Assortative matching and risk sharing

Citation for published version:
Li, S, Sun, H, Wang, T & Yu, J 2016, 'Assortative matching and risk sharing', *Journal of Economic Theory*, vol. 163, pp. 248-275. https://doi.org/10.1016/j.jet.2016.01.008

Digital Object Identifier (DOI):
10.1016/j.jet.2016.01.008

Assortative Matching and Risk Sharing*

Sanxi Li[†], Hailin Sun[‡], Tong Wang[§], Jun Yu[¶]

September 16, 2015

Abstract

This paper explores the sorting patterns in a two-sided matching market where agents facing different risks match to share them. When preference belongs to the class of harmonic absolute risk aversion (HARA), the risk premium is perfectly transferable within each partnership; thus a stable match minimizes the social cost of risk. In the systematic risk model, where agents are ranked by their holdings of a common risky asset, the convexity of the joint risk premium in joint risk size leads to negative assortative matching (NAM). In the idiosyncratic risk model, where agents are ranked by their independent riskiness in the sense of second-order stochastic dominance (SSD), NAM arises when preference exhibits decreasing absolute risk aversion (DARA) in the sense of Ross and riskier background risk leads to more risk-averse behavior. However, it may fail to arise when riskier background risk leads to more risk-tolerant behavior.

Keywords Assortative matching, efficient risk sharing, transferable utility, systematic risk, idiosyncratic risk, background risk, risk vulnerability

JEL Classification C78, D31, D81, J12

*We are grateful to Christian Gollier who introduced us to this topic and gave us enlightening guidance. We also thank the editor Marciano Siniscalchi and two anonymous referees for their helpful comments and constructive advices.

†School of Economics, Renmin University of China, Beijing 100872, China. Email: sanxi@ruc.edu.cn

‡Institute of Finance and Banking, the People's Bank of China, 32 Chengfang Street, 100080, Beijing, China; Email: hailinsun@gmail.com

§Business School, University of Edinburgh, EH8 9JS, Scotland, UK. Email: tong.wang@ed.ac.uk

¶School of Economics, Shanghai University of Finance and Economics; Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, 777 Guoding Road, Shanghai 200433, China. Email: yu.jun@mail.shufe.edu.cn

1 Introduction

When insurance and financial markets are incomplete, individuals often form partnerships to diversify their risks. For instance, families - mainly in developing countries - often arrange for long-distance marriages for the purpose of sharing production shocks, manufacturing employers often cushion temporary shocks on profit by sharing with their workers, and different parties in related businesses sometimes develop joint ventures to share resources and revenues for mutual benefit (Townsend, 1994; Fafchamps and Lund, 2003; Bigsten et al, 2003). When risk sharing is a primary concern in forming partnerships, it is legitimate to ask how the agents should match to insure against risks. Do the evidences in the marriage market or the financial market reflect the mitigation of an incomplete insurance market, or are they boosted by other concerns at the cost of efficiency in risk sharing?

In this paper, we examine the sorting patterns in a two-sided matching market where agents facing different risks match to share them. It is known that when agents have different degrees of risk aversion, negative assortative matching (NAM) arises because risk bearings are generally substitutes: a very risk-averse female is a demanding buyer for insurance and a very risk-tolerant male is a ready seller for it (Chiappori and Reny, 2006; Schulhofer-Wohl, 2006). Rather than employing different degrees of risk aversion, our paper focuses on different risks that each agent faces. Since the Pareto frontier in a given match does not have constant slope, standard type-complementarity conditions (Becker, 1973) cannot be used in general. However, with respect to risk-sharing problems, it is known that when preference belongs to the class of harmonic absolute risk aversion (HARA), the Pareto frontier in the monetary-equivalent space is a straight line, or, in other words, the total surplus summarized by the certainty equivalent is independent of how risk sharing is performed. In this case, the matching game permits a transferable expected utility representation and the type-complementarity condition translates into minimizing social risk premium.

We then consider two applications: one where risks are perfectly correlated and one where risks are independent. In the systematic risk model, agents are ranked by their percentage of ownership of a common risky asset. Because joint risk premium is a convex function of the joint size of the common risk, it is extremely costly to pair two highly risky agents together. Hence, negative sorting is socially preferable and stable. One may wonder to what extent the result of negative sorting depends on the HARA assumption. As a robustness check, we show that, with general utility functions, NAM still arises if the supports of all risks are not too large compared with agents' risk-free incomes and/or if risk tolerance is sufficiently linear.

In the idiosyncratic risk model, agents are ranked by their independent riskiness in the sense of second-order stochastic dominance (SSD). NAM arises if the preference exhibits DARA and if riskier background risk leads to more risk-averse behavior, but may fail to arise when riskier background risk leads to more risk-tolerant behavior. There are four key points to note here. First, the conditions for NAM have clear economic implications and are supported by empirical evidence. Guiso et al. (1996) concluded from Italian survey data that a consumer's perception of a riskier distribution of uninsurable human-capital wealth is negatively related to the proportion of risky assets held in his/her investment portfolio. Second, the seemingly strong conditions for NAM to arise come from the fact that we are looking for the equilibrium sorting patterns for *any* SSD-ordered risks. For a special case of the SSD order where risks are ranked in the sense of SSD by taking the form of adding independent noise, we only need HARA and DARA to guarantee NAM. Third, when risks are large with respect to agents' risk-free incomes, an SSD deterioration in the background risk may lead to more risk-tolerant behavior, and thus, NAM may fail to arise in equilibrium. Fourth, the different results in the two applications suggest that one should investigate carefully whether agents are sharing highly correlated risks or independent risks.

The results of this paper may help us to understand the composition of risk-sharing groups in developing countries. Ghatak (1999) argued that PAM should arise because similar people will find it easier to monitor and enforce informal contracts. Empirical evidence, however, is mixed: on one hand, Bacon et al. (2014) found evidence that individuals were more likely to positive assortative mate on their risk attitude; and Arcand and Fafchamps (2012) also found solid evidence of positive sorting for peers with respect to physical or ethnic proximity as well as wealth or household size. On the other hand, Dercon et al. (2006) found little evidence of positive sorting in group-based funeral insurance. Our results from the idiosyncratic model suggest that the risk-sharing effect might drive matching to be negative assortative and, therefore, offset the monitoring and enforcing effects; however, when risks are large compared with individuals' risk-free incomes, it is possible that the two effects might work in the same direction and drive matching to be positive assortative.

Our work contributes to the recent literature on the risk-sharing matching game. Since the efficient risk sharing rule is typically nonlinear, the risk-sharing matching game permits non-transferable utilities, and thus, standard type-complementarity conditions cannot be used. Legros and Newman (2007) noticed that the risk-sharing matching game admits a transferable utility representation when agents have logarithmic or exponential utility functions. Schulhofer-Wohl (2006) generalized their findings, showing that the game admits a transferable utility representation when preferences are in the harmonic absolute risk aversion class with identical shape (ISHARA). Both Legros and Newman (2007) and Schulhofer-Wohl (2006) proved that the equilibrium sorting pattern is negative assortative on risk preferences. Chiappori and Reny (2006) further showed that negative sorting over risk preferences is robust under general utility functions. The key difference between our work and the existing literature is that our paper focuses on *different risks that each agent faces* rather than different degrees of risk aversion. Among the papers on risk-sharing matching games, ours is one of the first to investigate sorting over agents' risk exposure.[1]

There are two reasons we think examining riskiness is important. First, individual risk preferences have not proved to be stable across different stimulus domains and situations. For example, the predictive power of investors' risk taking heavily depends on whether their risk attitudes are elicited in an

[1] In their recent paper, Jaramillo, Kempf and Moizeau (2013) studied the formation of risk-sharing coalitions where individuals differ with respect to their risky exposure. The insurance scheme considered in their work is limited to equal sharing regardless of agents' initial incomes, while in our model, there is no barrier to efficient insurance within the household.

investment-related context (Slovic, 1964; MacCrimmon and Wehrung, 1986, 1990; Schoemaker, 1990, 1993; Weber and Milliman, 1997). Second, because income riskiness presumably is easier to observe than attitudes toward risk, one might expect to drive testable predictions concerning the role of risk-sharing in the formation of partnerships much more easier if agents are ranked on the basis of riskiness.

Moreover, the results of our paper differ from those in the literature. Chiappori and Reny (2006) rigorously proved that NAM arises if agents hold the same exogenous risky assets but differ in their risk attitude. Following their results, Li, Sun and Chen (2013) showed that PAM may arise if agents' incomes are endogenous (also see Wang, 2013a). Wang (2013b) showed that the presence of moral hazard may also lead to PAM. Our results show that without any other confounding factors such as endogenous income or moral hazard, the counter-intuitive PAM may arise if agents differ in their idiosyncratic risks instead of their risk preferences: while agents with highly risky assets always try to avoid matching with other large, perfectly-correlated risks, they might prefer to match with other large, independent risks.

The rest of this paper is organized as follows. Section 2 presents the risk-sharing matching game. Section 3 applies a monotonic transformation to this game and characterizes the stable match. Section 4 and 5 consider two applications, one where risks are perfectly correlated and the other where risks are independent. Section 6 extends the model to allow individuals to have different incomes and face different risks. Section 7 concludes the paper.

2 Risk-Sharing Matching Game

Consider a one-to-one matching market with two lines of agents. We denote them as N males $\{i = 1, ..., N\}$ and N females $\{j = 1, ..., N\}$. Each agent is endowed with an exogenous risky income, denoted by $\tilde{w}_i$ for male i and $\tilde{w}_j$ for female j. All agents are expected-utility maximizers with respect to the homogeneous probabilistic belief, and identically risk-averse with vNM utility function $u(c)$, which is bounded and continuously differentiable in consumption c, with $u'(c) > 0$ and $u''(c) < 0$.

Agents match in order to share risks. At period 0, each agent voluntarily matches with a mate from the opposite side. Each partnership (i, j) will commit to rules for sharing their joint income, which depends on the state of the world. At period 1, the value of all shocks are realized, and agents consume according to the prior sharing rules. We rule out any search or coordination frictions, and there is no limited commitment or asymmetric information. Denote $\tilde{z}_{ij} \equiv \tilde{w}_i + \tilde{w}_j$ as the ***joint income*** received by the matched pair (i, j). Division $(z_{ij} - c_{ij}, c_{ij})$, which is associated with partnership (i, j) prior to the realization of shocks, specifies individual consumptions to i and j under each realization of $\tilde{z}_{ij}$. Under this agreement, i's expected utility is $Eu(\tilde{z}_{ij} - c_{ij}(\tilde{z}_{ij}))$ and j's is $Eu(c_{ij}(\tilde{z}_{ij}))$.

Assume that risk is shared within each partnership in a Pareto-efficient way, a situation in which no agent's expected utility can be strictly increased without decreasing his/her partner's. A ***risk-sharing rule*** $c_{ij}(\cdot)$ is a deterministic function that maps each realized value of $\tilde{z}_{ij}$ to an individual consumption level for j. Given the random joint income $\tilde{z}_{ij}$ associated with partnership (i, j), a risk-sharing rule $c_{ij}(\cdot)$ is Pareto optimal if and only if there exists a scalar $\lambda \in \mathbb{R}_{++}$ such that $c_{ij}(\cdot)$ solves the following maximization problem:

$$\max_{\{c_{ij}(\cdot)\}} \{Eu(\tilde{z}_{ij} - c_{ij}(\tilde{z}_{ij})) + \lambda Eu(c_{ij}(\tilde{z}_{ij}))\} \quad (1)$$

The set of Pareto optimal risk-sharing rules is called Pareto efficient frontier.

Definition 1 *A* ***matching correspondence*** *is an assignment of males to females. A* ***stable match*** *specifies a matching correspondence and the associated risk-sharing rules for each partnership, which is immune to coalitional deviations. That is, there does not exist a risk-sharing rule under which a male and a female, who are not matched to one another, prefer each other to their current assignments.*

Assume that incomes strictly differ within each side of the population; further assume that the marginal utility of consumption is bounded at autarky. The existence of stable matches has been established by Legros and Newman (2007). Then there is a one-to-one matching of i to j. Under a positive/negative assortative matching (PAM/NAM), the most risky male is matched with the most/least risky female, the second-most risky male is matched with the second-most/least risky female, and so on. The formal definition of the equilibrium matching pattern is stated as follows:

Definition 2 *A stable match is* ***positive (negative) assortative*** *if and only if for any i, i', j and j', such that i and i' are matched with j and j' respectively, we have*

$$i' \geq i \Longleftrightarrow j' \geq (\leq) j.$$

3 Stable Match and Social Risk Premium

Becker's (1973) seminal paper provided a foundation for analyzing the competitive assignments of partners with transferable utility. But in our risk-sharing matching game, the Pareto efficient frontier in the utility space within a given partnership does not necessarily have a constant slope, and thus standard type-complementarity conditions cannot OKbe used in general. However, a simpler case arises when it is possible to apply a monotonic transformation to the expected utility levels such that the transformed Pareto efficient frontier has a constant slope. In this case, the matching game permits a transferable expected utility representation as defined below:

Definition 3 *The risk-sharing matching game has a transferable expected utility representation if for any random joint income $\tilde{z}_{ij}$, there exists a constant C_{ij} such that $u^{-1}[Eu(c_{ij}(\tilde{z}_{ij}))] + u^{-1}[Eu(\tilde{z}_{ij} - c_{ij}(\tilde{z}_{ij}))] = C_{ij}$ for all Pareto optimal risk sharing rules $c_{ij}(\cdot)$.*

A central implication of the above definition is that, if the risk-sharing matching game has a transferable expected utility representation, the joint output, C_{ij}, in terms of the certainty equivalent, depends only on the characteristics of the members' joint income distribution $\tilde{z}_{ij}$. This output measure allows agents to compare gains from potential partnerships they may acquire. In other words, C_{ij} can be treated as the joint monetary output associated with partnership (i, j). Similar to Becker (1973), the condition for a stable match is to maximize the social output $\sum_{i,j} C_{ij}$, which is the sum of the outputs over all partnerships for a given matching correspondence. Denote the ***Joint Risk Premium*** as $\pi_{ij} = E\tilde{z}_{ij} - C_{ij}$, and the associated ***Social Risk Premium*** as $\sum_{i,j} \pi_{ij}$, that is, the sum of the joint risk premium over all partnerships for a given matching correspondence. Then, the maximization of the social output will be equivalent to the minimization the social risk premium.

