Institute
and Faculty
of Actuaries

ORIGINAL RESEARCH PAPER

Understanding the correlation risk premium

Jan Dhaene[1], Daniël Linders[2] , Biwen Ling[1] and Qian Wang[3]

[1]Katholieke Universiteit Leuven, Leuven, Belgium; [2]University of Amsterdam, Amsterdam, Netherlands;
and [3]Beijing Technology and Business University, Beijing, China
Corresponding author: Biwen Ling; Email: biwen.ling@kuleuven.be

Abstract

In this paper, we provide a theoretical framework justifying the existence of a correlation risk premium in a market with two traded assets. We prove that risk-neutral dependence can differ substantially from real-world dependence by characterizing the set of risk-neutral martingale measures. This implies that implied correlation can be significantly different with the realized correlation. Depending on the choice of the market regarding the pricing measure, implied correlation can be high or low. We label the difference between risk-neutral and real-world correlation the "correlation gap" and make the connection with correlation risk premium. We show how dispersion trading can be used to exploit this correlation gap and demonstrate how there can exist a negative correlation risk premium in the financial market.

1. Introduction

Correlation among assets plays an important role in the financial market. Empirical evidence shows that market returns are lower when correlations among assets are increasing, since higher correlations reduce the diversification effect and increase the market volatility. Roughly stated, the correlation risk premium (CRP) is the difference between the realized and the option-implied, i.e., risk-neutral correlation.[1] The CRP can be interpreted as an insurance premium paid for assets that hedge against unanticipated rises in correlation. Academic research has provided more empirical evidence to show that the CRP in equity markets is economically and statistically significant and long-term market returns can be predicted using the option-implied correlation; see, e.g., Driessen et al. (2009), Faria & Kosowski (2014), Buss et al. (2017), and Faria et al. (2022). In this paper, we will provide a theoretical framework that allows to better understand the CRP.

Stocks and market indices are modeled as random variables on the probability space $\left(\Omega, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$. Under the assumption of no-arbitrage, the prices of traded derivatives can be expressed as discounted expected payoffs under a risk-neutral probability measure $\mathbb{Q}$. Implied correlations (i.e., the correlation between assets under the risk neutral measure) can be determined from traded derivative prices, as was shown in Skintzi & Refenes (2004). Indeed, implied correlation provides a measure of the relative cheapness/richness of index options in relation to the index components, see, e.g., Chicago Board Options Exchange (2022). It reveals in the first place information about the degree of the comovement under the probability measure $\mathbb{Q}$, see also Dhaene et al. (2012), Linders et al. (2015), and Madan & Schoutens (2013) for alternative

---

[1]we will provide an exact definition of the correlation risk premium later in this paper.

implied dependence measures. The discrepancy between the realized correlation (i.e., the correlation between assets under the real-world measure $\mathbb{P}$) and the implied correlation gives rise to the existence of the CRP.

The increasing complexity of insurance products has introduced the need to understand the difference between implied and real-world correlation when dealing with insurance problems. Indeed, modern insurance products combine actuarial and financial risks. Therefore, their valuation and risk management rely on real-world and risk-neutral probabilities. For example, Solvency II requires that insurance companies value their liabilities in a "fair'" way, which implies that risk-neutral valuation has to be used for financial risks and real-world valuation for the actuarial risks. See, for example, Pelsser & Stadje (2014), Ghalehjooghi & Pelsser (2020), Dhaene et al. (2017), Barigou et al. (2019), and Linders (2023) for various methodologies for pricing insurance products based on combinations of real-world and risk-neutral information. Examples of complex insurance products combining financial and actuarial risks are variable annuities. In Bauer et al. (2008) and Bacinello et al. (2011), a valuation framework for a general class of variable annuities based on risk-neutral and real-world expectations was introduced. Concrete examples are then provided in Coleman et al. (2007), Feng & Jing (2017), and MacKay et al. (2023). Therefore, understanding the difference between risk-neutral and real-world information is important when considering variable annuities.

In this paper, we consider a discrete market setting. For stochastic finance in discrete time, one can refer to Föllmer & Schied (2004). We investigate to what extent the implied correlation reveals information about the degree of the comovement under the probability measure $\mathbb{P}$. We illustrate how statements which hold true in the risk-neutral world do not necessarily hold in the real world. For example, stock prices can be strongly negative dependent in the risk-neutral world (under $\mathbb{Q}$) while being positive dependent in the real world (under $\mathbb{P}$), this leads to a large difference between the real-world and the risk-neutral correlation, which is quantified by the correlation gap. Moreover, we introduce a new derivative called dispersion swap to trade the correlation gap and demonstrate that the correlation gap does not converge to zero for market equilibrium, i.e., the realized correlation can be different with the implied correlation in the case of market equilibrium.

In Section 2, we introduce a simple discrete financial market with two traded stocks. We apply the multivariate binomial tree model for the underlying stock prices. At time t, $t = 0, 1, 2 \ldots$, the price of the stock at the next valuation moment $t + 1$ can only take two possible outcomes. This financial market is arbitrage-free and incomplete, we characterize the set of feasible risk-neutral probability measures and demonstrate that it is feasible to determine a risk-neutral measure $\mathbb{Q}$ by explicitly specifying the implied correlation. Note that we limit our analysis to two stocks in the financial market. This choice is motivated by the fact that a market with two assets offers an intuitive setting for both theoretical and numerical exploration of the correlation gap.[2]

In Section 3, we consider the situation where a pricing measure is chosen by the market from the set of feasible risk-neutral probability measures, see Section 3.1. We show that the pricing measure chosen by the market can differ substantially from the real-world probability measure without introducing arbitrage opportunities. We give an example where the dependence structure used to price multivariate derivatives is different from the real-world dependence structure. In such a situation there can be a significantly large correlation gap in the financial market, i.e., the difference between risk-neutral and real-world correlations is substantial.

In Section 4, we consider the sale of a unit-linked insurance product to a group of N policyholders. The payoff of this unit-linked contract is contingent upon the performance of a stock market fund comprising two stocks within the financial market. We demonstrate how disparities between real-world dependence and risk-neutral dependence contribute to determining the expected excess return above the risk-free return for each policyholder. To be more specific, an

---

[2]In a future research paper, we generalize this market situation to n assets, where $n \geq 2$.

example is presented in Section 4.1 to illustrate that the expected excess return for the purchase of the unit-linked insurance is determined by the correlation gap. A nonzero correlation gap leads to a nonzero expected excess return for buying the unit-linked insurance product. Therefore, policyholders of the unit-linked insurance product are also facing the correlation risk in the financial market.

At first, a large correlation gap may look like a dysfunction of the financial market. However, we propose in Section 5 that, in our simple market model, one can use a new derivative, the dispersion swap, to exploit the correlation gap. This strategy is not an arbitrage strategy, but as dependence under $\mathbb{Q}$ is moving further and further away from its $\mathbb{P}$ counterpart, buying the dispersion swap becomes more attractive in terms of larger expected profit. To be more precise, a large positive correlation gap, corresponding to the situation that the realized correlation is expected to exceed the implied correlation, results in a significant positive expected profit for longing the dispersion swap. Furthermore, we show how one can combine the floating leg of individual variance swaps and the floating leg of the index variance swap to approximate the floating leg of the dispersion swap, which is called the realized dispersion. The expected profit for a buyer of the dispersion swap is directly related to the correlation gap, the higher this gap, the higher the expected profit. The idea of setting up trading strategies to exploit the difference between realized and implied dependence was also discussed in Laurence (2008), Laurence & Wang (2008), Bossu (2014), and Meissner (2015).

Apart from trading the dispersion swap to make a profit, investors can also long the dispersion swap to hedge against unanticipated correlation spikes in the financial market. Because the investors may have a high degree of risk aversion towards the correlation risk, they are willing to buy the dispersion swap, even with a negative expected profit. The expected profit of a dispersion swap is also called the CRP, as it represents the price that market participants are willing to pay to sell correlation risk. On the other hand, we show that the dispersion swap with a strictly negative (or positive) expected return is not an arbitrage strategy. Therefore, we conclude in Section 6 that a market in equilibrium can accommodate a negative CRP.

2. The financial market

2.1 A discrete financial market

We consider a discrete financial market with two nondividend paying stocks over a finite time horizon. Today is time 0, the price of stock i ($i = 1$ or 2), at the future time t ($t = 1, 2, \ldots, n$), is denoted by $S_i(t)$.[3] Given the price of stock i at time $t - 1$, the future stock price at time t can only increase to $e^{u_i} S_i(t - 1)$ or decrease to $e^{d_i} S_i(t - 1)$. The *forward return* of stock i at time t is denoted by $R_i(t)$ and defined as:

$$R_i(t) = \log \frac{S_i(t)}{S_i(t-1)}, \quad i = 1 \text{ or } 2 \text{ and } t = 1, 2, \ldots, n. \tag{1}$$

The financial market is also home to a bank account, which allows borrowing and lending at a constant, risk-free interest rate r. The time 0 value of the risk-free asset is $B(0)$, and its time t value is given by $B(t) = e^{rt} B(0)$. We assume that e^{u_i} and e^{d_i} are symmetric with respect to the forward rate e^r:

$$e^{u_i} - e^r = -(e^{d_i} - e^r), \quad \text{for } i = 1, 2. \tag{2}$$

---

[3]All random variables we encounter are assumed to be defined on the common probability space $\left(\Omega, (\mathcal{F}_t)_{t=1,2,\ldots,n}, \mathbb{P}\right)$. The filtration $(\mathcal{F}_t)_{t=1,2,\ldots,n}$ is the natural filtration generated by the stock price processes.

Under the real-world probability measure $\mathbb{P}$, we denote the joint probabilities of the random vector $(R_1(t), R_2(t))$ as follows:

$$\mathbb{P}\left[R_1(t) = d_1, R_2(t) = d_2\right] = p_{dd}(t),$$
$$\mathbb{P}\left[R_1(t) = u_1, R_2(t) = d_2\right] = p_{ud}(t),$$
$$\mathbb{P}\left[R_1(t) = d_1, R_2(t) = u_2\right] = p_{du}(t),$$
$$\mathbb{P}\left[R_1(t) = u_1, R_2(t) = u_2\right] = p_{uu}(t). \tag{3}$$

We assume that the joint probabilities under real-world measure $\mathbb{P}$ are strictly positive. The distribution of the random vector $(R_1(t), R_2(t))$ is determined by the marginal distributions of $R_i(t)$, and the dependence structure connecting $R_1(t)$ and $R_2(t)$. We assume that for $i = 1, 2$, the random variables $R_i(1), R_i(2), \ldots, R_i(n)$ are independent of each other. The probability of stock 1 moving up and down from $t - 1$ to t is denoted by $p_{u\cdot}(t)$ and $p_{d\cdot}(t)$, respectively. For stock 2, these probabilities are denoted by $p_{\cdot u}(t)$ and $p_{\cdot d}(t)$.

Assume the financial market is arbitrage free, hence, there exists at least one probability measure $\mathbb{Q}$, called a risk-neutral probability measure, satisfying the following conditions:

1. $\mathbb{Q}$ and $\mathbb{P}$ are equivalent probability measures;
2. For any traded asset, its future payoff discounted at the risk-free rate r is a martingale with respect to $\mathbb{Q}$:

$$e^{-r}\mathbb{E}_{\mathbb{Q}}\left[S_i(t)|\mathcal{F}_{t-1}\right] = S_i(t-1), \text{ for } i = 1, 2, \text{ and } t = 1, 2, .., n. \tag{4}$$

If $\mathbb{Q}$ is a probability measure satisfying the above-stated conditions, we say that it is a feasible risk-neutral probability measure. Under $\mathbb{Q}$, the joint probabilities of the random vector $(R_1(t), R_2(t))$ are denoted by $q_{uu}(t)$, $q_{ud}(t)$, $q_{du}(t)$, and $q_{dd}(t)$. Additionally, the risk-neutral marginal probabilities in $[t-1, t]$ are denoted by $q_{u\cdot}(t)$, $q_{d\cdot}(t)$, $q_{\cdot u}(t)$, and $q_{\cdot d}(t)$. We show in the next subsection that this simple market model is incomplete by characterizing the set of feasible risk-neutral probability measures.

