpha \in (1 + \sigma_1, 1 + \sigma_2]$. The increase from σ_1 to $\sigma_2 \geq \sigma_1$ thus weakly decreases the weighted average, which proves Lemma 2. **Q.E.D.**

Given (i) that w is strictly decreasing in t (Lemma 1) and (ii) the sign difference between the marginal rate of substitution definitions in the (T, t) space (Definition 1) and the (T, w) space (Definition 4), it follows that $\Delta(T, t|\sigma_2, v) \leq \Delta(T, t|\sigma_1, v)$ (for all T, t), which completes the proof of part a of Proposition 1. **Q.E.D.**

Proof of Proposition 1 part b. In the main text of the paper, the proposition is stated for uniformly distributed α . Here we prove the general case in which α is distributed according to some $f \in \mathcal{F}$.

Because v_2 is more concave than v_1 , there is a concave function u such that $\forall x \in \mathbb{R} : v_2(x) = u(v_1(x))$. Given that both v_1 and v_2 are strictly increasing functions, u is strictly increasing as well. Thus, we have the following two equations:

$$\Delta(T, w|f, v_1) = \frac{\int_0^\infty \alpha v_1'(\alpha w - T) f(\alpha) d\alpha}{\int_0^\infty v_1'(\alpha w - T) f(\alpha) d\alpha}, \text{ and}$$

$$\Delta(T, w|f, v_2) = \frac{\int_0^\infty \alpha u'(v_1(\alpha w - T)) v_1'(\alpha w - T) f(\alpha) d\alpha}{\int_0^\infty u'(v_1(\alpha w - T)) v_1'(\alpha w - T) f(\alpha) d\alpha}.$$

Now let:

$$g_1(\alpha) = \frac{v_1'(\alpha w - T) f(\alpha)}{\int_0^\infty v_1'(\beta w - T) f(\beta) d\beta}, \text{ and}$$

$$g_2(\alpha) = \frac{u'(v_1(\alpha w - T)) v_1'(\alpha w - T) f(\alpha)}{\int_0^\infty u'(v_1(\beta w - T)) v_1'(\beta w - T) f(\beta) d\beta}.$$

It is straightforward to verify that g_1 and g_2 are probability density functions over the range $[0, \infty]$. Thus, part b of Proposition 1 is essentially equivalent to showing that the expected value of α is lower if it comes from probability distribution g_2 than if it comes from g_1 . To see that the expected value of α under g_2 is indeed smaller than under g_1 , divide g_2 by g_1 , which yields:

$$\frac{g_2(\alpha)}{g_1(\alpha)} = \frac{\int_0^\infty v_1'(\beta w - T) f(\beta) d\beta}{\int_0^\infty u'(v_1(\beta w - T)) v_1'(\beta w - T) f(\beta) d\beta} \times u'(v_1(\alpha w - T)).$$

Because u' is a decreasing function, this expression is decreasing in α . Thus, g_2 is dominated by g_1 in the likelihood ratio order. This implies that the expected value of α under g_2 is smaller than under g_1 , which yields $\Delta(T, w|f, v_2) \leq \Delta(T, w|f, v_1)$.

Again given (i) that w is strictly decreasing in t and (ii) the sign difference between the marginal rate of substitution definitions in the (T, t) space and the (T, w) space, it follows that $\Delta(T, t|f, v_2) \leq \Delta(T, t|f, v_1)$, which completes the proof of part b of Proposition 1.

Q.E.D.

Proof of Proposition 2 part a. The proof, which applies to general $f \in \mathcal{F}$, proceeds by contradiction. Suppose that (T, t) is an equilibrium and that $\mathbf{E}v_D(T, t|f, v) \neq 0$. There are two cases. First, suppose that $\mathbf{E}v_D(T, t|f, v) < 0$. This cannot be an equilibrium because D can reject the U 's offer at stage 2 of the game and attain a payoff of 0. Second, suppose that $\mathbf{E}v_D(T, t|f, v) > 0$. This cannot be an equilibrium because U can slightly increase T such

that (i) U 's payoff is higher and (ii) the offer is still acceptable to D . Thus, $\mathbf{E}v_D(T, t|f, v) = 0$ in any equilibrium. **Q.E.D.**

Proof of Proposition 2 part b. It is straightforward to verify that when $\sigma_2 \geq \sigma_1$, a uniform distribution over $[1 - \sigma_1, 1 + \sigma_1]$ second-order stochastically dominates a uniform distribution over $[1 - \sigma_2, 1 + \sigma_2]$. Therefore, by definition, part b of Proposition 2 holds. **Q.E.D.**

Proof of Proposition 2 part c. This result is an immediate implication of the Arrow-Pratt theorem. **Q.E.D.**

Proof of Proposition 3. First note that for any $t \geq a$, D will set the downstream price such that there will be no demand, and hence no total profit to be made by the channel. In equilibrium, we should therefore expect that $t < a$. We solve for the equilibrium by backward induction. Examination of D 's first order condition implies the pricing policy of $p(t) = \frac{a+t}{2}$. Plugging $p(t)$ into U 's profit function yields $\mathbf{E}\pi_U(T, t) = (\frac{a-t}{2})(t - c) + T$. U will maximize $\mathbf{E}\pi_U(T, t)$ subject to D 's zero-expected-value condition. If D is risk-neutral or faces no risk (as in part a of the proposition), then $\mathbf{E}v_D(T, t) = (\frac{a-t}{2})^2 - T$. The optimal fixed fee as a function of t is thus $(\frac{a-t}{2})^2$. Plugging this fixed fee into $\mathbf{E}\pi_U(T, t)$ and solving for the optimal linear fee yields $t^* = c$.