The existence of transferable expected utility representation is subject to certain regularity conditions. With respect to risk-sharing problems, it is known that when preference belongs to the HARA class, the Pareto frontier in the monetary-equivalent space is a straight line, or, in other words, the total surplus summarized by the certainty equivalent is independent of the way risk sharing is performed (Schulhofer-Wohl, 2006).

Definition 4 *Preference belongs to the* ***HARA*** *class if and only if absolute risk tolerance is a linear function of consumption:*

$$T(c) = \frac{1}{\gamma}c + \frac{1}{\alpha} \tag{2}$$

where risk tolerance $T(c) = -u'(c)/u''(c) > 0$ is the reciprocal of the Arrow-Pratt measure of absolute risk aversion.

In particular, preference exhibits ***decreasing/increasing absolute risk aversion (DARA/IARA)*** *if $\gamma > 0$ ($\gamma < 0$), it exhibits* ***constant absolute risk aversion (CARA)*** *if $\gamma \to \infty$, it exhibits* ***constant relative risk aversion (CRRA)*** *if $\alpha \to \infty$, and it exhibits* ***risk neutral*** *if $\gamma \to 0$.*

The results for HARA preference can be stated as follows:

Lemma 1 *If the preference belongs to the HARA class, then the risk-sharing matching game has a transferable expected utility representation .*[2]

Proof. The proof can be found in Mazzocco's (2004) and Schulhofer-Wohl's (2006) studies. ■

Because all agents have identical utility function, the solution to (1) when $\lambda = 1$ is $c_{ij}(z_{ij}) = \frac{z_{ij}}{2}$. That is, sharing the joint income equally is one of the Pareto optimal risk-sharing rules. According to

[2] Schulhofer-Wohl (2006) showed that the risk-sharing matching games admit a transferable expected utility representation if and only if preferences are in the class of identical shape harmonic absolute risk aversion (ISHARA). In other words, agents can have different utility functions, but the slope of their risk tolerance must be the same: $T_i(c) = \frac{1}{\gamma}c + \frac{1}{\alpha_i}$. This is equivalent to saying that all agents have the same HARA utility function, but with different initial wealth. We discuss the case in which agents differ in both initial wealth and riskiness of their assets in Section 6.

Definition 3, if the risk-sharing matching game has a transferable expected utility representation, the joint output C_{ij}, which is associated with partnership (i, j), does not depend on the risk-sharing rules. Thus, without loss of generality, we can derive C_{ij} by applying the particular risk-sharing rule $c_{ij}(z_{ij}) = \frac{z_{ij}}{2}$, which gives that $C_{ij} = 2u^{-1}Eu\left(\frac{\tilde{z}_{ij}}{2}\right)$. If we define a new utility function $v(c) \equiv u\left(\frac{c}{2}\right)$, then it follows that $C_{ij} = v^{-1}[Ev(\tilde{z}_{ij})]$. Hence, $v(\cdot)$ can be interpreted as the utility function of a representative agent for any matched pair (i, j). The joint output C_{ij} and the joint risk premium $\pi_{ij} = E\tilde{z}_{ij} - C_{ij}$ are simply the certainty equivalent and risk premium of the representative agent, respectively. Finally, using the definition of $v(\cdot)$ and condition (2), one can quickly confirm that preference $u(\cdot)$ belongs to the HARA class if and only if $v(\cdot)$ also belongs to the same class.

The results for the HARA preferences immediately follow:

Lemma 2 *If the preference belongs to the HARA class and if the joint risk premium π_{ij} is sub(super)modular in (i, j), then any stable matching of the risk-sharing matching game will be positive (negative) assortative on the partners' income riskiness.*

Proof. See Appendix. ■

4 Sorting over Systematic Risk

In this section, we consider the application in which risks are perfectly correlated. Agents are ranked by their holdings of a common risky asset. That is, male i's income is $\tilde{w}_i = w_0 + k_i\tilde{x}$ and female j's income is $\tilde{w}_j = w_0 + k_j\tilde{x}$, with $k_i < k_{i+1}$ and $k_j < k_{j+1}$. Define $k_{ij} \equiv k_i + k_j$. With $\pi_{ij} = E\tilde{z}_{ij} - v^{-1}[Ev(\tilde{z}_{ij})]$ and $\tilde{z}_{ij} = 2w_0 + k_{ij}\tilde{x}$, we have π_{ij} as a function of k_{ij} and w_0: $\pi_{ij} = \pi(k_{ij}, w_0)$. As a result of market competition, stable match guarantees the minimization of the social cost of risk. According to Lemma 2, stable match will be positive (negative) assortative on risk sizes (k_i, k_j) if

$$\frac{\partial^2 \pi(k_{ij}, w_0)}{\partial k_{ij}^2} \leq 0 (\geq 0),$$

i.e., the joint risk premium is concave (convex) in the size of the joint risk exposure. As suggested by Eeckhoudt and Gollier (2001), multiplicative risk is self-aggravating in the sense that the cost curve of risk $\pi(k_{ij}, w_0)$ is convex in the unit holdings of such risk k_{ij}. Here, if the joint risk premium is a convex function of the joint size of the common risk, it is extremely costly to pair two highly risky agents together, and thus NAM is socially preferable and stable. Thus, we have the following proposition:

Proposition 1 *If the preference belongs to the HARA class, then the joint risk premium is convex in the size of joint risk exposure and, therefore, the stable match of the risk-sharing matching game is negative assortative over the riskiness of the agents' income.*

Proof. Because $\pi_{ij} = E\tilde{z}_{ij} - v^{-1}[Ev(2w_0 + k_{ij}\tilde{x})]$, we have

$$\frac{\partial^2 \pi_{ij}}{\partial k_{ij}^2} = -\frac{v'(v^{-1}(Ev(2w_0 + k_{ij}\tilde{x})))E\left(v''(2w_0 + k_{ij}\tilde{x})\tilde{x}^2\right) - (Ev'(2w_0 + k_{ij}\tilde{x}))^2 \frac{v''(v^{-1}(Ev(2w_0+k_{ij}\tilde{x})))}{v'(v^{-1}(Ev(2w_0+k_{ij}\tilde{x})))}}{[v'(v^{-1}(Ev(2w_0 + k_{ij}\tilde{x})))]^2}$$

Therefore, $\frac{\partial^2 \pi_{ij}}{\partial k_{ij}^2} \geq 0$ iff

$$-\frac{v''(v^{-1}(Ev(2w_0 + k_{ij}\tilde{x})))}{[v'(v^{-1}(Ev(2w_0 + k_{ij}\tilde{x})))]^2} \leq -\frac{E(v''(2w_0 + k_{ij}\tilde{x})\tilde{x}^2)}{[E(v'(2w_0 + k_{ij}\tilde{x})\tilde{x})]^2} \tag{3}$$

Solving for $v(c)$ from (2) and substituting into (3), we find that there will be NAM iff

$$[E(T(2w_0 + k_{ij}\tilde{x})^{-\gamma}\tilde{x})]^2 \leq E(T(2w_0 + k_{ij}\tilde{x})^{-(1+\gamma)}\tilde{x}^2)ET(2w_0 + k_{ij}\tilde{x})^{1-\gamma}$$

which holds as a direct application of the Cauchy-Schwarz inequality. ■

One may wonder to what extent the result of negative sorting depends on the HARA assumption. Notice that the result of Proposition 1 immediately follows from the fact that the risk premium is convex in the risk size. Without the HARA assumption, Eeckhoubt and Gollier (2001) show that the risk premium may not be convex in the size of risk and thus, Proposition 1 may fail. However, as a robustness check, we are able to show that with general utility functions, NAM still arises if the supports of all risks are not too large compared with the agents' risk-free incomes and/or if the risk tolerance is sufficiently linear.[3]

In their paper, Chiappori and Reny (2006) show that competitive forces will lead risk-sharing groups to be composed of individuals who are rather different in their risk preferences. Here, consistent with their result, we show that it will lead risk-sharing groups to consist of agents with rather different risk sizes. There are two reasons we believe riskiness is an important factor to explore. First, in practice, individual risk preferences have not proven to be stable across different stimulus domains and situations. This creates a difficulty in assessing agents' risk attitudes because different methods and procedures often result in different classifications. Second, because risk sizes are much easier to track down, one may expect to drive testable predictions much more easily.

In the next section, which concerns the two factors that determine agents' risk-taking behavior, i.e., agents' risk preferences and risk exposures, we will show that there is a fundamental difference in their effects: while a highly risk-averse agent always prefers to match with a less risk-averse agent for better insurance, an agent with a very risky asset may prefer to match with another agent with a very risky asset for the purpose of risk sharing.

[3] With general preferences, the utility (under any monotone transformation) is not fully transferable between partners. In this case, Legros and Newman (2007) presented sufficient conditions for monotone matching. Applying their "generalized difference conditions", we are able to show that NAM still arises under fairly reasonable assumptions.The proof can be found in the Appendix.

5 Sorting over Idiosyncratic Risks

In this section, we consider the application when risks are idiosyncratic. Agents are ranked by their independent riskiness in the sense of second-order stochastic dominance (SSD). That is, male i's income is $w_i^m = w_0 + \tilde{\varepsilon}_i^m$ and female j's income is $w_j^f = w_0 + \tilde{\varepsilon}_j^f$, where $\tilde{\varepsilon}_{i+1}^m \overset{SSD}{\precsim} \tilde{\varepsilon}_i^m$ and $\tilde{\varepsilon}_{j+1}^f \overset{SSD}{\precsim} \tilde{\varepsilon}_j^f$. Again, the joint risk premium is given by $\pi_{ij} = E\tilde{z}_{ij} - v^{-1}[Ev(\tilde{z}_{ij})]$ with $\tilde{z}_{ij} = 2w_0 + \tilde{\varepsilon}_i^m + \tilde{\varepsilon}_j^f$. Thus, in this case, we have π_{ij} as a function of the joint risk $\tilde{\varepsilon}_i^m + \tilde{\varepsilon}_j^f$ and the initial wage w_0: $\pi_{ij} = \pi\left(\tilde{\varepsilon}_i^m + \tilde{\varepsilon}_j^f, w_0\right)$. Before proceeding, we show through examples that sorting in either direction is possible without further restrictions other than HARA preference.

5.1 Preliminary Examples

Example 1 *(CARA utility). Suppose there are two males m_1 and m_2 endowed with $w_0 + \tilde{\varepsilon}_i^m$, $i = 1, 2$ and two females f_1 and f_2 endowed with $w_0 + \tilde{\varepsilon}_j^f$, $j = 1, 2$, where $\tilde{\varepsilon}_i^m$ and $\tilde{\varepsilon}_j^f$ are independent. Because all agents have identical CARA utilities, given the initial wage level w_0, π_{ij} is additive over $(\tilde{\varepsilon}_i^m, \tilde{\varepsilon}_j^f)$: $\pi\left(\tilde{\varepsilon}_i^m + \tilde{\varepsilon}_j^f, w_0\right) = \pi\left(\tilde{\varepsilon}_i^m, w_0\right) + \pi\left(\tilde{\varepsilon}_j^f, w_0\right)$. Therefore, π_{ij} is both (but not strictly) supermodular and submodular in (i, j), which leads to arbitrary matching.*

Example 2 *(IARA utility). Suppose instead that all agents have quadratic utility $u(c) = c - \frac{c^2}{2}$. For the sake of simplicity, we assume that m_1 and f_1 are endowed with certain income w_0, and that m_2 and f_2 are endowed with risky incomes $w_0 + \tilde{\varepsilon}^m$ and $w_0 + \tilde{\varepsilon}^f$, respectively, where $\tilde{\varepsilon}^m$ and $\tilde{\varepsilon}^f$ are independently distributed with zero mean and variance σ_m^2 and σ_f^2, respectively. Notice that for quadratic utility, the mean-variance approach is exact. We have joint risk premium $\pi_{ij} = w_0 - (1 - \sqrt{(1 - w_0)^2 + Var(\tilde{z}_{ij})})$. We can easily show that, for any given initial wage w_0, π_{ij} is concave in $Var(\tilde{z}_{ij})$, which implies that PAM is stable.*

Example 3 *(DARA utility). Suppose alternatively that all agents have logarithm utility function $u(c) = \ln c$. m_1 and f_1 are endowed with certain income $w_0 = 3$, and m_2 and f_2 are endowed with risky income $3 + \tilde{\varepsilon}^m$ and $3 + \tilde{\varepsilon}^f$, respectively, where $\tilde{\varepsilon}^m$ and $\tilde{\varepsilon}^f$ are i.i.d., and $\Pr(\tilde{\varepsilon}^m = 1) = \Pr(\tilde{\varepsilon}^m = -1) = \frac{1}{2}$. A simple calculation gives $\pi_{21} = \pi_{12} = 0.17$ and $\pi_{22} = 0.41$. Therefore, we have $\pi_{11} + \pi_{22} > \pi_{12} + \pi_{21}$, and thus, NAM is stable.*

The key insights from these examples are that all sorting patterns are possible with idiosyncratic risks and that DARA may be necessary for NAM to arise.