2.2 The set of equivalent martingale measures

Let us now characterize the set of all feasible risk-neutral probability measures. This set is denoted by $\mathcal{M}$. Each $\mathbb{Q} \in \mathcal{M}$ is characterized by the joint probabilities $q_{dd}(t)$, $q_{ud}(t)$, $q_{du}(t)$, $q_{uu}(t)$, $t = 1, 2, \ldots, n$. In Theorem 1, we characterize the risk-neutral pricing measure $\mathbb{Q}$ by the correlation coefficients $\rho_{\mathbb{Q}}(t) = \text{Corr}_{\mathbb{Q}}\left[R_1(t), R_2(t)\right]$, $t = 1, 2, \ldots, n$. A proof of this Theorem can be found in Appendix A.1.

Theorem 1. *Consider the stock price model (3) satisfying the conditions (2). The set $\mathcal{M}$ of risk-neutral probability measures can be characterized as follows:*

$$\mathbb{Q} \in \mathcal{M} \Leftrightarrow \exists \, \rho_{\mathbb{Q}}(t) \in (-1, 1), t = 1, 2, \ldots, n,$$

$$\text{such that } \begin{cases} q_{uu}(t) = q_{dd}(t) = \frac{1}{4}(1 + \rho_{\mathbb{Q}}(t)), \\ q_{ud}(t) = q_{du}(t) = \frac{1}{4}(1 - \rho_{\mathbb{Q}}(t)). \end{cases} \tag{5}$$

The risk-neutral measure $\mathbb{Q}$ is not unique and the market is incomplete. It is clear to see from Theorem 1 that the risk-neutral marginal probabilities are all equal to $\frac{1}{2}$. Note, however, that our model can be generalized to situations where the marginal risk neutral probabilities are different from $\frac{1}{2}$. Specifying a feasible risk-neutral measure $\mathbb{Q}$ under the stock price model (3), requires specifying the correlation efficient $\rho_{\mathbb{Q}}(t), t = 1, 2, \ldots, n$. Each risk-neutral probability measure $\mathbb{Q}$ in $\mathcal{M}$ has the same marginal distributions but different dependence structures. For instance, take $\rho_{\mathbb{Q}}(t) \equiv 0$, then we find the risk-neutral measure $\mathbb{Q}^{\perp} \in \mathcal{M}$ where the marginals are independent.

The situations characterized by the minimal correlation coefficient $\rho_{\mathbb{Q}^{\min}}(t) \equiv -1$ and the maximal correlation coefficient $\rho_{\mathbb{Q}^{\max}}(t) \equiv 1$ correspond with the probability measure $\mathbb{Q}^{\min}$ and the probability measure $\mathbb{Q}^{\max}$, respectively. The random vector $(R_1(t), R_2(t))$ under the probability measure $\mathbb{Q}^{\min}$ is counter-monotonic. In this situation, the components of the random vectors are maximum negative dependent. Inversely, the random vector $(R_1(t), R_2(t))$ under the probability measure $\mathbb{Q}^{\max}$ is comonotonic and the components of the random vector are maximum positive dependent.

The real-world joint probabilities are assumed to be strictly positive, hence all the risk-neutral probabilities specified by (5) are strictly positive. Using (5), we can directly find that for each $\mathbb{Q} \in \mathcal{M}$, $\rho_{\mathbb{Q}}(t) \in (-1, 1)$, which means that the comonotonic and the counter-monotonic cases are not reachable in $\mathcal{M}$. The larger $\rho_{\mathbb{Q}}(t)$, the "closer" the risk-neutral probability measure $\mathbb{Q}$ is to the maximum measure $\mathbb{Q}^{\max}$.

The set $\mathcal{M}$ contains a wide range of dependence structures. Each element $\mathbb{Q}$ in the set $\mathcal{M}$ of risk-neutral probability measures can be expressed as a linear combination of $\mathbb{Q}^{\min}$ and $\mathbb{Q}^{\max}$:

$$\mathbb{Q} \in \mathcal{M} \Leftrightarrow \exists \, \rho_{\mathbb{Q}}(t) \in (-1, 1) \text{ such that } \mathbb{Q} = \frac{(1 - \rho_{\mathbb{Q}}(t))}{2} \mathbb{Q}^{\min} + \frac{(1 + \rho_{\mathbb{Q}}(t))}{2} \mathbb{Q}^{\max}. \quad (6)$$

By increasing the correlation coefficient $\rho_{\mathbb{Q}}(t)$, we can gradually increase the dependence of the components $R_1(t)$ and $R_2(t)$.

It directly follows from (5) that the joint risk-neutral cdf $F_t^{\mathbb{Q}}$ of the forward return vector $(R_1(t), R_2(t))$ can be given by:

$$F_t^{\mathbb{Q}}(x_1, x_2) = \begin{cases} 0, & \text{if } x_1 < d_1 \text{ or } x_2 < d_2, \\ \frac{\rho_{\mathbb{Q}}(t) + 1}{4}, & \text{if } x_1 \in [d_1, u_1) \text{ and } x_2 \in [d_2, u_2), \\ \frac{1}{2}, & \text{if } x_1 \in [d_1, u_1) \text{ and } x_2 \geq u_2, \\ \frac{1}{2}, & \text{if } x_1 \geq u_1 \text{ and } x_2 \in [d_2, u_2), \\ 1, & \text{if } x_1 \geq u_1 \text{ and } x_2 \geq u_2. \end{cases} \quad (7)$$

It follows from (7) that the joint risk-neutral cdf of forward return $F_t^{\mathbb{Q}}(x_1, x_2)$ can be unambiguously determined by the correlation coefficient $\rho_{\mathbb{Q}}(t)$.

2.3 Comparing different risk-neutral measures

The set $\mathcal{M}$ contains different risk-neutral measures such that the marginals R_1 and R_2 are always the same. Therefore, the difference between multivariate risk-neutral measures is in dependence structure. To compare different probability measures in $\mathcal{M}$ and identify under which one the dependence is stronger or weaker, we can use multivariate stochastic orders. The notion of multivariate stochastic orders in actuarial science goes back to Yanagimoto & Okamoto (1969).

We first introduce the *Positive Quadrant Dependence* and the *Negative Quadrant Dependence*. The notions of PQD and NQD were introduced in Lehmann (1966).

Definition 2 (Quadrant Dependence). *The random vector $(R_1(t), R_2(t))$ is said to be Positive Quadrant Dependent under the probability measure $\mathbb{Q}$, notation $(R_1(t), R_2(t)) \sim \mathbb{Q}$-PQD, in case the vector $(R_1(t), R_2(t))$ satisfies:*

$$\mathbb{Q}[R_1(t) \leq x_1] \, \mathbb{Q}[R_2(t) \leq x_2] \leq F_t^{\mathbb{Q}}(x_1, x_2), \text{ for all } (x_1, x_2) \in \mathbb{R}^2.$$

The vector $(R_1(t), R_2(t))$ is said to be Negative Quadrant Dependent *under the probability measure $\mathbb{Q}$, notation $\mathbb{Q}$-NQD, in case the vector satisfies:*

$$\mathbb{Q}[R_1(t) \leq x_1] \, \mathbb{Q}[R_2(t) \leq x_2] \geq F_t^{\mathbb{Q}}(x_1, x_2), \text{ for all } (x_1, x_2) \in \mathbb{R}^2,$$

where $F_t^{\mathbb{Q}}$ is the joint cdf of $(R_1(t), R_2(t))$ under the probability measure $\mathbb{Q}$.

Using Expression (7) for $F_t^{\mathbb{Q}}$ results in the following implications:

$$\rho_{\mathbb{Q}}(t) \geq 0 \Longleftrightarrow \rho_{\mathbb{Q}}(t) \geq \rho_{\mathbb{Q}^\perp}(t) \Longleftrightarrow (R_1(t), R_2(t)) \text{ is } \mathbb{Q}\text{-PQD}, \tag{8}$$
$$\rho_{\mathbb{Q}}(t) \leq 0 \Longleftrightarrow \rho_{\mathbb{Q}}(t) \leq \rho_{\mathbb{Q}^\perp}(t) \Longleftrightarrow (R_1(t), R_2(t)) \text{ is } \mathbb{Q}\text{-NQD},$$

where $\rho_{\mathbb{Q}^\perp}(t)$ is the correlation of the independent copy of the random vector $(R_1(t), R_2(t))$, it is straightforward to show that $\rho_{\mathbb{Q}^\perp}(t)$ equals 0. Using the notion of quadrant dependence, we can measure the joint behavior of two random variables. If the risk-neutral correlation $\rho_{\mathbb{Q}}(t)$ is larger than 0, then the random vector $(R_1(t), R_2(t))$ is $\mathbb{Q}$-PQD, which means that the two random variables $R_1(t)$ and $R_2(t)$ are likely to assume small or large values simultaneously. Conversely, if $\rho_{\mathbb{Q}}(t)$ is less than 0, the random vector is $\mathbb{Q}$-NQD, implying an inverse relationship where $R_1(t)$ and $R_2(t)$ may move in different directions.

The concept of quadrant dependence measures the association between two random variables. However, under different risk-neutral measures $\mathbb{Q}^{(1)}$ and $\mathbb{Q}^{(2)}$, the marginals and quadrant dependence of the random vector $(R_1(t), R_2(t))$ may be identical. To distinguish the difference under such two different risk-neutral measures, we can use the correlation order introduced in Dhaene & Goovaerts (1996), since the joint distribution of $(R_1(t), R_2(t))$ is different when using two different risk-neutral measures.

Definition 3 (Correlation order). *Consider the risk-neutral probability measures $\mathbb{Q}^{(1)}$ and $\mathbb{Q}^{(2)}$ with correlation parameters $\rho_{\mathbb{Q}^{(1)}}(t)$ and $\rho_{\mathbb{Q}^{(2)}}(t)$, respectively. We say that the cdf's $F_t^{\mathbb{Q}^{(1)}}$ and $F_t^{\mathbb{Q}^{(2)}}$ are ordered in the correlation order, notation $F_t^{\mathbb{Q}^{(1)}} \preceq_{Corr} F_t^{\mathbb{Q}^{(2)}}$ if the following holds:*

$$F_t^{\mathbb{Q}^{(1)}}(x_1, x_2) \leq F_t^{\mathbb{Q}^{(2)}}(x_1, x_2), \text{ for all } (x_1, x_2) \in \mathbb{R}^2. \tag{9}$$

Intuitively, the inequality $F_t^{\mathbb{Q}^{(1)}} \preceq_{\mathrm{Corr}} F_t^{\mathbb{Q}^{(2)}}$ implies that the two stock prices in $[t-1, t]$ move stronger together under the probability measure $\mathbb{Q}^{(2)}$ than under the probability measure $\mathbb{Q}^{(1)}$, i.e., the probability of having simultaneously large/small realizations in $[t-1, t]$ is larger under $\mathbb{Q}^{(2)}$, compared to $\mathbb{Q}^{(1)}$. Moreover, comparing the correlations $\rho_{\mathbb{Q}^{(1)}}(t)$ and $\rho_{\mathbb{Q}^{(2)}}(t)$ gives information about the correlation order between the probability measures $\mathbb{Q}^{(1)}$ and $\mathbb{Q}^{(2)}$. Indeed, it follows directly from Expression (7) that the following equivalence relation holds:

$$\rho_{\mathbb{Q}^{(1)}}(t) \leq \rho_{\mathbb{Q}^{(2)}}(t) \Longleftrightarrow F_t^{\mathbb{Q}^{(1)}} \preceq_{\mathrm{Corr}} F_t^{\mathbb{Q}^{(2)}}. \tag{10}$$

There are infinitely many risk-neutral probability measures and each risk-neutral measure models the stock prices using a different dependence structure. If a contingent claim has to be priced, the market will pick a suitable pricing measure.

