Université de Montréal

Asymmetric Dependence Modeling and Implications for International Diversification and Risk Management

par

Georges D. Tsafack K.

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences économiques

Septembre 2007

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Asymmetric Dependence Modeling and Implications for International Diversification and Risk Management

présentée par

Georges D. Tsafack K.

a été évaluée par un jury composé des personnes suivantes:

[illegible]
président -rapporteur

[illegible]
directeur de recherche

[illegible]
membre du jury

[illegible]
examinateur externe

[illegible]
....
représentant du doyen de la FES

Thèse acceptée le: 11 janvier 2008

Sommaire

Nous traitons dans cette thèse les problèmes de modélisation de la dépendance asymétrique et nonlinéaire avec ses implications sur le choix de portefeuille et la gestion de risque. L'asymétrie de la dépendance désigne un fait stylisé bien établi selon lequel les marchés sont plus corrélés dans les mauvais moments que dans les bons. Dans le premier chapitre nous proposons un nouveau modèle bien adapté à ce phénomène. Nous établissons ensuite un lien entre cette dépendance asymétrique et la coskewness et montrons comment elle peut expliquer la très faible diversification internationale observée et la grande tendance a investir dans les actifs moins risques par rapport à ce que le modèle moyenne-variance prévoit. Le chapitre 2 prouve que le fait d'ignorer cette asymétrie conduit à la sous-estimation du risque mesuré par la Valeur à Risque (VaR) ou par la Perte moyenne au-delà de la VaR. Le chapitre 3 restaure la crédibilité de la VaR comme mesure cohérente du risque dans un contexte pratique.[1]

Dans le premier Chapitre, nous examinons les problèmes liés à la modélisation du phénomène d'asymétrie de la dépendance selon lequel les rendements négatifs des actifs financiers sont plus dépendants que les rendements positifs. Premièrement, nous montrons analytiquement qu'un modèle multivarié du type GARCH ou a changement de régime avec des innovations normales ne peut pas reproduire la dépendance extrême. Nous proposons un modèle alternatif qui permet de la dépendance pour les rendements négatifs tout en gardant l'indépendance pour les rendements positifs. Ce modèle est appliqué aux marchés d'actions et d'obligations internationaux pour analyser leur structure de dépendance. Il est constitué d'un état normal dans lequel la dépendance est symétrique et d'un état asymétrique. Les résultats empiriques suggèrent que la dépendance entre les actifs internationaux du même type est plus large dans les deux états tandis que la dépendance entre les actions et les obligations reste assez faible même dans le même pays. La volatilité du taux d'échange apparaît comme un facteur influant sur la dépendance extrême. Nous analysons ensuite dans le cadre de l'asymétrie, le problème de la faible diversification du portefeuille international encore appelé le « home bias puzzle ». Les explications principales de ce phénomènes reposent sur

[1] Le premier chapitre de cette thèse a été écrit en collaboration avec René Garcia. Le troisième chapitre a été écrit en collaboration avec René Garcia et Eric Renault.

les deux premiers moments : les coûts de transactions affectent le rendement espéré pendant que le risque du taux de change et la corrélation entre actifs internationaux affectent la volatilité des rendements. Ang et Bekaert (2002) utilisent un modèle avec changement d'état markovien pour analyser l'effet de la corrélation asymétrique sur la diversification internationale et concluent qu'il est assez faible. Nous proposons ici une explication basée sur le troisième moment est proposée en montrant qu'il est lié à l'asymétrie de la dépendance. Nous montrons aussi que la même intuition explique le « flight to safety » : le fait d'aller un peu plus que prévu vers les actifs moins risqués.

Le deuxième chapitre est consacré à l'analyse des implications de l'asymétrie de la dépendance sur la gestion du risque extrême. Nous montrons qu'en présence de dépendance asymétrique, un modèle de portefeuille basé sur les modèles classiques de GARCH avec des innovations normales ou plus généralement symétrique conduit à une sous-estimation de la valeur à risque (VaR) du portefeuille de même que celle de la perte moyenne au-delà de la VaR. Ces mesures du risque ont tendance à croître dès lors que les rendement négatifs deviennent de plus en plus dépendants par rapport aux rendements positifs, les distributions marginales restants inchangées. Pour la précision dans l'estimation de ces mesures de risques extrêmes en situation de dépendance asymétrique, nous constatons de manière générale une supériorité de la copule de Gumbel qui prend en compte l'asymétrie observée dans la structure de dépendance.

Le but du troisième chapitre est de restaurer la crédibilité de la VaR comme mesure cohérente du risque dans un contexte pratique. Artzner et al (1999) mettent en exergue l'absence de la sous-additivité de la VaR requise pour être une mesure cohérente du risque. L'idée clef dans ce chapitre est que si l'épaisseur des queues de distributions est responsable de la violation de la sous-additivité, une utilisation appropriée de l'information conditionnelle pourrait rendre la VaR plus rationnelle pour la gestion décentralisée du risque. L'argumentation est triple. Premièrement, dès lors que les traders sont embauchés sur la base qu'ils possèdent sur leur segment de marché une information plus riche que le gestionnaire central, Ils doivent simplement respecter les contraintes prudentielles imposées par celui-ci pour que le contrôle de la VaR décentralisée fonctionne de façon cohérente. Deuxièmement, dans ce contexte de décentralisation, nous montrons que si le gestionnaire central a accès ex-post à la composition du portefeuille des traders individuels, il

pourra récupérer une bonne part de leur information privée. Cette composition du portefeuille peut être utilisée pour améliorer le backtesting dans le but de vérifier que la contrainte prudentielle a été respectée par les traders. Finalement, nous montrons que l'épaisseur exigée des queues pour violer la sous-additivité même pour les petits niveaux de probabilités, induit une situation tellement extrême qu'elle correspond à une information tellement faible que la perte espérée est infinie. Nous concluons dont que l'incohérence de la gestion décentralisée par la VaR caractérisée par l'absence de sous-additivité avec une information assez riche, est une exception et non une règle.

Mots clés: corrélation asymétrique, dépendance asymétrique, copule, dépendance des queues, GARCH, modèles à changement de régime, finance internationale, valeur à risque, « expected shortfall », DCC, mesures cohérente de risque, sous-additivité, distributions à queues épaisses, distributions stables.

Summary

We address in this dissertation the issue of asymmetric and nonlinear dependence modeling with its implications for international portfolio choice and risk management. We speak of asymmetric dependence when downside market returns are more dependent than upside market returns, a well established stylized fact. In the first chapter, we propose a new multivariate model to capture this fact and establish the link between this asymmetric dependence and coskewness to demonstrate how it can explain the lack of international diversification, while chapter 2 shows that by ignoring this asymmetry, one underestimates the risk measured with Value at Risk (VaR) or Expected Shortfall (ES). Chapter 3 restores the credibility of VaR as a coherent risk measure in a practical context.

In the first chapter, we examine the problems associated with modeling the stylized fact that asset returns are more dependent in bear markets than in bull markets, called asymmetric dependence. First, we analytically show that a multivariate GARCH or regime switching model with Gaussian innovations cannot reproduce extreme dependence. We propose an alternative model which allows tail dependence for lower returns and keeps tail independence for upper returns. This model is applied to international equity and bond markets to investigate their dependence structure. It includes one normal regime in which dependence is symmetric and a second regime characterized by asymmetric dependence. Empirical results suggest that the dependence between international markets in the same class of assets is large in both regimes, while equity and bond markets exhibit little dependence, even in the same country. Exchange rate volatility appears to be related to asymmetric extreme dependence. We also use this model to analyze the lack of international portfolio diversification known as the home bias puzzle. The previous explanations for this phenomenon are relied on the first two moments: transaction costs affect the expected return, while exchange rate risk and correlation between international assets affect the volatility. Ang and Bekaert (2002) with a regime changing correlation investigate the effect of asymmetric correlation on international diversification and conclude that the costs of ignoring regimes are small. We propose an explanation based on the third moment and using a stochastic dominance argument, we prove its link with dependence asymmetry. Using the same framework, we show that asymmetric

dependence amplifies the investment in the bonds, while reducing the investment in equities. This is another diversification phenomenon known as « flight to safety ».

We analyze in the second chapter the implications of the asymmetric dependence on the management of extreme risks. We show that in the presence of asymmetric dependence, a portfolio model based on a multivariate symmetric GARCH with Gaussian or Student-t innovations will lead to an underestimation of the portfolio value at risk (VaR) or expected shortfall. The latter will increase when negative returns become more dependent and positive returns less dependent, while the marginal distributions are left unchanged. In fact, we show that the strong dependence for low returns increases the downside risk and this additional risk cannot be captured by the Gaussian distribution. By introducing lower tail dependence, the Student-t distribution corrects this shortcoming of the Gaussian distribution. However, the symmetric property of the Student-t means also the same dependence in the upper tail and this will reduce the downside risk. The risk model that takes into account asymmetric dependence should allow lower tail dependence and upper tail independence as put forward by Longin and Solnik (2001). The Gumbel copula captures this asymmetry and shows superiority compared to Gaussian and student-t while combined with DCC in terms of accuracy of extreme risk measures.

The third chapter addresses the problem of credibility of VaR as a risk measure in a practical context. As stressed by Artzner et al. (1999), VaR may not possess the subadditivity property required to be a coherent measure of risk. The key idea of this chapter is that, when tail thickness is responsible for violation of subadditivity, eliciting proper conditioning information may restore a VaR rationale for decentralized risk management. The argument is threefold. First, since individual traders are hired because they possess richer information on their specific market segment than senior management, they just have to follow consistently the prudential targets set by senior management to ensure that decentralized VaR control will work in a coherent way. Second, in this decentralization context, we show that if senior management has access ex-post to the portfolio shares of the individual traders, it amounts to recovering some of their private information. These shares can be used to improve backtesting in order to check that the prudential targets have been enforced by the traders. Finally, we stress that tail thickness required to violate subadditivity, even for small probabilities, remains an extreme situation since it corresponds to such poor conditioning information that expected loss appears to be infinite. We then conclude that lack of coherency of decentralized VaR management, that is VaR non-subadditivity at the richest level of information, should be an exception rather than a rule. .

Key words: asymmetric correlation, asymmetric dependence, copula, tail dependence measures, GARCH, regime switching models, international finance, Value at Risk, expected shortfall, DCC, coherent measures of risk, subadditivity, heavy-tail distributions, stable distributions.

Contents

List of Tables

List of Figures

À ma femme, Yollande,

à mon fils, William,

à ma fille, Elka,

à mon père, David,

à ma mère, Marie,

à mes frères et soeurs

et à tous mes amis et camarades.

Remerciements

Je tiens ici à remercier de nombreuses personnes et institutions pour leur soutient et leur aide dans l'accomplissement de cette thèse. Tout d'abord mon directeur de recherche René Garcia qui a su orienter et guider mes recherches de manière fructueuse. Durant les années de préparation de cette thèse, il a su m'encourager, me prodiguer des conseils et surtout me soutenir dans les moments les plus difficiles. Sans son soutien, j'aurais sûrement abandonné.

Je voudrais remercier mon co-auteur Éric Renault pour sa collaboration significative dans le troisième chapitre de cette thèse. Ce dernier m'a beaucoup appris et je lui suis très reconnaissant.

Je voudrais aussi remercier les nombreux organismes qui m'ont soutenu par leur financement: Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ).

Je voudrais surtout remercier mon épouse Yollande pour tout son soutient, sa comprehension et ses encouragements tout au long de la redaction de cette thèse. Je remercie également mon fils William, ma fille Elka, mon père David, ma mère Marie, mes frères et sœurs sans oublier mes amis et camarades.

Je ne saurais terminer sans rendre grâce à l'éternel Dieu tout puissant sans qui cette thèse n'aurait jamais pu aboutir.

Introduction générale

La spécification de la distribution des rendements des actifs financiers revêt une importance capitale dans la gestion de portefeuilles et la gestion du risque. Dès lors, il est impératif de mieux comprendre le comportement stochastique de ces rendements, leurs caractéristiques et surtout la structure de leur dépendance.

Des travaux récents notamment de Longin et Solnik (2001) et Ang et Chen (2002) ont révélé l'existence d'une asymétrie dans la corrélation des rendements des actifs internationaux. En effet, ils ont constaté une forte corrélation entre des rendements pendant les périodes de détresse et une faible corrélation entre les rendements lors des booms sur les marchés internationaux. Poussant leur analyse plus loin au moyen de la théorie des valeurs extrêmes, Longin et Solnik (2001) se sont rendu compte que la corrélation était nulle pour les rendements extrêmement élevés, alors qu'elle restait strictement positive pour des rendements extrêmement bas. Bien que la théorie des valeurs extrêmes permette d'établir de façon précise l'existence de cette asymétrie, elle ne permet ni de la modéliser dans une distribution globale, ni de déterminer quel modèle est capable de la reproduire.

Le premier chapitre de cette thèse s'attaque d'abord à l'analyse des modèles classiques dans le but de voir s'ils sont capables ou non de reproduire cette asymétrie dans la dépendance. En établissant un lien entre la corrélation extrême et la dépendance des queues, nous montrons que les modèles à changement de régime et les modèles GARCH classiques la reproduisent partiellement ou ne la reproduisent pas du tout. En fait, les modèles GARCH classiques utilisent la loi normale ou la loi t de Student qui sont des lois symétriques et de ce fait, bien que la dynamique dans la volatilité soit capable de changer la dépendance dans les queues elle ne peut modifier la symétrie de cette dépendance tant que la moyenne reste constante dans le temps. Par contre, comme le constate Ang et Bekaert (2002), le modèle a changement de régime quant à lui reproduit de l'asymétrie dans la dépendance, mais cette asymétrie disparait lorsqu'on va plus loin dans les queues. Nous proposons alors un modèle basé sur les copules qui permet de prendre en compte l'asymétrie dans la dépendance.

Récemment introduites et aujourd'hui très utilisées dans la modélisation, les copules servent à relier les distributions univariées pour en construire des distributions multivariées. Elles définissent complètement la structure de dépendance et permettent d'aller au delà de la dépendance linéaire (corrélation) pour prendre en compte les nonlinéarités observées dans les faits.

En utilisant ces outils, nous avons pu construire un modèle à quatre variables, qui nous permet non seulement d'analyser le comportement de la corrélation mais aussi la dépendance dans les queues. Notre modèle comprend deux régimes de dépendance : la dépendance normale (ou Gaussienne) qui est caractérisée par une structure symétrique, et une dépendance dite « rotated Gumbel » qui présente une asymétrie avec une dépendance dans la queue basse et une indépendance dans la queue supérieure. Ces deux régimes alternent dans le temps selon un processus dit « markovien » qui n'est cependant pas observable. L'utilisation des copules nous permet d'autre part de modéliser séparément les distributions marginales et la structure de la dépendance. Ainsi, nous avons utilisé les modèles GARCH univariés pour modéliser chaque rendement pris individuellement.

Ce modèle nous a permis d'analyser le comportement des rendements des actions et des obligations sur les marchés internationaux. Pour cela, deux paires de pays ont été considérées. Le Canada et les Etats-Unis d'une part, la France et l'Allemagne d'autre part. Les résultats empiriques ont révélé une plus forte dépendance internationale entre les actifs du même type par rapport à la dépendance entre les obligations et les actions même lorsqu'on considère le même pays. D'autre part, nous avons mis en exergue une relation entre la volatilité du taux d'échange et l'asymétrie dans la dépendance. Ainsi, la dépendance entre la France et l'Allemagne s'est révélée très asymétrique avant l'introduction de la monnaie unique européenne (Euro), alors que cette asymétrie a considérablement diminué après l'introduction de cette monnaie.

L'application de ce modèle dans la diversification internationale de portefeuille a permis d'apporter un élément d'explication additionnelle aux phénomènes tels que la faible diversification internationale et la tendance des investisseurs à se rabattre sur les actifs moins risqués comme les obligations au détriment des actions. L'intuition derrière cette explication est qu'en présence d'une forte dépendance dans les moments de baisse sur les marchés, les bénéfices liés à la diversification diminuent du fait que l'effet de la diversification s'estompe au moment ou les investisseurs en ont le plus besoin.

Dans le second chapitre, nous analysons les implications de cette asymétrie de la dépendance sur les mesures de risques extrêmes. Notamment la valeur-à-risque (VaR) qui représente le niveau de perte maximale encourue sur une période déterminée et avec un seuil de confiance donné. Mais aussi sur la « Expected Shortfall » (ES) qui représente la perte moyenne lorsqu'elle dépasse la VaR. Nous montrons que lorsque cette asymétrie n'est pas prise en compte dans un modèle et que la loi Gaussienne est utilisée comme c'est souvent le cas dans la pratique, ces mesures de risques extrêmes sont sous-estimées.

Dans une analyse empirique, en utilisant le cadre du modèle DCC de Engle (2002) dans lequel nous introduisons différente structure de dépendance et en l'appliquant à des portefeuilles équipondérés d'actions américaines et canadiennes d'une part et d'obligations de ces deux mêmes pays d'autre part. Nous nous sommes rendu compte que bien que ce cadre général estime assez bien la VaR au niveau 5% pour toutes les structures de dépendance, il sous-estime cette mesure pour des niveaux beaucoup plus prudentiels (1%, 0.5%) lorsque l'asymétrie présente dans les données n'est pas prise en compte. Nous avons obtenu par contre un résultat inattendu bien que explicable pour ce qui est de l'estimation de la ES. En effet, les modèles de dépendance asymétrique comme la Gaussienne et la t de Student sous-estiment cette seconde mesure au niveau 5%, bien qu'ils aient donné plutôt une bonne estimation de la VaR à ce même niveau de couverture. Une explication peu être liée au fait que si ces distributions génèrent des portefeuilles avec des queues moins épaisses qu'elles devraient l'être, alors elles auraient tendance à sous-estimer la moyenne dans les queues. Pour la précision dans l'estimation des mesures de risques extrêmes en situation de dépendance asymétrique, nous avons constaté de manière générale une supériorité de la copule de Gumbel qui prend en compte l'asymétrie observée dans la structure de dépendance.

Le problème de la cohérence de la VaR comme mesure de risque est aborde dans le troisième chapitre. En effet Artzner et al (1999) ont formulé quatre propriétés qu'une mesure de risque doit satisfaire pour être considérée comme cohérente. L'une de ses propriétés notamment la sous-additivité n'est pas toujours vérifiée par la VaR. La pertinence de cette propriété repose sur le fait que la diversification est considérée comme un moyen de réduction du risque, cependant pour que ce soit le cas, il faudrait que la mesure de risque soit sous-additive. L'idée clef de notre démarche est que si l'épaisseur des queues de distributions est responsable de la violation de la sous-additivité, une utilisation appropriée de l'information conditionnelle pourrait rendre la VaR plus rationnelle pour la gestion décentralisée du risque. Nous développons une triple argumentation. Premièrement, partant du fait que les traders possèdent sur leur segment de marché une information dite privée plus riche que le gestionnaire central, Ils doivent simplement respecter les contraintes prudentielles imposées par celui-ci pour que le contrôle de la VaR décentralisée fonctionne de façon cohérente. Nous montrons par la suite, que dans ce contexte de décentralisation, si le gestionnaire central a accès ex-post à la composition du portefeuille des traders individuels, il pourra récupérer une bonne part de leur information privée. Cette composition du portefeuille peut être utilisée pour améliorer le backtesting dans le but de vérifier que la contrainte prudentielle a été respectée par les traders. Finalement, en utilisant les distributions à queues épaisses telles que les lois stables et les lois de type Pareto, nous montrons que l'épaisseur exigée des queues pour violer la sous-additivité même pour les petits niveaux de probabilités, induit une situation tellement extrême qu'elle correspond à une information tellement faible que la perte espérée est infinie. Nous concluons dont que l'incohérence de la gestion décentralisée par la VaR caractérisée par l'absence de sous-additivité avec une information assez riche, est une exception et non une règle, d'autant plus que dans la pratique, la moyenne conditionnelle ou inconditionnelle est en général supposée finie. En d'autres termes, sans remettre en cause la validité mathématique du résultat selon lequel la VaR peut dans certaines conditions violer la sous-additivité, nous recentrons le débat sur le plan pratique en montrant qu'elle est cohérente non seulement dans le cadre des lois normales, mais aussi dans un contexte plus général des distributions fréquemment utilisées dans la modélisation des valeurs extrêmes telles que les lois de type Pareto et les lois stables tant que la moyenne est finie.

Chapter 1

Dependence Structure and Extreme Comovements in International Equity and Bond Markets with Portfolio Diversification Effects

1. Introduction

Negative returns are more dependent than positive returns in financial markets, especially in international asset markets. This phenomenon known as asymmetric dependence has been reported by many previous studies including Erb et al (1994), Longin and Solnik (2001), Ang and Bekaert (2002), Ang and Chen (2002), Das and Uppal (2003), Patton (2004), and references therein. This asymmetric dependence has important implications for portfolio choice and risk management.[1] However, measuring and modeling asymmetric dependence remains a challenge.

Previous studies commonly use simple, dynamic or exceedance correlation to investigate the dependence structure between financial returns.[2] These measures are adequate for linear and especially when the returns are jointly normal or conditionally normal, a property which is rarely verified empirically, especially at high frequency. Boyer et al (1999) and Forbes and Rigobon (2002) remark that correlations estimated conditionally on high or low returns or volatility suffer from some conditioning bias. Correlation asymmetry may therefore appear spuriously if these biases are not accounted for. To avoid these problems, Longin and Solnik (2001) use extreme value theory (EVT) by focusing on the asymptotic value of exceedance correlation.[3] The benefit of EVT resides in the fact that the asymptotic result holds regardless of the whole distribution of returns. However, as emphasized by Longin and Solnik (2001), EVT cannot help to determine if a given return-generating process is able to reproduce the extreme asymmetric exceedance correlation observed in the data.

---

[1] Patton (2004) finds that the knowledge of asymmetric dependence leads to gains that are economically significant, while Ang and Bekaert (2002), in a regime switching setup, argue that the costs of ignoring the difference between regimes of high and low dependence are small, but increase with the possibility to invest in a risk-free asset.