5.2 General Results

In order to find general results, we can simply look at the 2×2 case in which two males are matched with two females[4]. Via Lemma 2, NAM is stable if the following supermodular condition holds:

$$\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m, w_0) \geq \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f, w_0) \tag{4}$$

Because $\tilde{\varepsilon}_2^m \overset{SSD}{\precsim} \tilde{\varepsilon}_1^m$, $\tilde{\varepsilon}_2^f \overset{SSD}{\precsim} \tilde{\varepsilon}_1^f$ and $\tilde{\varepsilon}_i^m$s, $\tilde{\varepsilon}_i^f$s are independent, we have the orders: $\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m \overset{SSD}{\precsim} \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \overset{SSD}{\precsim} \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f$ and $\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m \overset{SSD}{\precsim} \tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \overset{SSD}{\precsim} \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f$. Thus, condition (4) holds if the joint risk premium is "convex" in the riskiness of idiosyncratic risks.[5] The first lemma in this section is given as follows.

Lemma 3 *If the preference belongs to the HARA class, then the joint risk premium $\pi(\tilde{x}, w_0)$ is convex in w_0.*

Proof. Similar to the proof of Proposition 1. ■

One implication of Lemma 3 is that if the agents all face the same risks but differ in their initial wealth levels, NAM will arise.

Before proceeding to the general conditions for negative or positive sorting, we introduce the following definition:

Definition 5 *A utility function u_1 is more risk-averse than another utility function u_2 **in the sense of Ross** if there exist a positive constant λ and a differentiable function g with $g' \leq 0$ and $g'' \leq 0$ such that*

$$u_1 = \lambda u_2 + g.$$

Risk aversion in the sense of Ross is a stronger concept than risk aversion in the sense of Pratt. It is easy to verify that $\frac{-u_1''}{u_1'} \geq \frac{-u_2''}{u_2'}$ always holds in this case. To help in further understanding the concept of "risk aversion in the sense of Ross", we denote $\pi_i(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w)$ as the price that agent i is ready to pay to replace lottery $\tilde{\varepsilon}_2$ with lottery $\tilde{\varepsilon}_1$ at wealth level w, i.e.,

$$Eu_i(w + \tilde{\varepsilon}_2) = Eu_i(w - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w) + \tilde{\varepsilon}_1). \tag{5}$$

[4] The reason for discussing only the 2×2 case is merely for expositional purposes. If there is a complete order of agents' risks, i.e., $\varepsilon_1^m \succeq \varepsilon_2^m \succeq \ldots \succeq \varepsilon_N^m$ and $\varepsilon_1^f \succeq \varepsilon_2^f \succeq \ldots \succeq \varepsilon_N^f$, then our conditions for NAM/PAM can immediately apply in the case in which there are equal numbers of males and females, as well as in the case of the matched agents when there are unequal numbers of males and females, although in the latter case, the identities of the agents who are left unmatched depend on the distribution of the population.

[5] Suppose f is a convex function with one variable. Then for any $x_1 \leq \min(x_2, x_3) \leq \max(x_2, x_3) \leq x_4$ such that $x_1 + x_4 = x_2 + x_3$ we must have $f(x_1) + f(x_4) \geq f(x_2) + f(x_3)$.

Ross (1981) showed that agent u_1 is more risk-averse than agent u_2 in the sense of Ross if and only if agent u_1 is ready to pay more than agent u_2 for any SSD reduction in risk (i.e., $\pi_1(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w) \geq \pi_2(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w)$, $\forall w$, $\tilde{\varepsilon}_2$ and $\tilde{\varepsilon}_1$, with $\tilde{\varepsilon}_2 \overset{SSD}{\precsim} \tilde{\varepsilon}_1$).

Definition 6 *A utility function u exhibits DARA in the sense of Ross if* $\pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_1) \geq \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_2)$, $\forall w_1, w_2$, $\tilde{\varepsilon}_2$ *and* $\tilde{\varepsilon}_1$, *with* $\tilde{\varepsilon}_2 \overset{SSD}{\precsim} \tilde{\varepsilon}_1$ *and* $w_1 \leq w_2$.

Definition 7 *A utility function u satisfies the property that any SSD deterioration in the background risk increasing risk aversion in the sense of Ross if for any* $\tilde{\varepsilon}_2 \overset{SSD}{\precsim} \tilde{\varepsilon}_1$, U_2 *is more risk-averse than* U_1 *in the sense of Ross, where*

$$U_i(x) \equiv Eu(x + \tilde{\varepsilon}_i), \text{ for } i = 1, 2.$$

The above notation and results allow us to derive the following proposition:

Proposition 2 *If the preference belongs to the HARA class and exhibits DARA in the sense of Ross and any SSD deterioration in the background risk increases risk aversion in the sense of Ross, then the risk-sharing matching game will be negative assortative on agents' income riskiness.*

Proof. Define $\pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w)$ as the price that agent v is ready to pay to replace lottery $\tilde{\varepsilon}_2$ with lottery $\tilde{\varepsilon}_1$ at wealth level w. From the concept of risk premium and the definition of $\pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w)$, we have

$$\begin{aligned} v(w_0 - \pi(\tilde{\varepsilon}_2, w_0)) &= Ev(w_0 + \tilde{\varepsilon}_2) \\ &= Ev(w_0 - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0) + \tilde{\varepsilon}_1) \\ &= v(w_0 - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0) - \pi(\tilde{\varepsilon}_1, w_0 - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0))) \end{aligned}$$

from which we obtain

$$\pi(\tilde{\varepsilon}_2, w_0) = \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0) + \pi(\tilde{\varepsilon}_1, w_0 - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0)) \tag{6}$$

The two sides of the above equation represent two equivalent ways of eliminating risk $\tilde{\varepsilon}_2$. One is to eliminate $\tilde{\varepsilon}_2$ once and for all, and agent v is willing to pay $\pi(\tilde{\varepsilon}_2, w_0)$ for this. The other is to eliminate $\tilde{\varepsilon}_2$ step by step, first replacing $\tilde{\varepsilon}_2$ with a smaller risk $\tilde{\varepsilon}_1$ at the price of $\pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0)$ and then eliminating $\tilde{\varepsilon}_1$ at the price of $\pi(\tilde{\varepsilon}_1, w_0 - \pi(\tilde{\varepsilon}_2 \to \tilde{\varepsilon}_1, w_0))$.

A stable match is NAM if

$$\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_2^f, w_0) \geq \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f, w_0) \tag{7}$$

Applying (6), we can rewrite the above inequality as

$$\begin{aligned} &\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_1)\right] \\ \geq \ &\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_2)\right] \end{aligned} \tag{8}$$

where $w_1 = w_0 - \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0)$ and $w_2 = w_0 - \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$.
Now we prove

$$\pi(\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) \geq \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) \tag{9}$$

Consider agent 1 with utility function $V_1(x) = Ev(x + \tilde{\varepsilon}_1^f)$ and agent 2 with $V_2(x) = Ev(x + \tilde{\varepsilon}_2^f)$. As in Gollier (2001), we define the risk premium $\pi_1(\tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m)$ as the price that agent 1 is willing to pay to replace $\tilde{\varepsilon}_2^m$ with $\tilde{\varepsilon}_1^m$, and we define $\pi_2(\tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m)$ as the counterpart for agent 2. Then $\pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) = \pi_1(\tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m)$ and $\pi(\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) = \pi_2(\tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m)$. (9) holds iff V_2 is more risk averse than V_1 in the sense of Ross, i.e., an SSD deterioration in the background risk makes the agents more risk averse in the sense of Ross.
Given (9), (8) holds if

$$\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_1)\right] \geq \left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_2)\right] \tag{10}$$

which, by applying (6), can be rewritten as

$$\begin{aligned} &\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_4) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_1)\right] \\ \geq\ &\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_3) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_2)\right] \end{aligned}$$

where $w_4 = w_1 - \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_1)$, $w_3 = w_0 - \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$. If utility exhibits DARA in the sense of Ross, we have $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_1) \geq \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$. Thus, we only need to show:

$$\left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_4)\right] \geq \left[\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_2) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_3)\right] \tag{11}$$

To prove (11), we first show that

$$w_0 + w_4 \leq w_2 + w_3 \tag{12}$$

Using the expressions to substitute for w_0, w_2, w_3 and w_4, (12) can be rewritten as

$$\begin{aligned} &\left[\pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_1)\right] \\ \geq\ &\left[\pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)\right] \end{aligned}$$

Via (9), we know that $\pi(\tilde{\varepsilon}_2^f + \tilde{\varepsilon}_2^m \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f, w_0) \geq \pi(\tilde{\varepsilon}_2^m + \tilde{\varepsilon}_1^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$. Moreover, the fact that the utility is DARA in the sense of Ross and $w_0 > w_1$ implies that $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_1) \geq \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_2^f \to \tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$. Hence, the above inequality and therefore (12) hold.
Now, inequality (11) follows by noticing that (i) π is decreasing and convex in w_0; (ii) $w_0 > \max(w_2, w_3) > \min(w_2, w_3) > w_4$; and (iii) inequality (12) holds. ■

Proposition 2 provides a sufficient condition for NAM: if the preference belongs to the HARA class and exhibits DARA in the sense of Ross and a higher background risk leads to more risk-averse behavior, then negative sorting is stable. Thus, in facing risks from the male side, female 2 (taking $\tilde{\varepsilon}_2^f$ as her background risk) behaves in a more risk-averse way than female 1 (taking $\tilde{\varepsilon}_1^f$ as her background risk). Therefore, in order to match with the less risky male 1, female 2 is ready to offer male 1 a higher premium over male 2 than female 1 is. Similarly, in order to match with the less risky female 1, male 2 is ready to offer female 1 a higher premium over female 2 than male 1 is. Hence, negative sorting is stable.

One may wonder how restrictive the condition is in Proposition 2. On the theoretical front, Gollier (2001)[6] proved that u_1 is more risk averse than u_2 in the sense of Ross if and only if there exists a scalar η such that:

$$\forall x_1, x_2 : \frac{u_1''(x_1)}{u_2''(x_1)} \geq \eta \geq \frac{u_1'(x_1)}{u_2'(x_1)}.$$

Applying the above condition, one can easily show that u exhibits DARA in the sense of Ross if there exists a scalar λ, such that

$$p(w+y) \geq \lambda \geq r(w+y'), \; \forall y, y', \tag{13}$$

where $p(w) = \frac{-u'''(w)}{u''(w)}$ denotes the measure of absolute prudence and $r(w) = \frac{-u''(w)}{u'(w)}$ denotes the measure of absolute risk aversion.

Under our HARA assumption regarding preference, the utility function can be written as $u(c) = \left(c + \frac{\gamma}{\alpha}\right)^{1-\gamma}$. Then, (13) becomes

$$\frac{\gamma+1}{(w+y)+\frac{\gamma}{\alpha}} \geq \lambda \geq \frac{\gamma}{(w+y')+\frac{\gamma}{\alpha}}, \; \forall y, y'.$$

Suppose the relevant range of wealth is bounded on the interval $[a, b]$. Then, the above inequality becomes

$$\frac{\gamma+1}{b+\frac{\gamma}{\alpha}} \geq \frac{\gamma}{a+\frac{\gamma}{\alpha}},$$

which can be simplified to

$$b - a \leq \frac{1}{\alpha} + \frac{1}{\gamma}a. \tag{14}$$

When the range of the relevant wealth is not too large, the utility exhibits DARA in the sense of Ross.

Now, we derive the conditions under which any SSD deterioration in the background risk increases risk aversion in the sense of Ross. Consider agent 1 with utility function $V_1(x) = Ev(x + \tilde{\varepsilon}_1^f)$ and agent 2 with $V_2(x) = Ev(x + \tilde{\varepsilon}_2^f)$. The condition of Proposition 2 requires that V_1 be more risk averse than V_2 in the sense of Ross. That is, there exists a scalar η such that ,

$$\frac{Ev''(w_1 + \tilde{\varepsilon}_2)}{Ev''(w_1 + \tilde{\varepsilon}_1)} \geq \eta \geq \frac{Ev'(w_2 + \tilde{\varepsilon}_2)}{Ev'(w_2 + \tilde{\varepsilon}_1)}, \forall w_1, w_2. \tag{15}$$

For small risk $\tilde{\varepsilon}_i$, we have:

$$\begin{array}{l} Ev''(w + \tilde{\varepsilon}_i) \approx v''(w) + \frac{1}{2}v''''(w)\,\sigma_i^2 \\ Ev'(w + \tilde{\varepsilon}) \approx v'(w) + \frac{1}{2}v'''(w)\,\sigma_i^2 \end{array}.$$

[6] Proposition 28, Ch 8, page 122.