3. Real world vs risk-neutral measures

Section 2 shows that the market is incomplete and a whole set of risk-neutral measures exists. Market prices are determined by supply and demand, and we assume they do not allow for arbitrage. We also assume that sufficiently many derivatives are traded and these prices are publicly available. All market participants can observe these prices. Having these prices at our disposal allows to back out the choice of the market concerning the risk-neutral measure.

If we can obtain the risk-neutral pricing measure used to price traded derivatives, we can back out the view of the market about future dependencies between the stock prices. However, the market can choose from a wide range of possibilities for $\rho_{\mathbb{Q}}(t)$ and our results show that there is, mathematically, no reason why the market should take $\mathbb{Q}$ such that $\rho_{\mathbb{Q}}(t)$ is close to the real-world correlation $\rho_{\mathbb{P}}(t)$ for all t.

3.1 A market with a combined asset

From Theorem 1, the spot price of stock i and its potential outcomes d_i and u_i, fully specify the distribution of R_i. However, individual derivatives written on each stock do not give additional information about the joint distributions. To determine the multivariate distribution, a combined asset and derivative prices on this combined asset are necessary.

Consider the market described above, but now also assume that a stock market index is traded. Its price at time 0 is denoted by $S(0)$ and its price at time t is denoted by $S(t)$:

$$S(t) = S_1(t) + S_2(t), \text{ for } t = 0, 1, 2, \ldots, n. \tag{11}$$

We have that $\mathbb{E}_{\mathbb{Q}}[S(t)] = \mathbb{E}_{\mathbb{Q}}[S_1(t) + S_2(t)] = e^{rt}S(0)$. The forward return of the stock market index at time t is denoted by $R(t)$ and defined as follows:

$$R(t) = \log \frac{S(t)}{S(t-1)}, \text{ for } t = 1, 2, \ldots, n. \tag{12}$$

Call options on the index with maturity $T = 1, 2, \ldots, n$ are also traded. The payoff of an index call option is given by $(S(T) - K)_+$ where $(x)_+ = \max\{x, 0\}$. The price of an index call with maturity T and strike K is then denoted by $C_{\mathbb{Q}}[K, T]$ and can be expressed as follows:

$$C_{\mathbb{Q}}[K, T] = e^{-rT}\mathbb{E}_{\mathbb{Q}}\left[(S(T) - K)_+\right].$$

The cumulative distribution function of the random variable $S(T)$ under the risk-neutral measure $\mathbb{Q}$ is denoted by $F_{S(T)}^{\mathbb{Q}}$. We introduce the stop-loss order between two risk-neutral probability measures in terms of their call option curves.

Definition 4 (Stop-loss order). *Consider the stock price model described in (3) satisfying the conditions (2) and the stock market index defined in (11). Consider the risk-neutral probability measures $\mathbb{Q}^{(1)}$ and $\mathbb{Q}^{(2)}$. We say that $F_{S(T)}^{\mathbb{Q}^{(1)}}$ and $F_{S(T)}^{\mathbb{Q}^{(2)}}$ are ordered in the stop-loss order, notation $F_{S(T)}^{\mathbb{Q}^{(1)}} \preceq_{sl} F_{S(T)}^{\mathbb{Q}^{(2)}}$ if:*

$$C_{\mathbb{Q}^{(1)}}[K, T] \le C_{\mathbb{Q}^{(2)}}[K, T] \text{ for all } K \ge 0. \tag{13}$$

Intuitively, the stop-loss order relation $F_{S(T)}^{\mathbb{Q}^{(1)}} \preceq_{sl} F_{S(T)}^{\mathbb{Q}^{(2)}}$ implies that the stock market index $S(T)$ is more volatile under the pricing measure $\mathbb{Q}^{(2)}$ than under the pricing measure $\mathbb{Q}^{(1)}$. Indeed, one can prove the following implication:

$$F_{S(T)}^{\mathbb{Q}^{(1)}} \preceq_{sl} F_{S(T)}^{\mathbb{Q}^{(2)}} \implies \text{Var}_{\mathbb{Q}^{(1)}}[S(T)] \le \text{Var}_{\mathbb{Q}^{(2)}}[S(T)].$$

Therefore, call options are more expensive under the measure $\mathbb{Q}^{(2)}$ than under $\mathbb{Q}^{(1)}$. The following theorem shows that the correlation $\rho_{\mathbb{Q}}(1)$ determines the variability of the stock market index $S(1)$.

Theorem 5. *Consider the stock price model described in (3) satisfying the conditions (2) and the stock market index defined in (11). Then, the following equivalence relation holds:*

$$\rho_{\mathbb{Q}^{(1)}}(1) \le \rho_{\mathbb{Q}^{(2)}}(1) \iff F_1^{\mathbb{Q}^{(1)}} \preceq_{Corr} F_1^{\mathbb{Q}^{(2)}} \iff F_{S(1)}^{\mathbb{Q}^{(1)}} \preceq_{sl} F_{S(1)}^{\mathbb{Q}^{(2)}}.$$

Proof. From (10), we find that $\rho_{\mathbb{Q}^{(1)}}(1) \le \rho_{\mathbb{Q}^{(2)}}(1) \iff F_1^{\mathbb{Q}^{(1)}} \preceq_{Corr} F_1^{\mathbb{Q}^{(2)}}$. The call option price $C_{\mathbb{Q}}[K, 1]$ can be expressed as follows:

$$C_{\mathbb{Q}}[K, 1] = e^{-r}\left(\frac{1}{4}(P_1 + P_2 + P_3 + P_4) + (P_1 - P_2 - P_3 + P_4)\frac{\rho_{\mathbb{Q}}(1)}{4}\right), \tag{14}$$

where,

$$P_1 = \left(S_1\,(0)\,e^{u_1} + S_2\,(0)\,e^{u_2} - K\right)_+ \,; P_2 = \left(S_1\,(0)\,e^{u_1} + S_2\,(0)\,e^{d_2} - K\right)_+ \,;$$

$$P_3 = \left(S_1\,(0)\,e^{d_1} + S_2\,(0)\,e^{u_2} - K\right)_+ \,; P_4 = \left(S_1\,(0)\,e^{d_1} + S_2\,(0)\,e^{d_2} - K\right)_+ \,.$$

It follows from Expression (14) that $C_{\mathbb{Q}}\,[K, 1]$ is an increasing linear function of $\rho_{\mathbb{Q}}(1)$, since $(P_1 - P_2 - P_3 + P_4) \geq 0$. □

The time 0 market price of a call option with strike K and maturity 1 year, denoted by $C_{\mathbb{Q}}\,[K, 1]$, can then be used to extract the correlation $\rho_{\mathbb{Q}}\,(1)$ associated with the risk-neutral measure $\mathbb{Q}$, as demonstrated in (14). Indeed, we have that:

$$\rho_{\mathbb{Q}}\,(1) = \frac{4e^r C_{\mathbb{Q}}\,[K, 1] - (P_1 + P_2 + P_3 + P_4)}{(P_1 - P_2 - P_3 + P_4)}. \tag{15}$$

From Theorem 1, the joint probabilities $q_{dd}(1), q_{du}(1), q_{ud}(1), q_{uu}(1)$ in this case are fully specified. For the single period case, Theorem 5 also indicates the equivalence relation between the correlation of two stocks and the volatility of the stock index. Moreover, we provide the following Theorem 6 to show that having available the option prices $C_{\mathbb{Q}}\,[K, T]$, for $T = 1, 2, \ldots, n$, one can back out the pricing measure $\mathbb{Q}$.

Theorem 6. *Assume the index call option prices $C_{\mathbb{Q}}\,[K, T]$, $T = 1, 2, \ldots, n$ are all available, then the correlation coefficient $\rho_{\mathbb{Q}}(t), t = 1, 2, \ldots, n$ can unambiguously be determined.*

Proof. First, we already determined $\rho_{\mathbb{Q}}(1)$ from expression (15). Next, knowledge about the price $C_{\mathbb{Q}}[K, 2]$ enables us to back out the correlation $\rho_{\mathbb{Q}}(2)$. Indeed, we can write:

$$\begin{aligned}
C_{\mathbb{Q}}\,[K, 2] &= e^{-2r} \mathbb{E}_{\mathbb{Q}} \left[\mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid \mathcal{F}_1 \right] \right] \\
&= e^{-2r} \Big[q_{uu}(1) \mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid S_1(1) = S_1\,(0)\,e^{u_1}, S_2(1) = S_2\,(0)\,e^{u_2} \right] \\
&\quad + q_{ud}(1) \mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid S_1(1) = S_1\,(0)\,e^{u_1}, S_2(1) = S_2\,(0)\,e^{d_2} \right] \\
&\quad + q_{du}(1) \mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid S_1(1) = S_1\,(0)\,e^{d_1}, S_2(1) = S_2\,(0)\,e^{u_2} \right] \\
&\quad + q_{dd}(1) \mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid S_1(1) = S_1\,(0)\,e^{d_1}, S_2(1) = S_2\,(0)\,e^{d_2} \right] \Big]. \tag{16}
\end{aligned}$$

From Expression (14), we find that the conditional expectations are given by:

$$\mathbb{E}_{\mathbb{Q}} \left[(S(2) - K)_+ \mid S_1(1) = s_1, S_2(1) = s_2 \right] = \frac{(P_1 + P_2 + P_3 + P_4) + (P_1 - P_2 - P_3 + P_4)\,\rho_{\mathbb{Q}(2)}}{4},$$

where the constants $P_j, j = 1, 2, 3, 4$, are now defined as follows:

$$P_1 = \left(s_1 e^{u_1} + s_2 e^{u_2} - K\right)_+ \,; P_2 = \left(s_1 e^{u_1} + s_2 e^{d_2} - K\right)_+ \,;$$

$$P_3 = \left(s_1 e^{d_1} + s_2 e^{u_2} - K\right)_+ \,; P_4 = \left(s_1 e^{d_1} + s_2 e^{d_2} - K\right)_+ \,.$$

The probabilities $q_{uu}(1), q_{ud}(1), q_{du}(1), q_{dd}(1)$ are already determined from the call option price $C_{\mathbb{Q}}[K, 1]$. As a result, the only unknown parameter in the Expression (16) is $\rho_{\mathbb{Q}}(2)$ and observing the option price $C_{\mathbb{Q}}\,[K, 2]$ allows to solve for $\rho_{\mathbb{Q}}(2)$. Suppose we have derived the correlations $\rho_{\mathbb{Q}}(t), t = 1, 2, \ldots, i$, where $i < n$, we can then apply the same approach to acquire the correlation $\rho_{\mathbb{Q}}\,(i + 1)$ from the correlations $\rho_{\mathbb{Q}}(1), \rho_{\mathbb{Q}}(2), \ldots, \rho_{\mathbb{Q}}\,(i)$ and the index option price $C_{\mathbb{Q}}\,[K, i + 1]$. The probability distributions of $S_1(i)$ and $S_2(i)$ can be specified using $\rho_{\mathbb{Q}}\,(1), \rho_{\mathbb{Q}}\,(2), \ldots, \rho_{\mathbb{Q}}\,(i)$. Hence, by using the option price $C_{\mathbb{Q}}\,[K, i + 1]$, we can solve for $\rho_{\mathbb{Q}}\,(i + 1)$. Therefore, we conclude that having available the option prices $C_{\mathbb{Q}}\,[K, T]$, for $T = 1, 2, \ldots, n$, one can back out the pricing measure $\mathbb{Q}$. □

Theorem 6 showed that adding index call options with maturities $t = 1, 2, \ldots, T$, completes the market described in (3). Indeed, one can back out the risk-neutral correlation using index call options, and therefore the risk-neutral probability measure is unique in this market setting. The idea of extracting the implied correlation from available multivariate option prices is widely discussed in the literature. For example, Linders & Schoutens (2014) used basket option prices, Ballotta et al. (2017) used quanto options, and Garcia et al. (2009) derived implied correlations using CDO spreads.