[2] The exceedance correlation between two series of returns is defined as the correlation for a sub-sample in which the returns of both series are simultaneously lower (or greater) than the corresponding thresholds θ_1 and θ_2. Formally, exceedance correlation of variables X and Y at thresholds θ_1 and θ_2 is expressed by

$$Ex\_corr\,(Y,X;\theta_1,\theta_2) = \begin{cases} corr\,(X,Y\,|X \leq \theta_1, Y \leq \theta_2)\,, & \text{for } \theta_1 \leq 0 \text{ and } \theta_2 \leq 0 \\ corr\,(X,Y\,|X \geq \theta_1, Y \geq \theta_2)\,, & \text{for } \theta_1 \geq 0 \text{ and } \theta_2 \geq 0 \end{cases}$$

. Longin and Solnik (2001) use $\theta_1 = \theta_2 = \theta$, while Ang and Chen (2002) use $\theta_1 = (1+\theta)\overline{X}$ and $\theta_2 = (1+\theta)\overline{Y}$, where $\overline{X}$ and $\overline{Y}$ are the means of Y and X respectively.

[3] Extreme Value Theory (EVT) is used to characterize the distribution of a variable conditionally to the fact that its values are beyond a threshold, and the asymptotic distribution is obtained when this threshold tends to infinity.

This paper provides a first solution to this shortcoming. By using the concept of tail dependence instead of exceedance corelation, we are able to investigate which model can reproduce these empirical facts. The tail dependence coefficient can be seen as the probability of the worst event in one market given that the worst event occurs in another market. Contrary to exceedance correlation, the estimation of the tail dependence coefficient is not subject to the problem of choosing an appropriate threshold and the use of extreme value distributions such as the Pareto distribution. Another difference is that tail dependence is completely defined by the dependence structure and is not affected by variations in marginal distributions.

Thanks to the tail dependence formulation of asymptotic dependence, we establish important analytical results. We show that the multivariate GARCH or regime switching (RS) models with Gaussian innovations that have been used to address asymmetric dependence issues (see Ang and Bekaert, 2002, Ang and Chen, 2002, and Patton, 2004) cannot reproduce an asymptotic exceedance correlation. The key point is that these classes of models can be seen as mixtures of symmetric distributions and cannot produce asymptotically an asymmetric dependence. Of course this does not mean that at finite distance a mixture of these classes cannot produce some asymmetric dependence. The RS model of Ang and Chen (2002) offers a good example. However, the asymmetry put forward disappears asymptotically. When we go far in the tails, we obtain a similar dependence for the upper and lower tails. In RS models, extreme positive (or negative) returns are independent. Moreover, the asymmetry in this RS model comes from the asymmetry created in the marginal distributions with regime switching in the mean. Hence it is not separable from the marginal asymmetry or skewness. [4]

We propose an alternative model based on copulas that allows tail dependence for lower returns and keeps tail independence for upper returns as suggested by the findings of Longin and Solnik (2001). We apply this model to international equity and bond markets to investigate their dependence structure. It includes one normal regime in which dependence is symmetric and a second

[4] Ang and Chen (2002) conclude that even if regime-switching models perform best in explaining the amount of correlation asymmetry reflected in the data, these models still leave a significant amount of correlation asymmetry in the data unexplained.

regime characterized by asymmetric dependence. We separately analyze dependence between the two leading markets in North-America (US and Canada) and two major markets of the Euro zone (France and Germany). We further investigate the implications of this asymmetric dependence on international portfolio choice especially its ability to explain the home bias investment and flight to safety.

Copulas are functions that build multivariate distribution functions from their unidimensional marginal distributions. The theory of this useful tool dates back to Sklar (1959) and a clear presentation can be found in Nelsen (1999). Well designed to analyze nonlinear dependence, copulas were initially used by statisticians for nonparametric estimation and measure of dependence of random variables (see Genest and Rivest, 1993 and references therein). Their application to financial and economic problems is a new and fast-growing field of interest. Here, the use of this concept is essentially motivated by the fact that it allows to separate the features due to each marginal distribution from the dependence effect between all variables. This helps overcoming the curse of dimensionality associated with the estimation of models with several variables. For example, in multivariate GARCH models, the estimation becomes intractable when the number of series being modeled is high. The CCC of Bollerslev (1990), the DCC of Engle (2002), and the RSDC of Pelletier (2004) deal with this problem by separating the variance-covariance matrix in two parts, one part for the univariate variances of the different marginal distributions, another part for the correlation coefficients. This separation allows them to estimate the model in two steps. In the first step, they estimate the marginal parameters and use them in the estimation of the correlation parameters in a second step. Copulas offer a tool to generalize this separation while extending the linear concept of correlation to nonlinear dependence.

The empirical investigation shows that the dependence between equities and bonds is low even in the same country, while the dependence between international assets of the same type is large in both regimes. Extreme dependence appears across countries in both the bond and equity markets, but it is nonexistent across the bond and the equity markets, even in the same country. Another finding is that the correlation in the normal regime differs from the unconditional correlation.

This may be due to nonlinear dependence of international returns characterized by the presence of extreme dependence that is absent in the tail of a multivariate normal distribution. Exchange rate volatility seems to be a factor contributing to asymmetric dependence. With the introduction of a fixed exchange rate the dependence between France and Germany becomes less asymmetric and more normal than before. High exchange rate volatility is associated with a high level of asymmetry. These results are consistent with those of Cappiello, Engle and Sheppard (2003) who find an increase in correlation after the introduction of the Euro currency.

We use this model in a simple portfolio choice framework with a CRRA utility function involving the same categories of assets. We explore the implications in terms of diversification, both internationally (the home bias phenomenon) and domestically (the flight to safety phenomenon). The main result is that asymmetric dependence increases the downside risk and therefore, very risk averse investors tend to switch toward less risky assets when downside dependence increases. So, for a Canadian investor who holds US and Canadian bonds and equities, the share invested in Canada increases with the asymmetric dependence since the Canadian market in our sample is less risky. A similar behavior is observed for the bond and equity trade-off. In the asymmetric dependence regime, the very risk averse agent increases the fraction of its wealth in bonds.

The rest of this paper is organized as follows. Section 2 reformulates the empirical facts about exceedance correlation in terms of tail dependence and shows how classical GARCH or regime switching models fail to capture these facts. In section 3 we develop a model with two regimes that clearly disentangles dependence from marginal distributional features and allows asymmetry in extreme dependence. As a result, we obtain a model with four variables that features asymmetry and a flexible dependence structure. Empirical evidence on the dependence structure is examined in section 4, while section 5 analyses the implications of asymmetric dependence on international and domestic diversification. Conclusions are drawn in section 6.

2. Extreme Asymmetric Dependence and Modeling Issues

In this section we present empirical facts about exceedance correlation in international equity market returns put forward by Longin and Solnik (2001) and the related literature. We next argue that these facts can be equivalently reformulated in terms of tail dependence. The latter formulation will allow us to explain why classical return-generating processes such as GARCH and regime switching models based on a multivariate normal distribution fail to reproduce these empirical facts.

2.1 Empirical Facts

Longin and Solnik (2001) investigate the structure of correlation between various equity markets in extreme situations. Their main finding is that equity returns exhibit a high correlation in extreme bear markets and no correlation in extreme bull periods. They arrive at this conclusion by testing the equality of exceedance correlations, one obtained under a joint normality assumption and the other one computed using EVT. For the latter distribution, they model the marginal distributions of equity index returns with a generalized Pareto distribution (GPD) and capture dependence through a logistic function. Their analysis brings forward two important facts. First, there exists asymmetry in exceedance correlation, that is large negative returns are more correlated than large positive returns. Ang and Chen (2002) who develop a test statistic based on the difference between exceedance correlations computed from the data and those obtained from a given model.[5] They find as in Ang and Bekaert (2002) that regime switching models can reproduce the above fact. However, in their regime switching model, it is difficult to separate asymmetric dependence from marginal asymmetries or skewness in the marginal distributions.

[5] Ang and Chen (2002) define a test statistic $H = \left[\sum_{i=1}^{N} \frac{1}{N} \left(\rho(\vartheta_i) - \widehat{\rho}(\vartheta_i)\right)^2\right]^{1/2}$ which is the distance between exceedance correlations obtained from the normal distribution $(\rho(\vartheta_1), ..., \rho(\vartheta_N))$ and exceedance correlations estimated from the data $(\widehat{\rho}(\vartheta_1), ..., \widehat{\rho}(\vartheta_N))$ for a set of N selected thresholds $\{\vartheta_1, ..., \vartheta_N\}$. In the same way they define H^- and H^+ by considering negative points for H^- and nonnegative points for H^+ such that $H^2 = (H^-)^2 + (H^+)^2$. They can therefore conclude to asymmetry if H^- differs from H^+. Their results depend on the choice of the set of thresholds and can only account for asymmetry at finite distance but not asymptotically.

The second fact relates to exceedance correlation in the limit. Longin and Solnik find that exceedance correlation is positive and statistically different from zero for very large negative returns and not different from zero for very large positive returns.

We illustrate these facts and the capacity of models to reproduce them in Figure 1 with US and Canadian returns. We specify thresholds in term of quantiles: $\theta_1 = F_X^{-1}(\alpha)$ and $\theta_2 = F_Y^{-1}(\alpha)$ where F_X and F_Y are the cumulative distribution functions of Y and X respectively. Following Longin and Solnik (2001) and Ang and Chen (2002) exceedance correlations are symmetric if $Ex\_corr\,(Y,X;\theta_1) = Ex\_corr\,(Y,X;\theta_2)$; $\alpha \in (0,1)$. Correlations of return exceedances exhibit the typical shape put forward in Longin and Solnik (2001) for the US equity market with various European equity markets. For the models, we chose to retain the multivariate normal, as a benchmark case to show that correlations go to zero as we move further in the tails, as well a normal regime switching model, as in Ang and Chen (2002). The last model produces some asymmetry in correlations for positive and negative returns but not nearly as much as in the data. We also exhibit the exceedance correlations estimated from the model used by Longin and Solnik (2001). It is evidently much closer to the data. Finally, we also report the correlations obtained from a rotated Gumbel copula for the dependence function (see Appendix for a definition), with Gaussian marginal distributions. The graph is very close to the Longin and Solnik (2001) one.

Since asymptotic exceedance correlation is zero for both sides of a bivariate normal distribution, Longin and Solnik (2001) interpreted these findings as rejection of normality for large negative returns and non-rejection for large positive returns. In the conclusion of their article, Longin and Solnik stress that their approach has the disavantage of not explicitly specifying the class of return-generating processes that fail to reproduce these two facts.

We provide a first answer to this concern by characterizing some classes of models which cannot reproduce these asymmetries in extreme dependence. The difficulty in telling which model can reproduce these facts is the lack of analytical expressions for the asymptotic exceedance correlation and its intractability even for classical models such as Gaussian GARCH or regime switching models. In order to investigate this issue, we introduce the concept of tail dependence.

2.2 Tail Dependence

To measure the dependence between an extreme event on one market and a similar event on another market, we define two dependence functions one for the lower tail and one for the upper tail, with their corresponding asymptotic tail dependence coefficients. For two random variables X and Y with cumulative distribution functions F_X and F_Y respectively, we call the lower tail dependence function (TDF) the conditional probability $\tau^L(\alpha) \equiv \Pr\left[X \leq F_X^{-1}(\alpha) \,|Y \leq F_Y^{-1}(\alpha)\right]$ and similarly, the upper tail dependence function is $\tau^U(\alpha) \equiv \Pr\left[X \geq F_X^{-1}(1-\alpha) \,|Y \geq F_Y^{-1}(1-\alpha)\right]$ for $\alpha \in (0, 1/2]$.[6] The tail dependence coefficient (TDC) is simply the limit (when it exists) of this function when α tends to zero. More precisely *lower TDC* is $\tau^L = \lim\limits_{\alpha \to 0} \tau^L(\alpha)$ and *upper TDC* is $\tau^U = \lim\limits_{\alpha \to 0} \tau^U(\alpha)$. As in the case of joint normality, we have lower tail-independence when $\tau^L = 0$ and upper tail-independence for $\tau^U = 0$.

Compared to exceedance correlation used by Longin and Solnik (2001), Ang and Chen (2002), Ang and Bekaert (2002), and Patton (2004), one of the advantages of TDF and corresponding TDCs is their invariance to modifications of marginal distributions that do not affect the dependence structure. Figure 2 gives an illustration of this invariance. We simulate a bivariate Gaussian distribution $N(0, I_\rho)$, where I_ρ is the two dimensional matrix with standard deviations equal to one in all elements of the diagonal and $\rho = 0.5$ is the correlation coefficient outside the diagonal elements. Both exceedance correlation and tail dependence measures show a symmetric behavior of dependence in extreme returns. However, when we replace one of the marginal distributions $N(0,1)$ by a mixture of normals one $N(0,1)$ and one $N(4,4)$ with equal weights and let the other marginal distribution and the dependence structure unchanged, the TDF remains the same while the exceedance correlation is affected. In fact, the correlation coefficient and the exceedance correlation are a function of the dependence structure and of the marginal distributions while the tail dependence is a sole function of the dependence structure, regardless of the marginal distributions.

[6] In the literature (see Rodriguez, 2004 and references therein), only the limit of this function is considered. Here, we define the TDF for every $\alpha \in (0, 1over2]$ to make a comparison with conditional correlation, which is also a function of a threshold. The tail dependence measure is also related to the concept of lower (upper) orthant dependence concept (see Denuit and Scaillet, 2004).

Another disadvantage of exceedance correlation is that asymptotic exceedance correlation cannot be estimated without sample bias since fewer data points are available when we move further into the tails of the distribution. Longin and Solnik (2001) determine by simulation an optimal threshold and use the subsample beyond this threshold to estimate the asymptotic exceedance correlation. But this shortcoming does not compromise the results of Longin and Solnik (2001) since they choose different levels of threshold and still obtain the same result. With tail dependence, the estimation is done using all data points in the sample and the estimators of the tail coefficients are unbiased.

By observing that for the logistic function used by Longin and Solnik (2001), the zero value for the asymptotic correlation coefficient is exactly equivalent to tail independence, we can reformulate their asymptotic result as follow : lower extreme returns are tail-dependent, while upper extreme returns are tail-independent.[7]

This reformulation presents at least two main advantages. Compared to exceedance correlation, the tail dependence coefficient is generally easier to compute and analytical expressions can be obtained for almost all distributions. This is not the case for exceedance correlation even for usual distributions. Moreover, we can easily derive the tail dependence of a mixture from the tail dependence of the different components of the mixture. The last property will be used below to investigate which model can or cannot reproduce the results of Longin and Solnik (2001).

2.3 Why classical multivariate GARCH and RS model cannot reproduce asymptotic asymmetries?

Ang and Chen (2002) and Ang and Bekaert (2002) try to reproduce asymmetric correlations facts with classical models such as GARCH and RS based on a multivariate normal distribution. After examining a number of models, they found that GARCH with constant correlation and fairly asymmetric GARCH cannot reproduce the asymmetric correlations documented by Longin and Solnik. However, they found that a RS model with Gaussian innovations is better at reproducing

[7] For the logistic function with parameter α, the correlation coefficient of extremes is $1-\alpha^2$ (see Longin and Solnik, 2001). We find that the upper tail dependence coefficient is $2-2^{\alpha}$. Then, both coefficients are zero when α equals 1 and different from zero when α is different from 1.

asymmetries in exceedance correlation. They clearly reproduce asymmetric correlations at finite distance, however the asymptotic asymmetric dependence put forward in Longin and Solnik (2001). Their finite distance asymmetric correlation comes from the asymmetries produced in the marginal distributions with a regime switching in means.[8] Therefore it becomes difficult to distinguish asymmetries in dependence from asymmetry in marginal distributions.

By reinterpreting Longin and Solnik (2001) results in term of TDC instead of asymptotic exceedance correlation, we show analytically that all these models cannot reproduce asymptotic asymmetry even if some can reproduce finite distance asymmetry. These results are extended to the rejection of more general classes of return-generating processes. The key point of this result is the fact that many classes of models including Gaussian(or Student) GARCH and RS can be seen as mixtures of symmetric distributions. We establish the following result.

Proposition 2.1:

(i) *Any GARCH model with constant mean and symmetric conditional distribution has a symmetric unconditional distribution and hence has a symmetric TDC.*

(ii) *If the conditional distribution of a RS model has a zero TDC, then the unconditional distribution also has a zero TDC.*

(iii) *From a multivariate distribution with symmetric TDC, it is impossible to construct an asymmetric TDC with a mixture procedure (as GARCH, RS or any other) by keeping all marginal distributions unchanged across mixture components.*

Proof: see Appendix A.

This proposition allows us to argue that the classical GARCH or RS models cannot reproduce asymmetries in asymptotic tail dependence. Therefore, the classical GARCH models (BEKK, CCC

[8] Ang and Bekaert (2002) note that the ability of RS model (compared to GARCH model) to reproduce asymmetries, derives from the fact that it accounts the persistence in both first and second moments while the GARCH accounts this persistence only in second moments. We give analytical arguments to this intuition.

or DCC) with constant mean can be seen as a mixture of symmetric distributions with the same first moments and therefore exhibit a symmetric tail dependence function as well as a symmetric TDC.[9] When the mean becomes time-varying as in the GARCH-M model the unconditional distribution can allow asymmetry in correlation (Ang and Chen, 2002), but this asymmetry comes from the mixture of the marginal distributions. The resulting skewness cannot be completely disentangled from the asymmetric correlation, since correlations are affected by marginal changes. Similarly, the classical RS model with Gaussian innovations is a discrete mixture of normal distributions which has a TDC equal to zero on both sides. Therefore, by (ii) we argue that both its TDCs are zero. However, at finite distance, when the mean changes with regimes, the exceedance correlation is not symmetric. This asymmetry is found by Ang and Chen (2002) and Ang and Bekaert (2002) in their RS model, but it disappears asymptotically and it comes from the asymmetry created in the marginal distributions by regime switching in means. Hence, the asymmetries in correlation are not separable from the marginal asymmetry, exactly like in the GARCH-M case. The part (iii) of proposition 2.1 extends this intuition in terms of more general multivariate mixture models based on symmetric innovations. Actually when the marginal distributions are the same across all symmetric TDC components of a mixture, it is impossible to create asymmetry in TDCs.

Two relevant questions arise from the above discussion. First, how can we separate the marginal asymmetries from the asymmetry in dependence? Second, how can we account not only for asymmetries at finite distance but also for asymptotic dependence? In the next section, we propose a flexible model based on copulas that addresses these two issues.

3. A Copula Model for asymmetric dependence

Our model aims at capturing the type of asymmetric dependence found in international equity returns. Our discussion in the last section showed that it is important to disentangle the marginal

[9]The BEKK proposed by Engle and Kroner (1995) is the straightforward generalization of the GARCH model to a multivariate case which guarantees positive definiteness of the conditional variance-covariance matrix. In the CCC model proposed by Bollerslev (1990) the conditional variance-covariance matrix is assumed constant, while in the DCC of Engle (2002) this matrix is dynamic.

distributions from the dependence structure. Therefore, we need to allow for asymmetry in tail dependence, regardless of the possible marginal asymmetry or skewness. Copulas, also known as dependence functions, are an adequate tool to achieve this aim.

3.1 Disentangling the marginal distributions from dependence with copulas

Estimation of multivariate models is difficult because of the large number of parameters involved. Multivariate GARCH models are a good example since the estimation becomes intractable when the number of series being modeled is high. The CCC of Bollerslev (1990), the DCC of Engle (2002), and the RSDC of Pelletier (2004) deal with this problem by separating the variance-covariance matrix into two parts, one for the univariate variances of the different marginal distributions, the other for the correlation coefficients. This separation allows them to estimate the model in two steps. In the first step, they estimate the marginal parameters and use them in the estimation of the correlation parameters in a second step. Copulas offer a tool to generalize this separation while extending the linear concept of correlation to nonlinear dependence.

Copulas are functions that build multivariate distribution functions from their unidimensional margins. Let $X \equiv (X_1, ..., X_n)$ be a vector of n univariate variables. Denoting F the joint n-dimensional distribution function and $F_1, ..., F_n$ the respective margins of $X_1, ..., X_n$. Then the Sklar theorem states that there exists a function C called copula which joins F to $F_1, ..., F_n$ as follows.[10]

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \tag{3.1}$$

This relation can be expressed in term of densities by differentiating with respect to all arguments. We can therefore write (2.1) equivalently as

$$f(x_1, ..., x_n) = c(F_1(x_1), ..., F_n(x_n)) \times \prod_{i=1}^{n} f_i(x_i) \tag{3.2}$$

where f represents the joint density function of the n-dimensional variable X and f_i the density function of the variable X_i for $i = 1, ..., n$. The copula density function is naturally defined by

[10]See Nelsen (1999) for a general presentation. Note that if F_i is continuous for any $i = 1, ..., n$ then the copula C is unique.

$c(u_1,...,u_n) \equiv \frac{\partial^n}{\partial u_1 ... \partial u_n} C(u_1,...,u_n)$. Writing the joint distribution density in the above form, we understand why it can be said that copula contains all information about the dependence structure.[11]

We now suppose that our joint distribution function is parametric and we separate the marginal parameters from the copula parameters. So the relation (3.2) can be expressed as:

$$\begin{aligned} f(x_1,...,x_n;\delta,\theta) &= c(u_1,...,u_n;\theta) \times \prod_{i=1}^{n} f_i(x_i;\delta_i); \\ u_i &= F_i(x_i;\delta_i) \quad \text{for } i=1,...,n \end{aligned} \tag{3.3}$$

where $\delta = (\delta_1,...,\delta_n)$ are the parameters of the different margins and θ denotes the vector of all parameters that describe dependence through the copula. Therefore, copulas offer a way to separate margins from the dependence structure and to build more flexible multivariate distributions.

More recent work allow some dynamics in dependence. In a bivariate context, Rodriguez (2004) introduces regime switching in both theparameters of marginal distributions and the copula function.[12] Ang and Bekaert (2002; 2004) allow all parameters of the multivariate normal distribution to change with the regime. The extension of these models to a large number of series faces the above-mentioned curse of dimensionality. Since the switching variable is present in both the margins and the dependence function, separation of the likelihood function into two parts is not possible and the two-step estimation cannot be performed. Pelletier (2004) uses the same separation as in the CCC or DCC and introduces the regime switching variable only in the correlation coefficients. By doing so, he can proceed with the two-step procedure to estimate the model while limiting the number of parameters to be estimated.[13] We carry out a similar idea but for nonlinear dependence.