For part b of the proposition, due to D 's extreme risk aversion, no fixed fee can be charged because D only cares about the case where demand is zero. If any positive fixed fee is charged, then D will reject the contract. Thus, in part b, U maximizes $(\frac{a-t}{2})(t - c)$. Solving for the optimal linear fee yields $t^* = \frac{a+c}{2}$. **Q.E.D.**

7.3 Proofs of main results (section 3)

Proof of Theorem 1. The first step of the proof (Lemma 3 below) is to show that an inequality similar to (6) in the main text governs the response of the equilibrium linear fee to an increase in the CARA risk coefficient. In words, if an increase in the CARA risk coefficient makes D more sensitive to changes in the fixed fee (T) than changes in the profits subject to the demand risk (w) – even after restoring D to zero utility by reducing the fixed fee – then the equilibrium linear fee will increase. The second step is to verify that the conditions of Lemma 3 hold as the CARA risk coefficient increases.

Lemma 3. Denote the equilibrium contract with demand risk distribution f and CARA risk coefficient r_1 by (T^*, w^*) . Theorem 1 holds if we can show that for all $r_2 \geq r_1$:

$$\Delta(T(r_2), w^*|f, r_2) \leq \Delta(T^*, w^*|f, r_1), \quad (14)$$

where $T(r)$ is as outlined in Definition 5.

Proof of Lemma 3. Denote U 's marginal rate of substitution at (T, w) by

$$\Delta_U(T, w|f, v) = -\frac{\frac{\partial \mathbf{E}\pi_U(T, w)}{\partial w}}{\frac{\partial \mathbf{E}\pi_U(T, w)}{\partial T}}.$$

At the equilibrium contract for f , r_1 , the marginal rates of substitution for U and D are equal to one another. That is:

$$\Delta(T^*, w^*|f, r_1) = \Delta_U(T^*, w^*|f, r_1). \quad (15)$$

Given that U is risk-neutral, $\frac{\partial \mathbf{E}\pi_U(T, w)}{\partial T} = 1$, which implies that $\Delta_U(T, w|f, v)$ is always independent of the value of T . Therefore:

$$\Delta_U(T^*, w^*|f, r_1) = \Delta_U(T(r_2), w^*|f, r_2). \quad (16)$$

If the inequality (14) holds, using (15) and (16) we thus have that:

$$\Delta(T(r_2), w^*|f, r_2) \leq \Delta_U(T(r_2), w^*|f, r_2). \quad (17)$$

In words, at $(T(r_2), w^*)$, D is more sensitive to changes in T than U . Therefore, U can achieve (weakly) higher profits while keeping D indifferent by (weakly) reducing both T and w . These incentives are further illustrated in Figure 4. Since the equilibrium value of w under r_2 weakly decreases compared to r_1 , the equilibrium value of t weakly increases, as required by the statement of the theorem. **Q.E.D.**

The second step of the proof is to show that the conditions of Lemma 3 always hold as the CARA risk coefficient increases. That is, $\Delta(T(r_2), w^*|f, r_2) \leq \Delta(T^*, w^*|f, r_1)$ for all $r_2 \geq r_1$. To begin, note that by part b of Proposition 1, we have:

$$\Delta(T^*, w^*|f, r_2) \leq \Delta(T^*, w^*|f, r_1). \quad (18)$$

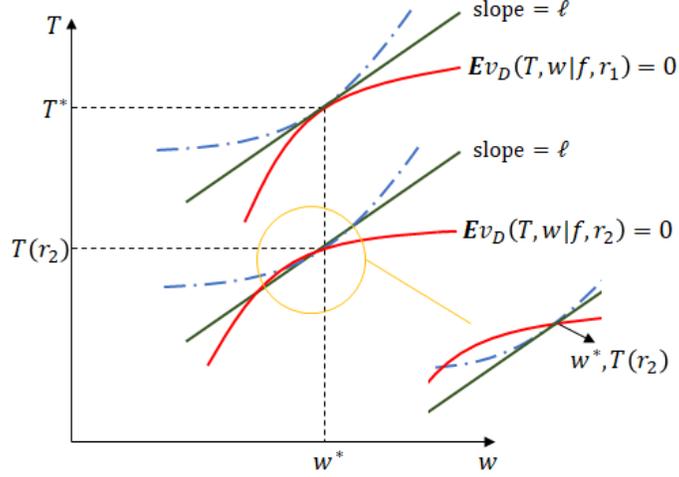


Figure 4: Illustration of Lemma 3 (w^*, T^*) is the equilibrium contract under f and r_1 . At this point, U and D 's marginal rates of substitution are equal to one another ($= \ell$). After compensating D for the increase in risk aversion by moving to the point $(w^*, T(r_2))$, the slope of U 's indifference curve passing through this point is still ℓ , while the slope of D 's indifference curve is (weakly) less than ℓ . Therefore, U can offer a contract with a (weakly) smaller T and w that still delivers zero utility to D and yields U (weakly) higher profits.