Prices of traded stocks and options can be used to back out the corresponding risk-neutral pricing measure, the correlation of the stock returns and the volatility of the stock market index. We refer to these quantities as implied measures, implied correlation, implied volatility. Note that market prices are expectations under the pricing measure $\mathbb{Q}$ and therefore the implied correlation and the implied volatility have to be understood as correlation and volatility levels with respect to the risk-neutral probability measure $\mathbb{Q}$. Correlation and volatility under the probability $\mathbb{P}$ are referred to as real-world correlation and volatility, respectively.

3.2 Example: A single period model

We consider a one-period financial market as described in Section 2. The risk-free rate r is assumed to be 0 and consider $e^{u_1} = 1.4$, $e^{u_2} = 1.7$. The time 0 spot prices of the traded assets are given by $S_1(0) = 100$ and $S_2(0) = 200$. The dynamics of the financial market under the real-world measure $\mathbb{P}$ are described by the following equations:

$$p_{dd}(1) = 0.3, p_{ud}(1) = 0.2, \tag{17}$$
$$p_{du}(1) = 0.2, p_{uu}(1) = 0.3.$$

Notice that we choose the marginals such that the $\mathbb{P}$-marginals are the same as the $\mathbb{Q}$-marginals, this is to make the example simpler and is not required in general case.

Consider the stock market index $S(t) = S_1(t) + S_2(t)$, $t = 0, 1$. We find that the real-world correlation $\rho_{\mathbb{P}}(1) = \text{Corr}[R_1(1), R_2(1)] = 0.2$, and the real-world volatility $\sigma_{\mathbb{P}}(1) = \sqrt{\text{Var}_{\mathbb{P}}[R(1)]} = 0.585$. Since $\rho_{\mathbb{P}}(1) = 0.2 > 0$, the return vector $(R_1(1), R_2(1))$ is Positive Quadrant Dependent under the real-world probability measure $\mathbb{P}$:

$$(R_1(1), R_2(1)) \text{ is } \mathbb{P} - PQD. \tag{18}$$

We conclude from (18) that stocks are positively dependent under the real-world measure $\mathbb{P}$ specified by (17). However, the dependence structure under the risk-neutral measure $\mathbb{Q}$ can be different, even opposite from the dependence under $\mathbb{P}$. The following Proposition 1 is presented to show that the stock prices can be negatively dependent under the risk-neutral measure.

Proposition 1. *A call option written on the stock market index $S(1)$, with strike $K = 300$, is traded and its time 0 price $\hat{C}_{\mathbb{Q}}$ can be observed in the market. Then we have the following equivalence relations:*

$$\hat{C}_{\mathbb{Q}} \geq 70 \Longleftrightarrow \mathbb{Q} - PQD, \tag{19}$$
$$\hat{C}_{\mathbb{Q}} \leq 70 \Longleftrightarrow \mathbb{Q} - NQD$$

Proof. From (14), we have $\hat{C}_{\mathbb{Q}} = 70 + 20\rho_{\mathbb{Q}}(1)$. Hence, it follows from (8) directly to derive (19). $\square$

Proposition 1 shows that the market decides if the stocks are positive or negative dependent under the risk-neutral measure. Indeed, different market situations result in different risk-neutral measures, which can be different with the real-world measure $\mathbb{P}$. Here we provide three different

Table 1. Three different market situations given that $\rho_{\mathbb{P}} = 0.2$ and $\sigma_{\mathbb{P}} = 0.585$

Market	$\hat{c}_{\mathbb{Q}}$	$\rho_{\mathbb{Q}}$	$\sigma_{\mathbb{Q}}$	Quadrant Dependence
$\mathbb{Q}^{(1)}$	52	-0.9	0.373	$\mathbb{Q}^{(1)} - $ NQD
$\mathbb{Q}^{(2)}$	73	0.15	0.578	$\mathbb{Q}^{(2)} - $ PQD
$\mathbb{Q}^{(3)}$	86	0.8	0.668	$\mathbb{Q}^{(2)} - $ PQD

Table 2. Real-world and risk-neutral correlations

T	$\rho_{\mathbb{P}}(T)$	$\widehat{C}[300, T]$	$\rho_{\widehat{\mathbb{Q}}}(T)$
1	0.84	71.72	0.086
2	0.76	98.03	0.258
3	0.68	109.89	0.348
4	0.60	127.91	0.428
5	0.52	144.79	0.496
6	0.44	154.04	0.553
7	0.36	162.56	0.600
8	0.28	174.16	0.634
9	0.20	182.13	0.651
10	0.12	187.13	0.663

market situations in Table 1.

From Table 1, we can also conclude that the risk-neutral pricing measures $\mathbb{Q}^{(1)}$, $\mathbb{Q}^{(2)}$, and $\mathbb{Q}^{(3)}$ are ordered in the correlation order:

$$F_t^{\mathbb{Q}^{(1)}} \preceq_{\text{Corr}} F_t^{\mathbb{Q}^{(2)}} \preceq_{\text{Corr}} F_t^{\mathbb{Q}^{(3)}}.$$

This example shows that deriving the implied correlation and volatility to learn about the future dynamics of the stock prices is only part of the story. Implied measures are giving information about the risk-neutral dynamics and these statements cannot be directly translated to statements under the real-world probability measure. If the risk-neutral measure $\mathbb{Q}^{(1)}$ is chosen by the market, then implied volatility and correlation are substantially different with the real-world volatility and correlation. It might feel counter-intuitive to use a negative dependence structure to price the index option while under the real-world probability measure, the stocks are positive dependent. Notice that, however, the corresponding index option price does not lead to arbitrage and is consistent with other derivative prices.

3.3 Example: A multiperiod model

Consider the financial market described above, but now we consider the future times $t = 1, 2, \ldots, 10$. Assume that the real-world correlation $\rho_{\mathbb{P}}(t)$ is given by $\rho_{\mathbb{P}}(t) = 0.92 - 0.08t$, for $t = 1, 2, \ldots 10$. Similar to the one-period case, we still assume the real-world marginals are the same as the risk-neutral marginals, then the real-world dynamics can be expressed as:

$$p_{dd}(t) = \frac{1}{4}(1 + \rho_{\mathbb{P}}(t)), \, p_{ud}(t) = \frac{1}{4}(1 - \rho_{\mathbb{P}}(t)), \tag{20}$$

$$p_{du}(t) = \frac{1}{4}(1 - \rho_{\mathbb{P}}(t)), \, p_{uu}(t) = \frac{1}{4}(1 + \rho_{\mathbb{P}}(t)).$$

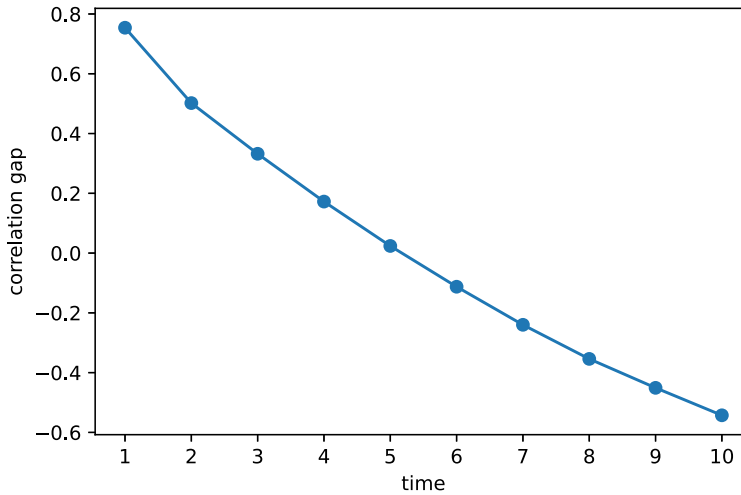We show real-world correlations in the second column of Table 2.

Figure 1. The correlation gap $\rho_{\mathbb{P}}(t) - \rho_{\widehat{\mathbb{Q}}}(t)$ with respect to time t.

Assume at time 0, at-the-money call options with maturities $T = 1, 2, \ldots, 10$ are traded. The market prices are denoted by $\widehat{C}[300, T]$, where $T = 1, 2, \ldots, 10$, and are given in the third column of Table 2. The market prices can be used to determine the unique risk-neutral measure $\widehat{\mathbb{Q}}$. Indeed, Theorem 6 shows that implied correlations $\rho_{\widehat{\mathbb{Q}}}(t)$ can be uniquely determined by index call option prices $\widehat{C}[300, t]$ for $t = 1, 2, \ldots, n$. We present implied correlations $\rho_{\widehat{\mathbb{Q}}}(t)$ in the fourth column of Table 2. We can find from Table 2 that there always exists a gap between real-world and implied correlations. We call this gap the correlation gap. The following Fig. 1 shows the plot of the correlation gap, $\rho_{\mathbb{P}}(t) - \rho_{\widehat{\mathbb{Q}}}(t)$ for time $t = 1, 2, \ldots, 10$.

Implied correlations capture the aggregate view of the market about future risk-neutral correlations. Table 2 and Fig. 1 clearly show that these views cannot always be directly translated to real-world statements. Indeed, in our example, the risk-neutral correlation is increasing: market participants anticipate a increase in correlation. However, the real-world correlation is moving in the opposite direction and decreases over time. Moreover, we can see from Fig. 1 that the correlation gap $\rho_{\mathbb{P}}(t) - \rho_{\widehat{\mathbb{Q}}}(t)$ is decreasing over time. The correlation gap is highly positive at time 1 and becoming strongly negative at time 10. Notably, a nonzero correlation gap may persist in the financial market over time. In the next Section 4, we demonstrate that this nonzero correlation gap not only manifests in the financial market but also carries implications for the purchase of unit-linked insurance products.

4. The correlation gap: Market-consistent valuation for insurance products

Consider an insurer selling a unit-linked product to a group of N policyholders of the same age x. For $i = 1, 2, \ldots, N$,, we denote by I_i the indicator random variable modeling the survival of the policyholder i,

$$I_i = \begin{cases} 0, & \text{if policyholder } i \text{ dies before time } T, \\ 1, & \text{if policyholder } i \text{ survives to time } T. \end{cases} \tag{21}$$

The payoff of this unit-linked contract depends on the performance of a stock market fund. We assume here, for simplicity, the fund consists of two stocks $S_1(t)$ and $S_2(t)$. However, if the fund value drops below a threshold K at time T, the insurer will compensate such that the policy holder

still receives the payoff K. To be more specific, policyholder i receives upon survival at time T the payoff $h_i(T)$ given by:

$$h_i(T) = \max(S_1(T) + S_2(T), K) \times I_i, \text{ for } i = 1, 2 \ldots, N. \tag{22}$$

Then the per-policy payout of the insurer is denoted by $h(T)$ and given by:

$$h(T) = \max(S_1(T) + S_2(T), K) \times \frac{1}{N} \sum_{i=1}^{N} I_i. \tag{23}$$

We assume the survival random variables $I_1, I_2, \ldots, I_n$ are i.i.d. and each I_i is independent from the stock prices $S_1(T)$ and $S_2(T)$. Moreover, we assume that $\mathbb{P}[I_i = 1] = {}_Tp_x$.