Therefore, we separate the modeling of marginal distributions from the modeling of dependence

[11] The tail dependence coefficients are easily defined through copula as $\tau^L = \lim_{\alpha \to 0} \frac{C(\alpha,\alpha)}{\alpha}$ and $\tau^U = \lim_{\alpha \to 0} \frac{2\alpha - 1 + C(1-\alpha, 1-\alpha)}{\alpha}$

[12] The models proposed by Rodriguez (2004) in his analysis of contagion can reproduce asymmetric dependence but it cannot distinguish between skewness and asymmetry in the dependence structure. In fact, a change in regime produces both skewness and asymmetric dependence, two different features that must be characterized separately.

[13] Since Pelletier (2004) uses the normal distribution with constant mean, the resulting unconditional distribution is symmetric and cannot reproduce asymmetric dependence.

by using univariate GARCH models for the marginal distributions and introducing changes in regime in the copula dependence structure. The pattern of the model with four variables (two countries, two markets in our following application) is illustrated in Figure 3a.

3.2 Specification of the Marginal Distributions

For marginal distributions, we use a M-GARCH (1,1) model similar to Heston-Nandi (2000):

$$x_{i,t} = \mu_i + \lambda_i \sigma_{i,t}^2 + \sigma_{i,t} z_{i,t}; \qquad z_{i,t} \sim N(0,1); \quad i = 1, \cdots, 4 \tag{3.4}$$

$$\sigma_{i,t}^2 = \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i \left(z_{i,t-1} - \gamma_i \sigma_{i,t-1} \right)^2. \tag{3.5}$$

The variables $x_{1,t}$ and $x_{2,t}$ represent the log returns of equities and bonds respectively for the first country while $x_{3,t}$ and $x_{4,t}$ are the corresponding series for the second country; $\sigma_{i,t}^2$ denotes the conditional variance of $x_{i,t}$, λ_i can be interpreted as the price of risk and γ_i captures potential asymmetries in the volatility effect.[14] In the Heston-Nandi (2000) interpretation, μ_i represents the interest rate.[15] The parameters of the marginal distributions are grouped into one vector $\delta \equiv (\delta_1, \cdots, \delta_4)$, with $\delta_i = (\mu_i, \lambda_i, \omega_i, \beta_i, \alpha_i, \gamma_i)$.

3.3 Specification of the Dependence Structure

Our dependence model is characterized by two regimes, one Gaussian regime in which the dependence is symmetric (C_N) and a second regime that can capture the asymmetry in extreme dependence (C_A). The conditional copula is given by:

$$C\left(u_{1,t}, ..., u_{4,t}; \rho^N, \rho^A \,|s_t\right) = s_t C_N \left(u_{1,t}, ..., u_{4,t}; \rho^N\right) + (1 - s_t)\, C_A \left(u_{1,t}, ..., u_{4,t}; \rho^A\right), \tag{3.6}$$

where $u_{i,t} = F_{it}(x_{i,t}; \delta_i)$, with F_{it} denoting the conditional cumulative distribution function of $x_{i,t}$ given the past observations. The variable s_t follows a Markov chain with a time-invarying

[14] The condition $\beta_i + \alpha_i \gamma_i^2 < 1$ is sufficient to have the stationarity of the process $x_{i,t}$ with finite unconditional mean and variance (see Heston and Nandi, 2000).

[15] Here we keep μ_i as a free parameter to give more flexibility to our model.

transitional probability matrix

$$M = \begin{pmatrix} P & 1-P \\ 1-Q & Q \end{pmatrix}; P = \Pr\left(s_t = 1 \,|s_{t-1} = 1\right) \text{ and } Q = \Pr\left(s_t = 0 \,|s_{t-1} = 0\right) \tag{3.7}$$

The normal regime ($s_t = 1$) corresponds to the symmetric regime where the conditional joint normality can be supported and the asymmetric regime ($s_t = 0$) corresponds to the asymmetric regime in which markets are strongly more dependent for large asymptotic negative returns than for large positive returns.

The Gaussian copula C_N is defined straightforwardly by (2.1) where the joint distribution $F = \Phi_{\rho^N}$ is the 4-dimensional normal cumulative distribution function with all diagonal elements of the covariance matrix equal to one, i.e. $C_N\left(u_1, ..., u_4; \rho^N\right) = \Phi_{\rho^N}\left(\Phi^{-1}\left(u_1\right), ..., \Phi^{-1}\left(u_4\right)\right)$, where Φ is the univariate standard normal cumulative distribution function.

The asymmetric components of the copula are illustrated in figure 3b. The first one is characterized by independence between the two countries, but possibly extreme dependence between equities and bonds for each country. The second one is characterized by independence between equity and bond markets but allows for extreme dependence between equity returns and bond returns separately. The third one allows for possible extreme dependence between bonds in one country and equities in another country but supposes independence for the rest.

Formally the asymmetric copula is the mixture of these three components and is expressed as follows

$$\begin{aligned} C_A\left(u_1, ..., u_4; \rho^A\right) \equiv \; & \pi_1 C_{GS}\left(u_1, u_2; \tau_1^L\right) \times C_{GS}\left(u_3, u_4; \tau_2^L\right) \\ & + \pi_2 C_{GS}\left(u_1, u_3; \tau_3^L\right) \times C_{GS}\left(u_2, u_4; \tau_4^L\right) \\ & + \left(1 - \pi_1 - \pi_2\right) C_{GS}\left(u_1, u_4; \tau_5^L\right) \times C_{GS}\left(u_2, u_3; \tau_6^L\right) \end{aligned} \tag{3.8}$$

with $\rho^A = \left(\pi_1, \pi_2, \tau_1^L, \tau_2^L, \tau_3^L, \tau_4^L, \tau_5^L, \tau_6^L\right)$, and the bivariate component is the Gumbel survival copula given by

$$C_{GS}\left(u,v;\tau^{L}\right)=u+v-1+\exp\left[-\left(\left(-\log\left(1-u\right)\right)^{\theta\left(\tau^{L}\right)}+\left(-\log\left(1-v\right)\right)^{\theta\left(\tau^{L}\right)}\right)^{1/\theta\left(\tau^{L}\right)}\right], \quad (3.9)$$

where $\theta\left(\tau^{L}\right)=\dfrac{\log(2)}{\log\left(2-\tau^{L}\right)}$, $\tau^{L}\in[0,1)$ is the lower TDC and the upper TDC is zero.[16]

One can notice that our asymmetric copula specification assumes some constraints in the dependence structure. For three different couples from different components of this copula, the sum of their TDC is lower than one.[17] Without any constraints this sum may reach 3. Such constraints are dictated by some copula limitations.[18] A major problem in multivariate distributions' construction today and perhaps the most important open question concerning copulas as mentioned by Nelsen (1999, page 86) is how to construct multivariate copulas with specific bivariate marginal distributions. A theorem by Genest et al. (1995) states that it is not always possible to construct multivariate copulas with given bivariate margins. Therefore, even if in the bivariate case we can have a nice asymmetric copula with lower tail dependence and upper tail independence as Longin and Solnik (2001) suggest, some problems remain when we contemplate more than two series. Most existing asymmetric tail dependent copulas are in the family of archimedian copulas and the usual straightforward generalization in multivariate copulas constrains all bivariate marginal copulas to be the same. This is clearly not admissible in the context of our analysis. In the above model, we allow each of the six couples of interest to have different levels of lower TDC. As C_A is constructed, it is easy to check that it is a copula since each component of the mixture is a copula and the mixture of copulas is a copula.[19]

It is important to notice that, in this model, the labeling of each regime is defined ex-ante. The

[16]The Longin and Solnik (2001) result implies that lower tails are dependent while upper tails are independent. Hence, the Gumbel survival copula is designed to model this feature since it has this tail dependence structure.

[17]For example, the TDC between bonds and equities in the first country is $\pi_1\tau_1^L$, between equities of two countries $\pi_2\tau_3^L$, and between equities in the first country and bonds in the second country $(1-\pi_1-\pi_2)\,\tau_5^L$. Therefore, the sum is $\pi_1\tau_1^L+\pi_2\tau_3^L+(1-\pi_1-\pi_2)\,\tau_5^L\leq 1$, since $\tau_1^L\leq 1$, $\tau_3^L\leq 1$, and $\tau_5^L\leq 1$.

[18]This model can be generalized in the same way to a copula of any dimension. The same type of restrictions are applied, but we obtain a copula with a more flexible dependence structure.

[19]A copula can be seen as the cdf of a multidimensional variable with uniform $[0,1]$ margins. If we consider two bivariate independent variables with uniform margins the copula linking the four variables is simply the product of the corresponding bivariate copulas. Hence, such a product is always a copula.

normal regime ($s_t = 1$) corresponds to the symmetric regime where the conditional joint normality can be supported and the asymmetric regime ($s_t = 0$) corresponds to the asymmetric regime in which markets are strongly more dependent for large negative returns than for large positive returns.

3.4 An adapted parsimonious model

Given our application, we impose an additional constraint: $\pi_1 + \pi_2 = 1$. This means that we neglect the asymmetric cross-dependence between equities in one country and bonds in another country, which seems like an economically reasonable assumption given that we maintain cross-dependence through the normal regime. The mixed copula becomes.

$$\begin{aligned} C_A\left(u_1, ..., u_4; \rho^A\right) \equiv \; & \pi C_{GS}\left(u_1, u_2; \tau_1^L\right) \times C_{GS}\left(u_3, u_4; \tau_2^L\right) \\ & +(1-\pi) C_{GS}\left(u_1, u_3; \tau_3^L\right) \times C_{GS}\left(u_2, u_4; \tau_4^L\right) \end{aligned} \tag{3.10}$$

Therefore, the asymmetry copula is now characterized by just five parameters $\rho^A = \left(\pi, \tau_1^L, \tau_2^L, \tau_3^L, \tau_4^L\right)$.

3.5 Estimation

As already mentioned, our structure allows for a two-step estimation procedure. The likelihood function must be evaluated unconditionally to the unobservable regime variable s_t and decomposed in two parts. Let us denote the sample of observed data by $\underline{X}_T = \{X_1, \cdots, X_T\}$ where $X_t \equiv \{x_{1,t}, \cdots, x_{4,t}\}$. The log likelihood function is given by:

$$L\left(\delta, \theta; \underline{X}_T\right) = \sum_{t=1}^{T} \log f\left(X_t; \delta, \theta | \underline{X}_{t-1}\right) \tag{3.11}$$

where $\underline{X}_{t-1} = \{X_1, ...X_{t-1}\}$ and θ is a vector including the parameters of the copula and the transition matrix. Hamilton (1989) describes a procedure to perform this type of evaluation[20]. With $\xi_t = (s_t, 1 - s_t)'$ and denoting

[20] A general presentation can be found in Hamilton (1994, chapter 22).

$$\eta_t = \begin{bmatrix} f\left(X_t;\delta,\theta|\,\underline{X}_{t-1},s_t=1\right) \\ f\left(X_t;\delta,\theta|\,\underline{X}_{t-1},s_t=0\right) \end{bmatrix} \tag{3.12}$$

the density function conditionally to the regime variable s_t and the past returns can be written as:

$$f\left(X_t;\delta,\theta|\,\underline{X}_{t-1},s_t\right) = \xi_t'\eta_t \tag{3.13}$$

Since s_t (or ξ_t) is unobservable, we integrate on s_t and obtain the unconditional density function:

$$\begin{aligned} f\left(X_t;\delta,\theta|\,\underline{X}_{t-1}\right) = \quad & \Pr\left[s_t=1|\underline{X}_{t-1};\delta,\theta\right] \times f\left(X_t;\delta,\theta|\,\underline{X}_{t-1},s_t=1\right) \\ & +\Pr\left[s_t=0|\underline{X}_{t-1};\delta,\theta\right] \times f\left(X_t;\delta,\theta|\,\underline{X}_{t-1},s_t=0\right) \end{aligned} \tag{3.14}$$

The conditional probabilities of being in different regimes at time t conditional on observations up to time $t-1$, denoted by $\widehat{\xi}_{t|t-1} \equiv (\Pr\left[s_t=1|\underline{X}_{t-1};\delta,\theta\right], \Pr\left[s_t=0|\underline{X}_{t-1};\delta,\theta\right])'$, are computed through the Hamilton filter. Starting with the initial value $\widehat{\xi}_{1|0}$, the optimal inference and forecast for each date in the sample is given by the iterative equations:

$$\widehat{\xi}_{t/t} = \left[\widehat{\xi}'_{t|t-1}\eta_t\right]^{-1}\left(\widehat{\xi}_{t|t-1}\odot\eta_t\right) \tag{3.15}$$

$$\widehat{\xi}_{t+1/t} = M'.\widehat{\xi}_{t|t} \tag{3.16}$$

where $\odot$ denotes element-by-element multiplication. Finally, the unconditional density can be evaluated with the observed data as $f\left(X_t;\delta,\theta|\,\underline{X}_{t-1}\right) = \widehat{\xi}'_{t|t-1}\eta_t$ and the log likelihood becomes:

$$L\left(\delta,\theta;\underline{X}_T\right) = \sum_{t=1}^{T}\log\left(\widehat{\xi}'_{t|t-1}\eta_t\right) \tag{3.17}$$

To perform the two-step procedure, we decompose the log likelihood function into two parts: the first part includes the likelihood functions of all margins, while the second part represents the likelihood function of the copula.

Proposition 3.2 (Decomposition of the log likelihood function) *The log likelihood function can be decomposed into two parts including the margins and the copula*

$$L(\delta, \theta; \underline{X}_T) = \sum_{i=1}^{4} L_i(\delta_i; \underline{X}_{i,T}) + L_C(\delta, \theta; \underline{X}_T) \tag{3.18}$$

where

$$\begin{aligned} \underline{X}_{i,t} &= \{x_{i,1}, ..., x_{i,t}\}; \\ L_i(\delta_i; \underline{X}_{i,T}) &= \sum_{t=1}^{T} \log f_i(x_{i,t}; \delta_i | \underline{X}_{i,t-1}) \\ L_C(\delta, \theta; X) &= \sum_{t=1}^{T} \log\left(\widehat{\xi}'_{t|t-1} \eta_{ct}\right) \end{aligned}$$

with

$$\eta_{ct} = \begin{bmatrix} c(u_{1,t}(\delta_1), ..., u_{n,t}(\delta_n); \theta | s_t = 1) \\ c(u_{1,t}(\delta_1), ..., u_{n,t}(\delta_n); \theta | s_t = 0) \end{bmatrix}; \quad u_{i,t}(\delta_i) = F_i(x_{i,t}; \delta_i | \underline{X}_{i,t-1})$$

and $\widehat{\xi}'_{t|t-1}$ *filtered from* η_{ct} *as*

$$\begin{aligned} \widehat{\xi}_{t/t} &= \left[\widehat{\xi}'_{t|t-1} \eta_{ct}\right]^{-1} \left(\widehat{\xi}_{t|t-1} \odot \eta_{ct}\right) \\ \widehat{\xi}_{t+1/t} &= M'.\widehat{\xi}_{t|t} \end{aligned}$$

Proof: see Appendix A.

Several options are available for the estimation of the initial value $\widehat{\xi}_{1|0}$. One approach is to set it equal to the vector of unconditional probabilities, which is the stationary transitional probability of the Markov chain. Another simple option is to set $\widehat{\xi}_{1|0} = N^{-1}1_N$. Alternatively it could be considered as another parameter, which will be estimated subject to the constraint that $1'_N\widehat{\xi}_{1|0} = 1$. We will use the first option here.

Through the above decomposition, we notice that each marginal log likelihood function is separable from the others. Therefore, even if the estimation of all margins is performed in a first step, we can estimate each set of marginal parameters separately into this step. The first step is then equivalent to n single estimations of univariate distributions. The two-step estimation is formally written as follows:

$$\widehat{\delta} = \underset{\delta=(\delta_1,...,\delta_4)\in\Delta}{\arg\max} \sum_{i=1}^{4} L_i\left(\delta_i; X_{i,.}\right) \tag{3.19}$$

$$\widehat{\theta} = \underset{\theta\in\Theta}{\arg\max} L_C\left(\widehat{\delta},\theta; X\right) \tag{3.20}$$

The estimator for the parameters of the marginal distributions is then $\widehat{\delta} = \left(\widehat{\delta}_1, ..., \widehat{\delta}_4\right)$, with $\widehat{\delta}_i = \left(\widehat{\mu}_i, \widehat{\lambda}_i, \widehat{\omega}_i, \widehat{\beta}_i, \widehat{\alpha}_i, \widehat{\gamma}_i\right)'$; and $\widehat{\theta} = \left(\widehat{\rho}^N; \widehat{\rho}^A; \widehat{P}; \widehat{Q}\right)$ includes all the estimators of the parameters involved in the dependence structure. Δ and Θ represent the sets of all possible values of δ and θ respectively.

3.6 Testing asymmetry in dependence

We want to test the hypothesis $H_0 : (P = 1$ and $Q = 0)$ where P and Q are the parameters of the transition probability matrix. The natural way to evaluate whether dependence is asymmetric is to test the null hypothesis of one normal copula regime against the alternative hypothesis of two-copula regimes including the normal one and the asymmetric one. This test faces many irregularity problems. Under the null hypothesis, some nuisance parameters are unidentified and the scores are identically zero. These are the general problems of testing in RS models. Hansen (1996) describes the asymptotic distributions of standard test statistics in the context of regression models with additive nonlinearity. Garcia (1998) and Hansen (1992) provide the asymptotic null distribution of the likelihood ratio test. Andrews and Ploberger (1993) address the first problem in a general context and derive an optimal test. The above procedures solve the problem of unidentified nuisance parameters under the null and the identically zero scores. However, there is an additional problem of testing parameter on the boundary. Andrews (2001) deals with this boundary problem but in the absence of the first two problems.

Maximized Monte Carlo (MMC) tests of Dufour (2005), which are a generalization of classical Monte Carlo (MC) tests of Dwass (1957) and Barnard (1963), are adapted for tests facing all these problems. The MC tests of Dwass (1957) and Barnard (1963) are performed by doing many replications (with the same size as the sample data) under the null hypothesis, and compute the test statistic for each replication. The distribution of the test statistic is therefore approximated by the distribution of the obtained values. One can therefore compute the value of the test statistic with the data and deduce from the MC distribution the p-value of the test. The classical MC test does not deal with the presence of nuisance parameters under the null hypothesis. The MMC of Dufour (2005) addresses the problem of nuisance parameters under the null. When the tests statistic involve the nuisance parameters as in the case of the likelihood ratio test under the alternative, the values of these parameters are needed to compute the test statistic on simulated data. The MMC technique is the maximization of the p-values given all the possible values of the nuisance parameters. This test is computationally very demanding. However, Dufour (2005) proposes a simplified version which focus on the estimated values of the nuisance parameters and shows that it works under the assumptions of uniform continuity, and convergence over the nuisance parameter space. Our model satisfies these assumptions of uniform continuity and convergence. Therefore, we can apply this simpler version also known as parametric bootstrap test.

4. Dependence structure in international bond and equity markets: an empirical investigation

4.1 Data

We will consider the same model for two pairs of two countries. First, we model the equity and bond markets in the United States and Canada. The US equity returns are based on the SP 500 index, while the Canadian equity returns are computed with the Datastream index. The bond series are indices of five-year government bonds computed by Datastream. These bond indices are available daily and are chain linked allowing the addition and removal of bonds without affecting the value of the index.

We also consider France and Germany as a pair of countries. An additional interest here will be to see how the introduction of the European common currency changed the dependence structure between the asset markets in these two countries. The bond indices are the Datastream five-year government bond indices, while the equity indices are the MSCI series.

All returns are total returns and are expressed in US dollars on a weekly basis from January 01, 1985 to December 21, 2004, which corresponds to a sample of 1044 observations. Descriptive statistics for these bond and equity series are reported in Table 1.

Sharpe ratios appear to be of the same magnitude for both equities and bonds, around 0.6 in average for the first and slightly above 1 for the second. The United States exhibits the highest ratios among the four countries. All return series present negative skewness except for the French bond index. Both mean returns and return volatility are higher in France and Germany than in the US and Canada. The volatility of returns in France and Germany is more than 23%, while it is only 18% in the US and Canada.

Unconditional for all eight series are reported in Table 2. The US and Canadian markets exhibit relatively high correlations, 0.72 for equities and 0.5 for bonds. The same is true for the France-Germany pair, although the bond markets are tightly linked, with a correlation of 0.94. The North-American equity markets are less correlated with European equity markets (around 0.2) than their bond counterparts (around 0.32). The cross-correlations between equity and bond markets vary from one country to the other. In average the two markets seem to move independently in the United States, while they are more closely related in Canada (0.44) and in Europe (around 0.3 for both France and Germany). Cross-correlations between equities and bonds in two different countries are negligible, justifying our model specification.

4.2 Marginal distributions

The estimates of the marginal parameters are reported in Table 3. The large values for the β_i parameters (arround 90%) capture the high persistence in volatility. The values of the parameters α are close to zero and not significantly different from zero at the 5% level. However, the high degree of significance for the parameter λ indicates that asset returns are skewed.

One assumption for these GARCH models is that the error terms are i.i.d. Therefore, to verify if the assumption is fulfilled, we perform some tests of independence on the residuals. The test results in Table 4 suggest that the independence assumption of residuals cannot be rejected for all series with a good degree of confidence.

4.3 Dependence structure in bond and equity markets

Three main conclusions emerge from the empirical results. First, there appears to be a large extreme cross-country dependence in both markets, while there is little dependence between equities and bonds in the same country. Second, the dependence structure exhibits a strong nonlinearity. Third, there seems to be a link between exchange rate volatility and asymmetry of dependence.

4.3.1 US-Canada Dependence Structure

In Table 5, we report the results of estimating the dependence model described in section (3.4). The cross-country extreme dependence is large in both equity and bond markets, but the dependence across the two markets is relatively low in both countries. In the asymmetric regime, the TDCs are larger than 54 % in both bond-bond and equity-equity markets, while both equity-bond TDCs in US and Canada are lower than 2%. This observation has an important implication for international diversification. The fact that extreme dependence in international equity and bond markets is larger than national bond-equity dependence can have a negative effect on the gain of international diversification and encourage the switching from equity to the domestic bond or risk-free asset in case of bear markets.