Substituting into the condition (15), we have:

$$\frac{v''''(w_1)}{v''(w_1)} \geq \eta \geq \frac{v'''(w_2)}{v'(w_2)}, \forall w_1, w_2.$$

Or equivalently:

$$t(w_1)\, p(w_1) \geq \eta \geq p(w_2)\, r(w_2), \forall w_1, w_2 \tag{16}$$

where $t(w) = \frac{-u''''(w)}{u'''(w)}$ denotes the measure of absolute temperance, $p(w) = \frac{-u'''(w)}{u''(w)}$ denotes the measure of absolute prudence, and $r(w) = \frac{-u''(w)}{u'(w)}$ denotes the measure of absolute risk aversion. Under our specification for the utility function, (16) becomes:

$$\frac{\gamma+2}{\left(w_1 + \frac{\gamma}{\alpha}\right)^2} \geq \frac{\gamma}{\left(w_2 + \frac{\gamma}{\alpha}\right)^2}, \forall w_1, w_2$$

which holds if the support of income realizations is sufficiently narrow. In general, for large risks, deriving the conditions for equation (15) is quite complicated and we leave it for future work.

Although the conditions in Proposition 2 impose strict restrictions on preference, as well as risk size, the economic implications are clear and supported by the empirical evidence. Guiso et al. (1996) concluded from Italian survey data that a consumer's perception of a riskier distribution of uninsurable human-capital wealth is negatively related to the proportion of risky assets held in his/her investment portfolio. It is also worthwhile to point out that the seemingly strong conditions for NAM come from the fact that we are looking for sorting patterns for *any* SSD ordered risks. In this sense, the conditions for PAM could be equally if not more restrictive. In the next subsection, we will show that HARA and DARA are sufficient to guarantee NAM if risks are ranked in the sense of SSD taking the form of adding independent noise.

5.3 SSD Risks with Independent Noise

To see a less restrictive condition for monotone sorting, we consider a special case in which $\tilde{\varepsilon}_2^m$ is an increase in risk of $\tilde{\varepsilon}_1^m$ in the sense of SSD taking the form of adding independent noise $\tilde{\varepsilon}^m$, and similarly for $\tilde{\varepsilon}_2^f$ and $\tilde{\varepsilon}_1^f$. That is, assume male 1 and female 1 are endowed with $\tilde{w}_1^m = w_0 + \tilde{\varepsilon}_1^m$ and $\tilde{w}_1^f = w_0 + \tilde{\varepsilon}_1^f$, respectively; and assume male 2 and female 2 are endowed with $\tilde{w}_2^m = w_0 + \tilde{\varepsilon}_1^m + \tilde{\varepsilon}^m$ and $\tilde{w}_2^f = w_0 + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^f$, respectively. All idiosyncratic risks and noises are independently distributed with $E\tilde{\varepsilon}^m = E\tilde{\varepsilon}^f = 0$. Here, in order to characterize the equilibrium sorting pattern, we need to compare $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0)$ and $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^m, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^f, w_0)$.

The following proposition characterizes the equilibrium sorting pattern:

Proposition 3 *If the preference belongs to the HARA class and exhibits DARA, and agents are ranked by their independent riskiness in the sense of SSD taking the form of adding independent noise, then the risk-sharing matching game will be negative assortative on the agents' income riskiness.*

Proof. See Appendix. ■

The concept of risk vulnerability is important in understanding the results of negative sorting. In their seminal paper, Gollier and Pratt (1996) introduced the concept of risk vulnerability as a basic tool for examining the effect of an unfair background risk on an agent's attitude towards other independent risks. In particular, utility is risk vulnerable if and only if the introduction of an unfair risk increases the risk premium of every independent risk.[7] Gollier and Pratt (1996) have listed several necessary and sufficient conditions for risk vulnerability, among which, under HARA, one sufficient condition is DARA[8]. Risk vulnerability guarantees that SSD deterioration taking the form of adding independent noise increases the cost of existing independent risks. DARA implies that SSD deterioration taking the form of adding independent noise increases the cost of the deterioration itself. Multiple risks are, in this sense, self-aggravating, and thus negative sorting is socially preferable and stable.

A special example in this case occurs when m_1 and f_1 are endowed with certain income $w_1^m = w_1^f = w_0$ (i.e., $\tilde{\varepsilon}_1^m = \tilde{\varepsilon}_1^f = 0$) and m_2 and f_2 are endowed with risky income $\tilde{w}_2^m = w_0 + \tilde{\varepsilon}^m$ and $\tilde{w}_2^f = w_0 + \tilde{\varepsilon}^f$. Then NAM arises if $\pi(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0) \geq \pi(\tilde{\varepsilon}^m, w_0) + \pi(\tilde{\varepsilon}^f, w_0)$: the risk premium of the sum of risks is larger than the sum of the risk premiums of the risks, which is indeed the case if utility is risk vulnerable and exhibits DARA.

5.4 An Example of PAM under HARA and DARA

The following example helps us to understand, in general, why we need restrictions beyond HARA and DARA for NAM to arise. In particular, if SSD ordered risks do not take the form of adding independent noise, NAM may fail to arise, even when the preference belongs to the HARA class and exhibits DARA.

Example 4 *Consider the utility function $v(c) = \sqrt{c}$, and assume $w_0 = 0$, $\tilde{\varepsilon}_1^f = (0, \frac{1}{2}; 1, \frac{1}{2})$[9], $\tilde{\varepsilon}_1^m = (0, \frac{1}{2}; x, \frac{1}{2})$, $\tilde{\varepsilon}_2^f = (0, \frac{1}{2}; 0.5, \frac{1}{4}; 1.5, \frac{1}{4})$, $\tilde{\varepsilon}_2^m = (0, \frac{1}{2}; \frac{x}{2}, \frac{1}{4}; \frac{3x}{2}, \frac{1}{4})$, where $x > 0$. $\tilde{\varepsilon}_2^f$ is SSD-dominated by $\tilde{\varepsilon}_1^f$ by introducing a zero-mean risk to $\varepsilon_1^f = 1$; $\tilde{\varepsilon}_2^m$ is SSD-dominated by $\tilde{\varepsilon}_1^m$ by introducing a zero-mean risk to $\varepsilon_1^m = x$. Recall that $C_{ij} = v^{-1}Ev(\tilde{z}_{ij})$. After careful calculation, we obtain C_{11}, C_{12}, C_{21}, and C_{22} as functions of x.[10] Define $f(x) = (C_{11} + C_{22}) - (C_{12} + C_{21})$. Then, NAM arises if $f(x) < 0$, while PAM arises if $f(x) > 0$. It can be shown that there exists a threshold $\widehat{x}$, such that $f(x) < 0$ for $x <$*

[7] The mathematical definition of risk vulnerability is as follows. Define the generalized risk premium $\Pi_{\tilde{\varepsilon}}(\tilde{x}, w)$ of risk $\tilde{x}$ in the presence of initial wealth w and background risk $\tilde{\varepsilon}$ as the price that an agent with utility function u would be willing to pay to avoid risk $\tilde{x}$ at an uncertain position $w + \tilde{\varepsilon}$: $Eu(w + \tilde{\varepsilon} + \tilde{x}) = Eu(w + \tilde{\varepsilon} - \Pi_{\varepsilon}(\tilde{x}, w))$. Define $\pi(\tilde{x}, w)$ as the standard risk premium of risk $\tilde{x}$, which is determined by the following equation: $Eu(w + \tilde{x}) = Eu(w + \tilde{\varepsilon} - \pi(\tilde{x}, w))$. We say that u is risk vulnerable if and only if $\Pi_{\tilde{\varepsilon}}(\tilde{x}, w) \geq \pi(\tilde{x}, w)$ for all w and unfair $\tilde{\varepsilon}$ ($E\tilde{\varepsilon} \leq 0$).

[8] See Gollier and Pratt (1996) Corollary 1, page 117.

[9] This formula means that $\Pr(\tilde{\varepsilon}_1^f = 0) = \Pr(\tilde{\varepsilon}_1^f = 1) = \frac{1}{2}$. Similar explanations apply to other random variables.

[10] $C_{11} = \left(\frac{1}{4}\sqrt{x} + \frac{1}{4} + \frac{1}{4}\sqrt{x+1}\right)^2$; $C_{12} = \left(\frac{1}{8}\sqrt{\frac{x}{2}} + \frac{1}{8}\sqrt{\frac{3}{2}x} + \frac{1}{4}\sqrt{1} + \frac{1}{8}\sqrt{1 + \frac{x}{2}} + \frac{1}{8}\sqrt{\frac{3}{2}x + 1}\right)^2$;

$C_{21} = \left(\frac{1}{8}\sqrt{0.5} + \frac{1}{8}\sqrt{1.5} + \frac{1}{4}\sqrt{x} + \frac{1}{8}\sqrt{x + 0.5} + \frac{1}{8}\sqrt{x + 1.5}\right)^2$;

$C_{22} = \left(\frac{1}{8}\sqrt{\frac{x}{2}} + \frac{1}{8}\sqrt{\frac{3}{2}x} + \frac{1}{8}\sqrt{0.5} + \frac{1}{16}\sqrt{\frac{x}{2} + 0.5} + \frac{1}{16}\sqrt{\frac{3x}{2} + 0.5} + \frac{1}{8}\sqrt{1.5} + \frac{1}{16}\sqrt{\frac{x}{2} + 1.5} + \frac{1}{16}\sqrt{\frac{3x}{2} + 1.5}\right)^2$

$\widehat{x}$ and $f(x) > 0$ for $x > \widehat{x}$. This suggests that NAM is more likely to arise if the support of risks is sufficiently narrow, while PAM may arise if the support of risks is sufficiently large.

In this example, we show that for large risks (x being sufficiently large), HARA and DARA are not sufficient to guarantee NAM. The key point to note here is that an SSD deterioration in the background risk may reduce an agent's degree of risk aversion. Define $r_j(w) = \frac{-Ev''\left(w+\tilde{\varepsilon}_j^f\right)}{Ev'\left(w+\tilde{\varepsilon}_j^f\right)}$ as the Arrow-Pratt coefficient of risk aversion for an agent with utility function v in the presence of background risk $\tilde{\varepsilon}_j^f$. It can be shown that, there exists a threshold value $\hat{w}$, such that $r_1(w) > r_2(w)$ for $w < \hat{w}$ and $r_1(w) < r_2(w)$ for $w > \hat{w}$[11]. This suggests that the agent is more locally risk averse at $w < \hat{w}$ in the presence of background risk $\tilde{\varepsilon}_1^f$ than in the presence of background risk $\tilde{\varepsilon}_2^f$. So, in facing risks from the male side, female 1 (taking $\tilde{\varepsilon}_1^f$ as her background risk) may behave in a more risk-averse way than female 2 (taking $\tilde{\varepsilon}_2^f$ as her background risk)[12]. Therefore, in order to be matched with the less risky male 1, female 1 may be willing to offer male 1 a higher premium over male 2 than female 2 is. Similarly, in order to match with the less risky female 1, male 1 may be willing to offer female 1 a higher premium over female 2 than male 2 is. As a result, PAM may arise in equilibrium.

In their paper, Chiappori and Reny (2006) showed that NAM always arises if agents differ only in their risk attitude. Following their results, Li, Sun, and Chen (2013) showed that PAM may arise if agents can make an effort to reduce their income riskiness. Wang (2013b) showed that the presence of moral hazard may also lead to PAM. Here, without any other confounding factor such as endogenous income or moral hazard, counter intuitive PAM may arise if agents differ in terms of their idiosyncratic risks: while agents with highly risky assets always attempt to avoid being matched with other large perfectly correlated risks, they may prefer to be matched with other large independent risks.

This result helps us to understand the composition of risk-sharing groups in developing countries. Ghatak (1999) argued that PAM should arise because similar people will find it easier to monitor and enforce informal contracts. Empirical evidence, however, is mixed: on one hand, Bacon et al. (2014) found evidence that individuals were more likely to positive assortative mate on their risk attitude; and Arcand and Fafchamps (2012) also found solid evidence of positive sorting for peers with respect to physical or ethnic proximity as well as wealth or household size. On the other hand, Dercon et al. (2006) found little evidence of positive sorting in group-based funeral insurance. Our results suggest that, the risk-sharing effect may drive matching to be negative assortative and therefore offset the monitoring and enforcing effects; however, when risks are large compared with individuals' risk-free incomes, it is possible that the two effects may work in the same direction and drive matching to be positive assortative.

[11] Notice that $r_1(w) = \frac{1}{2}\frac{\frac{1}{2}w^{-\frac{3}{2}}+\frac{1}{2}(1+w)^{-\frac{3}{2}}}{\frac{1}{2}w^{-\frac{1}{2}}+\frac{1}{2}(1+w)^{-\frac{1}{2}}}$ and $r_2(w) = \frac{\frac{1}{2}w^{-\frac{3}{2}}+\frac{1}{4}(0.5+w)^{-\frac{3}{2}}+\frac{1}{4}(1.5+w)^{-\frac{3}{2}}}{\frac{1}{2}w^{-\frac{1}{2}}+\frac{1}{4}(0.5+w)^{-\frac{1}{2}}+\frac{1}{4}(1.5+w)^{-\frac{1}{2}}}$.