Market-consistent valuations are considered extensively for the valuation of unit-linked insurance products, see, e.g., Malamud et al. (2008), Artzner & Eisele (2010), Pelsser & Stadje (2014), Dhaene et al. (2017), and Linders (2023). A market-consistent valuation can be interpreted in the sense that the valuation of any hedgeable part within a claim aligns with the price of its hedge. Now let us consider a market-consistent (MC) valuation for the unit-linked insurance product with payout per policy given by (23). Denote this MC valuation by $\rho[h(T)]$ and it is given by:

$$\rho[h(T)] = e^{-rT} \mathbb{E}^{\mathbb{Q}}[\max(S_1(T) + S_2(T), K)]_Tp_x. \tag{24}$$

Assume that N is large and the insurance risk can be diversified. For simplicity, here, we use the best-estimate for the valuation of the actuarial risk within $h(T)$. Hence, we can regard the valuation $\rho(h(T))$ as a MC valuation based on perfect actuarial diversification. By using the fact that $e^{-rT} \mathbb{E}^{\mathbb{Q}}[\max(S_1(T) + S_2(T), K)]$ is equal to $C_{\mathbb{Q}}[K, T] + Ke^{-rT}$, the value $\rho[h(T)]$ can be rewritten as:

$$\rho[h(T)] = C_{\mathbb{Q}}[K, T]_Tp_x + Ke^{-rT}{}_Tp_x. \tag{25}$$

In a broader context, market-consistent valuations can fulfill different purposes, which can be employed not only in determining technical provisions but also in setting premiums. Assume that $\rho[h(T)]$ is the premium charged by the insurance company for each policyholder. The expected payoff for policyholder i at time T is given by $\mathbb{E}^{\mathbb{P}}[h_i(T)]$ and can be expressed as:

$$\mathbb{E}^{\mathbb{P}}[h_i(T)] = \mathbb{E}^{\mathbb{P}}[\max(S_1(T) + S_2(T), K)]_Tp_x, \text{ for } i = 1, 2, \ldots, N. \tag{26}$$

It directly follows (24) and (26) that the time T expected excess return above the risk-free return (Expected ER) for each policyholder to buy this unit-linked insurance product is given by:

$$\text{Expected ER} = \mathbb{E}^{\mathbb{P}}[h_i(T)] - e^{rT}\rho[h(T)] \tag{27}$$
$$= {}_Tp_x \left(\mathbb{E}^{\mathbb{P}}[\max(S_1(T) + S_2(T), K)] - \mathbb{E}^{\mathbb{Q}}[\max(S_1(T) + S_2(T), K)] \right).$$

By recognizing that $\mathbb{E}^{\mathbb{P}}[\max(S_1(T) + S_2(T), K)]$ can also be written as $K + \mathbb{E}^{\mathbb{P}}[(S_1(T) + S_2(T) - K)_+]$, the expected excess return for purchasing this unit-linked insurance contract can be rewritten as:

$$\text{Expected ER} = {}_Tp_x \left(\mathbb{E}^{\mathbb{P}}[(S_1(T) + S_2(T) - K)_+] - e^{rT}C_{\mathbb{Q}}[K, T] \right). \tag{28}$$

From (28), it is evident that the expected excess return for buying this unit-linked insurance contract depends on the expected excess return for investing in the index call option with a strike K and a maturity T. In Section 3.1, it is shown that implied correlations can be uniquely determined from the index call option prices. Hence, the expected excess return for each policyholder is associated with implied correlations. In the following subsection, we examine the same market setting in Example 3.2 to illustrate that the expected excess return for the policyholder of this unit-linked insurance can be determined by the correlation gap in the financial market.

Table 3. Expected excess return for each policyholder in three different market situations

Market	$\hat{C}_{\mathbb{Q}}$	$\rho_{\mathbb{Q}}$	$\rho_{\mathbb{P}} - \rho_{\mathbb{Q}}$	$\rho[h(1)]$	Expected ER
$\mathbb{Q}^{(1)}$	52	-0.9	1.1	$352p_x$	$22p_x$
$\mathbb{Q}^{(2)}$	73	0.15	0.05	$373p_x$	$3p_x$
$\mathbb{Q}^{(3)}$	86	0.8	-0.6	$386p_x$	$-12p_x$

4.1 Illustration: A single period model

Example 3.2 illustrates that there exists a correlation gap in the financial market. Consider the same market setting as in Example 3.2, we show how the correlation gap determines the expected excess return for buying the unit-linked insurance product.

In the single period market with $r = 0$, as in Example 3.2, let us consider a one-period unit-linked insurance product with $T = 1$. The payout per policy for this unit-linked insurance contract, given by (23), is denoted as $h(1)$. Assume the threshold $K = 300$, from (24), a market-consistent valuation of this one-period unit-linked insurance product is denoted as $\rho[h(1)]$ and given by:

$$\rho[h(1)] = \mathbb{E}^{\mathbb{Q}}[\max(S_1(1) + S_2(1), 300)]p_x = (\hat{C}_{\mathbb{Q}} + 300)p_x,$$

where $\hat{C}_{\mathbb{Q}}$ represents the time-0 index call option price with strike $K = 300$ and maturity $T = 1$, and p_x stands for the probability that a policyholder aged exactly x survives to time 1. Subsequently, the expected excess return for each policyholder at time-1 can be written as:

$$\text{Expected ER} = p_x \left(\mathbb{E}^{P}[(S(1) - 300)_+] - \hat{C}_{\mathbb{Q}} \right), \tag{29}$$

where $S(1)$ is the market index defined in (11).

Using the real-world dynamics in (17), the expectation $\mathbb{E}^{P}[(S(1) - 300)_+]$ can be expressed as follows:

$$\mathbb{E}^{P}[(S(1) - 300)_+] = 180 \times \frac{1}{4}(1 + \rho_{\mathbb{P}}(1)) + 100 \times \frac{1}{4}(1 - \rho_{\mathbb{P}}(1)) = 70 + 20\rho_{\mathbb{P}}(1).$$

Additionally, Proposition 1 implies that $\hat{C}_{\mathbb{Q}} = 70 + 20\rho_{\mathbb{Q}}(1)$. Hence in this market scenario, we can find that:

$$\text{Expected ER} = 20 \cdot p_x(\rho_{\mathbb{P}}(1) - \rho_{\mathbb{Q}}(1)). \tag{30}$$

From (30), we can clearly see that different correlation gaps give rise to different expected excess returns. Indeed, a positive (negative) correlation gap implies a positive (negative) expected excess return for buying the unit-linked insurance. Table 3 compares the expected excess return of the unit-linked insurance policyholder across the three different market scenarios presented in Example 3.2.

Table 3 clearly reveals that the presence of a nonzero correlation gap results in a nonzero expected excess return for the purchase of the unit-linked insurance product. Therefore, a policyholder of the unit-linked insurance contract is also exposed to a correlation risk in the financial market. In the following Section 5, we show how one can trade and hedge the correlation risk in the financial market.

5. Trading the correlation gap

We demonstrated that volatility and correlation under $\mathbb{P}$ and $\mathbb{Q}$ can differ substantially. In this section, we show that the existence of the correlation gap gives rise to trading strategies which have a large expected profit. We will introduce a new financial derivative called dispersion swap,

and show that the expected profit for longing the dispersion swap is related to the CRP in the financial market.

5.1 Variance swaps

We first introduce a new derivative in our financial market: A *variance swap* on the stock market index. At maturity T, the holder of the floating leg of a variance swap receives the difference $RV[T] - SR[T]$, where $RV[T]$ is defined as:

$$RV[T] = \frac{1}{T} \sum_{t=1}^{T} R(t)^2. \tag{31}$$

The quantity $RV[T]$ is called the annualized realized variance and represents the floating leg of the contract. The fixed leg of the contract is $SR[T]$, also called the swap rate. The swap rate of the contract is a constant which is determined at inception such that the contract is fair and no money has to be exchanged at inception. Denote by $\mathbb{Q}$ the pricing measure used by the market to price derivatives. The swap rate is then given by:

$$SR[T] = \mathbb{E}_{\mathbb{Q}} [RV[T]]. \tag{32}$$

The holder of a variance swap will pay the swap rate $SR[T]$ at maturity and will receive the realized variance $RV[T]$. At time 0, the swap rate is known but the realized variance is random. Meanwhile, investors are willing to pay a premium to hedge against the variance risk in the financial market, called the *variance risk premium (VRP)*; see, e.g., Carr & Wu (2009), Bollerslev & Zhou (2007), and Alibeiki & Lotfaliei (2021). Indeed, the VRP is the expected profit for the holder of the variance swap, which is given by:

$$VRP = \text{Expected Profit} = \mathbb{E}_{\mathbb{P}} [RV[T]] - SR[T].$$

The VRP can be expressed as follows:

$$VRP = \frac{1}{T} \sum_{t=1}^{T} \left(\sigma_{\mathbb{P}}^2(t) - \sigma_{\mathbb{Q}}^2(t) \right) + \frac{1}{T} \sum_{t=1}^{T} \left((\mathbb{E}_{\mathbb{P}} [R(t)])^2 - (\mathbb{E}_{\mathbb{Q}} [R(t)])^2 \right), \tag{33}$$

where $\sigma_{\mathbb{Q}}(t)$ is the $t-$year forward volatility. Expression (33) clearly shows that the holder of the variance swap longs the real-world volatility and shorts the risk-neutral volatility. The summation in the expression of the VRP arises because a variance swap is actually trading the second moment, rather than the variance. Investing in a variance swap becomes attractive if the gap between real and risk-neutral volatility is substantially large. If this gap is positive, it may lead to a positive VRP, while in the other case, it may yield a negative risk premium. Investors who are exposed to variance risk may be willing to buy the variance swap, even if the risk premium is negative, since it can provide insurance against adverse variance risk scenarios. Indeed, investors who buy the variance swap receive a hedge for their exposure to variance risk and are willing to accept a negative expected profit in return.

5.2 Dispersion swaps

In Example 3.2, it is clear that there exists a correlation gap. However, variance swaps do not allow us to directly trade the difference between the real-world correlation and the implied correlation. Indeed, two prevalent approaches have been proposed for exploiting or hedging the correlation risk within the financial market. First, investors can implement a dispersion trading strategy, wherein they construct a portfolio by taking a position in an index option and the opposite position in the options written on the constituents of the index. Such dispersion strategies have been established using calls, puts, straddles, and variance swaps, see, e.g., Jacquier & Slaoui

(2007), Bossu ([2014](#)), and Meissner ([2015](#)). In the second approach, one directly trades in correlation derivatives such as correlation swaps or covariance swaps. Correlation swaps were first introduced in the early 2000s as a derivative to hedge against the correlation risk exposure inherent in exotic derivatives trading. However, unlike implied volatility, implied correlations are not directly observable from option markets, making the pricing and hedging of correlation swaps a challenging task. For literature on the pricing of the correlation swap, we refer to Bossu ([2005](#), [2007](#)). In addition to correlation swaps, covariance swaps are a generalization of the variance swap, providing an alternative method to hedge correlation risk within a two-asset framework. The floating leg of the covariance swap is the realized covariance between the log returns of two underlying assets. Discussions on the pricing and hedging strategies of covariance swaps can be explored in Carr & Corso ([2001](#)) and Salvi & Swishchuk ([2014](#)).

Instead of employing existing approaches to capitalize on the correlation risk between several assets, we propose a new derivative called the dispersion swap. The floating leg of the contract relies solely on the product of forward returns of the two underlying stocks, establishing an intrinsic link to the correlation between the assets, and simplifying the determination of the fixed leg. Theoretical analysis and numerical results will be presented to demonstrate that the dispersion swap provides an opportunity to trade the correlation gap. Moreover, we also show that the existence of a CRP in the financial market is justified.

The holder of the floating leg of a *dispersion swap* receives at maturity the following payoff:

$$RD[T] - P[T],$$

where

$$RD[T] = \frac{1}{T} \sum_{t=1}^{T} R_1(t) R_2(t). \tag{34}$$

We call $RD[T]$ the realized dispersion. $P[T]$ is the fixed leg of the contract, determined at inception such that the contract is fair and with zero initial payment. The following proposition is presented to show the realized dispersion can be approximated by the realized variance of stock index and the realized variances of individual stocks. A proof of this proposition can be found in Appendix [A.2](#).