The average absolute value of correlation in the normal regime is larger than 39% for cross-country dependence and lower than 41% for equity-bond dependence. In the last case the correlation between bonds and equities in Canada is unusually high. The results underline the differences between unconditional correlation and the correlation in the normal regime. In fact, the presence of extreme dependence in the negative returns explains this difference since the multivariate Gaussian distribution has independence in the tails of returns regardless of the level of correlation.

The separation of the distribution into two parts, including the normal regime and the asymmetric regime, allows to capture the strong nonlinear pattern in the dependence structure. Moreover, it is interesting to see that for a high unconditional correlated couple such as the US and Canada equity markets, this separation gives not only an extreme dependence for the asymmetric regime, but also a high correlation in the normal regime (87 %) that appears larger than the unconditional correlation (72 %). This result may seem counter-intuitive if we take the unconditional correlation as a "mean" of the correlations in the two regimes. Of course, one must realize that the asymmetric regime can be characterized by a low correlation but by a large TDC. This demonstrates the importance of distinguishing between correlation and extreme dependence. The mixture model is better able to capture this distinction in fitting the data. A normal distribution may be a good approximation for measuring finite distance dependence, but an appropriate copula structure is necessary for characterizing extreme dependence.

4.3.2 France-Germany Dependence Structure

The estimation results are shown in Table 6. Due to a high cross-country unconditional correlation in both markets, the results for France and Germany are more eloquent. The dependence between equities and bonds is low, while the dependence between assets of the same type is large in both regimes. For France and Germany, equity-equity correlation and bond-bond correlation are larger than 90% while bond-equity correlations are lower than 21% in the same country as well as between the two countries. In the asymmetric regime, the TDC are larger than 67% between assets of the same type and lower than 2% between bond and equities in both France and Germany.

To analyze the effect of the Euro on the dependence structure, we split the observation period in two sub-periods, before and after the introduction of the currency. Tables 7 and 8 contain the results for the respective subperiods. We find that the introduction of the Euro increases the correlation in the normal regime between the French and German markets. Before the introduction of the Euro, in the normal regime, the cross-country correlation between assets of the same type is in average 80%, against more than 96% after the introduction. The cross-asset correlations exhibit a similar pattern since all correlations increase after the introduction of the Euro. This result is consistent with those of Cappiello, Engle and Sheppard (2003) who find that the introduction of a fixed exchange rate leads to a structural break characterized by a high correlation.[21] For the asymmetric

[21]The goal of Cappiello, Engle and Sheppard (2003) was to investigate the asymmetric effect of past news on the

regime, the results are more surprising since the extreme dependence between the French and German equity markets drastically decreases from 87% to 26%. All the other extreme dependence coefficients increase, but only the TDC of the FR bond-DE bond pair increases significantly. Since this change in the level of dependence suggests a relationship between the dependence structure and the exchange rate, we investigate in the next section for both pairs of countries.

To conclude, let us mention that the results of the Monte Carlo tests shown in Table 9 confirm the presence of asymmetry in the dependence structure in both pairs of countries.

4.3.3 Link between asymmetric dependence and the exchange rate

The filtered probabilities to be in asymmetric regime for France and Germany show a clear break after the introduction of the Euro (see figure 8). Before its introduction, the dependence is more likely asymmetric and becomes more Gaussian after the event.

To confirm this graphical observation, we perform a logistic regression of the conditional probabilities to be in the asymmetric regime on the volatility of the exchange rate.[22]

For France and Germany, we have:

$$\widehat{P}_t = \underset{\substack{\text{-1.26e+0} \\ \text{(6.81e-2)}}}{a} + \underset{\substack{\text{5.06e+2} \\ \text{(2.29e+1)}}}{b} \times Vol_t + e_t$$

The $R-square$ of the regression is 0.86. The explained variable $\widehat{P}_t = \log(P_t/(1-P_t))$, P_t is the conditional probability to be in the asymmetric regime given the time-t available information, and Vol_t is the exchange rate volatility between the two countries obtained by a M-GARCH(1,1) filter.

We run the same regression for US and Canada to investigate if the relation holds when no structural change occurs. The results are similar to the European results.

correlation. Since it is well documented that the negative shocks have a larger effect on volatility than the positive shocks of the same magnitude, they try to see if the result is similar for correlation.

[22] Since the probability P_t to be in a regime is between 0 and 1, the logistic regression allows us to keep this constraint by proceeding as follows $P_t = \exp(a+Vol_t+\varepsilon_t)/(1+\exp(a+Vol_t+\varepsilon_t))$ or equivalently $\log(P_t/(1-P_t)) = a+bVol_t+\varepsilon_t$ and we can perform the usual regression.

$$\underset{(1.76e\text{-}1)}{\underset{-7.71e\text{-}1}{\widehat{P}_t = a}} + \underset{(2.36e+1)}{\underset{9.30e+1}{b}} \times Vol_t + e_t$$

The $R - square$ of the regression remains high at 0.75.

These results suggest that high exchange rate volatility is associated with asymmetric dependence. With the introduction of the European currency the dependence between France and Germany becomes more normal. This result is coherent with the literature, which finds asymmetric dependence mainly in the international markets (see Longin and Solnik, 2001). We find the same asymmetric dependence in international bond markets as well. This evidence is reflected in the fact that in the normal regime the correlation is higher than the unconditional correlation. Moreover, since the introduction of the Euro reduces the volatility of the exchange rate, it increases the correlation due to the link between a fixed exchange rate and the normal distribution regime.

5. Asymmetric Dependence Effect on International Diversification

The benefits of international diversification are well documented in the literature (see Solnik, 1974, DeSantis and Gerard, 1997 and reference therein). However, investors tend to invest mainly in their country despite these alleged international diversification benefits. In fact, the share invested by home investors in domestic assets is much larger than the share predicted by the Mean-Variance (MV) model. Two main explanations have been put forward. Transaction costs for international assets reduce the expected gain on foreign assets, while information asymmetry between local and foreign investors increases the risk of foreign assets. These explanations affect the first two moments of asset returns. The transaction costs affect the first moment by reducing the expected return and the asymmetric information affects the second moment since it increases the risk of foreign assets.

Glassman and Riddick (2001) perform an empirical assessment of these potential explanations. Using data for six developed countries, they find that to explain the deviations, transaction costs must be in excess of 1% per month, 14–19% per year,[23] against the actual estimation of 1–4% per annum, with some variation across countries (see, e.g., Perold and Sirri, 1994; Solnik, 1996). Moreover, Glassman and Riddick (2001) find that the implied volatility that matches the portfolio data is greater than twice the historical volatility and therefore is unreasonable.

We go beyond the two first moments to investigate the effect of skewness and specially co-skewness on cross-country diversification and also on bond against equity diversification. We show how strong dependence for lower returns in two markets can reduce co-skewness and therefore reduce skewness in a portfolio with long positions on both markets. Since the reduction of co-skewness lowers the gains of diversification, investors tend to hold a higher share of low risk assets than in a MV portfolio.

Two recent studies have examined the portfolio allocation effects of asymmetric correlation or dependence between equities and cash. In a two-regime correlation model, Ang and Bekaert (2004) find that the investor tends to switch to cash when a persistent bear market hits, while Patton (2004) notices a significant gain when an investor takes into account the existence of the asymmetric dependence structure. Here we examine the effects of asymmetric dependence on cross-country diversification and on domestic diversification between bonds and equities.

The agent's wealth at time t invested in domestic and foreign bonds and equities is described by the following equation

$$W_t = W_{t-1}\left[w_t\eta_t^h R_t^{h,b} + w_t\left(1-\eta_t^h\right)R_t^{h,e} + \left(1-w_t\right)\eta_t^f R_t^{f,b} + \left(1-w_t\right)\left(1-\eta_t^f\right)R_t^{f,e}\right],$$

where $R_t^{h,b}$, $R_t^{h,e}$, $R_t^{f,b}$, and $R_t^{f,e}$ are the returns of domestic bond, domestic equity, foreign bond, and foreign equity respectively. We adopt a specification which simplified the analysis of two mentioned effects, cross-country and domestic diversification. So, w_t is the share invested in domestic assets, the remaining $(1-w_t)$ being invested in foreign assets, while η_t^h and η_t^f are the shares invested in domestic and foreign bonds respectively.

[23] France, Germany, Japan, UK, Canada, and the US.

5.1 Investor Problem

To analyze the effects of asymmetric dependence on cross-country and domestic diversification, we assume that the investor has to choose the share w_t invested in domestic assets, and the bond shares η_t^h and η_t^f. Therefore, the return on his domestic portfolio is $R_t^h = w_t \eta_t^h R_t^{h,b} + w_t \left(1 - \eta_t^h\right) R_t^{h,e}$, while the return on the foreign portfolio is $R_t^f = (1 - w_t)\, \eta_t^f R_t^{f,b} + (1 - w_t) \left(1 - \eta_t^f\right) R_t^{f,e}$. His portfolio wealth for one period is then $W_t = W_{t-1} \left[w_t R_t^h + (1 - w_t) R_t^f\right]$. The investor is assumed to maximize his expected utility function $EU(W_t)$.

Going back to Samuelson (1970), we can consider that a cubic expansion provides a reasonable approximation of the expected utility function, especially for distributions with low volatility. In order to take into account the third moments, we consider a cubic Taylor expansion of expected utility around the average wealth.

$$E\left(U\left(W_t\right)\right) = U\left(\overline{W_t}\right) + \frac{U''\left(\overline{W_t}\right)}{2} E\left(W_t - \overline{W_t}\right)^2 + \frac{U'''\left(\overline{W_t}\right)}{3!} E\left(W_t - \overline{W_t}\right)^3 + o^{(4)},$$

where $\overline{W_t} = E(W_t)$, and $o^{(4)}$ represents the terms of order larger than three that are supposed to be negligible compared to the terms of smaller order. We also made the usual assumptions regarding the properties of the investor's utility function, that is positive marginal utility ($U' \geq 0$), risk aversion ($U'' \leq 0$), and non-increasing absolute risk aversion ($U''' \geq 0$).

The third centered moment of the investor portfolio is given by

$$\begin{aligned} E\left(W_t - \overline{W_t}\right)^3 = \; & W_{t-1}^3 \left[w_t^3 \sigma_{ht}^3 s_{ht} + (1 - w_t)^3 \sigma_{ft}^3 s_{ft} \right. \\ & \left. + 3w_t^2 (1 - w_t) \sigma_{ht}^2 \sigma_{ft} c_{hft} + 3w_t (1 - w_t)^2 \sigma_{ht} \sigma_{ft}^2 c_{hft}\right] \end{aligned}$$

where

$$\sigma_{it}^2 = var\left(R_t^i\right);$$

$$s_{it} = E\left(\frac{R_t^i - E\left(R_t^i\right)}{\sigma_{it}}\right)^3 \equiv Skew\left(R_t^i\right);$$

$$c_{ijt} = E\left(\left(\frac{R_t^i - E\left(R_t^i\right)}{\sigma_{it}}\right)^2 \left(\frac{R_t^j - E\left(R_t^j\right)}{\sigma_{jt}}\right)\right) \equiv CoSkew\left(R_t^i, R_t^j\right)$$

When a representative international investor has positive shares of foreign and domestic assets in his portfolio, skewness and co-skewness affect positively investor expected utility. Intuitively, when skewness (or co-skewness) decreases, the investor is less likely to diversify. In presence of negative skewness, investor will diversify less than he does for the MV portfolio which corresponds to a case of zero skewness. The results below formalize this intuition.

5.2 Asymmetric Dependence and Cross-Country Portfolio Diversification: *Home Bias Investment*

The importance of skewness in asset pricing and portfolio choice is well documented by Harvey and Siddique (2000) and the references therein. They find a negative trade-off between expected returns and skewness. In a portfolio with a long position in two assets, co-skewness has a similar effect since it is positively related to the portfolio skewness. In a MV trade-off behavior, for a portfolio of two identically distributed assets, we allocate one half of the portfolio to each asset. When the variance of one asset increases, its share decreases. The issue here is to investigate what is the effect of asymmetric dependence through co-skewness when we consider the third moment for expected utility.

To characterize asymmetric dependence, Longin and Solnik (2001) use exceedance correlation. This characterization does not allow us to make a link with the portfolio third moment. With the copula model we developed in the previous sections, it is possible to establish a link between co-skewness and asymmetric dependence.

Proposition 5.1: *For F and F' with the same marginal distributions and the same correlation coefficient, let $(X_1, X_2) \rightsquigarrow F \equiv (F_1, F_2, C_{rG})$ and $(X_1', X_2') \rightsquigarrow F' \equiv (F_1, F_2, C_N)$, where C_{rG} is a rotated Gumbel copula and C_N is a Gaussian copula such that $C_N \leq C_{rG}$. Therefore*

$$\begin{cases} CoSkew\,(X_1, X_2) \leq CoSkew(X_1', X_2') \\ CoSkew\,(X_2, X_1) \leq CoSkew(X_2', X_1') \end{cases}$$

Proof see Appendix

This result means that a strong dependence in lower returns creates a lower (or large negative) co-skewness. To analyze the effect of co-skewness on international diversification, we start from the MV optimal portfolio and then show that introducing skewness in the objective function, asymmetric dependence will reduce the portfolio share invested in the higher risk assets for very risk averse investors.

Proposition 5.2: *If the following conditions are satisfied*

i) $\left|\sum_{n\geq 4:} \frac{1}{n!} U^{(n)}\left(\overline{W}_t^*\right)\left[E\left(W_t^* - \overline{W}_t^*\right)^n\right]\right| \ll \left|\sum_{n=0:}^{3} \frac{1}{n!} U^{(n)}\left(\overline{W}_t^*\right)\left[E\left(W_t^* - \overline{W}_t^*\right)^n\right]\right|$ *: validity of the third order approximation of expected utility around W_t^*, the MV optimal portfolio final wealth*

ii) the optimal share invested in domestic assets in a MV behavior w_t^* is in the range $(1/3, 1)$, and is such that $\frac{\sigma_{ft}}{\sigma_{ht}} > \delta(w_t^*) \equiv \frac{w_t^*(2-3w_t^*)}{(1-w_t^*)(3w_t^*-1)}$: large (perceived) risk for foreign portfolio.

iii) $CoSkew\left(R_t^h, R_t^f\right) = CoSkew\left(R_t^f, R_t^h\right) \equiv c_t$,

then there exists a threshold $\overline{c}_t$ such that for $c_t \leq \overline{c}_t$ we have

$$\left.\frac{\partial}{\partial w_t} EU(W_t)\right|_{w_t = w_t^*} > 0$$

where $U^{(n)}$ is the n-order differential of U, and $U^{(0)} = U$.

Proof see Appendix

This proposition can be interpreted as follows. A strong downside market dependence which creates co-skewness combined with a large foreign risk implies that the share invested in the domestic portfolio will increase compared with the share invested in MV framework. This provides an additional explanation for the home bias phenomenon. We may notice that the lower threshold $\delta(.)$ for the ratio between foreign and domestic volatilities is a decreasing function of w_t^*, with $\delta(0.5) = 1$. It means that if in the MV framework less than half of the wealth is invested in the domestic portfolio, foreign volatility should be greater than domestic volatility to insure that strong downside dependence will increase the home investment.

5.3 Asymmetric Dependence Effect on Domestic Diversification: *Flight to Safety.*

Starting at the MV optimal point, we can also perform *local* analysis of the asymmetric dependence effect on the equity and bond diversification. Let $\eta_t^h\left(w_t^*, \eta_t^{f*}\right)$, the optimal share of bonds in the domestic portfolio, be a function of w_t^* the MV optimal share invested in domestic assets, and η_t^{f*} the MV optimal share of bonds in the foreign portfolio. As in the case of cross-country diversification, it can be similarly shown that asymmetric dependence will introduce a bond bias for a very risk averse investor. So, η_t^h will increase in the asymmetric regime, if its MV optimal solution η_t^{h*} belong to the range $(1/3, 1)$. A similar behavior will be observed for the share of bonds in the foreign portfolio.

The main intuition for the effect of asymmetric dependence on home bias is the increasing share invested in the asset with lower risk. The same intuition explains the fact that in the presence of asymmetric dependence, investors will increase the share of bonds in their portfolio relatively to equity.

For less risk averse agents, the bond share is lower in the asymmetric framework than the share in the normal regime, but it becomes larger for investors with higher risk aversion. These results are related to the downside risk premium found by Ang et al (2006). Actually, diversification beyond a certain level increases downside risk and due to the trade-off between this risk and the expected return, investors should adjust their portfolio according to their risk aversion level.

5.4 Monte Carlo Optimization Procedure for Portfolio Choice under each Regime

The aim of this exercise is to investigate the effect of asymmetric dependence on portfolio choice. So, we assume that the investor knows the regime of dependence. We perform one period ahead optimization for each regime, and for marginal distributions, we use the simple Gaussian distribution.

1. We estimate the parameters in our regime switching model with the Gaussian and asymmetric dependence structure. We get the correlation coefficients for Gaussian regime and the tail

dependence coefficients for the asymmetric regime.

2. For each regime of dependence, we generate n =10,000 independent draws $R_{t,i} = \left(R_{t,i}^{h,b}, R_{t,i}^{h,e}, R_{t,i}^{f,b}, R_{t,i}^{f,e}\right)$, $i = 1, ..., n$, 100 times using univariate unconditional Gaussian distribution for any single return.

3. We can therefore compute for each simulation the portfolio component weights

$$\left(w_t^{**}, \eta_t^{h**}, \eta_t^{f**}\right) = \underset{\left(w_t, \eta_t^h, \eta_t^f\right)}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} U\left(W_{t,i}\right),$$

where $W_{t,i} = \left[w_t \eta_t^h R_{t,i}^{h,b} + w_t \left(1 - \eta_t^h\right) R_{t,i}^{h,e} + \left(1 - w_t\right) \eta_t^f R_{t,i}^{f,b} + \left(1 - w_t\right) \left(1 - \eta_t^f\right) R_{t,i}^{f,e}\right]$, and the utility function is replaced by the third-order approximation to avoid explosive solution due to the discretization of expected utility.

Simulation Results

As expected, the empirical results show a positive link between the level of risk aversion and the share invested in lower risk assets (bonds). Given the lower level of Canadian market volatilities compared to the US in our sample period, a similar positive link is observed between the Canadian portfolio share and the risk aversion level. One important fact is the point where the share invested in the Canadian portfolio becomes larger in the asymmetric regime compared to the Gaussian regime. As expected from the theoretical analysis, this point is between 1/3 and 1.

Since the chosen home country (Canadian) market in our sample is less risky than the foreign (US) market, the only effect of the 10% adjustment for foreign perceived risk is the shift of the share invested in the Canadian portfolio. Of course, this share increases in both regimes.

6. Conclusion

In this paper, we present the abilities and the limitations of some classical models to reproduce asymmetric dependence and the need to disentangle marginal asymmetry from dependence asymmetry. Using copulas we provide a flexible model to achieve this aim. We build a two regimes dependence model, and by applying it to international bond and equity markets, we put forward some interesting facts about the dependence structure.

The dependence between the equity markets on one hand and the bond markets on the other were found to be much larger than the dependence between equities and bonds even in the same country. Extreme dependence appears especially large in cross-country bond markets and equity markets taken separately. The proposed model allows us to investigate the relationship between the filtered probabilities to be in the asymmetric regime with other factors. This was not possible with the Longin and Solnik (2001) model.

Using this model we analyse in a simple portfolio choice framework, the implications on international investment and national bond and equity diversification. We find that, to reduce the downside risk effect due to strong comovement of markets in bad situation, very risk averse investor when taking into account the asymmetric dependence, will increase the part invested in low risk country, and inside each country, he will increase the bond part. These results are in line with what is commonly called flight to safety and at the same time give an additional explanation to the lack of international diversification known as the home bias puzzle. International investors face high extreme dependence in bear markets and therefore lose the diversification gain when they most need it.

We find that the exchange rate volatility may be a factor behind the asymmetric behavior of international market dependence. Therefore, it will be interesting to use a model similar to the model explored in this paper, possibly incorporating exchange rate, to study the portfolio of an international investor with loss aversion in the spirit of Ang et al (2002).

Appendix A. Proofs

Proof of Proposition 2.1

To prove this proposition, we need the two following lemmas

Lemma 1: *(a) Let* $\{f^{(s)}\}_{s=1}^{n}$ *be a family of symmetric multivariate density functions of* $n\,(\leq\infty)$ *variables with same mean. The mixture* $f=\sum_{s=1}^{n}\pi_s f^{(s)}$*, where* $\sum_{s=1}^{n}\pi_s=1$*, and* $\pi_s\geq 0$ *for any* s*, is a symmetric multivariate density function. (b) Moreover for a continuum of symmetric multivariate density function* $\{f^{(\sigma)}\}_{\sigma\in A\subseteq\mathbb{R}}$ *with same mean, the mixture* $f=\int_A \pi_\sigma f^{(\sigma)}d\sigma$*, where* $\int_A \pi_\sigma d\sigma=1$*, is a symmetric multivariate density function.*

Proof: Let μ be the mean of all $f^{(s)}$ (and all $f^{(\sigma)}$)

$$f(\mu-x)=\sum_{s=1}^{n}\pi_s f^{(s)}(\mu-x)$$

by symmetry of all $f^{(s)}$, we have, $\sum_{s=1}^{n}\pi_s f^{(s)}(\mu-x)=\sum_{s=1}^{n}\pi_s f^{(s)}(\mu+x)=f(\mu+x)$
i.e. $f(\mu-x)=f(\mu+x)$ and the part (a) follows. Similarly for mixture of continuum,

$f(\mu-x)=\int_A \pi_\sigma f^{(\sigma)}(\mu-x)\,d\sigma=\int_A \pi_\sigma f^{(\sigma)}(\mu+x)\,d\sigma=f(\mu+x)$ and we have (b).