Due to the functional form of CARA utility functions, it can also be shown that:

$$\Delta(T(r_2), w^* | f, r_2) = \Delta(T^*, w^* | f, r_2). \quad (19)$$

To see that this equality holds, observe that when D has a CARA utility function, $v(x) = \frac{1 - \exp(-rx)}{r}$ and thus $v'(x) = \exp(-rx)$. Writing out $\Delta(T(r_2), w^* | f, r_2)$ according to equation (12), we have that:

$$\Delta(T(r_2), w^* | f, r_2) = \frac{\int_0^\infty \alpha \exp(-r_2(\alpha w^* - T(r_2))) f(\alpha) d\alpha}{\int_0^\infty \exp(-r_2(\alpha w^* - T(r_2))) f(\alpha) d\alpha}.$$

Multiplying both the numerator and the denominator of the above fraction by $\exp(-r_2(T(r_2) - T^*))$ and simplifying shows that equation (19) holds:

$$\Delta(T(r_2), w^* | f, r_2) = \frac{\int_0^\infty \alpha \exp(-r_2(\alpha w^* - T^*)) f(\alpha) d\alpha}{\int_0^\infty \exp(-r_2(\alpha w^* - T^*)) f(\alpha) d\alpha} = \Delta(T^*, w^* | f, r_2).$$

Combining inequality (18) with equation (19), we have that $\Delta(T(r_2), w^* | f, r_2) \leq \Delta(T^*, w^* | f, r_1)$, which is exactly inequality (14) (i.e., the conditions of Lemma 3). Therefore, Theorem 1 holds. **Q.E.D.**

Proof of Theorem 2. Similarly to Lemma 3 in the proof of Theorem 1, it can be shown that Theorem 2 holds if, for all $\sigma_2 \geq \sigma_1$:

$$\Delta(T(\sigma_2), w^*|\sigma_2, v) \leq \Delta(T^*, w^*|\sigma_1, v), \quad (20)$$

where (T^*, w^*) is the equilibrium contract under (σ_1, v) and $T(\sigma_2)$ is the fixed fee that restores D to zero utility (see Definition 5). We omit the proof that satisfying this condition implies Theorem 2 for parsimony: the proof follows the same logic as the proof of Lemma 3.

We begin with a lemma that provides an alternative characterization of $\Delta(T, w|\sigma, v)$.

Lemma 4. *For any $\sigma \in [0, 1]$ and $v \in \mathcal{V}$, D 's marginal rate of substitution at contract (T, w) can be written as:*

$$\Delta(T, w|\sigma, v) = 1 - \frac{1}{w} \times \frac{\int_a^b v(x)dx - (b-a)\frac{v(a)+v(b)}{2}}{v(b) - v(a)}, \quad (21)$$

where $a = (1 - \sigma)w - T$ and $b = (1 + \sigma)w - T$.

Proof of Lemma 4. Integrate the denominator of equation (13) and integrate the numerator by parts. The result is:

$$\begin{aligned} \Delta(T, w|\sigma, v) &= \frac{(1 + \sigma)v((1 + \sigma)w - T) - (1 - \sigma)v((1 - \sigma)w - T)}{v((1 + \sigma)w - T) - v((1 - \sigma)w - T)} \\ &\quad - \frac{\frac{1}{w} \int_{(1-\sigma)w}^{(1+\sigma)w} v(\xi - T)d\xi}{v((1 + \sigma)w - T) - v((1 - \sigma)w - T)}. \end{aligned}$$

Further rearranging and substituting the terms a and b , we arrive at formula (21). **Q.E.D.**

The formula (21) facilitates a geometric interpretation of $\Delta(T, w|\sigma, v)$. As depicted in Figure 5, the fraction $\frac{\int_a^b v(x)dx - (b-a)\frac{v(a)+v(b)}{2}}{v(b) - v(a)}$ can be interpreted as the area of surface S divided by the length of segment h .

We want to show that $\Delta(T^*, w^*|\sigma, v)$ decreases after: (i) σ increases from σ_1 to σ_2 ; and (ii) T decreases from T^* to $T(\sigma_2)$. Neither of these changes affect w . Therefore, verifying (20) is equivalent to verifying that $\frac{S}{h}$ increases as changes (i) and (ii) take place. The only channel through which S and h in Figure 5 respond to changes in σ and T is through the impact of σ and T on a and b . The following lemmas study this channel.

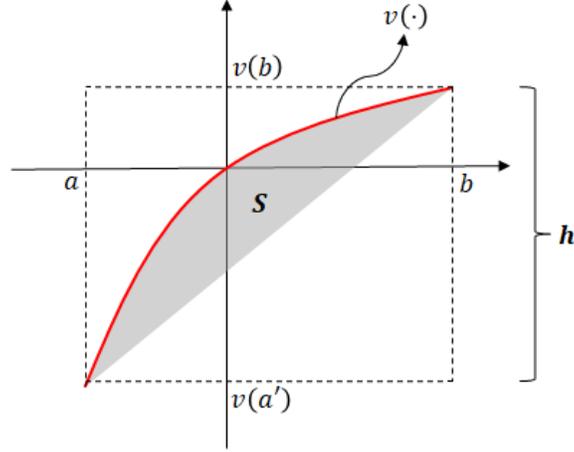


Figure 5: Illustration of Lemma 4 D 's marginal rate of substitution $\Delta(T, w|\sigma, v)$ is equal to $1 - \frac{1}{w} \times \frac{S}{h}$ where S and h are the area and segment depicted in the figure.

Lemma 5. Fix (σ_1, v) and the corresponding equilibrium contract (T^*, w^*) .

$$T'(\sigma)|_{\sigma=\sigma_1} = w^* \frac{v(b) - v(a)}{v(b) + v(a)}, \quad (22)$$

where $a = (1 - \sigma_1)w^* - T^*$ and $b = (1 + \sigma_1)w^* - T^*$.