[12] Notice that female 1 is not *uniformly* more risk averse than female 2. Indeed, she is locally more risk averse than female 2 for $w < \hat{w}$ but less risk averse than female 2 for $w > \hat{w}$. Therefore, there is no uniform prediction for the two agents' risk-taking behavior. Depending on the properties of the risk taken, female 1 *may* or *may not* behave in a more risk-averse way than female 2.

6 Extension: Multidimensional Matching

In general, individuals have different incomes and face different risks. When agents' types are multidimensional, there is not a complete order of types. We therefore only consider two cases with complete order, that is, the order of agents' riskiness goes in exactly the same or exactly the opposite direction as their risk-free incomes. We have shown via Lemma 3 that if agents all face the same risks but differ in their risk-free incomes, NAM will arise. We have also shown via Propositions 1 and 2 that if agents all have the same risk-free income but differ in the risks they face, NAM will arise under certain restrictions. Thus, if types are two dimensional, a natural guess would be that NAM will arise based on riskiness because higher risk-free incomes seem to go in the same direction as higher riskiness for DARA utilities. We now proceed to show that in fact, both NAM and PAM can arise in this case.

6.1 The Case of Systematic Risk

We first study a multidimensional matching game in which risks are perfectly correlated. Agents are characterized by a pair (w_i, k_i). Here we only consider two cases with complete order: (i) agents with lower risk-free incomes hold larger shares of the common asset, i.e., $k_i < k_{i+1}$ and $w_i > w_{i+1}$, $k_j < k_{j+1}$ and $w_j > w_{j+1}$ and (ii) agents with higher risk-free incomes hold larger shares of the common asset, i.e., $k_i < k_{i+1}$ and $w_i < w_{i+1}$, $k_j < k_{j+1}$ and $w_j < w_{j+1}$.

In the first case, consider male i, i' and female j, j', with $i < i'$, and $j < j'$. Remember that NAM arises if

$$\pi_{ij} + \pi_{i'j'} \geq \pi_{i'j} + \pi_{ij'} \tag{17}$$

where $\pi_{ij} = E\tilde{z}_{ij} - v^{-1}[Ev(w_{ij} + k_{ij}\tilde{x})]$ is the joint risk premium for (i, j); $w_{ij} = w_i + w_j$ and $k_{ij} = k_i + k_j$. Notice that π_{ij} can be written as a function of w_{ij} and k_{ij}: $\pi_{ij}(w_{ij}, k_{ij})$. To simplify, we drop the subscript "ij" and write the function of joint risk premium as $\pi(w, k) = E\tilde{z} - v^{-1}[Ev(w + k\tilde{x})]$, whose properties are listed below:

Lemma 4 $\frac{\partial^2 \pi(w,k)}{\partial k^2} \geq 0$, $\frac{\partial^2 \pi(w,k)}{\partial w^2} \geq 0$, $\frac{\partial^2 \pi(w,k)}{\partial w \partial k} \leq 0$.

Proof. See Appendix. ■

The first inequality is actually Proposition 1. The second inequality implies that if the agents all hold the same amount of the common asset, there will be NAM on the risk-free income. Higher income leads to higher tolerance for risk under DARA; thus if we take risk-free income as a proxy for agents' degree of risk aversion, then Lemma 4 coincides with Chiappori and Reny's (2006) argument that stable match is negative assortative on agents' risk attitude. The following lemma is also useful:

Lemma 5 *Let $f(x, y)$ be twice continuously differentiable in the domain $[0, \infty) \times [0, \infty)$, with $f_{11} \geq 0$, $f_{22} \geq 0$ and $f_{12} \leq 0$, where $f_{11} = \frac{\partial^2 f}{\partial x^2}$ $f_{22} = \frac{\partial^2 f}{\partial y^2}$ and $f_{12} = \frac{\partial^2 f}{\partial x \partial y}$. Then for any* $0 \leq x_1 \leq \min(x_2, x_3) \leq$

$\max(x_2, x_3) \le x_4$, *and* $y_1 \ge \max\{y_2, y_3\} \ge \min\{y_2, y_3\} \ge y_4 \ge 0$, *such that*

$$x_1 + x_4 = x_2 + x_3 \tag{18}$$

$$y_1 + y_4 = y_2 + y_3 \tag{19}$$

we must have

$$f(x_1, y_1) + f(x_4, y_4) \ge f(x_2, y_2) + f(x_3, y_3) \tag{20}$$

Proof. See Appendix. ■

The two lemmata immediately yield the following proposition:

Proposition 4 *If the preference belongs to HARA class and exhibits DARA, and agents with lower risk-free incomes hold larger sizes of common risky assets, then NAM is stable.*

Proof. Note that $\pi_{ij} + \pi_{i'j'} = \pi(w_{ij}, k_{ij}) + \pi(w_{i'j'}, k_{i'j'})$ and $\pi_{i'j} + \pi_{ij'} = \pi(w_{i'j}, k_{i'j}) + \pi(w_{ij'}, k_{ij'})$. Because $w_{ij} = w_i + w_j$ and $k_{ij} = k_i + k_j$, we have $w_{ij} + w_{i'j'} = w_{i'j} + w_{ij'}$ and $k_{ij} + k_{i'j'} = k_{i'j} + k_{ij'}$. Because in this case we must have $w_{ij} > \max(w_{i'j}, w_{ij'}) > \min(w_{i'j}, w_{ij'}) > w_{i'j'}$ and $k_{ij} > \max(k_{i'j}, k_{ij'}) > \min(w_{i'j}, w_{ij'}) > k_{i'j'}$, by Lemmata 4 and 5, the following inequality holds: $\pi_{ij} + \pi_{i'j'} \ge \pi_{i'j} + \pi_{ij'}$. Thus, NAM arises in equilibrium. ■

Remark: If agents with higher risk-free incomes hold larger sizes of common risky assets, then both NAM and PAM can arise in equilibrium. Through two examples, we show that both NAM and PAM are possible. (a) The NAM example. Suppose agent i's risk-free income is given by $w_i = w_0 + k_i a$, where k_i is the size of the risk held by agent i and a is a constant number. In this example, all agents' risk-free income and size of risks pairs (w_i, k_i)s lie on the same line. Define $\tilde{y} = a + \tilde{x}$. Then, agent i's income can be written as $w_i + k_i\tilde{x} = w_0 + k_i\tilde{y}$. This specification brings us back to the case in which all agents have the same risk-free income w_0 and hold different amounts of a common asset $\tilde{y}$. Thus, NAM is stable. (b) The PAM example. Consider a 2×2 case in which the utility function belonging to the HARA class takes the following form: $v(w) = \ln w$. Assume the common risk $\tilde{x}$ is a small risk with zero mean and variance σ^2. Applying Arrow-Pratt approximation, we have $\pi(w, k) \approx \frac{1}{2}\frac{1}{w}\ k^2\sigma^2$. Let $w_0^m = w_0^f = 1$, $k_1^m = k_1^f = 1$, $k_2^m = k_2^f = 10$, $w_1^m = w_0^m + k_1^m$, $w_2^m = w_0^m + k_2^m$, $w_1^f = w_0^f + k_1^f$ and $w_2^f = w_0^f + k_2^f * x$, with $x \ge 1$. In this example, the points (w_1^m, k_1^m), (w_2^m, k_2^m) and (w_1^f, k_1^f) lie on the same line $w_i = w_0 + k_i$, while the wealthier female's risk-free income and size of risk pair (w_2^f, k_2^f) lie off of the line. The parameter x measures how far away the point (w_2^f, k_2^f) is from the line $w_i = w_0 + k_i$. If $x = 1$, then the point (w_2^f, k_2^f) lies exactly on the line $w_i = w_0 + k_i$, and we are back in our example of NAM. A simple calculation gives $\pi_{11} + \pi_{22} - (\pi_{12} + \pi_{21}) = \frac{1}{2}\left[\frac{2^2}{4} + \frac{20^2}{(12+10x)} - \left(\frac{11^2}{13} + \frac{11^2}{13+10x}\right)\right] * \sigma^2$. It can be shown that there exists a threshold $\widehat{x} > 1$, such that $\pi_{11} + \pi_{22} - (\pi_{12} + \pi_{21}) > 0$ for $x < \widehat{x}$ and $\pi_{11} + \pi_{22} - (\pi_{12} + \pi_{21}) < 0$ for $x > \widehat{x}$. Thus, NAM arises if $x < \widehat{x}$ and PAM arises if $x > \widehat{x}$. The result suggests that if agents' risk-free income and size of risks pairs lies sufficiently away from the same line, then PAM arises.

6.2 The Case of Idiosyncratic Risks

We now study a multidimensional matching game in which the risks are independent. Suppose agent i's income is $w_i + \tilde{\varepsilon}_i$. Define $\tilde{\epsilon}_i = \tilde{\varepsilon}_i + w_i - w_0$. Then, agent i's income can be written as $w_0 + \tilde{\epsilon}_i$. If the newly defined risks $\tilde{\epsilon}_i$ can still be ranked in terms of SSD, i.e., $\tilde{\epsilon}_1 \overset{SSD}{\succsim} \tilde{\epsilon}_2 \overset{SSD}{\succsim} \ldots \overset{SSD}{\succsim} \tilde{\epsilon}_N$, then we are back to our idiosyncratic model seen in Section 5. The problem is that, even if the original risks can be ranked in the sense of SSD ($\tilde{\varepsilon}_1 \overset{SSD}{\succsim} \tilde{\varepsilon}_2 \overset{SSD}{\succsim} \ldots \overset{SSD}{\succsim} \tilde{\varepsilon}_N$), the newly defined risks $\tilde{\epsilon}_i$s may not have an SSD order. Here we consider two cases with a complete order: (i) agents with lower risk-free incomes face higher risks, i.e., $\tilde{\varepsilon}_i \overset{SSD}{\succsim} \tilde{\varepsilon}_{i+1}$ and $w_i \geq w_{i+1}$, $\tilde{\varepsilon}_j \overset{SSD}{\succsim} \tilde{\varepsilon}_{j+1}$ and $w_j \geq w_{j+1}$ and (ii) agents with lower risk-free incomes face lower risks, i.e., $\tilde{\varepsilon}_i \overset{SSD}{\precsim} \tilde{\varepsilon}_{i+1}$ and $w_i \geq w_{i+1}$, $\tilde{\varepsilon}_j \overset{SSD}{\precsim} \tilde{\varepsilon}_{j+1}$ and $w_j \geq w_{j+1}$.

Proposition 5 *If the preference belongs to the HARA class and exhibits DARA, any SSD deterioration in the background risk increases risk aversion in the sense of Ross, and if agents with lower risk-free incomes face higher idiosyncratic risks, then NAM is stable.*

Proof. The newly defined risks $\tilde{\epsilon}_i$ still have the SSD order $\tilde{\epsilon}_1 \overset{SSD}{\succsim} \tilde{\epsilon}_2 \overset{SSD}{\succsim} \ldots \overset{SSD}{\succsim} \tilde{\epsilon}_N$, and therefore the result follows immediately from Proposition 2. ■

Remark: The following example suggests that, if agents with lower risk-free incomes face smaller idiosyncratic risks, both NAM and PAM could be stable. Consider a 2×2 case in which utility function belonging to the HARA class takes the form $v(c) = \ln c$. Assume that the idiosyncratic risks $\tilde{\varepsilon}_i$s are small with zero mean and variance σ_i^2. Applying Arrow-Pratt approximation, we have $\pi_{ij} \approx \frac{1}{2} \frac{\sigma_i^2 + \sigma_j^2}{w_{ij}}$. Let $w_1^m = w_1^f = 1$, $w_2^m = w_2^f = 10$, σ_{1m}^2[13] $= \sigma_{1f}^2 = \sigma^2$, $\sigma_{2m}^2 = 10\sigma^2$ and $\sigma_{2f}^2 = x\sigma^2$, with σ^2 being arbitrarily small and $x > 1$. Note that larger x means that female 2, who has higher risk-free income than female 1, faces larger idiosyncratic risk. A simple calculation gives $\pi_{11} + \pi_{22} - (\pi_{12} + \pi_{21}) = \frac{1}{2}\left[\frac{9}{22} - \frac{9}{220}x\right] * \sigma^2$, which is positive if $x < 10$ and negative if $x > 10$. Thus, NAM arises if $x < 10$, and PAM arises if $x > 10$. The example illustrates that PAM is likely to arise in equilibrium if agents who have higher risk-free income face risks that are sufficiently large (x is sufficiently large).

7 Concluding Remarks

In this paper, we explore the sorting patterns in a two-sided matching market where agents facing different risks match to share them. We show that the competitive sorting pattern crucially depends on the interaction between risks. While negative sorting almost always arises when risks are perfectly correlated, the counter-intuitive positive sorting may arise when risks are independent. In the case where risks are independent, negative sorting tends to arise if a riskier background risk leads to more risk-averse behavior. Our findings enrich the literature on assortative matching, and to the best of our

[13] The formula σ_{1m}^2 means that the variance of male 1's income is σ_{1m}^2.

knowledge, are among the first attempts to investigate sorting over agents' risk exposure. Our results help in understanding the composition of risk-sharing groups in developing countries. Behind the mixed empirical evidence of sorting patterns, there might be a trade-off between the risk-sharing effect and the monitoring and enforcing effects.