Lemma 2: *Let* $\{F^{(s)}\}_{s=1}^{n}$ *be a family of bivariate cdf with zero lower (upper) TDC. The mixture* $F=\sum_{s=1}^{n}\pi_s F^{(s)}$*, where* $\sum_{s=1}^{n}\pi_s=1$*, and* $\pi_s\geq 0$*, for any* s*, is a bivariate density function with lower (upper) TDC.*

Proof: we do the proof for lower tail since by "rotation" we have the same result for upper tail.

Let τ_L^F be the lower TDC of F, we have

$$\begin{aligned}\tau_L^F &= \lim_{\alpha\to 0}\Pr\left[X\leq F_x^{-1}(\alpha)\,|Y\leq F_y^{-1}(\alpha)\right]\\ &= \lim_{\alpha\to 0}\frac{\Pr\left[X\leq F_x^{-1}(\alpha),Y\leq F_y^{-1}(\alpha)\right]}{\Pr\left[Y\leq F_y^{-1}(\alpha)\right]}\\ &= \lim_{\alpha\to 0}\frac{F\left(F_x^{-1}(\alpha),F_y^{-1}(\alpha)\right)}{F_y\left(F_y^{-1}(\alpha)\right)}\end{aligned}$$

and since $F=\sum_{s=1}^{n}\pi_s F^{(s)}$, we have

$$\begin{aligned}
\tau_L^F &= \lim_{\alpha\to 0}\frac{\sum_{s=1}^{n}\pi_s F^{(s)}\left(F_x^{-1}(\alpha), F_y^{-1}(\alpha)\right)}{\alpha} \\
&= \lim_{\alpha\to 0}\sum_{s=1}^{n}\pi_s\frac{F^{(s)}\left(F_x^{-1}(\alpha), F_y^{-1}(\alpha)\right)}{\alpha} \\
&= \sum_{s=1}^{n}\pi_s\lim_{\alpha\to 0}\frac{F^{(s)}\left(F_x^{-1}(\alpha), F_y^{-1}(\alpha)\right)}{\alpha}
\end{aligned}$$

by definition $F^{(s)}\left(F_x^{-1}(\alpha), F_y^{-1}(\alpha)\right) = C^{(s)}\left(F_x^{(s)}\left(F_x^{-1}(\alpha)\right), F_y^{(s)}\left(F_y^{-1}(\alpha)\right)\right)$

where $C^{(s)}$ is the copula and $F_x^{(s)}$, $F_y^{(s)}$ the marginal cdf corresponding to $F^{(s)}$, we have

$$\alpha = F_x\left(F_x^{-1}(\alpha)\right) = \sum_{s=1}^{n}\pi_s F_x^{(s)}\left(F_x^{-1}(\alpha)\right)$$

so

$F_x^{(s)}\left(F_x^{-1}(\alpha)\right) \le \alpha/\pi_s$ for all s and similarly $F_y^{(s)}\left(F_y^{-1}(\alpha)\right) \le \alpha/\pi_s$,

hence

$$\begin{aligned}
\lim_{\alpha\to 0}\tfrac{F^{(s)}\left(F_x^{-1}(\alpha), F_y^{-1}(\alpha)\right)}{\alpha} &= \lim_{\alpha\to 0}\frac{C^{(s)}\left(F_x^{(s)}\left(F_x^{-1}(\alpha)\right), F_y^{(s)}\left(F_y^{-1}(\alpha)\right)\right)}{\alpha} \\
&\le \lim_{\alpha\to 0}\frac{C^{(s)}\left(\alpha/\pi_s, \alpha/\pi_s\right)}{\alpha}, \text{ since copula is increasing function} \\
&= 1/\pi_s\lim_{\alpha'\to 0}\frac{C^{(s)}\left(\alpha', \alpha'\right)}{\alpha'} \text{ by setting } \alpha' = \alpha/\pi_s \\
&= 0, \text{ since } F^{(s)} \text{ and hence } C^{(s)} \text{ is zero lower TDC}
\end{aligned}$$

we therefore have $\tau_L^F = 0$

The part (i) and (ii) of the proposition is the straightforward application of above lemma

- For GARCH with constant mean and symmetric conditional distribution

$$X_t = \mu + \Sigma_{t-1}^{1/2}\varepsilon_t$$
$$(+\text{ any GARCH dynamic equation of } \Sigma_{t-1}\text{ })$$

where ε_t is stationary with symmetric distribution such that $E(\varepsilon_t) = 0$. The unconditional distribution of X_t is a mixture of distribution of symmetric variable with same mean μ but possibly different variance covariance matrix. By applying the lemma 1, we conclude that the unconditional distribution of X_t is symmetric and (i) follows.

- For RS model with zero TDC

$$X_t = \mu_{s_t} + \Sigma_{s_t}^{1/2} \varepsilon_t$$

where s_t takes a discrete value. Without loss of generality assume that X_t is bivariate and that $s_t = s$, $\mu + \Sigma^{1/2}\varepsilon_t$ is zero TDC such as in the normal case, therefore the unconditional distribution of X_t is a mixture of distribution with zero TDC. By applying the lemma 2, we conclude that the unconditional distribution of X_t has zero TDC. and (ii) follows

For (iii), with the same notations as lemma 1, keeping marginal distribution unchanged across mixture components means that. For discrete case

$f^{(s)}(x_1, ..., x_n; \delta, \rho) = c^{(s)}(u_1, ..., u_n; \theta) \times \prod_{i=1}^{n} f_i(x_i; \delta_i)$, with $u_i = F_i(x_i; \delta_i)$, hence

$$\begin{aligned} f(x_1, ..., x_n; \delta, \rho) &= \sum_{s=1}^{n} \pi_s f^{(s)}(x_1, ..., x_n; \delta, \rho) \\ &= \sum_{s=1}^{n} \pi_s c^{(s)}(u_1, ..., u_n; \theta) \times \prod_{i=1}^{n} f_i(x_i; \delta_i) \\ &= c(u_1, ..., u_n; \theta) \times \prod_{i=1}^{n} f_i(x_i; \delta_i) \end{aligned}$$

with $c(u_1, ..., u_n; \theta) = \sum_{s=1}^{n} \pi_s c^{(s)}(u_1, ..., u_n; \theta)$ is the copula of f and we can see that c is a mixture of copula with symmetric TDC and hence is a copula with symmetric TDC.

for the continuum case

$$\begin{aligned} f(x_1, ..., x_n; \delta, \rho) &= \int_A \pi_\sigma f^{(\sigma)}(x_1, ..., x_n; \delta, \rho)\, d\sigma \\ &= \int_A \pi_\sigma c^{(\sigma)}(u_1, ..., u_n; \theta)\, d\sigma \times \prod_{i=1}^{n} f_i(x_i; \delta_i) \\ &= c(u_1, ..., u_n; \theta) \times \prod_{i=1}^{n} f_i(x_i; \delta_i) \end{aligned}$$

with $c(u_1,...,u_n;\theta) = \int_A \pi_\sigma c^{(\sigma)}(u_1,...,u_n;\theta)\, d\sigma$ which is a copula with symmetric TDC for same the reasons mentioned above.

Q.E.D

Proof of Proposition 3.2.

By definition of a copula, we have

$$\begin{aligned}
\eta_t &= \begin{bmatrix} f(X_t;\delta,\theta|\,\underline{X}_{t-1}, s_t = 1) \\ f(X_t;\delta,\theta|\,\underline{X}_{t-1}, s_t = 0) \end{bmatrix} \\
&= \begin{bmatrix} c(u_{1,t}(\delta_1),...,u_{4,t}(\delta_4);\theta\,|s_t = 1) \times \prod_{i=1}^{4} f_i(x_{i,t};\delta_i) \\ c(u_{1,t}(\delta_1),...,u_{4,t}(\delta_4);\theta\,|s_t = 0) \times \prod_{i=1}^{4} f_i(x_{i,t};\delta_i) \end{bmatrix}
\end{aligned}$$

with $u_{i,t}(\delta_i) = F_i(x_{i,t};\delta_i)$

By denoting $\widehat{\xi}_{t|t-1} = \left(\widehat{\xi}^{(1)}_{t|t-1}, \widehat{\xi}^{(0)}_{t|t-1}\right)'$, the likelihood can be rewritten

$$\begin{aligned}
&L(\delta,\theta;\underline{X}_T) \\
&= \sum_{t=1}^{T} \log\left(\widehat{\xi}'_{t|t-1}\eta_t\right) \\
&= \sum_{t=1}^{T} \log\left(\sum_{k=0}^{1} \widehat{\xi}^{(k)}_{t|t-1} c(u_{1,t}(\delta_1),...,u_{4,t}(\delta_4);\theta\,|s_t = k) \times \prod_{i=1}^{4} f_i(x_{i,t};\delta_i)\right) \\
&= \sum_{t=1}^{T}\left[\sum_{i=1}^{4} \log(f_i(x_{i,t};\delta_i)) + \log\left(\sum_{k=0}^{1} \widehat{\xi}^{(k)}_{t|t-1} c(u_{1,t}(\delta_1),...,u_{4,t}(\delta_4);\theta\,|s_t = k)\right)\right]
\end{aligned}$$

it follows that

$$L(\delta,\theta;\underline{X}_T) = \sum_{i=1}^{4} L_i(\delta_i;\underline{X}_T) + L_C(\delta,\theta;\underline{X}_T)$$

where

$$\begin{aligned}
L_i\left(\delta_i;\underline{X}_{i,T}\right) &= \sum_{t=1}^{T} \log f_i\left(x_{i,t};\delta_i|\,\underline{X}_{i,t-1}\right) \\
L_C(\delta,\theta;X) &= \sum_{t=1}^{T} \log\left(\widehat{\xi}'_{t|t-1}\eta_{ct}\right)
\end{aligned}$$

with

$$\eta_{ct} = \begin{bmatrix} c\left(u_{1,t}(\delta_1), ..., u_{n,t}(\delta_n); \theta | \, s_t = 1\right) \\ c\left(u_{1,t}(\delta_1), ..., u_{n,t}(\delta_n); \theta | \, s_t = 0\right) \end{bmatrix}$$

by noticing that $\eta_t = \eta_{ct} \times \prod_{i=1}^{4} f_i\left(x_{i,t}; \delta_i\right)$ we have that

$$\widehat{\xi}_{t/t} = \left[\widetilde{\widehat{\xi}}_{t|t-1}\eta_t\right]^{-1}\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right) = \left[\widetilde{\widehat{\xi}}_{t|t-1}\eta_{ct}\right]^{-1}\left(\widehat{\xi}_{t|t-1} \odot \eta_{ct}\right)$$

Q.E.D

Proof of proposition 5.1

Let $(X_1, X_2) \rightsquigarrow F \equiv (F_1, F_2, C_{rG})$ and $(X_1', X_2') \rightsquigarrow F' \equiv (F_1, F_2, C_N)$.

for $w \in [0,1]$ let $X = wX_1 + (1-w)X_2$, and $X' = wX_1' + (1-w)X_2'$

$$E\left(U\left(X\right)\right) = U\left(\overline{X}\right) + \frac{U''\left(\overline{X}\right)}{2}E\left(X - \overline{X}\right)^2 + \frac{U'''\left(\overline{X}\right)}{3}E\left(X - \overline{X}\right)^3 + o^{(4)}$$

and $E\left(X'\right) = E\left(X\right) = \overline{X}$, we have

$$E\left(U\left(X'\right)\right) = U\left(\overline{X}\right) + \frac{U''\left(\overline{X}\right)}{2}E\left(X' - \overline{X}\right)^2 + \frac{U'''\left(\overline{X}\right)}{3}E\left(X' - \overline{X}\right)^3 + o^{(4)}$$

by assumption, we have $C_N \leq C_{rG}$ what by the below lemma, is equivalent to $F' \leq F$, and then $E\left(U\left(X'\right)\right) \geq E\left(U\left(X\right)\right)$ for any increasing function U. So for an utility function U that satisfies Arrow (1971) third main desirable property $U''' \geq 0$, we have $E\left(X' - \overline{X}\right)^3 \geq E\left(X - \overline{X}\right)^3$.

Since for any $w \in [0,1]$

$$\begin{aligned} E\left(X - \overline{X}\right)^3 = {} & w^3\sigma_1^3 s_1 + (1-w)^3\sigma_2^3 s_2 \\ & + 3w^2(1-w)\sigma_1^2\sigma_2 c_{12} + 3w(1-w)^2\sigma_1\sigma_2^2 c_{21} \end{aligned}$$

and

$$\begin{aligned} E\left(X' - \overline{X}\right)^3 = {} & w^3\sigma_1^3 s_1 + (1-w)^3\sigma_2^3 s_2 \\ & + 3w^2(1-w)\sigma_1^2\sigma_2 c_{12}' + 3w(1-w)^2\sigma_1\sigma_2^2 c_{21}' \end{aligned}$$

we have

$$\begin{cases} c_{12} \leq c'_{12} \\ c_{21} \leq c'_{21} \end{cases}$$

i.e.

$$\begin{cases} CoSkew\,(X_1, X_2) \leq CoSkew(X'_1, X'_2) \\ CoSkew\,(X_2, X_1) \leq CoSkew(X'_2, X'_1) \end{cases}$$

Q.E.D

Lemma: *Let* $F \equiv (F_1, F_2, C)$ *and* $(X'_1, X'_2) \rightsquigarrow F' \equiv (F_1, F_2, C')$.

$C' \leq C$ *is equivalent to* $F' \leq F$.

Proof

$$\begin{aligned}
&F' \leq F \\
&\Leftrightarrow F'(x,y) \leq F(x,y) \;\; forall\,(x,y) \in \mathbb{R}^2 \\
&\Leftrightarrow C\left(F_1(x), F_2(y)\right) \leq C'\left(F_1(x), F_2(y)\right) \;\; forall\,(x,y) \in \mathbb{R}^2 \\
&\Leftrightarrow C(u,v) \leq C'(u,v) \;\; forall\,(u,v) \in [0,1]^2 \\
&\Leftrightarrow C' \leq C.
\end{aligned}$$

Q.E.D

Proof of Proposition 5.2:

The exact expansion of the expected utility function is $EU(W_t) = \sum_{n=0}^{\infty} U^{(n)}\left(\overline{W_t}\right) E\left(W_t - \overline{W}_t\right)^{(n)}$.

Under assumption *i)*, and the third order validity of the Taylor expansion, the sign of $\left.\frac{\partial}{\partial w} EU(W_t)\right|_{w_t = w_t^*}$ depends on the sign of $\left.\frac{\partial}{\partial w_t} U'''\left(\overline{W}_t\right) E\left(W_t - \overline{W}_t\right)^3\right|_{w_t = w_t^*}$ since saying that w_t^* is the optimal part invested on home portfolio in a Mean-Variance behavior means that

$$\left.\frac{\partial}{\partial w_t}\left\{U\left(\overline{W_t}\right) + \frac{U''\left(\overline{W}_t\right)}{2} E\left(W_t - \overline{W}_t\right)^2\right\}\right|_{w_t = w_t^*} = 0.$$

and

$$\frac{\partial}{\partial w_t}U'''\left(\overline{W}_t\right)E\left(W_t-\overline{W}_t\right)^3\Big|_{w_t=w_t^*} = \left[\frac{\partial}{\partial w_t}U'''\left(\overline{W}_t\right)\right]\Big|_{w_t=w_t^*} E\left(W_t^*-\overline{W}_t^*\right)^3 + U'''\left(\overline{W}_t^*\right)\frac{\partial}{\partial w_t}E\left(W_t-\overline{W}_t\right)^3\Big|_{w_t=w_t^*}$$

we have

$$\begin{aligned}\frac{\partial}{\partial w_t}E\left(W_t-\overline{W}_t\right)^3\Big|_{w_t=w_t^*} &= 3w_t^{*2}\sigma_{ht}^3 s_{ht} - 3\left(1-w_t^*\right)^2\sigma_{ft}^3 s_{ft} \\ &\quad +3\left(2w_t^*-3w_t^{*2}\right)\sigma_{ht}^2\sigma_{ft}c_{12t}+3\left(1-4w_t^*+3w_t^{*2}\right)\sigma_{ht}\sigma_{ft}^2 c_{21t} \\ &= 3w_t^{*2}\sigma_{ht}^3 s_{ht} - 3\left(1-w_t^*\right)^2\sigma_{ft}^3 s_{ft} \\ &\quad +\left[3\left(2w_t^*-3w_t^{*2}\right)\sigma_{ht}^2\sigma_{ft}+3\left(1-4w_t^*+3w_t^{*2}\right)\sigma_{ht}\sigma_{ft}^2\right]c_t \\ &= B+A\overline{c}_t\end{aligned}$$

with

$$\begin{cases} B = 3\left[w_t^{*2}\sigma_{ht}^3 s_{ht} - \left(1-w_t^*\right)^2\sigma_{ft}^3 s_{ft}\right] \\ A = 3\left[\left(2w_t^*-3w_t^{*2}\right)\sigma_{ht}^2\sigma_{ft}+\left(1-4w_t^*+3w_t^{*2}\right)\sigma_{ht}\sigma_{ft}^2\right]\end{cases}$$

by assumptions ii) $A<0$ and by taking

$$\overline{c}_t = \left[\left[\frac{\partial}{\partial w_t}U'''\left(\overline{W}_t\right)\right]\Big|_{w_t=w_t^*}\frac{E\left(W_t^*-\overline{W}_t^*\right)^3}{U'''\left(\overline{W}_t^*\right)} - B\right]\Big/ A,$$

for $c_t \leq \overline{c}_t$, we have $\frac{\partial}{\partial w_t}U'''\left(\overline{W}_t\right)E\left(W_t-\overline{W}_t\right)^3\Big|_{w_t=w_t^*} > 0$, and the proposition 2 follows.

Q.E.D

Appendix B. Analytical expressions for various copulas

Normal copula

$$C_N\left(u_1,\ldots,u_n;\rho\right)=\Phi_\rho\left(\Phi^{-1}\left(u_1\right),\ldots,\Phi^{-1}\left(u_n\right)\right)$$

$$C_N\left(u_1,\ldots,u_n;\rho\right)=\int_{-\infty}^{\Phi^{-1}(u_1)}\ldots\int_{-\infty}^{\Phi^{-1}(u_n)}\left[(2\pi)^n\det\left(\rho\right)\right]^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}\left(z'\rho^{-1}z\right)\right]dz_1\ldots dz_n$$

where $z=(z_1,\ldots,z_n)'$, $\rho=\left(\rho_{ij}\right)_{i,j=1}^{n}$, with $\left|\rho_{ij}\right|\leq 1$, $\rho_{ii}=1$ and ρ positive defined matrix

$$c_N\left(u_1,\ldots,u_n;\rho\right)=\left(\det\left(\rho\right)\exp\left[x'\rho^{-1}x-x'x\right]\right)^{-1/2}$$

with $x=\left(\Phi^{-1}\left(u_1\right),\ldots,\Phi^{-1}\left(u_n\right)\right)'$,

Φ is *cdf* of standard normal distribution and Φ_ρ is *cdf* of multivariate normal distribution with correlation matrix ρ.

Tail dependence coefficients are

$$\tau^L=\tau^U=0$$

Bivariate Gumbel copula

$$C_G\left(u,v;\theta\right)=\exp\left[-\left(\left(-\log\left(u\right)\right)^\theta+\left(-\log\left(v\right)\right)^\theta\right)^{1/\theta}\right]$$

$$c_G\left(u,v;\theta\right)=\frac{C_G\left(u,v;\theta\right)\left(\log\left(u\right).\log\left(v\right)\right)^{\theta-1}}{uv\left(\left(-\log\left(u\right)\right)^\theta+\left(-\log\left(v\right)\right)^\theta\right)^{2-1/\theta}}\left(\left(\left(-\log\left(u\right)\right)^\theta+\left(-\log\left(v\right)\right)^\theta\right)^{1/\theta}+\theta-1\right)$$

Bivariate Rotated Gumbel (Survival) copula

$$C_{GS}\left(u,v;\theta\right)=u+v-1+C_G\left(1-u,1-v;\theta\right)$$

$$c_{GS}\left(u,v;\theta\right)=c_G\left(1-u,1-v;\theta\right)$$

The tail dependence coefficients of C_{GS} are

$\tau^L = 2 - 2^{\frac{1}{\theta}}$ and $\tau^U = 0$

so $\theta = \theta\left(\tau^L\right) = \dfrac{\log(2)}{\log\left(2 - \tau^L\right)}$

and we can re-parameterize the Copula $C_{GS}(u, v; \theta)$ with τ^L as $C_{GS}\left(u, v; \tau^L\right) = C_{GS}\left(u, v; \theta\left(\tau^L\right)\right)$

References

[1] Andrews, Donald W. K. (2001), "Testing when the Parameter is on the Boundary of the Maintained Hypothesis", *Econometrica* 69, 683-734

[2] Andrews Donald. W. K. and Werner Ploberger (1994), "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative", *Econometrica*, 62 1383-1414.

[3] Ang, Andrew and Geert Bekaert (2002), "International Asset Allocation with Regime Shifts", *Review of Financial Studies*, 11, 1137-1187

[4] Ang, Andrew and Geert Bekaert (2004), "How do Regimes Affect Asset Allocation?", *Financial Analysts Journal*, 60, 2, 86-99

[5] Ang, Andrew and Joseph Chen (2002), "Asymmetric Correlations of Equity Portfolios", *Journal of Financial Economics*, 63, 443-494

[6] Ang, Andrew, Joseph Chen and Yuhang Xing (2006), "Downside Risk", *Review of Financial Studies*, 19, 1191-1239

[7] Arrow, Kenneth (1971), *Essay in the theory of risk bearing*, Chicago: Markham Publishing Co.

[8] Barnard, G. A. (1963), "Comment on "The spectral analysis of point processes" by M. S. Bartlett", *Journal of the Royal Statistical Society*, Series B 25, 294

[9] Bollerslev, Tim. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327

[10] Bollerslev, Tim.. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model", *Review of Economics and Statistics*, 31, 307-327

[11] Bontemps, Christian and Nour Meddahi (2005), "Testing Normality: A GMM Approach", *Journal of Econometrics*, 124, 149-186.

[12] Boyer Brian H., Michael S. Gibson and Mico Loretan (1999), "Pitfalls in tests for changes in correlations", International Finance Discussion Paper 597, Board of the Governors of the Federal Reserve System.