Proposition 2. *The realized dispersion can be approximated by the difference between the realized variance of the stock market index and the sum of realized stock variances:*

$$RD[T] \approx \frac{1}{2} \left(RV[T] - RV_1[T] - RV_2[T] \right), \tag{35}$$

where $RV_i[T]$ is the realized variance for stock i and is given by:

$$RV_i[T] = \frac{1}{T} \sum_{t=1}^{T} R_i(t)^2, i = 1, 2.$$

From Proposition [2](#), the realized dispersion can be approximated by subtracting marginal variances from the basket variance. It implies that the realized dispersion is indeed linked to the dependence. Additionally, there is a more refined approximation approach based on Taylor series expansion, which can be found in Appendix [A.5](#). This approximation allows to determine the approximation error.

At time 0, the realized dispersion is unknown and random, while the strike is fixed and known. The fixed leg $P[T]$ is determined as follows:

$$P[T] = \mathbb{E}_{\mathbb{Q}} \left[RD[T] \right].$$

In our market setting, the fixed leg $P[T]$ can be expressed as follows:

$$P[T] = \frac{1}{4T} \sum_{t=1}^{T} \left[\rho_{\mathbb{Q}}(t)(u_1 - d_1)(u_2 - d_2) + (u_1 + d_1)(u_2 + d_2) \right]. \tag{36}$$

A proof of Expression (36) is given in Appendix A.3.

The expected profit for the holder of the floating leg of a dispersion swap is given by:

$$\text{Expected Profit} = \mathbb{E}_{\mathbb{P}} \left[RD[T] \right] - P[T].$$

Hence it is straightforward to show that the expected profit of the dispersion swap can be expressed as:

$$\text{Expected Profit} = \frac{1}{T} \sum_{t=1}^{T} \left(\sigma_{\mathbb{P}}[R_1(t)] \sigma_{\mathbb{P}}[R_2(t)] \rho_{\mathbb{P}}(t) - \sigma_{\mathbb{Q}}[R_1(t)] \sigma_{\mathbb{Q}}[R_2(t)] \rho_{\mathbb{Q}}(t) \right)$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \left(\mathbb{E}_{\mathbb{P}} [R_1(t)] \, \mathbb{E}_{\mathbb{P}} [R_2(t)] - \mathbb{E}_{\mathbb{Q}} [R_1(t)] \, \mathbb{E}_{\mathbb{Q}} [R_2(t)] \right). \tag{37}$$

Expression (37) clearly indicates that the buyer of the dispersion swap longs the real-world correlation and shorts the risk-neutral correlation. To be more specific, in Appendix A.4, it is proven that the expected profit of the dispersion swap can also be written as follows:

$$\text{Expected Profit} = \frac{(u_1 - d_1)(u_2 - d_2)}{T} \sum_{t=1}^{T} \left(\sqrt{p_{u \cdot}(t) p_{d \cdot}(t) p_{\cdot u}(t) p_{\cdot d}(t)} \rho_{\mathbb{P}}(t) - \frac{1}{4} \rho_{\mathbb{Q}}(t) \right)$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}} [R_1(t)] \, \mathbb{E}_{\mathbb{P}} [R_2(t)] - \frac{1}{4} \left(u_1 + d_1 \right) \left(u_2 + d_2 \right). \tag{38}$$

From Expressions (37) and (38), we can find that the expected profit of the dispersion swap is closely associated with the correlation gap. The *CRP* is also defined using the correlation gap. In our market model, the CRP is defined as follows:

$$CRP = \frac{1}{T} \sum_{t=1}^{T} (\rho_{\mathbb{P}}(t) - \rho_{\mathbb{Q}}(t)). \tag{39}$$

By assuming the real-world marginals are same as the risk-neutral marginals in this market setting, i.e., $p_{u \cdot}(t) = p_{\cdot u}(t) = \frac{1}{2}$, the expected profit of the dispersion swap given by (38) can be written as:

$$\text{Expected Profit Dispersion Swap} = \frac{(u_1 - d_1)(u_2 - d_2)}{4T} \sum_{t=1}^{T} \left(\rho_{\mathbb{P}}(t) - \rho_{\mathbb{Q}}(t) \right)$$

$$= \frac{(u_1 - d_1)(u_2 - d_2)}{4} CRP. \tag{40}$$

When the marginal distributions are the same under $\mathbb{P}$ and $\mathbb{Q}$, the expected profit of buying the dispersion swap is solely determined by the correlation gap. It is proportional to the CRP, which allows us to directly trade the weighted difference of the correlation gap $\rho_{\mathbb{P}}(t) - \rho_{\mathbb{Q}}(t)$. A larger positive correlation gap results in a higher expected profit, increasing the contract's appeal to potential investors.

Table 4. Variance risk premium (VRP) and correlation risk premium (CRP) in three different market situations

Market	$\rho_{\mathbb{Q}}$	$\sigma_{\mathbb{Q}}$	VRP	EPVS	CRP	EPDS
$\mathbb{Q}^{(1)}$	-0.9	0.373	0.224	0.224	1.1	0.404
$\mathbb{Q}^{(2)}$	0.15	0.578	0.010	0.010	0.05	0.060
$\mathbb{Q}^{(3)}$	0.8	0.668	-0.122	-0.122	-0.6	-0.220

5.3 Illustration: One-period example

In this subsection, we still consider the same market setting as in Example 3.2, and will demonstrate the potential use of dispersion swaps in exploiting the correlation gap. In this one-period market with $r = 0$, it follows directly from the Expression (33) that the expected profit of the variance swap (EPVS) is given by:

$$\text{EPVS} = VRP = \sigma_{\mathbb{P}}^2(1) - \sigma_{\mathbb{Q}}^2(1) + (\mathbb{E}_{\mathbb{P}}[R(1)])^2 - (\mathbb{E}_{\mathbb{Q}}[R(1)])^2. \tag{41}$$

Additionally, the expected profit of the dispersion swap (EPDS) given by Expression (40) can be expressed as follows:

$$\text{EPDS} = \frac{(u_1 - d_1)(u_2 - d_2)}{4} CRP = \frac{(u_1 - d_1)(u_2 - d_2)(\rho_{\mathbb{P}}(1) - \rho_{\mathbb{Q}}(1))}{4}. \tag{42}$$

We use Table 4 to compare the expected profits of the variance swap and the dispersion swap, the VRPs and the CRPs, concerning three different market situations in Example 3.2.

In Example 3.2, the first market situation featured a significant gap between the real-world and risk-neutral scenarios. Variance and dispersion swaps serve as tools to exploit this gap. As shown in Table 4, both the VRP and the CRP are positive, suggesting potential benefits from taking long positions in the variance swap and the dispersion swap. However, the gap between the real-world and the risk-neutral volatility and correlation is much smaller in the second market situation, which is reflected in the expected profit of the variance and the dispersion swap, as shown in Table 4. Expected profits turn out to be limited in case. Since these are only expectations, fluctuations can result in losses when stepping in a variance or dispersion trade, making these products less investment-worthy in this market situation than the first market situation. In the third market situation, both the volatility and correlation gaps are strongly negative, resulting in significantly negative variance and CRPs. Rational investors are willing to short the variance swap or the dispersion swap to gain a profit. But who is willing to long the dispersion swap or the variance swap? Market participants who are afraid that the correlation and volatility risk will have significant negative impact on their portfolios are willing to accept a negative expected profit to long the dispersion swap or the variance swap. Indeed, they are more afraid of the correlation and volatility risk than they are of the possible losses from the dispersion and volatility swap.

6. What are implied measures telling us?

We have shown that risk-neutral and real-world measures can be significantly different, see Example 3.2. Derivative prices reflect the aggregate view of the market about the future dynamics of the stocks involved. Example 3.3 shows that this view can be wrong: while stocks are becoming less dependent over time under the real-world measure, the risk-neutral dependence structure is evolving in the opposite direction; see Table 2 and Fig. 1. In Section 3.2, it seems irrational that the market is pricing a index option with the pricing measure $\mathbb{Q}^{(1)}$, since $\mathbb{Q}^{(1)}$ has a strong negative dependence structure while the real-world dependence is positive. Is this difference between real-world and risk-neutral dependence a dysfunctioning of the market? This seemingly irrational behavior does not lead to an arbitrage opportunity since the market is incomplete and the
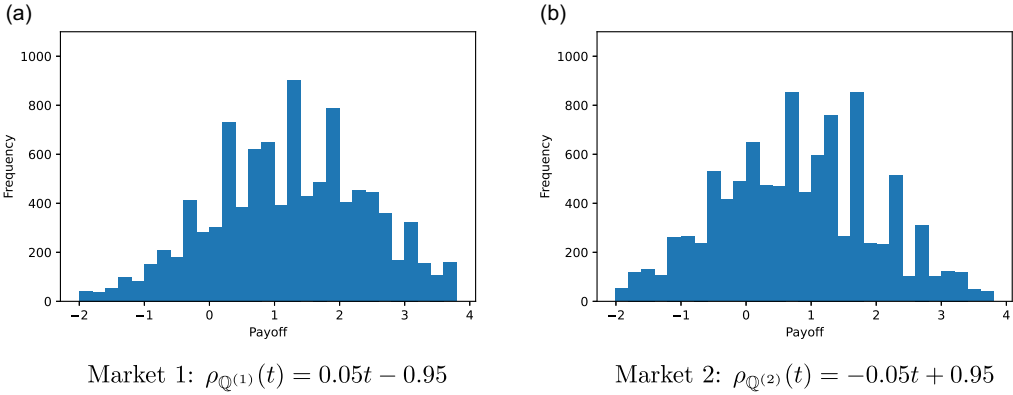
Figure 2. Comparison of payoff histograms for buying the dispersion swap in two market scenarios $\mathbb{Q}^{(1)}$ and $\mathbb{Q}^{(2)}$.

risk-neutral probability measure $\mathbb{Q}^{(1)}$ belongs to the set $\mathcal{M}$ of feasible risk-neutral measures; see Theorem 1. However, the discrepancy between real-world and risk-neutral correlation provides a trading opportunity. Indeed, the dispersion swap becomes more attractive if the correlation gap becomes larger; see Expression (37). A large correlation gap aligns with a large expected profit.

We showed that a positive correlation gap corresponds to a positive expected return for longing the dispersion swap, while a negative correlation gap yields a negative expected return. But a positive expected profit does not mean that there is no opportunity for the buyer of the dispersion swap to encounter a loss. Likewise, a negative correlation gap does not imply that investing in the dispersion swap cannot realize any profit at maturity. Consider the one-period setting in Example 3.2, the probability of having a loss for the holder of the floating leg of the dispersion swap is given by:

$$\mathbb{P}[\text{loss} > 0] = \mathbb{P}[RD[1] < P[1]]. \tag{43}$$

Expression (43) shows that if the fixed strike $P[1]$ exceeds the realized dispersion $RD[1]$, the buyer of the dispersion swap incurs a loss. In the one-period setting, realized dispersion $R_1(1)R_2(1)$ under the real-world measure $\mathbb{P}$ can take four possible values u_1u_2, u_1d_2, d_1u_2, or d_1d_2 with corresponding probabilities $p_{uu}(1)$, $p_{ud}(1)$, $p_{du}(1)$, and $p_{dd}(1)$. The fixed leg $P[1]$ is $\mathbb{E}_{\mathbb{Q}}[R_1(1)R_2(1)]$. Since all joint probabilities are strictly positive under $\mathbb{P}$, the probability of loss given by (43) is always positive and cannot reach 1.