In terms of notation, note that a and b are no longer defined in terms of an arbitrary contract (T, w) . Rather, they are defined in terms of the equilibrium contract under (σ_1, v) .

Proof of Lemma 5. By the definition of the function $T(\sigma)$, the fixed fee given by this function keeps the downstream firm D on its zero-expected-value curve. Thus, once we move from σ_1 to $\sigma = \sigma_1 + d\sigma$, D 's utility should stay at zero if the fixed fee changes to $T^* + d\sigma \times T'(\sigma)|_{\sigma=\sigma_1}$. Therefore:

$$\int_{\alpha=1-\sigma_1}^{1+\sigma_1} v(\alpha w^* - T^*) d\alpha = \int_{\alpha=1-\sigma_1-d\sigma}^{1+\sigma_1+d\sigma} v(\alpha w^* - T^* - d\sigma \times T'(\sigma)|_{\sigma=\sigma_1}) d\alpha.$$

Rearranging terms and simplifying, we get:

$$\int_a^b v(x) dx = \int_{a-w^* \times d\sigma - d\sigma \times T'(\sigma)|_{\sigma=\sigma_1}}^{b+w^* \times d\sigma - d\sigma \times T'(\sigma)|_{\sigma=\sigma_1}} v(x) dx$$

$$\Leftrightarrow d\sigma dx \times (v(b)(w^* - T'(\sigma)|_{\sigma=\sigma_1}) + v(a)(-w^* - T'(\sigma)|_{\sigma=\sigma_1})) = 0.$$

Given that dx and $d\sigma$ are not zero, we get:

$$v(b)(w^* - T'(\sigma)|_{\sigma=\sigma_1}) + v(a)(-w^* - T'(\sigma)|_{\sigma=\sigma_1}) = 0. \quad (23)$$

Rearranging the terms in equation (23) yields equation (22), thus completing the proof of Lemma 5. **Q.E.D.**

Given Lemma 5, it is straightforward to verify that:

$$\lim_{\sigma \rightarrow \sigma_1^+} \frac{b' - b}{a' - a} = \frac{v(a)}{v(b)} = \frac{\frac{1}{v(b)}}{\frac{1}{v(a)}}, \quad (24)$$

where $a' = (1 - \sigma)w^* - T(\sigma)$ and $b' = (1 + \sigma)w^* - T(\sigma)$.

Equation (24) sheds light on how a and b respond to changes in σ and T . To see the intuition behind what the equation captures, start from $T = T^*$ and $\sigma = \sigma_1$. Then slightly increase σ and set $T = T(\sigma)$. These two changes impact the ratio $\frac{S}{h}$ in Figure 5 by widening the interval $[a, b]$ to $[a', b']$. Equation (24) says that if $\sigma - \sigma_1$ is small enough, then the ratio between how much the rightmost point of the interval moves (i.e., $b' - b$) and how much the leftmost point moves (i.e., $a' - a$) is equal to $\frac{\frac{1}{v(b)}}{\frac{1}{v(a)}}$. We now show that such a change will lead to an increase in $\frac{S}{h}$. That is:

$$\frac{\int_{a'}^{b'} v(x)dx - (b' - a')\frac{v(a') + v(b')}{2}}{v(b') - v(a')} \geq \frac{\int_a^b v(x)dx - (b - a)\frac{v(a) + v(b)}{2}}{v(b) - v(a)}, \quad (25)$$

where $a = (1 - \sigma_1)w^* - T^*$, $b = (1 + \sigma_1)w^* - T^*$, $a' = (1 - \sigma)w^* - T(\sigma)$, and $b' = (1 + \sigma)w^* - T(\sigma)$.

Given (24), to show (25) it is sufficient to prove that the function A defined as follows satisfies $A'(0) \geq 0$:

$$A(y) \equiv \frac{\int_{a + \frac{y}{v(a)}}^{b + \frac{y}{v(b)}} v(x)dx - ((b + \frac{y}{v(b)}) - (a + \frac{y}{v(a)}))\frac{v(b + \frac{y}{v(b)}) + v(a + \frac{y}{v(a)})}{2}}{v(b + \frac{y}{v(b)}) - v(a + \frac{y}{v(a)})}. \quad (26)$$

The rest of the proof is an algebraic endeavor to demonstrate that $A'(0) \geq 0$. Let:

$$S_1 = \int_a^b v(x)dx - (b - a)\frac{v(a) + v(b)}{2} \quad \text{and} \quad S_2 = (v(b) - v(a))\frac{b - a}{2}.$$

Given these two definitions, $A'(0)$ can be written as:

$$A'(0) = \frac{S_1}{S_2} \times \left[\frac{1}{2v(a)} \left(\frac{v(b) - v(a) - (b - a)v'(a)}{S_1} + \frac{v'(a)(b - a)}{S_2} \right) + \right.$$

$$\frac{1}{2v(b)} \left(\frac{v(b) - v(a) - (b-a)v'(b)}{S_1} - \frac{v'(b)(b-a)}{S_2} \right)$$