The present research can be extended along several lines. Firstly, in many instances, the riskiness of income is not entirely exogenous but partially a choice variable. In the developed world, individuals usually choose their professions and investments as a function of their risk preferences and their abilities. There could then be a trade-off between competing for the most suitable partner for the purpose of risk sharing and for the motive of risk control. Li, Sun and Chen (2013) and Wang (2013a) studied endogenous risks and showed that PAM may arise in equilibrium. However, they only considered the case where preferences belong to the CARA class and incomes were subjected to normal distributions. It is therefore worthwhile to explore more general cases. Secondly, an interesting extension would permit agents to renegotiate sharing rules posterior to matching. Li, Sun, and Wang (2015) introduced a bargaining stage and showed that PAM may arise in equilibrium. Thirdly, the effect of risk factors on matching efficiency is also relevant for financial securities[14] or joint venture agreements. Our model indicates that it is costly to pair two highly risky assets together, which is associated with a high social cost of risk that the investors have to pay. The recent trend of overconcentration of risks in the subordinated debts raises our concern that, for the issuers, the main purpose of securitization is not to share risks with investors, but to keep the risk concentrated so that they can achieve as much leveraging as possible (Acharya and Richardson, 2010). Further studies are needed in the context of financial securities and institutions.

References

[1] Acharya, V., and Richardson, M. (2010), "Causes of the Financial Crisis," in *Critical Review*, vol. 21, 195-210 (Routledge Publishers).

[2] Arcand, J.-L., and Fafchamps, M. (2012), "Matching in Community-Based Organizations," *Journal of Development Economics*, 98, 203-221

[3] Bacon, M. et al. (2014), "Assortative mating on risk attitude ," *Theory and Decision* 77(3), 389-401.

[4] Becker, G. (1973), "A Theory of Marriage: Part I," *Journal of Political Economy* 81(4), 813-846.

[5] Bigsten et al, (2003), "Risk Sharing in Labour Markets," Tinbergen Institute Discussion Paper No. 03-077/2.

[6] Chiappori, P.-A. and Reny, P. (2006), "Matching to Share Risk." mimeo.

[7] Dercon et al. (2006), "Group-Based Funeral Insurance in Ethiopia and Tanzania", *World Development*, 34, 685-703.

[14] We thank an anonymous referee for pointing out the possibility of applying our model in this direction.

[8] Eeckhoudt, L. and Gollier, C. (2001), "Which Shape for the Cost Curve of Risk," *The Journal of Risk and Insurance*, vol. 68, No. 3, 387-401.

[9] Fafchamps, M. and Lund, S. (2003), "Risk Sharing Networks in Rural Philippines," *Journal of Development Economics*, 71:261-287.

[10] Ghatak, M. (1999), "Group Lending, Local Information and Peer Selection," *Journal of Development Economics*, 60, 27-53.

[11] Gollier, C. (2001), *The Economics of Risk and Time*, Cambridge, MA: The MIT Press.

[12] Gollier, C., and Pratt, J. W. (1996), "Risk Vulnerability and the Tempering Effect of Background Risk," *Econometrica*, vol. 64, No. 5, 1109-1123.

[13] Guiso, L., Jappelli, T., and Terilizzesse, D. (1996), "Income Risk, Borrowing Constraints and Portfolio Choice," *American Economic Review*, 86, 158–172.

[14] Jaramillo, F., Kempf, F. and Moizeau, H. (2013), "Heterogeneity and the Formation of Risk-Sharing Coalitions," working paper.

[15] Legros, P. and Newman, A. F. (2007), "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, Econometric Society, vol. 75(4), 1073-1102.

[16] Li, S., Sun, H., and Chen, P. (2013), "Assortative Matching of Risk-Averse Agents with Endogenous Risk", *Journal of Economics*, 109(1): 27-40.

[17] Li, S., Sun, H., and Wang, T. (2015), "Change in Risk and Bargaining Game", working paper.

[18] MacCrimmon, K. R. and Wehrung, D. A. (1986), *Taking Risks: The Management of Uncertainty*, Free Press, New York.

[19] MacCrimmon, K. R. and Wehrung, D. A. (1990), "Characteristics of Risk Taking Executives," *Management Science*, 36, 422-435.

[20] Mazzocco, M. (2004), "Saving, Risk Sharing, and Preferences for Risk," *American Economic Review*, 94(4), 1169-82.

[21] Rosenzweig, M. and Stark, O. (1989), "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India," *Journal of Political Economy,* 97: 905-926.

[22] Ross, S. A. (1981), "Some Stronger Measures of Risk Aversion in the Small and in the Large with Applications," *Econometrica* 3: 621-38.

[23] Schoemaker, P. J. (1990), "Are Risk-Preferences Related Across Payoff Domains and Response Modes?" *Management Science*, 36 1451-1463.

[24] Schoemaker, P. J. (1993), "Determinants of Risk-Taking: Behavioral and Economic Views," *Journal of Risk and Uncertainty*, 6, 49-73.

[25] Schulhofer-Wohl, S. (2006), "Negative Assortative Matching of Risk-Averse Agents with Transferable Expected Utility," *Economics Letters*, vol. 92, Issue 3, 383-388.

[26] Slovic, P. (1964), "Assessment of Risk Taking Behavior," *Psychological Bulletin*, 61, 330-333

[27] Stanton, R. and Wallace, C. (2010), "CMBS Subordination, Ratings Inflation, and the Crisis of 2007-2009," NBER Working Papers 16206, National Bureau of Economic Research, Inc.

[28] Townsend, R. (1994), "Risk and Insurance in Village India," *Econometrica,* 62: 539-592.

[29] Wang, X. (2013a), "Endogenous Insurance and Informal Relationships," working paper.

[30] Wang, X. (2013b), "Risk, Incentives, and Contracting Relationships," working paper.

[31] Weber, E. U. and Milliman, R. A. (1997), "Perceived Risk Attitudes: Relating Risk Perception to Risky Choice," *Management Science*, 43, 2, 123-144.

8 Appendix

8.1 Proof of Lemma 2

Proof. First, recall that π_{ij} being supermodular is equivalent to C_{ij} being submodular. Then we prove the lemma by contradiction. Suppose C_{ij} is submodular and NAM does not arise. This means that there exist $i < i'$ and $j < j'$, such that, in equilibrium, male i is matched with j and male i' is matched with female j'. Denote the equilibrium certainty equivalent of the four agents by C_i, C_j, $C_{i'}$ and $C_{j'}$ respectively. We have $C_i + C_j = C_{ij}$ and $C_{i'} + C_{j'} = C_{i'j'}$. Because C_{ij} is submodular, we must have $C_{i'j} + C_{ij'} > C_{ij} + C_{i'j'}$, which implies that either $C_{i'j} > C_{i'} + C_j$ or $C_{ij'} > C_i + C_{j'}$ holds but not both (which contradicts the fact that $C_{i'j} + C_{ij'} > C_{ij} + C_{i'j'}$). If $C_{i'j} > C_{i'} + C_j$, then i and j' are both better off if they deviate and are matched together; if $C_{ij'} > C_i + C_{j'}$, then i' and j are both better off if they deviate and are matched together. This contradicts our assumption that the matching is stable. Similarly, one can prove that PAM arises if π_{ij} is submodular. ■

8.2 Proof of Proposition 3

Proof. The following definition is useful for the proof of this proposition:

Definition 8 *(Gollier and Pratt, 1996) The* ***generalized risk premium*** $\Pi_{\tilde{\varepsilon}}(\tilde{x}, w)$ *of risk* $\tilde{x}$ *in the presence of initial wealth* w *and background risk* $\tilde{\varepsilon}$*, is the price that the representative agent would be willing to pay to avoid the risk* $\tilde{x}$ *at an uncertain position* $w + \tilde{\varepsilon}$*:* $Ev(w + \tilde{\varepsilon} + \tilde{x}) = Ev(w + \tilde{\varepsilon} - \Pi_{\varepsilon}(\tilde{x}, w))$*, where* $\tilde{x}$ *and* $\tilde{\varepsilon}$ *are independent.*

Gollier (2001)[15] proved that risk aversion in the sense of Ross is a sufficient condition for the comparative risk aversion to be preserved in the presence of a background risk. That is, if agent u_1 is more risk-averse than agent u_2 in the sense of Ross, then agent u_1 behaves in a more risk-averse way than agent u_2 in the presence of background risk. In technical terms, this means that if $u_1 = \lambda u_2 + g$, then $\Pi_{1\tilde{\varepsilon}}(\tilde{x}, w) \geq \Pi_{2\tilde{\varepsilon}}(\tilde{x}, w)\ \forall \tilde{x}, \tilde{\varepsilon}$, where $\Pi_{i\tilde{\varepsilon}}(\tilde{x}, w)$ is the generalized risk premium of agent u_i. We derive a useful equivalence for the generalized risk premium. For risks $\tilde{x}$, $\tilde{y}$ and $\tilde{z}$, by the above definition, we have

$$\begin{aligned} Ev(w - \Pi_{\tilde{x}}(\tilde{y} + \tilde{z}, w) + \tilde{x}) &= Ev(w + \tilde{x} + \tilde{y} + \tilde{z}) \\ &= Ev(w - \Pi_{\tilde{x}+\tilde{y}}(\tilde{z}, w) + \tilde{x} + \tilde{y}) \\ &= Ev(w - \Pi_{\tilde{x}+\tilde{y}}(\tilde{z}, w) - \Pi_{\tilde{x}}(\tilde{y}, w - \Pi_{\tilde{x}+\tilde{y}}(\tilde{z}, w)) + \tilde{x}) \end{aligned}$$

from which it follows that

$$\Pi_{\tilde{x}}(\tilde{y} + \tilde{z}, w) = \Pi_{\tilde{x}+\tilde{y}}(\tilde{z}, w) + \Pi_{\tilde{x}}(\tilde{y}, w - \Pi_{\tilde{x}+\tilde{y}}(\tilde{z}, w)) \tag{21}$$

In particular, when $\tilde{x} = 0$, the above equation is written as

$$\pi(\tilde{y} + \tilde{z}, w) = \Pi_{\tilde{y}}(\tilde{z}, w) + \pi(\tilde{y}, w - \Pi_{\tilde{y}}(\tilde{z}, w)) \tag{22}$$

That is, the costs of multiple risks can be decomposed into the cost of the first risk evaluated in the presence of the second risk and the cost of the second risk evaluated with a sure reduction in wealth due to the existence of the first risk.

A stable match is negative assortative if

$$\begin{aligned} & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0) \\ \geq\ & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^m, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f + \tilde{\varepsilon}^f, w_0) \end{aligned}$$

which under (22) is equivalent to

$$\begin{aligned} & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0) \\ +\ & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0)) \\ \geq\ & \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0)) \\ +\ & \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0)) \end{aligned} \tag{23}$$

By applying (21), we have

$$\begin{aligned} & \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0) \\ =\ & \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f+\tilde{\varepsilon}^f}(\tilde{\varepsilon}^m, w_0) + \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f+\tilde{\varepsilon}^f}(\tilde{\varepsilon}^m, w_0)) \end{aligned} \tag{24}$$

Under risk vulnerability we have

$$\Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f+\tilde{\varepsilon}^f}(\tilde{\varepsilon}^m, w_0) \geq \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0) \tag{25}$$

[15] See Chapter 8, proposition 25, page 118.

Since DARA is preserved under the generalized risk premium, we have

$$\Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f+\tilde{\varepsilon}^f}(\tilde{\varepsilon}^m, w_0)) \geq \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0) \tag{26}$$

Combining (24), (25), and (26), we have

$$\Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0) \geq \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0) + \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0). \tag{27}$$

Under DARA, $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$ is decreasing in w_0, which gives

$$\begin{aligned} & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0)) \\ \geq \quad & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0) - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0)) \end{aligned} \tag{28}$$

Combining (23) and (27), the stable match satisfies NAM if

$$\begin{aligned} & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m + \tilde{\varepsilon}^f, w_0)) \\ \geq \quad & \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^m, w_0)) + \pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0 - \Pi_{\tilde{\varepsilon}_1^m+\tilde{\varepsilon}_1^f}(\tilde{\varepsilon}^f, w_0)) \end{aligned}$$

Under (27) and (28), a sufficient condition is $\pi(\tilde{\varepsilon}_1^m + \tilde{\varepsilon}_1^f, w_0)$ being convex in w_0, which is indeed the case according to Lemma 3. ■

8.3 Proof of Lemma 4

Proof. The proof of $\frac{\partial^2 \pi(w,k)}{\partial k^2} \geq 0$ has already been given in the proof of Proposition 1. The proof of $\frac{\partial^2 \pi(w,k)}{\partial w^2} \geq 0$ is similar to the proof of $\frac{\partial^2 \pi(w,k)}{\partial k^2} \geq 0$. The proof of $\frac{\partial^2 \pi(w,k)}{\partial w \partial k} \leq 0$ is more complicated. Using the expression of $\pi(w,k)$, after careful calculation, we have $\frac{\partial^2 \pi(k\varepsilon, w_0)}{\partial k \partial w_0} \leq 0$ being equivalent to

$$E(T_v(w + k\tilde{x})^{-\gamma} \tilde{x}) E(T_v(w + k\tilde{x})^{-\gamma}) \geq E(T_v(w + k\tilde{x})^{-(1+\gamma)} \tilde{x}) E T_v(w + k\tilde{x})^{1-\gamma} \tag{29}$$

For $\gamma = 1$, the above inequality is equivalent to

$$Cov\left(T_v(w + k\tilde{x})^{-1}, T_v(w + k\tilde{x})^{-1} \tilde{x}\right) \leq 0$$

However, we know that $T_v(w + kx)^{-1}$ is decreasing and $T_v(w + kx)^{-1} x$ is increasing when $\gamma = 1$. Hence, the above inequality holds.