Similar to the single period case, the probability of loss when taking the floating leg of the dispersion swap is always positive and less than 1 in the multiperiod market. To illustrate this statement, consider $T = 10$. Assume the real-world joint probabilities are given by (20), and $\rho_{\mathbb{P}}(t) \equiv 0.2$ for $t = 1, 2, \ldots, T$. Two different market situations are specified by implied correlations $\rho_{\mathbb{Q}^{(1)}}(t) = 0.05t - 0.95$ and $\rho_{\mathbb{Q}^{(2)}}(t) = -0.05t + 0.95$ for $t = 1, 2, \ldots, T$. The following Fig. 2 is presented to show the frequency of different payoffs for the buyer of the dispersion swap in these two different market situations. We can see that in the first market situation, the correlation gap $\rho_{\mathbb{P}}(t) - \rho_{\mathbb{Q}}(t)$ is always assumed to be positive for $t = 1, 2, \ldots, T$, resulting in a positive CRP. Conversely, under $\mathbb{Q}^{(2)}$, the correlation gap remains negative, yielding a negative CRP. Compare Fig. 2a with b, we can see that the probability of having a loss is larger in the second market situation, and it is because there exists a negative correlation gap in the second market situation. Nevertheless, Fig. 2 shows that the probabilities of taking a loss in the first and the second market situations are both larger than 0 and smaller than 1. Due to the possibility of experiencing a loss being nonzero, trading the dispersion swap is not an arbitrage strategy.

Throughout the paper, we assumed the probability measure $\mathbb{P}$ to be *known*. The knowledge of the real-world dynamics of the stocks enables us to conclude in the example of Section 3.2 that real-world and risk-neutral correlation differ substantially and that entering a dispersion swap results in a positive expected profit. In reality, the probability measure $\mathbb{P}$ is *unknown* and subjective: different market participants may have different opinions about the real-world probability measure. The risk-neutral probability measure $\mathbb{Q}$ chosen by the market can be extracted unambiguously if sufficient derivative instruments are available. Therefore, it is reasonable to assume that the measure $\mathbb{Q}$ is objective in that all market participants agree on the choice of $\mathbb{Q}$ used for pricing derivatives. Consider now the first market situation in the example described in Section 3.2. All market participants observe the same option prices and thus agree that $\rho_{\mathbb{Q}^{(1)}} = -0.9$. In Section 5.3, we then showed that an investor using $\rho_{\mathbb{P}} = 0.2$ may want to invest in the floating leg of a dispersion swap since this gives the investor a positive expected profit. Consider now another investor, who uses another real-world measure $\mathbb{P}'$ and believes that $\rho_{\mathbb{P}'} = -0.95$. This investor may want to take the fixed leg of a dispersion swap since this produces a positive expected profit. Both investors are observing the same option prices, but have different opinions about the real-world measure. They both observe a correlation gap and consider a dispersion swap to exploit this gap. They do agree on the value of the fixed leg of the dispersion swap since the risk-neutral measure is assumed to be objective. However, they have different opinions about the sign of the correlation gap. As a result, they both agree to enter in a trade, where each one takes a side of the same dispersion swap.

Note that if the real-world measure is also objective, all market participants would see the same correlation gap. For example, consider market situation 1 in the example described in Section 3.2. All market participants agree that $\rho_{\mathbb{P}} = 0.2$, so most investors want to enter the floating leg of a dispersion swap, i.e. there is a larger number of individuals willing to receive the realized dispersion and pay the fixed leg than there are market participants willing to take up the opposite leg of the dispersion swap. Therefore, this imbalance between supply and demand will push the implied correlation to increase toward the realized correlation.

However, will the correlation gap converge to 0 eventually? To answer this question, we shift to the market situation 3 in Section 3.2. In this market situation, the implied correlation $\rho_{\mathbb{Q}^{(3)}}(t)$ equals 0.8, and the real-world correlation is still 0.2. It follows directly that the correlation gap in this market is negative. Moreover, the probability of taking a loss for buying the dispersion swap is equal to 0.7. Therefore, many investors are consequently willing to short the dispersion swap. However, there are still investors willing to long the dispersion swap, since they dislike the correlation risk in the financial market and would like to offset their correlation exposure. Investing in the dispersion swap can offer a hedge against the correlation increases.

Ultimately, the market may attain a state of equilibrium with a negative CRP. To be more precise, if there are more investors selling the dispersion swap than buying the dispersion swap in this market situation, the negative correlation gap will gradually tighten. It will reach an equilibrium when the number of sellers in the dispersion swap contract aligns with the number of buyers. Indeed, when the CRP comes closer to zero, a short position in the dispersion swap still has a positive expected profit, but becomes less attractive due to an increase in the probability of a loss. At the same time, hedging correlation risk through the dispersion swap becomes more attractive. Therefore, with a tightening of the correlation gap, the number of people who would like to short the dispersion swap decreases whereas the number of people who are inclined to hedge the correlation risk increases. At some point, the number of sellers of the dispersion swap will match the number of buyers to reach a market equilibrium. At this equilibrium, the CRP can still be strictly negative. Indeed, in such a situation, market participants looking for hedging their correlation risk are willing to pay a premium to offset this risk, whereas the increased probability of a loss when shorting the dispersion swap limits the supply of correlation hedges through the dispersion swap. Consequently, this situation explains the negative CRP which is often observed in the market.

7. Conclusion

This paper showed that risk-neutral dependence can significantly diverge from real-world dependence when working in an incomplete two-dimensional market model. We characterized the set of equivalent martingale measures and illustrated the correlation gap between $\mathbb{P}$ and $\mathbb{Q}$. A unit-linked insurance product was considered to illustrate that this nonzero correlation gap also exerts influence on the expected excess return for policyholders. We also introduced the dispersion swap to exploit the correlation gap. Both theoretical and numerical results were provided to illustrate that the dispersion swap serves not only as a tool to capitalize on the large correlation gap, but also as a hedging strategy for the correlation risk in the financial market.

We also demonstrated that investors with strong fear of correlation risk are willing to take long positions in the dispersion swap, even when the expected profit is negative. As a consequence, the market will reach an equilibrium with a negative CRP.

Data availability statement. The Python code for reproducing our numerical results can be accessed at the following GitHub repository: https://github.com/BiwenLing/Understanding_CRP.

Competing interests. The authors have no competing interests to declare.

References

Alibeiki, **H.**, & **Lotfaliei**, **B.** (2021). To expand and to abandon: Real options under asset variance risk premium. *European Journal of Operational Research*, **300**(2), 771–787.

Artzner, **P.**, & **Eisele**, **K.** (2010). Supervisory accounting: Comparison between solvency 2 and coherent risk measures. In: *Actuarial and financial mathematics conference: Interplay between finance and insurance*.

Bacinello, **A. R.**, **Millossovich**, **P.**, **Olivieri**, **A.**, & **Pitacco**, **E.** (2011). Variable annuities: A unifying valuation approach. *Insurance: Mathematics and Economics*, **49**(3), 285–297.

Ballotta, **L.**, **Deelstra**, **G.**, & **Rayée**, **G.** (2017). Multivariate fx models with jumps: Triangles, quantos and implied correlation. *European Journal of Operational Research*, **260**(3), 1181–1199.

Barigou, **K.**, **Chen**, **Z.**, & **Dhaene**, **J.** (2019). Fair dynamic valuation of insurance liabilities: Merging actuarial judgement with market- and time-consistency. *Insurance: Mathematics and Economics*, **88**, 19–29.

Bauer, **D.**, **Kling**, **A.**, & **Russ**, **J.** (2008). A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin: The Journal of the IAA*, **38**(2), 621–651.

Bollerslev, **T.**, & **Zhou**, **H.** (2007). Expeceted stock returns and variance risk premia. Finance and Economics(FEDS) Discussion Paper No. W 2007-11.

Bossu, **S.** (2005). Arbitrage pricing of equity correlation swaps, Technical report, JPMorgan Securities Ltd. – London.

Bossu, **S.** (2007). A new approach for modelling and pricing correlation swaps, Technical report, Dresdner Kleinwort.

Bossu, **S.** (2014). *Advanced equity derivatives: Volatility and correlation*. Wiley.

Buss, **A.**, **Schönleber**, **L.**, & **Vilkov**, **G.** (2017). Expected stock returns and the correlation risk premium, Technical report.

Carr, **P.**, & **Corso**, **A.** (2001). Covariance contracting for commodities, Technical report.

Carr, **P.**, & **Wu**, **L.** (2009). Variance risk premiums. *The Review of Financial Studies*, **22**(3), 1311–1341.

Chicago Board Options Exchange (2022). CBOE S&P 500 implied correlation index, Technical report.

Coleman, **T. F.**, **Kim**, **Y.**, **Li**, **Y.**, & **Patron**, **M.** (2007). Robustly hedging variable annuities with guarantees under jumps and volatility risks. *The Journal of Risk and Insurance*, **74**(2), 347–376.

Dhaene, **J.**, & **Goovaerts**, **M.** (1996). Dependency of risks and stop-loss order. *Astin Bulletin*, **26**(2), 201–212.

Dhaene, **J.**, **Linders**, **D.**, **Schoutens**, **W.** & **Vyncke**, **D.** (2012). The herd behavior index: A new measure for the implied degree of co-movement in stock markets. *Insurance: Mathematics and Economics*, **50**(3), 357–370.

Dhaene, **J.**, **Stassen**, **B.**, **Barigou**, **K.**, & **Chen**, **Z.** (2017). Fair valuation of insurance liabilities: Merging actuarial judgement and market-consistency. *Insurance: Mathematics and Economics*, **76**, 14–27.

Driessen, **J.**, **Maenhout**, **P. J.**, & **Vilkov**, **G.** (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, **64**(3), 1377–1406.

Faria, **G.**, & **Kosowski**, **R.** (2014). The correlation risk premium: Term structure and hedging, Technical report.

Faria, G., **Kosowski**, R., & **Wang**, T. (2022). The correlation risk premium: International evidence. *Journal of Banking & Finance*, **136**, 1–14.

Feng, R., & **Jing**, X. (2017). Analytical valuation and hedging of variable annuity guaranteed lifetime withdrawal benefits. *Insurance: Mathematics and Economics*, **72**, 36–48.

Föllmer, H., & **Schied**, A. (2004). *Stochastic finance: An introduction in discrete time*. De Gruyter.

Garcia, J., **Goossens**, S., **Masol**, V., & **Schoutens**, W. (2009). Levy base correlation. *Wilmott Journal*, **1**(2), 95–100.

Ghalehjooghi, A. S., & **Pelsser**, A. (2020). Time-consistent and market-consistent actuarial valuation of the participating pension contract. *Scandinavian Actuarial Journal*, **2021**(4), 266–294.

Jacquier, A., & **Slaoui**, S. (2007). Variance dispersion and correlation swaps, Working paper.

Laurence, P. (2008). A new tool for correlation risk management: The market implied comonotonicity gap, Global Derivatives, Paris, Invited Talk, May 2008.

Laurence, P., & **Wang**, T.-H. (2008). Distribution-free upper bounds for spread options and market-implied antimonotonicity gap. *The European Journal of Finance*, **14**(8), 717–734.

Lehmann, E. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, **37**(5), 1137–1153.

Linders, D. (2023). The 3-step hedge-based valuation: Fair valuation in the presence of systematic risks. *ASTIN Bulletin: The Journal of the IAA*, **53**(2), 418–442.

Linders, D., **Dhaene**, J., & **Schoutens**, W. (2015). Option prices and model-free measurement of implied herd behavior in stock markets. *International Journal of Financial Engineering*, **2**(2), 1–35.

Linders, D., & **Schoutens**, W. (2014). A framework for robust measurement of implied correlation. *Journal of Computational and Applied Mathematics*, **271**, 39–52.

MacKay, A., **Vachon**, M.-C., & **Cui**, Z. (2023). Analysis of VIX-linked fee incentives in variable annuities via continuous-time Markov chain approximation. *Quantitative Finance*, **23**(7-8), 1055–1078.

Madan, D. B., & **Schoutens**, W. (2013). Systemic risk tradeoffs and option prices. *Insurance: Mathematics and Economics*, **52**(2), 222–230.

Malamud, S., **Trubowitz**, E., & **Wüthrich**, M. (2008). Market consistent pricing of insurance products. *ASTIN Bulletin*, **38**(2), 483–526.

Meissner, G. (2015). Correlation trading strategies: limitations and pitfalls, Technical report, University of Hawai System.

Pelsser, A., & **Stadje**, M. (2014). Time-consistent and market-consistent evaluations. *Mathematical Finance*, **24**(1), 25–65.