[13] Cappiello, Lorenzo, Robert F. Engle and Kevin Sheppard (2003), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns", Forthcoming, *Journal of Financial Econometrics*

[14] Chen, Xiaohong, Yanqin Fan and Andrew Patton (2004), "Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates", working paper, London School of Economics.

[15] Das, Sanjiv, and Raman Uppal, (2003), "Systemic risk and international portfolio choice" *Journal of Finance*, Forthcoming

[16] Denuit, Michel and Olivier Scaillet (2004) "Nonparametric Tests for Positive Quadrant Dependence", *Journal of Financial Econometrics*, 2, 422-450

[17] Dufour Jean-Marie (2005) "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics", *Journal of Econometrics*, Forthcoming

[18] Dwass, M. (1957), "Modified randomization tests for nonparametric hypotheses", *Annals of Mathematical Statistics* 28, 181–187

[19] Engle, Robert F. (1982) "Autoregressive Conditional Heteroskedasticity Models with estimation of Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1007

[20] Engle, Robert F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics*, 20, 339-350

[21] Engle, Robert F and K. Kroner (1995), "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11, 122-150.

[22] Erb, Claude, Campbell Harvey, Tadas Viskanta (1994), "Forecasting international equity correlations", *Financial Analysts Journal* 50, 32–45

[23] Forbes, Kristin J. and Roberto Rigobon (2002), "No contagion, only interdependence: measuring stock market comovements", *Journal of Finance*, 57, 2223-2261

[24] Friend, Irwin, and Randolph Westerfield (1980) "Co-skewnes and Capital Asset Pricing", *Journal of Finance*, 35, 897-913.

[25] Garcia, René, (1998), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review*, 39, 763-788

[26] Genest, Christian, José Juan Quesada Molina and José Antonio Rodriguez Lallena (1995), "De l'impossibilité de construire des lois à marges multidimensionnelles données à partir des copules", *C. R. Acad. Sci. Paris Sér. I Math.* 320, 723-726.

[27] Genest, Christian and Louis Paul Rivest (1993), "Statiscal Inference Procedures for Bivariate Archimedean Copulas" *Journal of American Statistical Association*, 88, 1034-1043

[28] Glassman, Debra A., and Leigh A. Riddick (2001), "What causes home asset bias and how should it be measured", *Journal of Empirical Finance* 8, 35-54.

[29] Guidolin, M. and A., Timmermann, (2006), "International Asset Allocation under Regime Switching, Skew and Kurtosis Preferences" Working Paper, FED St Louis

[30] Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

[31] Hamilton, James D. (1994), *Time Series Analysis*,. Princeton University Press.

[32] Hansen, Bruce. E. (1992), "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 7, S61-S82

[33] Hansen, Bruce. E. (1996), "Inference When a nuisance Parameter Is not Identified Under the Null Hypothesis", *Econometrica*, 64, 413-430

[34] Harvey, Campbell, and Akhtar Siddique (2000) "Conditional Skewness in Asset Pricing Tests", *Journal of Finance*, 55, 1263-1296.

[35] Heston, Steven and Saikat Nandi (2000), "A Closed-Form GARCH Option Valuation Model", *Review of Financial Studies*, 13, 585-625

[36] Kraus, Alan, and Robert H. Litzenberger (1976) "Skewness Preference and The Valuation of Risk Assets", *Journal of Finance*, 31, 1085-1100.

[37] Longin, François and Bruno Solnik (2001), "Extreme correlations in international Equity Markets", *Journal of Finance*, 56, 649-676

[38] Nelsen, Roger B. (1998), *An Introduction to copula*, Springer-Verlag, New York

[39] Patton, Andrew J (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation", *Journal of Financial Econometrics*, 2, 130-168

[40] Pelletier, Denis (2004), "Regime Switching for Dynamic Correlations", *Journal of Econometrics*, Forthcoming

[41] Perold, A.F., Sirri, E.R., 1994. The Cost of International Equity Trading, Unpublished WP

[42] Rodriguez, Juan Carlos (2004), "Measuring Financial Contagion: A Copula Approach", working paper, EURANDOM

[43] Samuelson, Paul A. (1970) "The Fundamental Approximation Theorem of Portfolio Analysis in terms of Means, Variances and Higher Moments", *Review of Economics Studies*, 37, 537-542.

[44] Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges", *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229-231

[45] Solnik, B. (1996) *International Investments.* 3^{rd} ed. Addison-Wesley Publishing, Reading, MA.

Table 1: Summary statistics of weekly bond and equity index returns for the four countries. All returns are expressed in US dollars on a weekly base from January 01, 1985 to December 21, 2004, what corresponds to a sample of 1044 observations. ($^{\delta}$ Denotes annualized percent). Sharpe ratio represents the ratio of the mean over the standard deviation of return.

	Mean$^{\delta}$	Std$^{\delta}$	Kurtosis	Skewness	Min$^{\delta}$	Max$^{\delta}$	Sharpe ratio
US Equity	13.67	17.51	17.00	-1.55	-680.36	311.10	0.78
US Bond	7.57	4.69	0.67	-0.06	-66.91	58.81	1.61
CA Equity	11.24	16.72	13.62	-1.67	-610.87	225.15	0.67
CA Bond	8.81	8.15	1.13	-0.24	-130.55	118.07	1.08
FR Equity	14.72	23.43	7.18	-0.09	-582.12	512.16	0.63
FR Bond	11.52	11.16	0.92	0.04	-142.02	166.68	1.03
DE Equity	12.57	24.97	8.01	-0.46	-574.96	463.08	0.50
DE Bond	10.44	11.56	0.82	-0.01	-142.54	171.39	0.90

Table 2: Unconditional correlations between different assets (bond and equity) of four considered countries.

	US Equity	US Bond	CA Equity	CA Bond	FR Equity	FR Bond	DE Equity
US Bond	0.0576						
CA Equity	0.7182	0.0116					
CA Bond	0.1783	0.4706	0.4392				
FR Equity	0.1957	-0.0182	0.1974	0.1065			
FR Bond	-0.0499	0.3386	-0.0080	0.2433	0.3066		
DE Equity	0.2089	-0.0536	0.1995	0.1009	0.8099	0.2625	
DE Bond	-0.0832	0.3081	-0.0234	0.2143	0.3084	0.9403	0.2847

Table 3: Estimates of M-GARCH (1, 1) parameters for all bond and equity returns of four countries. The figures between brackets represent standard deviations of the parameters. L is the value of the log likelihood function.

	US		CA		FR		DE	
	Equity	Bond	Equity	Bond	Equity	Bond	Equity	Bond
β	7.94e-1	7.82e-1	8.09e-1	9.07e-1	9.68e-1	9.36e-1	9.24e-1	9.56e-1
	(3.49e-1)	(1.62e-1)	(4.06e-1)	(1.79e-1)	(3.61e-1)	(4.21e-1)	(1.54e-1)	(2.45e-1)
α	5.46e-5	2.63e-6	6.40e-5	7.30e-6	2.28e-5	1.51e-5	2.22e-5	1.08e-5
	(4.04e-5)	(6.36e-5)	(8.16e-5)	(2.94e-5)	(9.35e-6)	(2.17e-5)	(2.14e-4)	(1.88e-5)
γ	4.45e+1	3.84e+1	2.73e+1	3.28e+1	1.91e+1	6.53e+0	1.19e+1	3.26e+0
	(1.70e-2)	(6.11e-3)	(1.14e-2)	(1.22e-2)	(1.61e-2)	(1.85e-1)	(8.07e-2)	(2.45e-2)
λ	1.72e+0	1.37e+1	3.13e+0	1.01e+1	1.61e+0	5.61e+0	1.78e+0	6.13e+0
	(1.39e-2)	(1.05e-2)	(2.09e-2)	(7.59e-3)	(7.22e-3)	(1.96e-1)	(6.33e-2)	(7.86e-3)
ω	7.57e-6	6.49e-6	1.21e-5	3.49e-6	1.99e-6	1.51e-7	6.46e-5	4.79e-7
	(9.64e-5)	(1.90e-5)	(1.74e-5)	(2.52e-5)	(6.53e-5)	(4.33e-5)	(1.92e-4)	(3.25e-5)
μ	1.07e-3	7.18e-4	1.32e-3	4.73e-4	1.48e-3	5.37e-4	6.51e-4	1.35e-4
	(1.29e-4)	(6.74e-5)	(3.76e-5)	(5.26e-5)	(5.00e-4)	(1.45e-4)	(1.32e-4)	(3.38e-5)
L	2.49e+3	3.77e+3	2.50e+3	3.20e+3	2.10e+3	2.88e+3	2.04e+3	2.84e+3

Table 4: Box-Pierce and Ljung-Box statistics for tests of independence of residuals. For each series, the statistic is computed for different numbers of lags (1, 4, 6, and 12). * and ** means that we cannot reject independence at the 1 and 5 percent levels respectively

	US		CA		FR		DE	
	Equity	Bond	Equity	Bond	Equity	Bond	Equity	Bond
Box-Pierce								
12 lags	23.26*	18.57**	14.42**	9.27**	10.93**	9.88**	8.64**	12.19**
6 lags	14.85*	12.19**	10.26**	7.17**	10.70**	5.06**	4.55**	8.85**
4 lags	8.73**	10.49*	9.02**	6.34**	7.00**	3.7099**	3.39**	6.36**
1 lag	5.36*	0.01**	3.71**	0.45**	6.11*	1.33**	3.18**	2.78**
Ljung-Box								
12 lags	23.43*	18.71**	14.51**	9.32**	10.98**	9.97**	8.71**	12.28**
6 lags	14.93*	12.25**	10.31**	7.20**	10.74**	5.09**	4.57**	8.90**
4 lags	8.76**	10.55*	9.05**	6.37**	7.02**	3.7248**	3.40**	6.38**
1 lag	5.37*	0.01**	3.72**	0.45**	6.13*	1.33**	3.19**	2.79**

Table 5: Dependence structure between the United States and Canada in equity and bond markets. Correlation coefficients are reported for the normal regime, while tail dependence coefficients describe the asymmetric regime. The tail dependence coefficient is obtained as the product of parameter τ and the respective weight π for cross-asset dependence and 1-π for cross-country dependence. Standard deviations are reported between parentheses for all parameters estimated directly from the model. The last raw reports the diagonal elements of the transition probability matrix.

Cross-Country (US-CA) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
				τ	TDC($(1-\pi)\tau$)
US Equity - CA Equity		0.8739		0.9100	0.7917
		(0.1560)		(0.0185)	
US Bond - CA Bond		0.3870		0.6234	0.5424
		(0.0831)		(0.0124)	
			1-π	0.6897	
Cross-Asset (Equity-Bond) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
	US Bond	CA Bond		τ	TDC($\pi\tau$)
US Equity	-0.1101	0.1234	US Equity - US Bond	0.1300	0.0169
	(0.0416)	(0.0312)		(0.041)	
CA Equity	-0.0812	0.4085	CA Equity - CA Bond	0.1385	0.0180
	(0.0207)	(0.0103)		(0.0145)	
			π	0.3102	
				(0.0207)	
Parameters of transitional probability matrix					
	P	0.9020	Q	0.9586	
		(0.0207)		(0.0206)	

Table 6: Dependence structure between France and Germany in equity and bond markets. Correlation coefficients are reported for the normal regime, while tail dependence coefficients describe the asymmetric regime. The tail dependence coefficient is obtained as the product of parameter τ and the respective weight π for cross-asset dependence and 1-π for cross-country dependence. Standard deviations are reported between parentheses for all parameters estimated directly from the model. The last raw reports the diagonal elements of the transition probability matrix.

Cross-Country (FR-DE) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
				τ	TDC($(1-\pi)\tau$)
FR Equity - DE Equity	0.9083			0.9554	0.7787
	(0.0267)			(0.0603)	
FR Bond - DE Bond	0.9901			0.8261	0.6733
	(0.058)			(0.027)	
			1-π	0.8151	
Cross-Asset (Equity-Bond) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
	FR Bond	DE Bond		τ	TDC($\pi\tau$)
FR Equity	0.1893	0.2023	FR Equity - FR Bond	0.0923	0.0171
	(0.0170)	(0.0129)		(0.028)	
DE Equity	0.1175	0.1294	DE Equity - DE Bond	0.0969	0.0179
	(0.0214)	(0.030)		(0.029)	
			π	0.1849	
				(0.0294)	
Parameters of transitional probability matrix					
	P	0.8381	Q	0.9373	
		(0.0270)		(0.0373)	

Table 7: Subperiod I (period before the introduction of the Euro currency: from January 01, 1985 to December 29, 1998 for a sample of 731 observations). Dependence structure between France and Germany in equity and bond markets. Correlation coefficients are reported for the normal regime, while tail dependence coefficients describe the asymmetric regime. The tail dependence coefficient is obtained as the product of parameter τ and the respective weight π for cross-asset dependence and 1-π for cross-country dependence. Standard deviations are reported between parentheses for all parameters estimated directly from the model. The last raw reports the diagonal elements of the transition probability matrix.

Cross-Country (FR-DE) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
				τ	TDC((1-π)τ)
FR Equity - DE Equity	0.6924			0.9554	0.8663
	(0.0760)			(0.035)	
FR Bond - DE Bond	0.9082			0.8388	0.7606
	(0.038)			(0.061)	
			1-π	0.9067	

Cross-Asset (Equity-Bond) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
	FR Bond	DE Bond		τ	TDC($\pi\tau$)
FR Equity	0.2091	0.1641	FR Equity - FR Bond	0.1130	0.0105
	(0.0123)	(0.0151)		(0.021)	
DE Equity	0.1205	0.1519	DE Equity - DE Bond	0.0067	0.0006
	(0.0106)	(0.049)		(0.072)	
			π	0.0933	
				(0.010)	

Parameters of transitional probability matrix			
P	0.0651	Q	0.9438
	(0.0103)		(0.0102)

Table 8: Subperiod II (period after the introduction of the Euro currency: from January 05, 1999 to December 21, 2004 for a sample of 313 observations). Dependence structure between France and Germany in equity and bond markets. Correlation coefficients are reported for the normal regime, while tail dependence coefficients describe the asymmetric regime. The tail dependence coefficient is obtained as the product of parameter τ and the respective weight π for cross-asset dependence and 1-π for cross-country dependence. Standard deviations are reported between parentheses for all parameters estimated directly from the model. The last raw reports the diagonal elements of the transition probability matrix.

Cross-Country (FR-DE) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
				τ	TDC($(1-\pi)\tau$)
FR Equity - DE Equity	0.9426			0.2598	0.2582
	(0.0950)			(0.0106)	
FR Bond - DE Bond	0.9937			0.8946	0.8892
	(0.0382)			(0.071)	
			1-π	0.9940	
Cross-Asset (Equity-Bond) Dependence					
Normal Regime Correlation Coefficient			Asymmetric Regime Tail Dependence Coefficient		
	FR Bond	DE Bond		τ	TDC($\pi\tau$)
FR Equity	0.2272	0.2350	FR Equity - FR Bond	0.2249	0.0013
	(0.0241)	(0.0177)		(0.024)	
DE Equity	0.1516	0.1573	DE Equity - DE Bond	0.9760	0.0059
	(0.0118)	(0.059)		(0.082)	
			π	0.0060	
				(0.012)	
Parameters of transitional probability matrix					
P	0.9212		Q	0.2274	
	(0.0118)			(0.0117)	

Table 9: Monte Carlo Tests of Asymmetric Dependence. LR is the likelihood ratio statistic computed from the data. The $p-value$ is obtained from 1000 Monte Carlo repetitions with size 1043 (equal to the sample size) each.

	US-Canada	France-Germany
LR	0.0731	0.7889
$p-value$	0.0090	0.0000

Table 10: Longin and Solnik (2001) likelihood ratio test for extreme dependence correlation equal to zero at different thresholds. We apply this test on data, the regime switching model of Ang and Chen (2002), and the rotated Gumbel copula. we estimate the RS model and rotated Gumbel copula model and use estimates to simulate 10 000 Monte Carlo replications. We then perform the test on these replications.

	RS Model		Data		Rotated Gumbel copula	
Threshold	LR	p-value	LR	p-value	LR	p-value
0.10	0.7800	0.3771	1.5501	0.2131	0.4091	0.5224
0.20	2.2650	0.1323	1.5550	0.2124	8.6980	0.0032
0.30	16.7210	0.0000	8.0980	0.0044	14.4370	0.0001
0.40	22.3550	0.0000	30.9550	0.0000	27.6261	0.0000
0.60	15.5351	0.0001	285.1200	0.0000	258.9300	0.0000
0.70	10.8120	0.0010	168.6500	0.0000	219.2812	0.0000
0.80	7.2661	0.0070	69.1500	0.0000	71.2000	0.0000
0.90	3.4170	0.0645	20.3500	0.0000	29.7101	0.0000

Table 11: Results of Monte Carlo estimation of different weights for a canadian investor portfolio with bond and equity from Canada and US. The panel A represents the share of home assets in the whole portfolio, the panel B and C the bond share in the Canada and US portfolio respectively, while the panel D represents Canada portfolio share with 10 percent adjustment for foreign (US) perceived risk. for 100 simulations we present the median as the estimate and also the minimum and maximum.

	Gaussian			Asymmetric		
Gamma	Median	Minimum	Maximum	Median	Minimum	Maximum
Panel A: Canada portfolio share						
1	-2.8677	-5.0000	-0.3824	-3.5974	-5.0000	-0.7353
3	-0.2875	-1.1575	0.5314	-0.3384	-1.2916	0.5876
7	0.4468	0.0775	0.7916	0.5917	0.1788	0.9740
10	0.6051	0.3533	0.8496	0.8022	0.5119	1.0672
20	0.7905	0.6700	0.9158	1.0677	0.9146	1.1974
Panel B: bond share in the Canada portfolio						
1	-0.6066	-1.7036	0.5019	-0.7865	-2.3595	0.9861
3	0.4313	0.0705	0.8057	0.4234	-0.0934	1.0083
7	0.7276	0.5742	0.8924	0.7693	0.5486	1.0172
10	0.7931	0.6849	0.9117	0.8488	0.6935	1.0206
20	0.8684	0.8135	0.9341	0.9464	0.8656	1.0294
Panel C: bond share in the US portfolio						
1	-1.2634	-2.2837	-0.1876	-1.6277	-3.0289	-0.1749
3	0.1867	-0.1487	0.5422	0.1091	-0.3511	0.5857
7	0.5999	0.4562	0.7505	0.6024	0.4038	0.8030
10	0.6922	0.5919	0.7974	0.7136	0.5731	0.8519
20	0.7996	0.7495	0.8542	0.8416	0.7702	0.9095
Panel D: Canada portfolio share with 10 percent adjustment for perceived risk						
1	-1.8560	-4.1211	0.5242	-2.3029	-4.8884	-0.0394
3	0.1487	-0.5973	0.9319	0.1638	-0.6703	0.8984
7	0.7193	0.4023	1.0474	0.8688	0.5121	1.1736
10	0.8449	0.6256	1.0728	1.0312	0.7800	1.2406
20	0.9928	0.8821	1.1003	1.2334	1.1034	1.3353

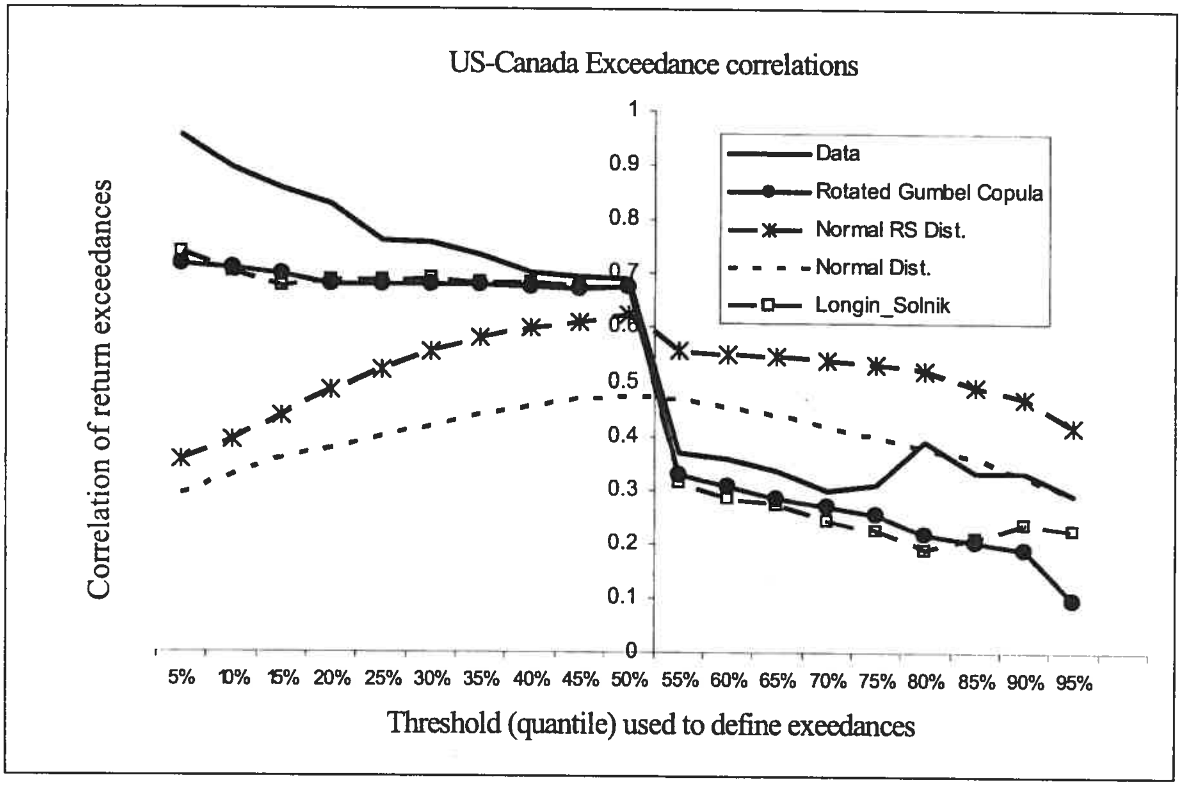


Figure 1: Calculates correlations from US-Canada equity returns data for different values of threshold θ, which is normalized. For θ less than 50% the correlation is calculated for left tail and for θ greater than 50%, the correlation is calculated for right tail. $\theta = 80\%$ means that we calculate the correlation conditional on 20% greatest observations for both U.S. and Canadian equity returns, and $\theta = 10\%$ means that we calculate the correlation conditional on 10% lowest observations for both U.S. and Canadian equity returns. Solid line represents the exceedance correlations calculated directly from data. For Rotated Gumbel Copula (with Gaussian Margins), Normal Regime Switching Distribution, and Normal Distribution, we first estimate the model and use estimates to generate 50 000 Monte Carlo simulations to calculate correlations. Longin_Solnik exceedances correlations are obtained by Longin and Solnik (2001) estimation method.