Let us first examine a case with discrete income distribution, in which continuous income distribution is a limiting case. Consider the probability distribution characterized by $p(x = x_k) = p_k$ with $x_1 < x_2 < ... < x_\infty$. Denote $t_k = T_v(w + kx_k)$. The necessary and sufficient condition for (29) is written as

$$\sum_k p_k t_k^{-\gamma} x_k \sum_k p_k t_k^{-\gamma} \geq \sum_k p_k t_k^{-(\gamma+1)} x_k \sum_k p_{kk} t_k^{1-\gamma}$$

which after rearranging is equivalent to

$$\sum_{k>l} \sum_l p_k p_l \left(t_k^{-\gamma} x_k t_l^{-\gamma} + t_l^{-\gamma} x_l t_k^{-\gamma} - t_k^{-(\gamma+1)} x_k t_l^{1-\gamma} - t_l^{-(\gamma+1)} x_l t_k^{1-\gamma}\right) \geq 0$$

The above holds for any discrete income distribution iff for all x_k, x_l,

$$t_k^{-\gamma} x_k t_l^{-\gamma} + t_l^{-\gamma} x_l t_k^{-\gamma} - t_k^{-(\gamma+1)} x_k t_l^{-\gamma} - t_l^{-(\gamma+1)} x_l t_k^{-\gamma} \geq 0$$

which, dividing both sides by $t_k^{-\gamma} t_l^{-\gamma}$, is equivalent to

$$x_k + x_l - t_k^{-1} x_k t_l - t_l^{-1} x_l t_k \geq 0$$

which is independent from γ. We already know that (29) holds for $\gamma = 1$; hence, the above inequality must hold. Consequently, (29) holds for all γ. ■

8.4 Proof of Lemma 5

Proof. Without the loss of generality, we assume that $0 \leq x_1 \leq x_2 \leq x_3 \leq x_4$ (otherwise, one can always change the subscripts of x_2 and x_3).

Suppose $y_2 < y_3$. Because $f_{12} \leq 0$ and $x_2 \leq x_3$, we have

$$f(x_2, y_3) - f(x_2, y_2) \geq f(x_3, y_3) - f(x_3, y_2)$$

or

$$f(x_2, y_3) + f(x_3, y_2) \geq f(x_2, y_2) + f(x_3, y_3)$$

Thus, to prove the lemma, it suffices to show that

$$f(x_1, y_1) + f(x_4, y_4) \geq f(x_2, y_3) + f(x_3, y_2)$$

which means that we only need to prove (20) for the case in which $y_2 \geq y_3$. That is, without a loss of generality, we can suppose that $y_1 \geq y_2 \geq y_3 \geq y_4$.

Because $f_{12} \leq 0$ and $y_4 \leq y_1$, we have

$$f(x_2, y_4) - f(x_1, y_4) \geq f(x_2, y_1) - f(x_1, y_1)$$

or

$$f(x_1, y_1) + f(x_2, y_4) \geq f(x_1, y_4) + f(x_2, y_1) \tag{30}$$

Because $f_{11} \geq 0$, f is convex with respect to x, for any given y. According to (18), we must have

$$f(x_1, y_4) + f(x_4, y_4) \geq f(x_2, y_4) + f(x_3, y_4) \tag{31}$$

Similarly, because $f_{12} \leq 0$ and $x_2 \leq x_3$, we have

$$f(x_2, y_1) - f(x_2, y_2) \geq f(x_3, y_1) - f(x_3, y_2)$$

or

$$f(x_2, y_1) + f(x_3, y_2) \geq f(x_2, y_2) + f(x_3, y_1) \tag{32}$$

Also, because $f_{22} \geq 0$, f is convex with respect to y, for any given x. According to (19)

$$f(x_3, y_1) + f(x_3, y_4) \geq f(x_3, y_2) + f(x_3, y_3) \tag{33}$$

Our result (20) immediately follows from summing (30), (31), (32), and (33). ■

8.5 Proposition 1 with General Preference

Consider the 2×2 case in which there are two males m_1, m_2 and two females f_1, f_2. Define (m_i, f_j)'s joint risk exposure as $k_{ij} \triangleq k_i + k_j$. We normalize $k_{11} = 1$, and thus $k_{12} > 1$, $k_{21} > 1$ and $k_{22} = k_{12} + k_{21} - 1$. Thus, $(m_i,\ f_j)$'s joint income is $\tilde{z}_{ij} = 2w_0 + k_{ij}\tilde{x}$. Using $V(k_{ij}, \bar{u})$, we can define the indirect utility function of the maximization problem as follows:

$$V(k_{ij}, \bar{u}) = \max_c E[u(z_{ij} - c)] \quad \text{s.t. } E[u(c)] \geq \bar{u}$$

and we have

$$V(k_{ij}, \bar{u}) = \max_{c,\lambda} E[u\left(z_{ij} - c\right) + \lambda\left(u\left(c\right) - \overline{u}\right)] \tag{34}$$

where $\lambda = \lambda(k_{ij}, \bar{u}) > 0$ is a function of the joint size of risk, k_{ij}, and the minimum expected utility level guaranteed for female $\bar{u}$. $V(k_{ij}, \bar{u})$ represents the maximum payoff m_i can get given f_j's payoff being no less than $\bar{u}$. The first-order conditions require perfectly correlated marginal utilities for the matched agents:

$$u'\left(z_{ij} - c\right) = \lambda u'\left(c\right), \forall z_{ij} \tag{35}$$

under the following constraint:

$$Eu\left(c\right) = \overline{u}. \tag{36}$$

Denote the solution to (35) and (36) as $c_{ij} = c\left(z_{ij}, \overline{u}\right) = c(x; k_{ij}, \bar{u})$[16] and $\lambda_{ij} = \lambda(k_{ij}, \bar{u})$. For the NTU matching game, Legros and Newman (2007) have established the "generalized difference condition" for monotone sorting. Applying their condition, we have the following lemma:

Lemma A1 *For the arbitrary distribution of risk sizes, the stable match of the risk-sharing matching game is negative assortative on agents' levels of systematic risk exposure if for* $\forall k_{12}$, $k_{21} > 1$ *and* $\forall \overline{u}$,

$$V\left(k_{12}, V\left(1, \overline{u}\right)\right) \geq V\left(k_{12} + k_{21} - 1, V\left(k_{21}, \overline{u}\right)\right) \tag{37}$$

The proof can be found in Legros and Newman (2007). The term $V\left(1, \overline{u}\right)$ represents the maximum expected utility for m_1 given that f_1 receives $\overline{u}$. Keeping m_1's payoff at the same level but matching him with f_2 would generate expected utility $V\left(k_{12}, V\left(1, \overline{u}\right)\right)$ for f_2. Thus, the LHS of (37) represents m_1's willingness to pay (in expected utility terms) to be matched with f_2 instead of f_1 given that f_1 receives $\overline{u}$, and the RHS is the counterpart for m_2. Hence, in competing for f_2 rather than f_1, m_1 can always outbid male m_2 and still leave more (compared to being matched with f_1) for himself. Before proceeding, the following lemma provides a condition equivalent to (37).

[16] The last equality is due to the fact that z_{ij} is a function of k_{ij} and x: $z_{ij} = 2w_0 + k_{ij}x$.

Lemma A2 *For the arbitrary distribution of risk sizes, the stable match of the risk-sharing matching game is negative assortative on agents' sizes of systematic risk exposure if for* $\forall k_{12},\, k_{21} > 1$ *and* $\forall \overline{u}$,

$$E[u'(c_{i1})\,\tilde{x}] \geq E[u'(c_{i2})\,\tilde{x}] \tag{38}$$

where $c_{i1} = c(x; k_{i1}, \bar{u})$ *and* $c_{i2} = c(x; k_{i2}, \bar{u})$ *are solutions to (35) and (36).*

Proof. We first prove that

$$E[u'(c_{21})\,\tilde{x}] \geq E[u'(c_{22})\,\tilde{x}] \tag{39}$$

Define $\phi(k_{21}) \triangleq V(k_{12} + k_{21} - 1, V(k_{21}, \overline{u}))$. From Lemma A1, a sufficient condition for NAM is that: for $\forall k_{12},\, k_{21} > 1$ and $\forall \overline{u}$,

$$\phi'(k_{21}) = V_1(k_{22}, V(k_{21}, \overline{u})) + V_2(k_{22}, V(k_{21}, \overline{u}))\, V_1(k_{21}, \overline{u}) \leq 0 \tag{40}$$

where $V_l()$ denotes the partial derivative of V w.r.t. the l-th argument, and $k_{22} = k_{12} + k_{21} - 1$. Note that (40) is the Legros and Newman General Differential Condition (2007: Proposition 3).
Because $V_1(k_{ij}, \overline{u}) = E[u'(z_{ij} - c_{ij})\,\tilde{x}]$ and $V_2(k_{ij}, \overline{u}) = -\lambda_{ij}$, a standard implication of the envelope theorem, (40) is equivalent to

$$E[u'(z_{22} - \widehat{c}_{22})\,\tilde{x}] - \lambda_{22} E[u'(z_{21} - c_{21})\,\tilde{x}] \leq 0$$

for $\forall k_{12},\, k_{21} > 1$ and $\forall \overline{u}$, where $\widehat{c}_{22}$ is the solution to (35) and (36) found by replacing $\bar{u}$ with $\widehat{u} = V(k_{21}, \overline{u})$. Taking the expectation of (35)$\times\tilde{x}$ yields

$$E[u'(z_{22} - c_{22})\,\tilde{x}] = \lambda_{22} E[u'(c_{22})\,\tilde{x}]$$

Also, recall that $V(k_{21}, \overline{u}) = \widehat{u}$, via which we obtain $z_{21} - c_{21} = \widehat{c}_{21}$, where $\widehat{c}_{21}$ is the solution to (35) and (36) by replacing $\bar{u}$ with $\widehat{u} = V(k_{21}, \overline{u})$. Hence, (40) is equivalent to:

$$E[u'(\widehat{c}_{22})\,\tilde{x}] - E[u'(\widehat{c}_{21})\,\tilde{x}] \leq 0$$

for $\forall k_{12},\, k_{21} > 1$ and $\forall \overline{u}$. Because the choice of $\overline{u}$ is arbitrary, so is $\widehat{u}$. The above inequality is equivalent to (39). Also, because (39) holds for any $k_{21} < k_{22}$, (38) holds via the same logic. ■

The intuition behind (38) is clear. A female who receives an expected utility level of $\bar{u}$ valuates the market stock $\tilde{x}$ by employing her marginal utility as a shadow price, which reflects the maximum price (in expected utility terms) she is willing to pay for an extra unit of joint risk exposure. Thus, the benefit of a less risky agent being matched to a highly risky partner must exceed the benefits conferred on a riskier agent for NAM to arise. In other words, (39) states that in competing for m_2, which will result in higher joint risk exposure as compared to with m_1, f_1 can always outbid f_2 as a consequence of the higher valuation of extra risk exposure.

Alternatively, the condition of (38) is equivalent to: for $\forall k_{12}$, $k_{21} > 1$ and $\forall \overline{u}$,[17]

$$E[u''(c_{ij}) \frac{\partial c_{ij}}{\partial k_{ij}} \tilde{x}] \leq 0 \tag{41}$$

Lemma A3 *Inequality (41) holds for $\forall k_{12}$, $k_{21} > 1$, $\overline{u}$ and $\tilde{x}$, if and only if the following inequality holds: for $\forall k_{12}$, $k_{21} > 1$, $\overline{u}$ and $\tilde{x}$,*

$$E(T_i T_j) E(\tilde{x}^2) \geq E(T_i \tilde{x}) E(T_j \tilde{x}) \tag{42}$$

where $T_i \triangleq T(z-c)$ and $T_j \triangleq T(c)$.