Salvi, G., & **Swishchuk**, A. V. (2014). Covariance and correlation swaps for financial markets with markov-modulated volatilities. *International Journal of Theoretical and Applied Finance*, **17**(1), 1450006.

Skintzi, V. D., & **Refenes**, A. N. (2004). Implied correlation index: A new measure of diversification. *Journal of Futures Markets*, **25**(2), 171–197. doi: 10.1002/fut.20137.

Yanagimoto, T., & **Okamoto**, M. (1969). Partial orderings of permutations and monotonicity of a rank correlation statistic. *Annals of the Institute of Statistical Mathematics*, **21**, 489–506.

A. Proofs and calculations

A.1 Proof of Theorem 1

We first take into account the martingale condition. For any t, we find that the risk-neutral marginal probabilities $\left(q_{d\cdot}(t), q_{u\cdot}(t), q_{\cdot u}(t), q_{\cdot d}(t)\right)$ satisfy:

$$\begin{cases} e^{-r} S_1(t-1)\left(e^{u_1} q_{u\cdot}(t) + e^{d_1} q_{d\cdot}(t)\right) = S_1(t-1), \\ e^{-r} S_2(t-1)\left(e^{u_2} q_{\cdot u}(t) + e^{d_2} q_{\cdot d}(t)\right) = S_2(t-1), \\ \qquad\qquad q_{d\cdot}(t) + q_{u\cdot}(t) = 1, \\ \qquad\qquad q_{\cdot d}(t) + q_{\cdot u}(t) = 1. \end{cases} \tag{44}$$

From (44), we find that the marginal probabilities are independent of the time, hence the risk-neutral marginal probabilities can be denoted as $(q_{d\cdot}, q_{u\cdot}, q_{\cdot d}, q_{\cdot u})$, and are given by:

$$q_{u\cdot} = q_{d\cdot} = q_{\cdot u} = q_{\cdot d} = \frac{1}{2}. \tag{45}$$

Notice we can write $q_{uu}(t) = \mathbb{Q}\left[R_2(t) = u_2 \mid R_1(t) = u_1\right] \times \mathbb{Q}\left[R_1(t) = u_1\right]$. We use $\alpha_{\mathbb{Q}}(t)$ to denote $\mathbb{Q}\left[R_2(t) = u_2 \mid R_1(t) = u_1\right]$, and taking into account Expressions (45), we find that $q_{uu}(t) = \frac{\alpha(t)}{2}$. Similarly, expressions can be derived for $q_{ud}(t)$, $q_{du}(t)$ and $q_{dd}(t)$. Hence the set $\mathcal{M}$ of feasible risk-neutral probability measure can be characterized as follows:

$$\mathbb{Q} \in \mathcal{M} \Leftrightarrow \exists\, \alpha(t) \in (0, 1)\,, t = 1, 2, \ldots, n,$$

$$\text{such that } \begin{cases} q_{uu}(t) = q_{dd}(t) = \frac{1}{2}\alpha(t) \\ q_{ud}(t) = q_{du}(t) = \frac{1}{2}(1 - \alpha(t)) \end{cases} \tag{46}$$

Since $\mathbb{Q}$ has to be equivalent to $\mathbb{P}$, we need that $p_{uu}(t), p_{ud}(t), p_{du}(t), p_{dd}(t) > 0$. Hence we find that $\alpha(t) \in (0, 1)\,.$

The risk-neutral covariance $\text{Cov}_{\mathbb{Q}}\,[R_1(t), R_2(t)]$ can be determined as follows:

$$\text{Cov}_{\mathbb{Q}}\,[R_1(t), R_2(t)] = \frac{2\alpha(t) - 1}{4}(u_1 - d_1)(u_2 - d_2),$$

We then find:

$$\rho_{\mathbb{Q}}(t) = \frac{\text{Cov}_{\mathbb{Q}}\,[R_1(t), R_2(t)]}{\sqrt{\text{Var}_{\mathbb{Q}}\,[R_1(t)]\,\text{Var}_{\mathbb{Q}}\,[R_2(t)]}} = 2\alpha(t) - 1.$$

Following (46), we can directly get the expression (5).

A.2 Proof of Proposition 2
First, the forward return of the stock market index at time t_i can be expressed as:

$$R(t) = \log \frac{S_1(t - 1)S_1(t)S_2(t - 1) + S_1(t - 1)S_2(t - 1)S_2(t)}{(S_1(t - 1) + S_2(t - 1))\,S_1(t - 1)S_2(t - 1)}$$

By replacing $S_1(t - 1)S_1(t)S_2(t - 1)$ by $S_1(t - 1)S_1(t)S_2(t)$ and replacing $S_1(t - 1)S_2(t - 1)S_2(t)$ by $S_1(t)S_2(t - 1)S_2(t)$, $R(t)$ can be approximated by:

$$R(t) \approx \log \frac{S_1(t - 1)S_1(t)S_2(t) + S_1(t)S_2(t - 1)S_2(t)}{(S_1(t - 1) + S_2(t - 1))\,S_1(t - 1)S_2(t - 1)} = \log \frac{S_1(t)S_2(t)}{S_1(t - 1)S_2(t - 1)}$$

It follows directly the realized dispersion can be approximated by (35).

A.3 Proof of expression (36)
It follows from (5) directly that the fixed leg $P[T]$ can be written as follows:

$$P[T] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\mathbb{Q}}\,[R_1(t)R_2(t)] = \frac{1}{4T}\sum_{t=1}^{T}\big(\rho_{\mathbb{Q}}(t)(u_1 - d_1)(u_2 - d_2) + (u_1 + d_1)(u_2 + d_2)\big)\,.$$

A.4 Proof of expression (37)
Note that $\mathbb{E}_{\mathbb{P}}\,[R_1(t)R_2(t)]$ can be expressed as:

$$\mathbb{E}_{\mathbb{P}}\,[R_1(t)R_2(t)] = \rho_{\mathbb{P}}(t) \times \sqrt{\text{Var}_{\mathbb{P}}\,[R_1(t)]\,\text{Var}_{\mathbb{P}}\,[R_2(t)]} + \mathbb{E}_{\mathbb{P}}\,[R_1(t)]\,\mathbb{E}_{\mathbb{P}}\,[R_2(t)]$$

$$= (u_1 - d_1)(u_2 - d_2)\sqrt{p_{u\cdot}(t)p_{d\cdot}(t)p_{\cdot u}(t)p_{\cdot d}(t)} \times \rho_{\mathbb{P}}(t) + \mathbb{E}_{\mathbb{P}}\,[R_1(t)]\,\mathbb{E}_{\mathbb{P}}\,[R_2(t)]\,.$$

Then we can derive the expected profit of the dispersion swap:

$$\text{Expected Profit} = \mathbb{E}_{\mathbb{P}} \left[\frac{1}{T} \sum_{t=1}^{T} R_1(t) R_2(t) \right] - P[T]$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left((u_1 - d_1)(u_2 - d_2) \left(\sqrt{p_{u\cdot}(t) p_{d\cdot}(t) p_{\cdot u}(t) p_{\cdot d}(t)} \rho_{\mathbb{P}}(t) - \frac{1}{4} \rho_{\mathbb{Q}}(t) \right) \right)$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}} \left[R_1(t) \right] \mathbb{E}_{\mathbb{P}} \left[R_2(t) \right] - \frac{1}{T} \sum_{t=1}^{T} \left(\frac{u_1 + d_1}{2} \right) \left(\frac{u_2 + d_2}{2} \right)$$

$$= \frac{(u_1 - d_1)(u_2 - d_2)}{T} \sum_{t=1}^{T} \left(\sqrt{p_{u\cdot}(t) p_{d\cdot}(t) p_{\cdot u}(t) p_{\cdot d}(t)} \rho_{\mathbb{P}}(t) - \frac{1}{4} \rho_{\mathbb{Q}}(t) \right)$$

$$+ \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}} \left[R_1(t) \right] \mathbb{E}_{\mathbb{P}} \left[R_2(t) \right] - \frac{1}{4} \left(u_1 + d_1 \right) \left(u_2 + d_2 \right).$$

A.5 Approximation of the realized dispersion: Taylor series expansion

Proposition 3. *The realized dispersion can be approximated by the difference between the realized variance of the stock market index and a weighted sum of realized stock variances:*

$$RD[T] = \frac{1}{T} \sum_{t=1}^{T} \frac{R(t)^2 - w_1(t-1)^2 R_1(t)^2 - w_2(t-1)^2 R_2(t)^2}{2 w_1(t-1) w_2(t-1)} + \epsilon, \qquad (47)$$

where

$$w_i(t-1) = \frac{S_i(t-1)}{S_1(t-1) + S_2(t-1)}, \quad i = 1, 2, \qquad (48)$$

and ϵ represents the approximation error given by:

$$\epsilon = \frac{1}{T} \sum_{t=1}^{T} \frac{\mathcal{O}\left(SR(t)^3\right) - w_1(t-1)^2 \mathcal{O}\left(SR_1(t)^3\right) - w_2(t-1)^2 \mathcal{O}\left(SR_2(t)^3\right)}{2 \omega_1(t-1) \omega_2(t-1)} \qquad (49)$$

$$- \mathcal{O}\left(SR_2(t)^2 SR_1(t)\right) - \mathcal{O}\left(SR_2(t) SR_1(t)^2\right),$$

and

$$SR(t) = \frac{S(t) - S(t-1)}{S(t-1)}$$

$$SR_i(t) = \frac{S_i(t) - S_i(t-1)}{S_i(t-1)}, \quad i = 1, 2.$$

Proof. First, the square of the index forward return at time t is given by:

$$R(t)^2 = \left(\log \frac{S(t)}{S(t-1)} \right)^2 = \left(\log \left(1 + \frac{S(t) - S(t-1)}{S(t-1)} \right) \right)^2.$$

Consider a function $f(x) = \left(\log \left(1 + x \right) \right)^2$, we use the first two terms of the Taylor series expansion to approximate $f(x)$ centered at $x = 0$ which gives $f(x) = x^2 + \mathcal{O}(x^3)$, where $\mathcal{O}(x^3)$ represents the

approximation error. Hence the approximation of $R(t)^2$ is given by:

$$R(t)^2 = \frac{(S_1(t) - S_1(t-1) + S_2(t) - S_2(t-1))^2}{(S_1(t-1) + S_2(t-1))^2} + \mathcal{O}\left(\left(\frac{S(t) - S(t-1)}{S(t-1)}\right)^3\right). \quad (50)$$

Similarly, we approximate the square of individual stock forward returns by the following expressions:

$$R_i(t)^2 = \frac{(S_i(t) - S_i(t-1))^2}{S_i(t-1)^2} + \mathcal{O}\left(\left(\frac{S_i(t) - S_i(t-1)}{S_i(t-1)}\right)^3\right), \quad i = 1, 2. \quad (51)$$

Meanwhile, the forward return of stock 1 at time t is given by:

$$R_1(t) = \log \frac{S_1(t)}{S_1(t-1)} = \log\left(1 + \frac{S_1(t) - S_1(t-1)}{S_1(t-1)}\right).$$

Consider another function $g(x) = \log(1 + x)$, we approximate $g(x)$ by using the first term of the Taylor series expansion of $g(x)$ centered at $x = 0$ to obtain $g(x) = x + \mathcal{O}(x^2)$. Then $R_1(t)R_2(t)$ can be approximated by:

$$R_1(t)R_2(t) = \prod_{i=1}^{2}\left(\frac{S_i(t) - S_i(t-1)}{S_i(t-1)} + \mathcal{O}\left(\left(\frac{S_i(t) - S_i(t-1)}{S_i(t-1)}\right)^2\right)\right). \quad (52)$$

From Expressions (50), (51), and (52), we find that the realized dispersion can be approximated by Expression (47). We can find that $w_1(t-1)$ and $w_2(t-1)$ are the relative importance of stock 1 and stock 2 in the market index $S(t-1)$, respectively. $\square$