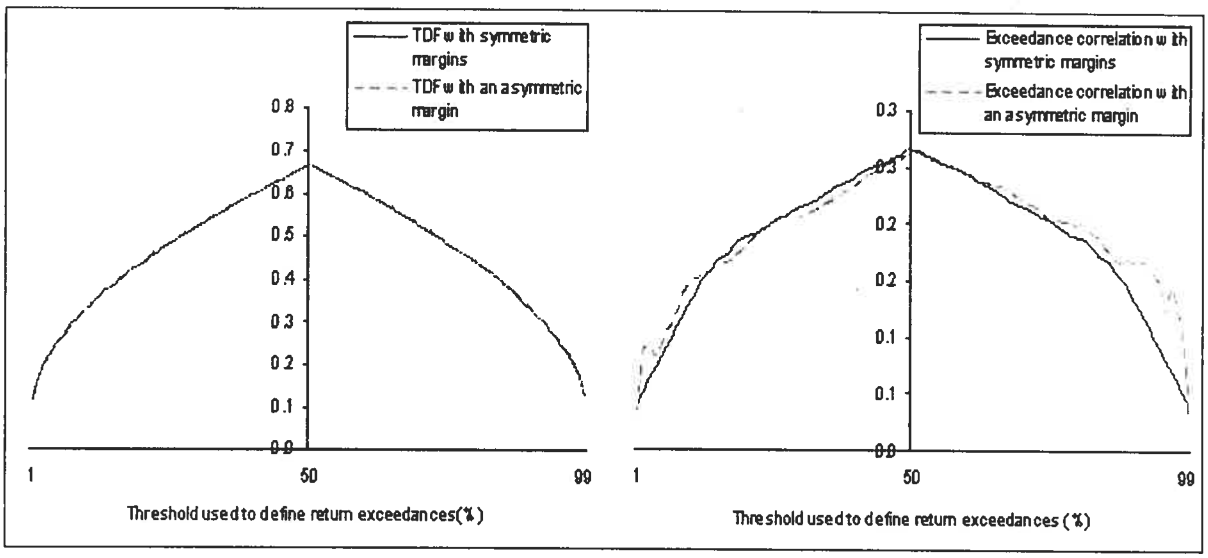


Figure 2: Effect of marginal distribution asymmetry on Tail Dependence function and Exceedance correlation: Firstly we simulate standard bivariate Gaussian distribution with correlation 0.5 and compute TDF and Exceedance correlation. Secondly, we create asymmetry in one marginal distribution by replacing the $N(0,1)$ by a mixture of $N(0,1)$ and $N(4,4)$ with equal weight.

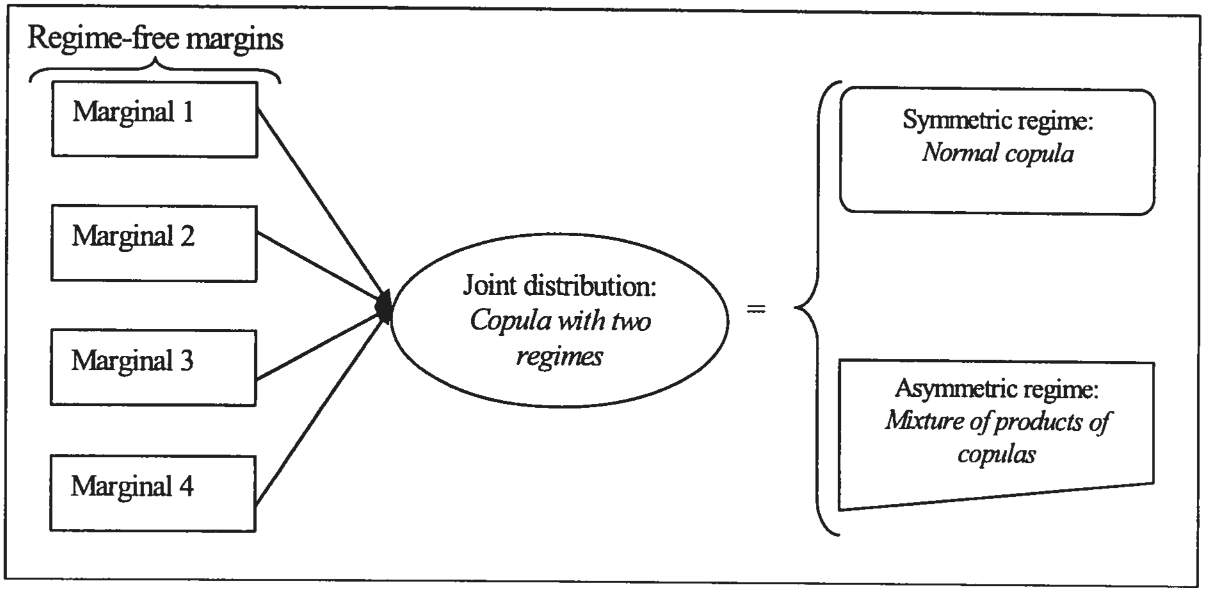


Figure 3: Model structure: Disentangling marginal distributions from the dependence structure with a two-regime copula, with one symmetric regime and one asymmetric regime. The marginal distributions are regime-free.

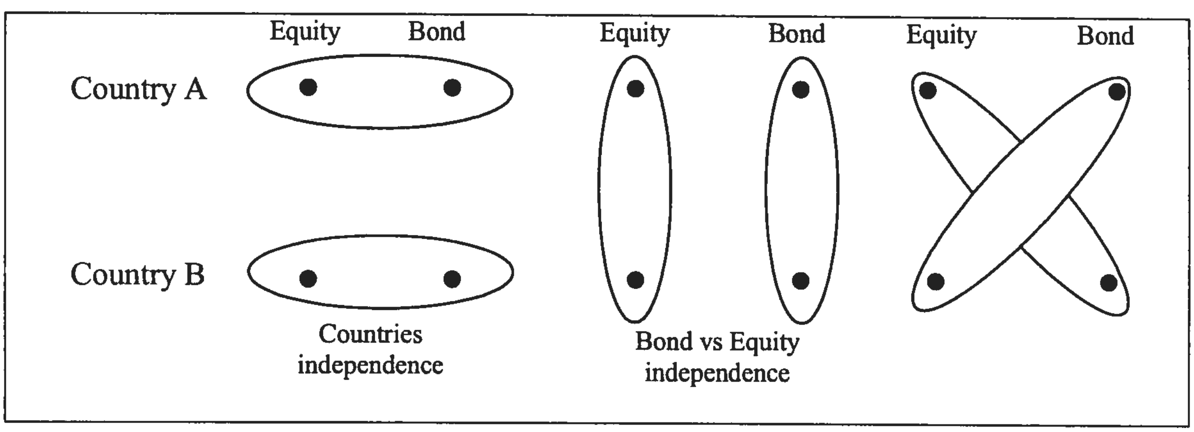


Figure 4: Illustration of the three components of asymmetric copula. Each component is the product of the two bivariate copulas representing the corresponding encircled couple of returns.

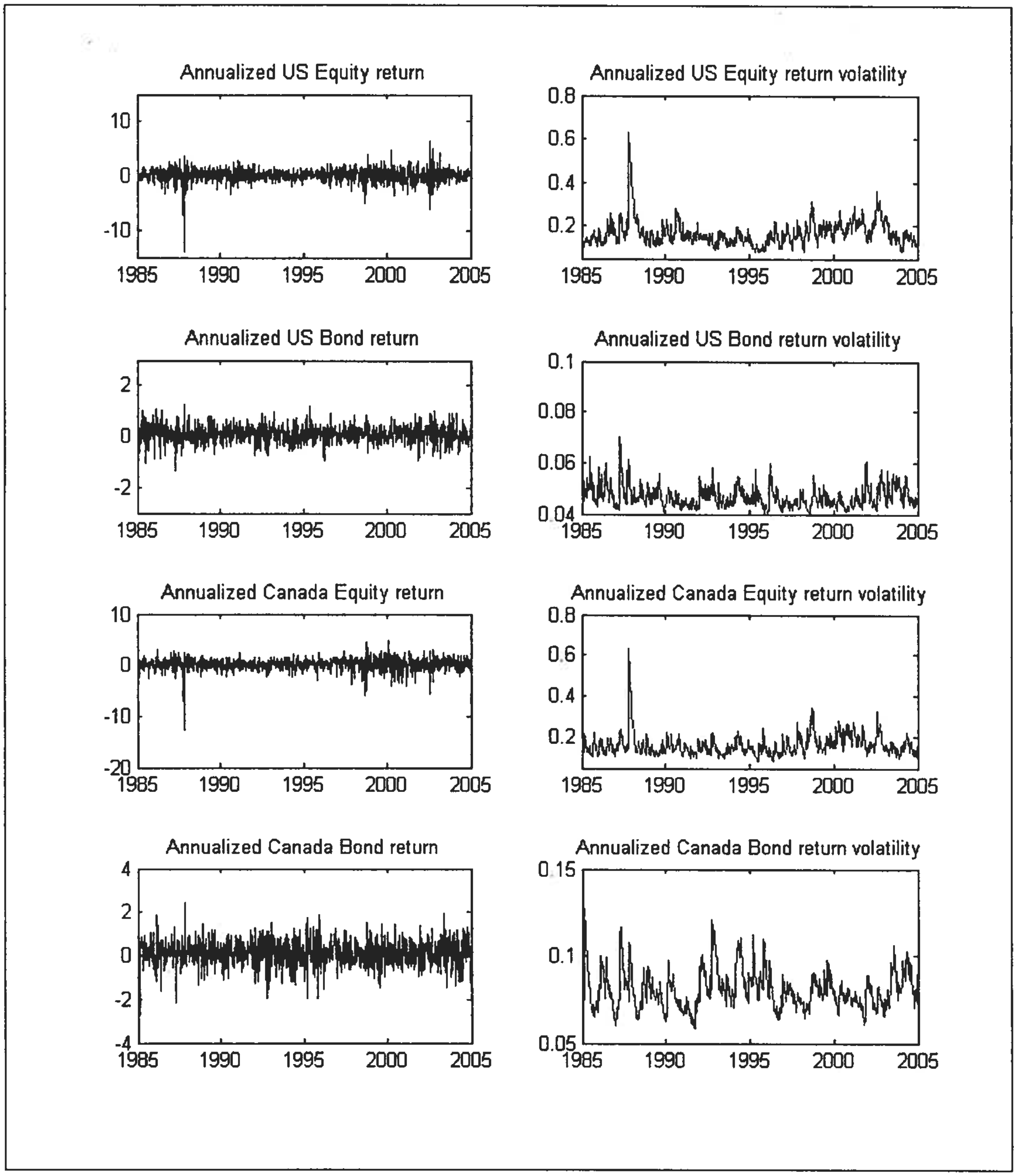


Figure 5: Annualized bond and equity returns time series for US and Canada, with their conditional volatilities obtained using the M-GARCH (1,1).

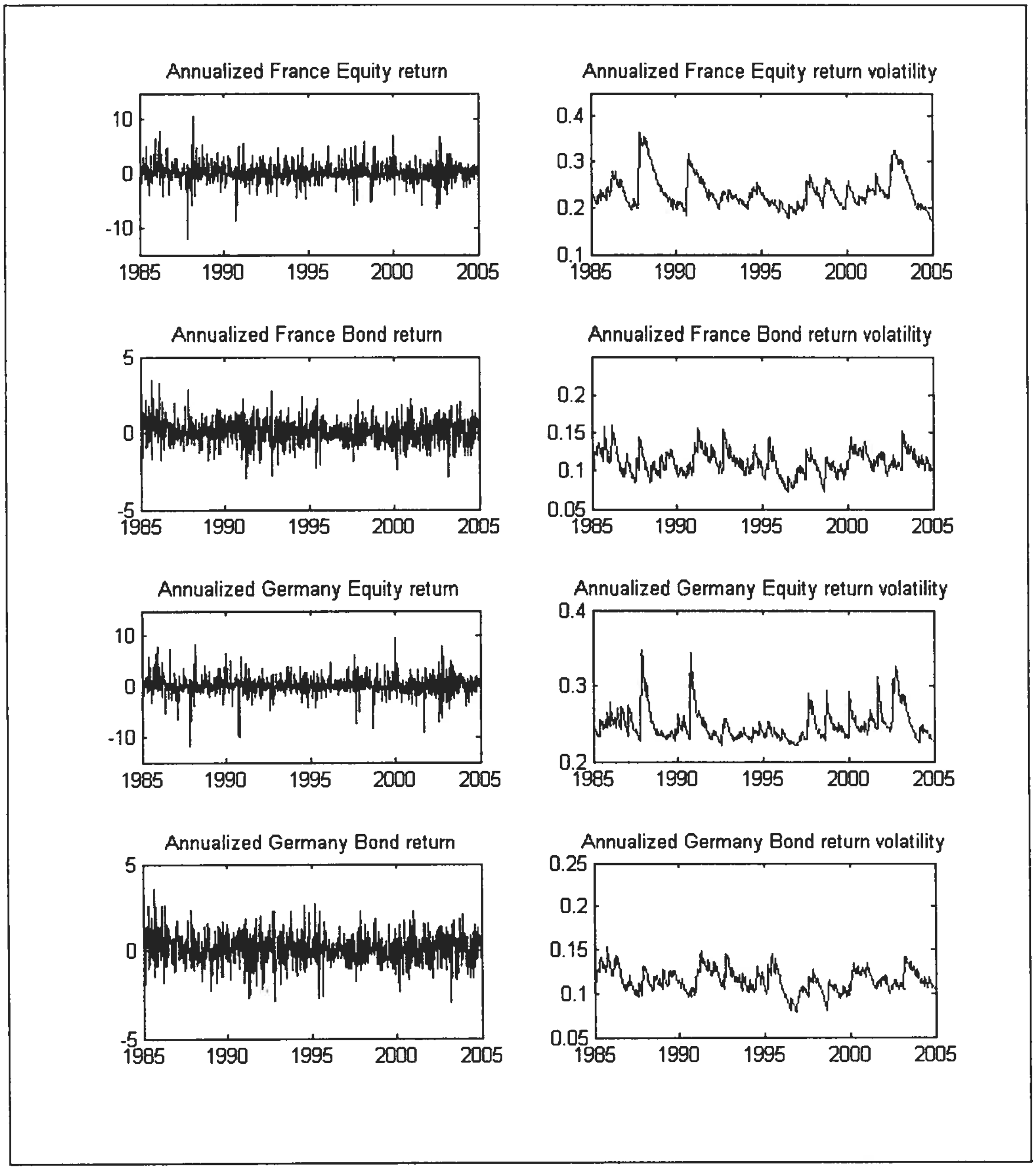


Figure 6: Annualized bond and equity returns time series for France and Germany, with their conditional volatilities obtained using the M-GARCH (1,1).

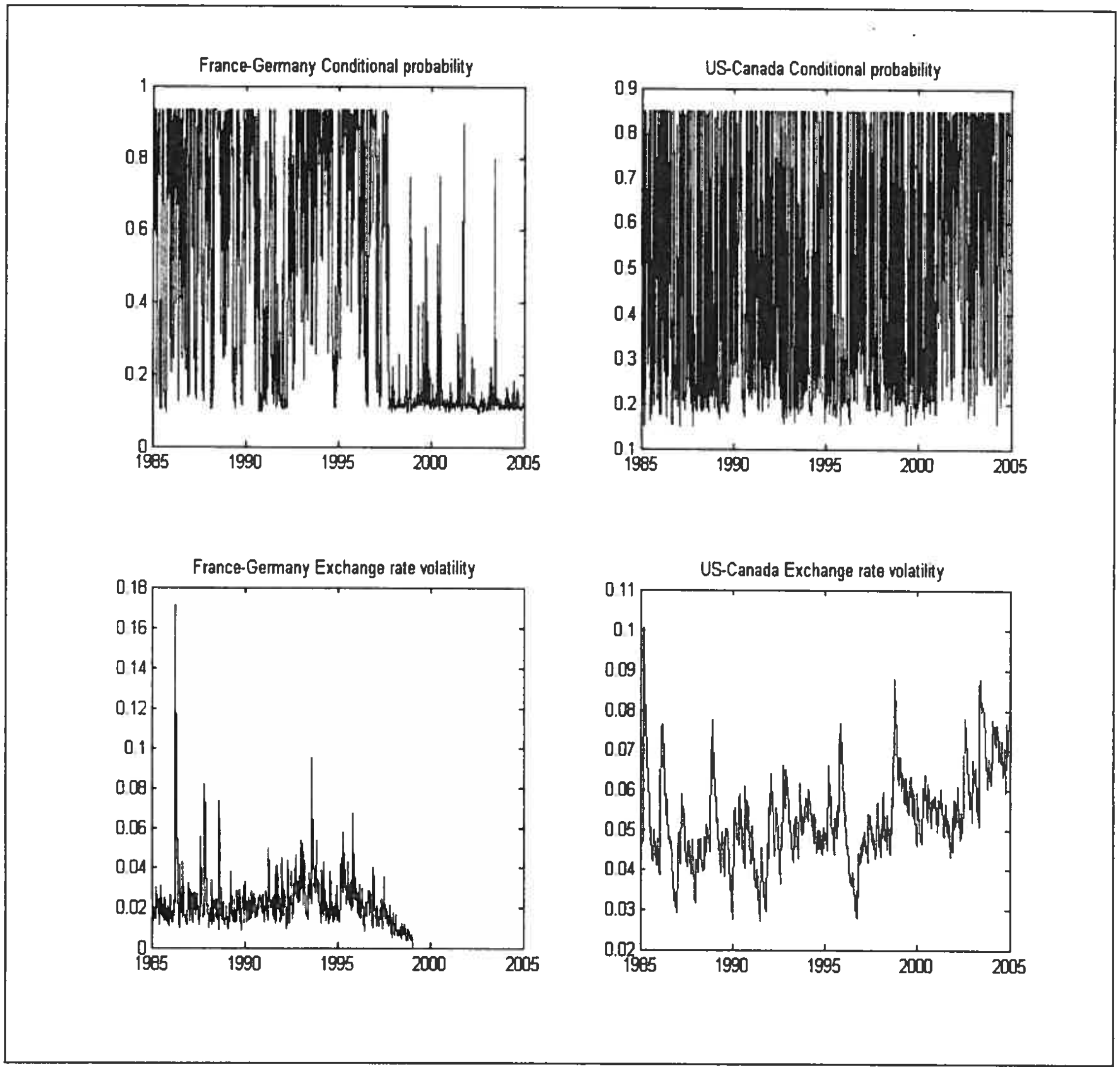


Figure 7: Conditional probability denotes the probability to be in asymmetric regime conditional to available information. Exchange rate volatility is the conditional volatility filtered with the M-GARCH (1,1) model.

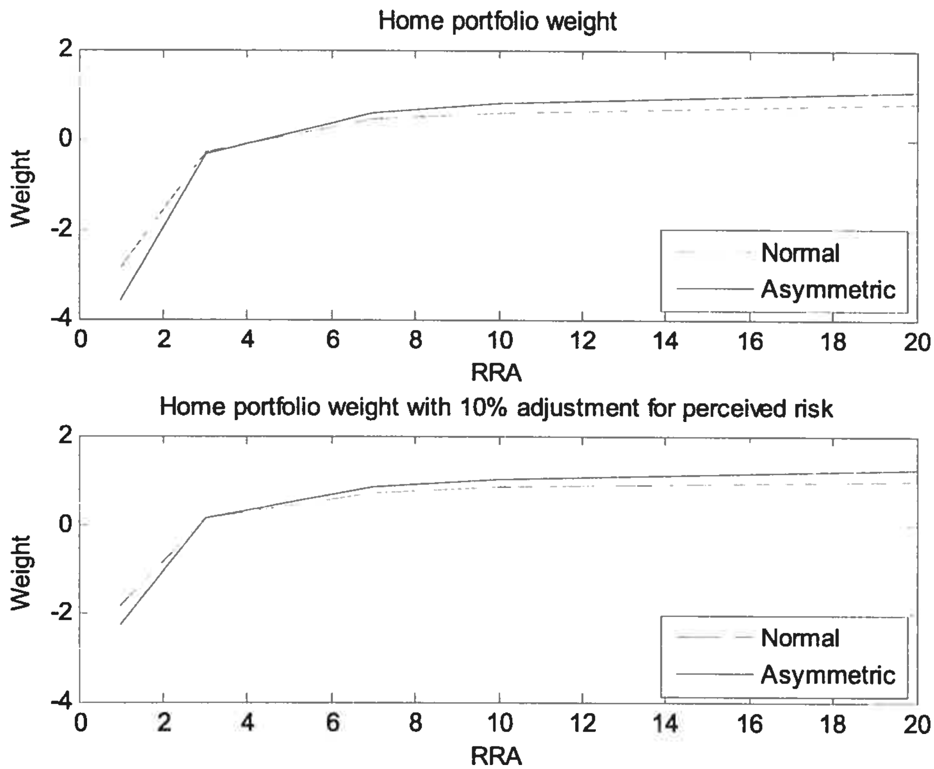


Figure 8: Canadian (home) portfolio share inside the international portfolio including both bonds and equities from Canada and US. The dash line represents optimal weight under normal dependence regime, while solid line represents optimal weight under asymmetric dependence structure. The second graph assume 10% more perceived risk for foreign (US) market. The standard deviations for US bond and equity are multiplied be 1.1.