Proof. Fixing $\bar{u}$ and solving from (35) yields c as a function of λ, k_{ij} and x, i.e., $c_{ij} = c(\lambda, k_{ij}, x)$. Substituting into (36) yields

$$Eu(c(\lambda(k_{ij})), k_{ij}, x) = \bar{u} \tag{43}$$

from which we can solve for λ as a function of k_{ij}, i.e., $\lambda_{ij} = \lambda(k_{ij})$. Hence, $\frac{\partial c_{ij}}{\partial k_{ij}} = c_1 \lambda' + c_2$, where c_l denotes the partial derivative of function c w.r.t. the l-th argument. Taking the log of both sides of (35) and taking the total differentiation yields $c_1 = \frac{T(c)T(z-c)}{\lambda(T(c)+T(z-c))}$, $c_2 = \frac{T(c)x}{T(c)+T(z-c)}$, $c_3 = \frac{T(c)k_{ij}}{T(c)+T(z-c)}$. Taking the total derivative of (43) w.r.t. k_{ij} yields $\lambda' = \frac{-Eu'(c)c_2}{Eu'(c)c_1}$. After substituting, we find that for $\forall k_{12}$, $k_{21} > 1$ and $\forall \overline{u}$, $E[u''(c_{ij}) \frac{\partial c_{ij}}{\partial k_{ij}} \tilde{x}] \leq 0$ holds iff for $\forall \lambda$, $\forall k_{ij}$ and any distribution of $\tilde{x}$,

$$E\frac{u'(c) T_i T_j}{T_i + T_j} E\frac{u'(c) \tilde{x}^2}{T_i + T_j} \geq E\frac{u'(c) T_j \tilde{x}}{T_i + T_j} E\frac{u'(c) T_i \tilde{x}}{T_i + T_j}$$

As the above inequality is expected to hold for the arbitrary distribution of $\tilde{x}$, we normalize $k_{ij} = 1$. For any distribution $\tilde{x}$ with p.d.f. $\mu(x)$, we can define a new distribution $\tilde{x}_\nu$ with p.d.f. $\nu(x) = \mu(x) \frac{u'(c)}{T_i+T_j} / \int \mu(x) \frac{u'(c)}{T_i+T_j} dx$, under which the above inequality with distribution $\tilde{x}$ can be rewritten as follows: for $\forall \lambda$, $\forall k_{ij}$ and any distribution of $\widetilde{x}_\nu$,

$$ET_i T_j E\tilde{x}_\nu^2 \geq ET_i \tilde{x}_\nu ET_j \tilde{x}_v$$

Because the distribution of $\tilde{x}_\nu$ is arbitrary, this inequality is equivalent to (42). ■

Note that if partners' absolute risk tolerance is linearly dependent, i.e., there exists a constant B, such that for $\forall x$, we have $T_i = BT_j$, then (42) can be rewritten as $ET_i^2 E\tilde{x}^2 \geq (ET_i \tilde{x})^2$, which holds as a direct implication of the Cauchy-Schwarz inequality. This leads to the following proposition:

Proposition A1 $T'' = 0$ *is sufficient for NAM to arise.*

Proof. We want to prove that $T'' = 0 \Rightarrow T_i = BT_j$. If $T'' = 0$, we can express tolerance as a linear function of consumption: $T(c) = \frac{1}{\gamma}c + \frac{1}{\alpha}$. Solving for $u'(c) = D_1 T(c)^{-\gamma}$, where D_1 is a constant,

[17] Note that C_{i1} and C_{i2} only differ in regard to the term k, with $k_{i1} < k_{i2}$.

combining with the F.O.C. of the Pareto optimization (35) yields

$$c_{ij}^{*}(\lambda)=\frac{\frac{\gamma}{\alpha}\left(1-\lambda^{\frac{-1}{\gamma}}\right)+z_{ij}}{1+\lambda^{\frac{-1}{\gamma}}}$$

Substituting the above into the expression of risk tolerance yields

$$T\left(c_{ij}^{*}\right)=\frac{\frac{\gamma}{\alpha}+z_{ij}}{\gamma\left(1+\lambda^{\frac{-1}{\gamma}}\right)};\ T\left(z_{ij}-c_{ij}^{*}\right)=\frac{\lambda^{\frac{-1}{\gamma}}\left(z_{ij}+\frac{\gamma}{\alpha}\right)}{\gamma\left(1+\lambda^{\frac{-1}{\gamma}}\right)} \tag{44}$$

Let $B=\lambda^{\frac{-1}{\gamma}}$, and we immediately have $T_i=BT_j$. ■

The above proposition suggests that when utility belongs to the HARA class, i.e., risk tolerance is linear, $T\left(c\right)=\frac{1}{\gamma}c+\frac{1}{\alpha}$, the matching pattern is negative assortative. It also suggests that (42) is more likely to hold when the relationships between potential partners' risk tolerances are sufficiently linear over the relevant range of wealth. Thus, we naturally require support of the relevant risky incomes being somewhat small and/or T'' being sufficiently close to zero. For instance, when $\tilde{x}$ is small risk w.r.t. w_0, partners' absolute risk tolerances can be approximated by a linear relationship, which is exact in the case of TU.

Before looking further for the necessary or sufficient conditions for (42) to hold, let us examine a case with discrete income distribution, in which continuous income distribution is a limiting case. Consider the probability distribution characterized by $p\left(x=x_k\right)=p_k$ with $x_1<x_2<...<x_\infty$. Denote $T_{ik}=T\left(z\left(x_k\right)-c\left(x_k\right)\right)$, $T_{jk}=T\left(c\left(x_k\right)\right)$. We can establish the following lemma:

Lemma A4 *An equivalent condition of (42) is that: for* $\forall x_k$, x_l,

$$\left(\frac{T_{ik}}{x_k}-\frac{T_{il}}{x_l}\right)\left(\frac{T_{jk}}{x_k}-\frac{T_{jl}}{x_l}\right)\geq 0 \tag{45}$$

Proof. The necessary and sufficient condition for NAM (42) is written as

$$\sum_k p_k T_{ik}T_{jk}\sum_k p_k x_k^2\geq\sum_k p_k T_{ik}x_k\sum_k p_k T_{jk}x_k$$

which after rearranging, is equivalent to

$$\sum_{k>l}\sum_l p_k p_l\left(T_{ik}T_{jk}x_l^2+T_{il}T_{jl}x_k^2-\left(T_{ik}T_{jl}+T_{il}T_{jk}\right)x_k x_l\right)\geq 0$$

The above holds for any discrete income distribution iff for all x_k, x_l, we have

$$\left(T_{ik}x_l-T_{il}x_k\right)\left(T_{jk}x_l-T_{jl}x_k\right)\geq 0 \tag{46}$$

Suppose that the above conditions are not met. Then, there must exist an x_k and an x_l such that $(T_{ik}x_l - T_{il}x_k)(T_{jk}x_l - T_{jl}x_k) < 0$. Let the distribution be such that $p(x = x_k) = p(x = x_l) = \frac{1}{2}$. We then have $ET_j^2 E\tilde{x}^2 = \frac{1}{4}(T_{ik}T_{jl} + T_{il}T_{jk}) x_k x_l$, and hence $ET_j^2 E\tilde{x}^2 < (ET_j\tilde{x})^2$, which contradicts (42). Dividing both sides by $(x_l x_k)^2$, (46) can be written as (45). ■

Obviously, (45) holds if $x_k > 0$ and $x_l < 0$. Hence, for NAM to arise, we only need conditions to guarantee that (45) holds for $x_k > 0$, $x_l > 0$ and $x_k < 0$, $x_l < 0$, where x_k, x_l belong to the support of relevant risks. A sufficient condition is that both the function $\frac{T(c(x))}{x}$ and the function $\frac{T(z(x)-c(x))}{x}$ are monotonely increasing or decreasing in x for $x > 0$ and for $x < 0$, where x belongs to the support of relevant risks. The following proposition states the fundamental results' disentangling effect from risk preference and from risk sizes for the equilibrium sorting pattern of the NTU risk-sharing matching game.

Proposition A2 *There exists an interval* $[\underline{x}, \overline{x}]$, *with* $-2w_0 \leq \underline{x} < 0$ *and* $\overline{x} > 0$, *such that NAM arises if all the supports of the risks are subsets of* $[\underline{x}, \overline{x}]$. *Moreover, 1) if utility exhibits DARA, then the interval is* $[-2w_0, \overline{x}]$, *and 2) if utility exhibits IARA, then the interval is* $[\underline{x}, +\infty]$.

Proof.

$$\frac{\partial}{\partial x}\frac{T(c(x))}{x} = \frac{T_j}{x^2(T_i + T_j)}(T'(c)x - (T_i + T_j)) \tag{47}$$

$$\frac{\partial}{\partial x}\frac{T(z(x) - c(x))}{x} = \frac{T_i}{x^2(T_i + T_j)}(T'(z - c)x - (T_i + T_j)) \tag{48}$$

When $x \to 0$, we have $\frac{\partial}{\partial x}\frac{T(c(x))}{x} \to -\infty$ and $\frac{\partial}{\partial x}\frac{T(z(x)-c(x))}{x} \to -\infty$. By continuity, there exist $\underline{x} < 0$ and $\overline{x} > 0$ such that $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are decreasing on the interval $[\underline{x}, 0)$ and on the interval $(0, \overline{x}]$, respectively.

If utility exhibits DARA, i.e., $T' \geq 0$, combining this with (47) and (48), we find that both $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are negative for $x < 0$. Hence, the only restriction on the lower bounds of the risks is to ensure that consumption is non negative, i.e. $\underline{x} = -2w_0$.

If utility exhibits IARA, i.e., $T' \leq 0$, combining this with (47) and (48) we have that both $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are negative for $x > 0$. Hence, there is no restriction on the upper bound: $\overline{x} = +\infty$. ■

As long as the relevant risks are not too large with respect to w_0, NAM will arise in equilibrium. In the case of DARA, NAM arises when the largest realizations of the risks are not too high. In particular, the sorting pattern will be unambiguously negative assortative for all downside-only risks such as bad weather, recession, war, etc. In the case of IARA, NAM arises as long as the lowest realizations of the risks are not too low. In particular, NAM is always the case for upside-only risks such as economic boom, technological progress, etc.

One can go even further. For example, for IARA utility function, the concavity of the risk tolerance is sufficient to guarantee NAM without imposing any restrictions on the support of the risks.[18] Also, for DARA and concave risk tolerance, which is commonly assumed in the literature when explaining the risk premium puzzle (see Gollier, 2001), one can derive a sufficient condition that provides the exact restriction on the upper bound as follows:

Proposition A3 *If preference exhibits DARA and the risk tolerance is concave, then NAM arises if*

$$\overline{x} \leq \frac{T(0) + T(2w_0 + \overline{x})}{T'(0)} \tag{49}$$

Proof. We must prove that both $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are decreasing on the interval $(0, \overline{x})$, where $\overline{x}$ is given by (49). Define $g(c,x) \triangleq T'(c)\,x - (T(c) + T(x + 2w_0 - c))$ and we have $\frac{T(c(x))}{x} \leq 0$ iff $g(c,x) \leq 0$. Taking the derivative of g w.r.t. c and x yields

$$\frac{\partial g}{\partial c} = T''(c)\,x - (T'(c) - T'(x + 2w_0 - c)) \tag{50}$$

$$\frac{\partial g}{\partial x} = T'(c) - T'(x + 2w_0 - c) \tag{51}$$

Suppose $x > 0$ and $c > z - c$. Combining (51) with $T'' \leq 0$, we have $\frac{\partial g}{\partial x} < 0$. Hence $g(c,x) \leq g(c,0) = -(T(c) + T(2w_0 - c)) < 0$. If $c < z - c$, combining with $T'' \leq 0$ and substituting into (50) and (51) yields $g(c,x) \leq g(0,\overline{x}) = T'(0)\,\overline{x} - (T(0) + T(2w_0 + \overline{x})) \leq 0$, where the last inequality holds if (49) holds. The proof of $\frac{T(z(x)-c(x))}{x}$ being decreasing is similar. ■

To see the role of the linearity of risk tolerance, let us consider $|T''| \equiv \varepsilon$. Then, (49) holds as long as $\overline{x} \leq \sqrt{\frac{4(T(0)+T'(0)w_0)}{\varepsilon}} - 2w_0$[19]. Notice that $\sqrt{\frac{4(T(0)+T'(0)w_0)}{\varepsilon}} - 2w_0$ is decreasing in ε and thus approaches infinity as ε goes to zero. This suggests that the more linear risk tolerance is, the fewer restrictions we need to impose on the risk supports for NAM to arise.

[18] For $x < 0$, $T'(c)\,x - (T_i + T_j) \leq T'(z)\,x - (T(c) + T(z-c)) \leq T'(z)\,x - (T(0) + T(z)) \leq -(T(0) + T(2w_0)) < 0$ where the first two inequalities are due to the fact that $T'' \leq 0$, and the last inequality is due to the fact that $T'(z)\,x - (T(0) + T(z))$ is a non-decreasing function of x for $x < 0$. Similarly $T'(c)\,x - (T_i + T_j) \leq 0$ for $x < 0$. Hence both $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are decreasing for $x < 0$. This, combined with the fact that $\frac{T(c(x))}{x}$ and $\frac{T(z(x)-c(x))}{x}$ are decreasing for $x > 0$ if the utility belongs to IARA, ensures that the sorting pattern will be negative assortative.

[19] $T(0) + T(2w_0 + \overline{x})$
$= 2T(0) + T'(0)(2w_0 + \overline{x}) + \int_0^{2w_0+\overline{x}} \int_0^s T''(t)\,dtds$
$= T'(0)\,\overline{x} + \left[2T(0) + 2T'(0)\,w_0 - \frac{\varepsilon}{2}(2w_0 + \overline{x})^2\right]$
$\geq T'(0)\,\overline{x}$ if $\overline{x} \leq \sqrt{\frac{4(T(0)+T'(0)w_0)}{\varepsilon}} - 2w_0$.