Note that $S_2 > 0$. Also $S_1 \geq 0$ with equality holding only for a linear v . In the case of a linear v , Theorem 2 is trivial. Therefore, assume that v has some points of strict concavity on $[a, b]$. Thus, $S_1 > 0$. Therefore, $A'(0) \geq 0$ if and only if $A'(0) \times \frac{S_1}{S_2} \geq 0$. Rearranging and simplifying terms, $A'(0) \geq 0$ if and only if:

$$\frac{-\frac{v(a)}{v(b)} + \frac{v(b)}{v(a)} - (b-a)\left(\frac{v'(b)}{v(b)} + \frac{v'(a)}{v(a)}\right)}{S_1} + \frac{(b-a)\left(\frac{v'(a)}{v(a)} - \frac{v'(b)}{v(b)}\right)}{S_2} \geq 0.$$

Further rearranging, $A'(0) \geq 0$ if and only if:

$$\frac{S_1 + S_2}{S_2} \leq \frac{-v(a)^2 + v(b)^2 - 2(b-a)v'(a)v(b)}{(b-a)[v'(b)v(a) - v'(a)v(b)]}.$$

Because $\int_a^b v(x)dx = 0$ and by the definitions of S_1 and S_2 , one can verify that:

$$\frac{S_1 + S_2}{S_2} = \frac{-2v(a)}{v(b) - v(a)}.$$

Substituting this equation into the prior inequality and further rearranging, $A'(0) \geq 0$ if and only if:

$$\frac{-2v(a)}{v(b) - v(a)} - \frac{2v'(a)v(b)}{v'(a)v(b) - v'(b)v(a)} \leq \frac{v(a)^2 - v(b)^2}{(b-a)[v'(a)v(b) - v'(b)v(a)]}.$$

Now note that $v'(a) \geq 1$, $v(b) > 0$, $v'(b) \geq 0$ and $v(a) < 0$. Therefore, $v'(a)v(b) - v'(b)v(a) > 0$. Given that v is concave but not always linear on $[a, b]$, $v(0) = 0$, and $v'(0) = 1$, we also have that $v(a)^2 > v(b)^2$. We also have $v(b) - v(a) > 0$. Thus, multiplying the two sides of the inequality by

$$\frac{[v'(a)v(b) - v'(b)v(a)](v(b) - v(a))}{v(a)^2 - v(b)^2}$$

will not change the direction of the inequality. Performing this multiplication and further rearranging, $A'(0) \geq 0$ if and only if:

$$\frac{2v'(b)v(a)^2 - 2v'(a)v(b)^2}{v(a)^2 - v(b)^2} \leq \frac{v(b) - v(a)}{b-a}. \quad (27)$$

Note that in inequality (27), only the numerator of the left-hand-side includes $v'(b)$ and $v'(a)$. Thus, showing that inequality (28) below holds is sufficient to prove the theorem:

$$\frac{2B(a, b, v(a), v(b))}{v(a)^2 - v(b)^2} \leq \frac{v(b) - v(a)}{b - a}, \quad (28)$$

where $B(a, b, v(a), v(b)) = \max_{v'(a), v'(b)} v'(b)v(a)^2 - v'(a)v(b)^2$ subject to the constraints that (i) $v \in \mathcal{V}$ and v is not always linear on $[a, b]$, (ii) $\int_a^b v(x)dx = 0$, and (iii) the values of v at points a and b are fixed.

To gain intuition on how the values of $v'(a)$ and $v'(b)$ interact with the constraints on the optimization problem that defines B , see Figure 6. The figure depicts a piece-wise linear function $z(x)$ whose values and derivatives are equal to those of $v(x)$ at three points: a , 0 , and b .

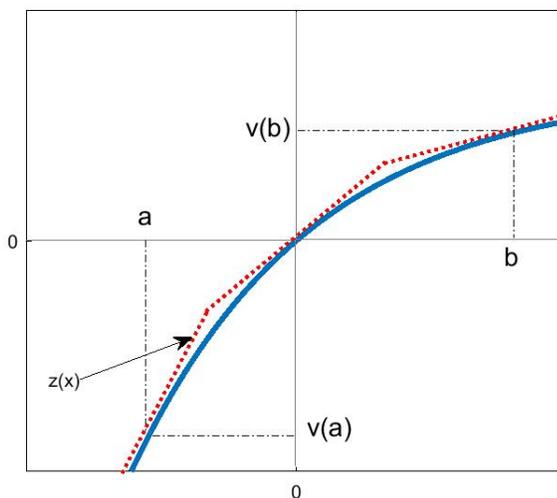


Figure 6: Piece-wise linear function $z(x)$ as an upper bound to $v(x)$

By the concavity of v , we know that $\forall x : z(x) \geq v(x)$. Therefore:

$$\int_a^b z(x)dx \geq \int_a^b v(x)dx = 0. \quad (29)$$

Thus, one can think of $\int_a^b z(x)dx \geq 0$ as a (weakly) weaker constraint on the optimization problem than $\int_a^b v(x)dx = 0$. Therefore, if we define

$$\bar{B}(a, b, v(a), v(b)) = \max_{v'(a), v'(b)} v'(b)v(a)^2 - v'(a)v(b)^2$$

subject to the same constraints as those for B except that $\int_a^b z(x)dx \geq 0$ replaces $\int_a^b v(x)dx =$

0, then $\bar{B}(a, b, v(a), v(b)) \geq B(a, b, v(a), v(b))$. Thus, the theorem will be proven if we show that equation (28) holds even if we replace $B(a, b, v(a), v(b))$ by $\bar{B}(a, b, v(a), v(b))$. To show this, we characterize \bar{B} in the following lemma.

Lemma 6. $\bar{B}(a, b, v(a), v(b)) = \frac{(bv(a) - av(b))^2}{a^2 - b^2} + v(a)^2 - v(b)^2$.

Proof of Lemma 6. As can be seen from definition of \bar{B} , it increases in $v'(b)$ and decreases in $v'(a)$. Nevertheless, it is clear from Figure 6 that if $v'(b)$ is chosen too high or $v'(a)$ is chosen too low, then the constraint (29) could be violated. Therefore, the maximum of \bar{B} is given by the highest $v'(b)$ and lowest $v'(a)$ that do not violate the inequality. We take a Lagrangian approach to solve the optimization problem. Writing out $\int_a^b z(x)dx$ and taking its derivatives with respect to $v'(a)$ and $v'(b)$ yields:

$$\frac{\partial}{\partial v'(a)} \left(\int_a^b z(x)dx \right) = \frac{(a - v(a))^2}{2} \frac{1}{(1 - v'(a))^2}$$

and

$$\frac{\partial}{\partial v'(b)} \left(\int_a^b z(x)dx \right) = \frac{(b - v(b))^2}{2} \frac{-1}{(1 - v'(b))^2}.$$

Taking the derivatives of $v'(b)v(a)^2 - v'(a)v(b)^2$ with respect to $v'(a)$ and $v'(b)$ yields $v(b)^2$ and $-v(a)^2$ (respectively). Thus, the first order condition implies:

$$\frac{(a - v(a))^2}{2v(b)^2} \frac{1}{(1 - v'(a))^2} = \frac{(b - v(b))^2}{2v(a)^2} \frac{1}{(1 - v'(b))^2}.$$

Given that $a - v(a) > 0$, $v(b) > 0$, $1 - v'(a) < 0$, $b - v(b) > 0$, $v(a) < 0$, and $1 - v'(b) > 0$, taking the square root of the above yields:

$$\frac{\frac{a-v(a)}{1-v'(a)}}{v(b)} = \frac{\frac{b-v(b)}{1-v'(b)}}{-v(a)} = \kappa. \quad (30)$$

Now the problem boils down to finding the “optimal” value of κ . The optimal value should set $\int_a^b z(x)dx$ to 0. This integral can be written out in terms of κ . By writing out the piece-wise linear integral, one can show:

$$\int_a^b z(x)dx = \frac{b^2 - a^2}{2} - \frac{\kappa}{2}(bv(a) - av(b)).$$

Thus, $\int_a^b z(x)dx = 0$ yields:

$$\kappa = \frac{a^2 - b^2}{bv(a) - av(b)}. \quad (31)$$

Equations (30) and (31) allow us write the optimal values of $v'(a)$ and $v'(b)$ based on the fixed values of a , b , $v(a)$, and $v(b)$. Substituting those expressions into $v'(b)v(a)^2 - v'(a)v(b)^2$ yields an expression for \bar{B} that does not contain $v'(a)$ or $v'(b)$. Performing the substitution and simplifying yields:

$$\bar{B}(a, b, v(a), v(b)) = \frac{(bv(a) - av(b))^2}{a^2 - b^2} + v(a)^2 - v(b)^2,$$

which is exactly what Lemma 6 states. **Q.E.D.**

Given this characterization, we can now show that (28) indeed holds, completing the proof of the theorem. Note that the right-hand-side of (28) is non-negative. Thus, it is sufficient to show that:

$$\frac{B(a, b, v(a), v(b))}{v(a)^2 - v(b)^2} \leq 0. \quad (32)$$

To see that this inequality holds, note that by Lemma 6 and by $B(a, b, v(a), v(b)) \leq \bar{B}(a, b, v(a), v(b))$, we have:

$$\frac{B(a, b, v(a), v(b))}{v(a)^2 - v(b)^2} \leq 2 \frac{\frac{(bv(a) - av(b))^2}{a^2 - b^2} + v(a)^2 - v(b)^2}{v(a)^2 - v(b)^2}.$$

In this last expression, multiplying the numerator and denominator by $a^2 - b^2$ and expanding all of the terms yields:

$$2 \frac{b^2v(a)^2 + a^2v(b)^2 - 2abv(a)v(b) + a^2v(a)^2 + b^2v(b)^2 - a^2v(b)^2 - b^2v(a)^2}{(a^2 - b^2)(v(a)^2 - v(b)^2)}.$$

Simplifying yields:

$$2 \frac{(av(a) - bv(b))^2}{(a^2 - b^2)(v(a)^2 - v(b)^2)}.$$

Note that $(av(a) - bv(b))^2 \geq 0$, $a^2 - b^2 < 0$ and $v(a)^2 - v(b)^2 > 0$. Thus, the whole expression is non-positive, which proves that (32) holds. This, in turn, implies that $A'(0) \geq 0$.

To summarize, after slightly increasing σ from σ_1 , D 's marginal rate of substitution $\Delta(T(\sigma), w^*|\sigma, v)$ will be smaller than $\Delta(T^*, w^*|\sigma_1, v)$ (i.e., inequality (20) holds). Thus, $t^*|\sigma, v \geq t^*|\sigma_1, v$, which proves Theorem 2. Note that the fact that the increase from σ_1 to σ needs to be small does not undermine the proof. We have proven that if σ slightly

increases, t^* will weakly increase. Thus, if we define t^* as a function of σ (keeping v fixed), that function is non-decreasing at every point, and hence non-decreasing globally. Thus even with large increases in σ , the linear fee t^* will increase. **Q.E.D.**

Proof of Theorem 3. See the main text (section 3.3) for specific counterexamples illustrating that increased risk or risk aversion can mitigate double marginalization.

Proof of Theorem 4. Similarly to Lemma 3 in the proof of Theorem 1, it can be shown that Theorem 4 holds if, for all $v_2 \in \mathcal{V}$ that are globally more risk averse than $v_1 \in \mathcal{V}$:

$$\Delta(T(v_2), w^* | \sigma, v_2) \leq \Delta(T^*, w^* | \sigma, v_1), \quad (33)$$

where (T^*, w^*) is the equilibrium contract under (σ, v_1) and $T(v_2)$ is the fixed fee that restores D to zero utility (see Definition 5). We again omit the proof that satisfying this condition implies Theorem 4 for parsimony: the proof follows the same logic as the proof of Lemma 3.

To show that inequality (33) indeed holds when v_2 is globally more risk averse than v_1 , we start by the premise of the theorem (i.e., inequality (10)). Let $L = 2w^*\sigma$. This implies that $a_{v_1, L} = w^*(1 - \sigma) - T^*$ and $a_{v_2, L} = w^*(1 - \sigma) - T(v_2)$. From (10), we have:

$$\frac{-v_2(a_{v_2, L})}{v_2(L + a_{v_2, L})} \geq \frac{-v_1(a_{v_1, L})}{v_1(L + a_{v_1, L})}.$$

Note that $-v_2(a_{v_2, L}) \geq v_2(L + a_{v_2, L}) > 0$. The same holds for v_1 . Rearranging thus yields:

$$\frac{v_2(a_{v_2, L}) + v_2(L + a_{v_2, L})}{v_2(L + a_{v_2, L}) - v_2(a_{v_2, L})} \geq \frac{v_1(a_{v_1, L}) + v_1(L + a_{v_1, L})}{v_1(L + a_{v_1, L}) - v_1(a_{v_1, L})}.$$

Now note that for $i \in \{1, 2\}$, we have $\int_{a_{v_i, L}}^{L+a_{v_i, L}} v(x)dx = 0$. Therefore, multiplying both sides of the inequality above by the positive number $\frac{L}{2}$ and then adding zero to both sides does not change the direction of the inequality:

$$\frac{\int_{a_{v_2, L}}^{L+a_{v_2, L}} v(x)dx - \frac{L}{2}(v_2(a_{v_2, L}) + v_2(L + a_{v_2, L}))}{v_2(L + a_{v_2, L}) - v_2(a_{v_2, L})} \geq \frac{\int_{a_{v_1, L}}^{L+a_{v_1, L}} v(x)dx - \frac{L}{2}(v_1(a_{v_1, L}) + v_1(L + a_{v_1, L}))}{v_1(L + a_{v_1, L}) - v_1(a_{v_1, L})}.$$

Combined with equation (21), this inequality directly implies (33), completing the proof of the theorem. **Q.E.D.**

Proof of Theorem 5 part a. For any $f \in \mathcal{F}$ and $v \in \mathcal{V}$, D 's marginal rate of substitution at contract (T, w) is given by equation (12), copied below:

$$\Delta(T, w|f, v) = \frac{\int_0^\infty \alpha v'(\alpha w - T) f(\alpha) d\alpha}{\int_0^\infty v'(\alpha w - T) f(\alpha) d\alpha}.$$

If f_1 is degenerate it will have its full mass at $\alpha = 1$ and thus $\Delta(T, w|f_1, v) = 1$ for all T, w . For the case of a non-degenerate $f_2 \in \mathcal{F}$ with an expected value of one, we can show $\Delta(T, w|f_2, v) < 1$ for all T, w . To see that $\Delta(T, w|f_2, v) < 1$, first observe that:

$$\Delta(T, w|f_2, v) = \int_0^\infty \alpha g(\alpha) d\alpha,$$

where $g(\alpha) = \frac{v'(\alpha w - T) f_2(\alpha)}{\int_0^\infty v'(\beta w - T) f_2(\beta) d\beta}$. It is straightforward to verify that g is itself a probability density function over the range $[0, \infty]$. Thus, $\Delta(T, w|f_2, v)$ is the expected value of α if it was drawn from probability distribution g instead of f_2 . Therefore, all we need to show is that the expected value of α when distributed according to g is strictly less than one. This result follows from the strict concavity of v , which implies that $v'(\alpha w - T)$ is positive and strictly decreasing in α . The probability distribution g is essentially a re-weighting of f_2 that shifts density from higher values of α to lower values. Since the expected value of α under f_2 is equal to one, the expected value under g is strictly less than one. Therefore, $\Delta(T', w'|f_2, v) < 1 = \Delta(T, w|f_1, v)$ for any contracts (T', w') and (T, w) . Following the same logic as Lemma 3, this condition ensures that the equilibrium linear fee under f_2 will be higher than under f_1 . **Q.E.D.**

Proof of Theorem 5 part b. For $v_1(x) = x$, we have that $v_1'(x) = 1$. Thus:

$$\Delta(T, w|f, v_1) = \frac{\int_0^\infty \alpha f(\alpha) d\alpha}{\int_0^\infty f(\alpha) d\alpha}.$$

In the fraction above, the numerator is the expected value of α according to f , which is equal to one for all $f \in \mathcal{F}$ (by definition). The denominator is also equal to one simply because f is a probability distribution. Thus, $\Delta(T, w|f, v_1) = 1$ for all T, w .

For a strictly concave $v_2 \in \mathcal{V}$, on the other hand, $\Delta(T, w|f, v_2) < 1$ for all T, w . To see that $\Delta(T, w|f, v_2) < 1$, construct the probability distribution g in the same way as in the proof of part a of the theorem. As explained there, g is essentially a re-weighting of f that shifts density from higher values of α to lower values. Therefore, the expected value

of α under g is strictly less than the expected value under f . Since the expected value of α under f is equal to one, the expected value under g is strictly less than one. Therefore, $\Delta(T', w'|f, v_2) < 1 = \Delta(T, w|f, v_1)$ for any contracts (T', w') and (T, w) . Following the same logic as Lemma 3, this condition ensures that the equilibrium linear fee under v_2 will be higher than under v_1 . **Q.E.D.**

7.4 Uniqueness of the equilibrium contract

In this section, we provide sufficient conditions for the equilibrium contract to be unique. To begin, it is worth noting that uniqueness is not the case in situations where there are no gains from trade. If the product is too costly for U to produce compared to the value that D 's customers assign to it, all contract offers by U that are not acceptable to D are equilibria. All such equilibria give U an expected profit of zero and D zero utility. In such situations, there may also be additional equilibria with $T = 0$ and a sufficiently large t so that $Q(p(t)) = 0$. D can accept this offer in equilibrium, though the consequences are the same as if the offer was rejected.

Setting aside cases in which there are no gains of trade, uniqueness follows from further assumptions on the shape of $\mathbf{E}\pi_U(T, w)$. To begin, it is helpful to show that D 's "better-than" sets are convex.

Lemma 7. *Fix the demand risk distribution f and D 's utility function v . For every real number k , define the set A_k as the set of (T, w) that deliver D utility of at least k . That is, $A_k = \{(T, w) : \mathbf{E}v_D(T, w) \geq k\}$. The set A_k is convex.*

Proof. Assume $\mathbf{E}v_D(T_1, w_1) = k_1 \geq k$ and $\mathbf{E}v_D(T_2, w_2) = k_2 \geq k$. Construct (T_3, w_3) with $T_3 = \theta T_1 + (1 - \theta)T_2$ and $w_3 = \theta w_1 + (1 - \theta)w_2$ for some $\theta \in (0, 1)$. It follows that for all α :

$$\pi_D(T_3, w_3|\alpha) = \theta\pi_D(T_1, w_1|\alpha) + (1 - \theta)\pi_D(T_2, w_2|\alpha).$$

By concavity of v , $\mathbf{E}v_D(T_3, w_3) \geq \theta k_1 + (1 - \theta)k_2 \geq k$. Therefore, $(T_3, w_3) \in A_k$, which implies that A_k is convex. **Q.E.D.**

We are now ready to demonstrate that an assumption on the shape of $\mathbf{E}\pi_U(T, w)$ guarantees that any equilibrium involving strict gains from trade is unique.

Proposition 4. *If $\mathbf{E}\pi_U(T, w)$ is strictly quasi-concave for all $w > 0$, then for any $f \in \mathcal{F}$ and $v \in \mathcal{V}$, the equilibrium two-part tariff contract $(T^*, t^*|f, v)$ is unique if $\mathbf{E}\pi_U(T^*, t^*) > 0$.*

Proof. In any such equilibrium, we have $w^* > 0$. To see that $w^* > 0$, first observe that $Q(p(t^*)) > 0$, otherwise there would be no gains from trade. Moreover, if $Q(p(t^*)) > 0$, then it must also be that $p(t^*) > t^*$, because D would never optimally set the downstream price less than or equal to the linear fee. ($p(t^*) = t^*$ cannot be optimal because a slight increase in the downstream price would strictly increase D 's profit given the continuity of Q .) Thus, $w^* = (p(t^*) - t) \cdot Q(p(t^*)) > 0$.

The rest of the proof proceeds by contradiction. In addition to (T^*, w^*) , suppose there is another equilibrium (\tilde{T}, \tilde{w}) . At these equilibria, $\mathbf{E}v_D(T^*, w^*) = 0$ and $\mathbf{E}v_D(\tilde{T}, \tilde{w}) = 0$ (part a of Proposition 2). Moreover, it must be that $\mathbf{E}\pi_U(T^*, w^*) = \mathbf{E}\pi_U(\tilde{T}, \tilde{w})$. If not, the contract with lower expected profit for U would not be an equilibrium.

Now construct (T', w') with $T' = \frac{T^* + \tilde{T}}{2}$ and $w' = \frac{w^* + \tilde{w}}{2}$. Since $w^* > 0$ and $\tilde{w} > 0$, and given the conditions of the proposition that $\mathbf{E}\pi_U(T, w)$ is strictly quasi-concave for all $w > 0$, we have:

$$\mathbf{E}\pi_U(T', w') > \mathbf{E}\pi_U(T^*, w^*) = \mathbf{E}\pi_U(\tilde{T}, \tilde{w}).$$

Moreover, Lemma 7 implies that $\mathbf{E}v_D(T', w') \geq 0$. Therefore, if U deviates to (T', w') , U will make a strictly higher expected profit and D will accept the offer, contradicting that (T^*, w^*) is an equilibrium. Thus, the equilibrium (T^*, w^*) must be unique. Finally, given $w^* > 0$ and that w is strictly decreasing in t when $w > 0$ (Lemma 1), this unique (T^*, w^*) translates to a unique (T^*, t^*) , completing the proof of the proposition. **Q.E.D.**