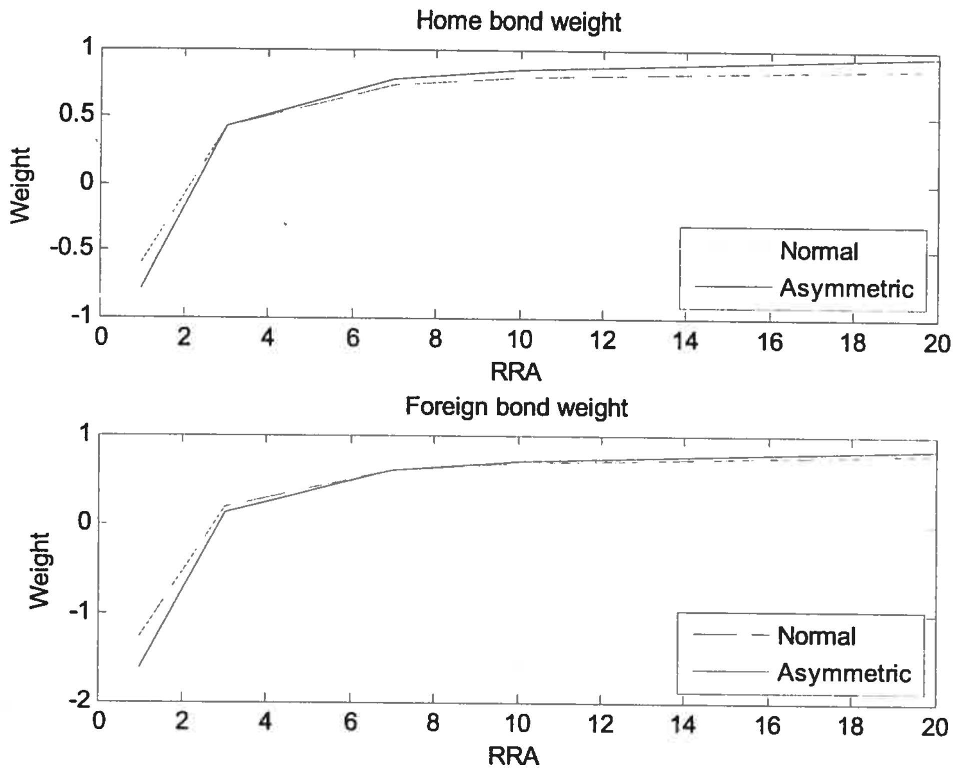


Figure 9: The bond share inside home (Canadian) portfolio and foreign (US) portfolio. The dash line represents optimal weight under normal dependence regime, while solid line represents optimal weight under asymmetric dependence structure.

Chapter 2

Asymmetric Dependence Implications for Extreme Risk Management

1. Introduction

The value-at-risk (VaR) which defines the maximum loss on an investment over a specified time horizon at a given confidence level is become the key measure for financial risk used by many banks and financial institutions. Despite its widespread adoption by the financial community, the VaR concept faces some criticisms[1]. To respond to these criticisms a complementary measure was proposed. Called expected shortfall (ES) it is the average loss when the VaR is violated. Both risk measures are very useful to manage extreme risk. However, the issue of their accurate estimation remains a challenge. In fact the estimates of VaR and ES depend on the distribution model. For the univariate models, Aas and Hobæk Haff (2006) find that GH skewed Student-t distributions have a better data fit for the univariate skewed financial returns.[2] In the multivariate parametric framework, the dynamic conditional correlation (DCC) of Engle (2002) is commonly used to model the dynamics of dependence. In this setup, the Gaussian or Student-t distribution assumption is usually made. Although these models are easy to implement and can give satisfactory results in many situations, they can be seriously misleading in the case where data exhibit strong asymmetric dependence.

Actually, one of the most important empirical facts observed in multivariate financial returns, is a much stronger correlation between equity returnsin bear markets than in normal or boom phases. This phenomenon known as asymmetric correlation or more generally asymmetric dependence has been analyzed by Longin and Solnik (2001), Ang and Bekaert (2002), Ang and Chen (2002), Garcia and Tsafack (2007), and references therein.

This paper examines the implications of ignoring this fact when implementing a model to estimate extreme risks such as VaR and ES. We show that by using multivariate Gaussian laws to characterize conditional distributions when a strong asymmetric dependence is present, one will

[1] Artzner et al. (1999) set four criteria that any coherent measure must satisfy and show that the VaR violates one of them. Basak and Shapiro (2001) also find that investors who just care about the VaR will take positions such that when the VaR is violated the loss can be very extreme

[2] The α-stable distributions, which are similar to skewed-t distributions, but with a more flexibility in the tails control should also be good candidates to modeling univariate skewed data (see Garcia Renault and Veredas, 2004).

underestimate VaR and ES. The strong dependence for low returns increases the downside risk and this additional risk cannot be captured by the Gaussian distribution. By introducing a lower tail dependence, the Student-t distribution corrects this shortcoming of the Gaussian distribution. However, the symmetric property of the Student-t means also the same dependence in the upper tail and this will reduce the downside risk. The risk model that takes into account asymmetric dependence should allow lower tail dependence and upper tail independence as put forward by Longin and Solnik (2001).

Following Lee and Long (2002), we introduce asymmetric dependence into the DCC framework. The idea is that for Gaussian distributions, the zero correlation means independence, what is not the case for asymmetric dependence. To keep the DCC setup, we perform a transformation of the innovation vector to obtain uncorrelated variables and use a Student-t copula to capture tail dependence or more importantly, a Gumbel copula to capture the asymmetry.

The empirical investigations using US and Canada equity and bond indices, shows that at the 5% level all three dependence specifications (Gaussian, Student-t, and Gumbel) provide a good estimation of VaR. However at this same level, Gaussian and Student-t seriously underestimate the ES. For a more prudential level (1% and 0.5%), symmetric specifications tend to underestimate VaR more than they do for ES. The asymmetric dependence specification namely the Gumbel copula works well at all levels.

In the next section of this paper, we deal with the estimation of extreme risk in the multivariate framework, and provide theoretical arguments to explain why symmetric dependence tend to underestimate VaR and ES. Section 3 presents backtesting and underestimation test procedures, while the empirical analysis is carried out in section 4. Section 5 concludes.

2. Portfolio Risk and Dependence Structure Modeling

When measuring the risk of a portfolio, a nice approach should be to express it as an analytical function of individual risk levels for different components of this portfolio and their dependence parameters. It is the case for a risk measure as the variance of a portfolio which is completely defined by the variances of different components and the correlation between them. For extreme risk measures like VaR and expected shortfall, it is not always possible to write the risk level of a portfolio as a function of the risk levels of its different components. Although these extreme risk measures are simple to define, their accurate estimation can be very challenging. In fact, the VaR of a portfolio depends on two main components. The distributions of different single assets and the dependence structure between all individual assets. Since the VaR is a quantile, it takes into account the distribution shape contrary to the variance which is completely defined by just the second moment.

Formally, the value at risk $VaR_p(X)$ of a variable X for a specified time horizon and a given level p is defined by: $\Pr[X \leq -VaR_p(X)] = p$, and the related Expected Shortfall (ES) which is the average loss beyond the VaR is $ES_p(X) = -E[X | X \leq -VaR_p(X)] = -p^{-1}E[X.I(X \leq -VaR_p(X))]$.

When dealing with multivariate modeling, Gaussian or conditional Gaussian multivariate distributions are usually used. In this situation, the correlation coefficients completely characterize the dependence among the variables and the estimation of both VaR and ES is straightforward.

2.1 Gaussian Case

The most common and easy approach to modeling multivariate asset returns is to assume normality. In this case, the value at risk is completely defined by the first two moments. If X_t is a normal variable, the analytical expression of the value at risk is given by: $VaR_p(X_t) = -\left(E(X_t) + \Phi^{-1}(p)[var(X_t)]^{1/2}\right)$, where Φ denotes the cumulative standard normal distribution. For a two-asset portfolio $X_t = w_t X_{1t} + (1 - w_t) X_{1t}$, if (X_{1t}, X_{2t}) is a joint Gaussian distribution, then

$$VaR_p(X_t) = -\left(\mu_{X_t} + \Phi^{-1}(p)\sigma_{X_t}\right)$$

with

$$\begin{cases} \mu_{X_t} = [w_t\mu_{1t} + (1 - w_t)\mu_{2t}] \\ \sigma_{X_t} = \left[w_t^2\sigma_{1t}^2 + (1 - w_t)^2\sigma_{2t}^2 + 2w_t(1 - w_t)\rho_t\sigma_{1t}\sigma_{2t}\right]^{1/2} \end{cases},$$

where μ_{it} is the mean and σ_{it} the standard deviation of X_{it} and ρ_t the correlation coefficient between X_{1t} and X_{2t}. So in the case of normal distributions, the VaR of portfolio is expressed in closed form as a function of the parameters of different single asset returns and the correlation coefficient between them. Moreover, both VaR and ES are related by the expression

$$ES_p(X_t) + \mu_{X_t} = \frac{\phi\left[\Phi^{-1}(p)\right]}{p\Phi^{-1}(p)}\left(VaR_p(X_t) + \mu_{X_t}\right).$$

So, the expected shortfall is

$$ES_p(X_t) = -\mu_{X_t} + \sigma_{X_t}\frac{\phi\left[\Phi^{-1}(p)\right]}{p}$$

where ϕ denotes the density of a standard normal distribution.

2.2 Multivariate GARCH

Volatility clustering is an important stylized fact that should be taken into account when dealing with conditional distributions of asset returns. For univariate distributions, GARCH models are commonly used to forecast the conditional volatility. The straightforward generalization of the univariate GARCH model brings some problems in the estimation process. Engle (2002) introduces a new class of multivariate GARCH models which is the generalization of the Bollerslev (1990) model. In fact, to allow tractability in multivariate GARCH, Bollerslev (1990) assumes a constant conditional correlation (CCC). The dynamic conditional correlation (DCC) of Engle (2002) extends this model by allowing time variation for correlation coefficients. As Engle (2002), we use a two-step estimation procedure for all models. In the first step, we estimate parameters of marginal distributions, and use them in a second step to estimate the parameters of the dependence structure.

2.2.1 Univariate Distribution Model

Lee and Long (2006) find that for multivariate models, the choice of copula functions is more important than the choice of the volatility models. Similarly, we will focus on the effect of the dependence structure on downside risk estimation. Therefore it is necessary to use for all different multivariate models the same marginal specification. For all single asset x_{it}, we use the simple GARCH(1,1) model.

$$\begin{cases} x_{it} = \mu_{it} + \sigma_{it}\varepsilon_{it}, i = 1,2 \\ \sigma_{it+1}^2 = \omega_i + \beta_i \sigma_{it}^2 + \alpha_i \left(x_{it} - \mu_{it}\right)^2 \end{cases}$$

2.2.2 Dynamic Conditional Correlation (DCC)

Recently proposed by Engle (2002) to capture the time dynamics in the correlation, the DCC model is become a benchmark model for multivariate specifications. One of the attractive points of this model is its flexibility in term of the specification of marginal distributions separately from the dependence structure. In our context of bivariate models, the GARCH(1,1)-type specification of conditional correlation coefficient $\rho_t = corr\left(x_{1t}, x_{2t}\right)$ is the following.

$$\rho_t = \frac{q_{12,t}}{\sqrt{q_{11,t}q_{22,t}}}$$

with the auxiliary variable $q_{ij,t}$ defined by the dynamic

$$q_{ij,t+1} = \overline{\rho}_{ij} + a\left(\varepsilon_{it}\varepsilon_{jt} - \overline{\rho}_{ij}\right) + b\left(q_{ij,t} - \overline{\rho}_{ij}\right).$$

The conditional value at risk $VaR_t^p(X_t)$ is defined by $\Pr[X_t \leq -VaR_t^p(X_t)|\,\mathcal{F}_{t-1}] = p$, where $\mathcal{F}_{t-1}$ is the information set available. For a Gaussian or a conditional Gaussian distribution, the importance of dependence is completely driven by the correlation coefficient. However, for a more complex dependence structure, even if all components are individually conditional Gaussian (but not jointly), the portfolio VaR becomes more complex and depends on the shape of the dependence structure. The portfolio return distribution can be very complex. In fact, the distribution of a linear combination involves convolution for which it is difficult to get an analytical expression. In a such context, the most common technique used is the Monte Carlo simulations and the accuracy of estimation will depend on how the dependence model fits the data.

By writing

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \rho_t & \sqrt{1-\rho_t^2} \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix},$$

we have $corr(\eta_{1t}, \eta_{2t}) = 0$. Then, if (x_{1t}, x_{2t}) is a bivariate normal vector, the zero correlation means that η_{1t} and η_{2t} are independent. However, if the joint normality is not valid as it is the case in many practical situations, the zero correlation does not necessary mean independence. This is the case of asymmetric dependence.

2.3 Asymmetric Dependence Distributions and Extreme Risk Measures

Lower returns are more dependent than upper returns in financial markets, especially in international asset markets. Longin and Solnik (2001) investigate the structure of correlation between various equity markets in extreme situations and find that equity markets exhibit a much higher correlation in extreme bear periods and zero correlation for asymptotic upper returns. Garcia and Tsafack (2007) by using tail dependence functions extend this analysis in terms of nonlinear dependence and find similar results.

2.3.1 Beyond Symmetric Dependence: Copula

Any bivariate distribution is defined by its marginal univariate distributions and its dependence structure between both variables. To completely characterize the dependence structure, we use copulas which are functions that build multivariate distribution functions from their unidimensional marginal distributions. Let $X \equiv (X_1, X_2)$ be a vector of two variables. Denoting F the joint distribution function and F_1 and F_2 the respective marginal distributions of X_1 and X_2. The Sklar theorem[3] states that there exists a function C called copula which joins F to F_1 and F_2 as follows.

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \qquad (2.1)$$

Equivalently the copula function C is directly defined as follows.

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right); \qquad u_1, u_2 \in [0, 1] \qquad (2.2)$$

[3] See Nelsen (1999) for a general presentation. Note that if F_i is continuous for any $i = 1, ..., n$ then the copula C is unique.

The nice thing about copulas is that we can model some stylized facts like asymmetric dependence without changing the DCC specification. By continuing to keep the zero correlation between η_{1t} and η_{2t}, we can use the Gumbel copula to model their dependence structure.

2.3.2 Underestimation of VaR and ES

Practitioners usually use the Gaussian (or symmetric) distributions in the computation of the extreme risks of their portfolios. In fact, if data exhibit asymmetric dependence, the use of the Gaussian distribution will be misleading. The issue of interest is to know on which side will the dependence be stronger. In other words, if the extreme risk will be overestimated or underestimated. In this section we will develop some theoretical arguments that give us intuition about why the Gaussian dependence structure underestimates these risks.

By using copulas, any bivariate cumulative distributional function can be represented with three elements as $F \equiv (F_1, F_2, C)$ such that $F(X_1, X_2) = C(F_1(X_1), F_2(X_2))$, with F_i be the cumulative distribution function of X_i, and C the copula function of (X_1, X_2). We want to analyze the effect of the third element which characterizes the dependence structure on the extreme risk measures. It is therefore relevant to keep the first and second elements which characterize the marginal distributions unchanged.

Definition 1. *(Stochastic ordering,: $\prec^{st}$ Joe, 1997). $F' \prec^{st} F$ if $\int g dF' \leq \int g dF$ for all increasing functions g for which the expectations exist*

The concept of stochastic ordering is equivalent to stochastic dominance in the case of univariate distribution.[4] The result below can be seen as an extension of the mean preserving spread to the multivariate case.[5]

Proposition 1. *Let $(X_1, X_2) \rightsquigarrow F \equiv (F_1, F_2, C)$ and $(X_1', X_2') \rightsquigarrow F' \equiv (F_1, F_2, C')$.*

[4] For univariate distributions F and F', by taking $g(v) = -I(v \leq x)$ which is an increasing function, we have $\int g dF' \leq \int g dF$ implies $F'(x) \geq F(x)$. So $F' \prec^{st} F$ is equivalent to $F' \geq F$.

[5] Here we are focusing on bivariate distributions for the sake of presentation. The proof of this result in the appendix is valid for any dimension.

Denote $X = wX_1 + (1-w)\,X_2$ *and* $X' = wX_1' + (1-w)\,X_2'$ $w \in [0,1]$

If $C' \prec^{st} C$ *then*

i) $VaR_p\,(X') \geq VaR_p\,(X)$

ii) $ES_p\,(X') \geq ES_p\,(X)$

Proof see Appendix

This result allows us to understand why the extreme risk measure in a left dependence asymmetric distribution should be larger than the one measured in a symmetric distribution with low dependence in the lower tail. The result below compares a normal copula with a rotated Gumbel copula.

Definition 2. (*Tail dependence*). *For a copula* C, *the lower tail dependence coefficient is*

$$\tau^L \equiv \lim_{u \to 0} \frac{C\,(u,u)}{u}$$

while the upper tail dependence coefficient is

$$\tau^U\,(\alpha) \equiv \lim_{u \to 1} \frac{1 - 2u + C\,(u,u)}{u}$$

The tail dependence coefficient measures the level of dependence between extreme events for both variables. For a Gaussian copula, regardless the level of correlation, both the upper and the lower tail dependences are zero. However, for a rotated Gumbel copula, the upper tail dependence is zero, while the lower tail dependence is strictly positive except in the case of independence. From this observation, we can compare both copulas in term of stochastic ordering.

Proposition 2. *Let* C_N *be a normal copula and* C_{rG} *be a rotated Gumbel copula. If they are stochastically ordered,*

then $C_{rG} \prec^{st} C_N$

Proof see Appendix

It is not always possible to compare a Gaussian copula and a rotated Gumbel copula, however when it is the case, the rotated Gumbel copula is lower ordered than the Gaussian one. This result can not be straightforwardly extended to the comparison of a Student-t copula with a rotated Gumbel copula, since the tail dependence coefficient of a Student-t copula is not necessary zero. But the fact that a Student-t copula is symmetric suggests that a rotated Gumbel copula should put more probability in the left tail than a Student-t does. In fact, the level of tail dependence for a Student-t copula depends on the correlation coefficient and the degree of freedom. For a zero correlation, the Student-t copula can have a tail dependence and this tail dependence decreases when the degree of freedom increases and tends to zero for an infinite degree of freedom, since it corresponds to the convergence of the Student-t distribution to the Gaussian distribution.

3. Testing and Comparison

It is important to perform a number of specification search to find the dependence functions which provide the best fit for a data set. Using a likelihood, Aikaike and Schwarz's Bayesian information criteria we compare the dependence model goodness-of-fit. The accuracy of distribution models to estimate risk is assessed furthermore by backtesting using unconditional and conditional coverage. Finally we test the assumption of underestimation of VaR and ES in three dependence structure specifications (Gaussian, Student-t and Gumbel).

3.1 Backtesting

We present below the backtesting procedure for unconditional and conditional coverage. The first one tests the accuracy of violation probability assuming independence between successive violations, while conditional coverage also tests the independence.[6]

3.1.1 Unconditional Coverage

Let $I_t = I\left(X_t \leq -VaR_p\left(X_t\right)\right)$ be the indicator of violation or not. The value 1 for I_t indicates that a violation occurred, while the value 0 means no violation. For a well specified risk model, the

[6] A complete presentation can be found in Christoffersen (2003).

time series $\{I_t\}_{t=1,\ldots T}$ should be time independent and follow identical Bernoulli distribution with the probability p for 1. So we test the null hypothesis

$$H_0 : I_t \rightsquigarrow i.i.d.\text{Bernoulli}(p)$$

against

$$H_{1uc} : I_t \rightsquigarrow i.i.d.\text{Bernoulli}(\pi)\,, \pi \neq p$$

The likelihood function is analytically derived and the likelihood ratio test can be easily performed. Denote

$$\begin{cases} T_1 = \sum_{t=1}^{T} I_t \\ T_0 = \sum_{t=1}^{T} (1 - I_t) = T - T_1 \end{cases},$$

the (log) likelihood ratio statistic is given by

$$LR_{uc} = -2\ln\left[(1-p)^{T_0} p^{T_1} \Big/ \left((1 - T_1/T)^{T_0} (T_1/T)^{T_1}\right)\right] \rightsquigarrow \chi_1^2$$

3.1.2 Conditional Coverage

The independence assumption can also be tested along with probability of violation accuracy. Therefore the alternative hypothesis should include the time dependence between I_t and I_{t+h}. We test

$$H_0 : I_t \rightsquigarrow i.i.d.\ \text{Bernoulli}(p)$$

against

$$H_{1cc} : I_t \rightsquigarrow \text{Markov}(\Pi_1)\,, \text{ with } \Pi_1 = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}$$

By denoting T_{ij} the number of observations of I_t with an i value followed by a j value. Formally,

$$\begin{cases} T_{11} = \sum_{t=1}^{T-1} I_t I_{t+1} \\ T_{10} = \sum_{t=1}^{T-1} I_t (1 - I_{t+1}) \\ T_{01} = \sum_{t=1}^{T-1} (1 - I_t) I_{t+1} \\ T_{00} = \sum_{t=1}^{T-1} (1 - I_t)(1 - I_{t+1}) \end{cases}$$

The likelihood ratio for this hypothesis is given by

$$LR_{cc} = \begin{cases} -2\ln\left((1-p)^{T_0} p^{T_1} \Big/ (1-\widehat{\pi}_{01})^{T_{00}} \widehat{\pi}_{01}^{T_{01}} (1-\widehat{\pi}_{11})^{T_{10}} \widehat{\pi}_{11}^{T_{11}}\right), & \text{if } \widehat{\pi}_{11} \neq 0 \\ \\ -2\ln\left((1-p)^{T_0} p^{T_1} \Big/ (1-\widehat{\pi}_{01})^{T_{00}} \widehat{\pi}_{01}^{T_{01}}\right), & \text{if } \widehat{\pi}_{11} = 0 \end{cases}$$

where $\widehat{\pi}_{ij} = T_{ij}/(T_{i0}+T_{i1})$ is the estimated probability of a j follows an i. The asymptotic distribution of LR_{cc} is a χ_2^2.

3.2 Testing the Extreme Risk Underestimation

For a risk manager, assessing the accuracy of the risk model is important and knowing if the model underestimates or not the risk is very important, since the underestimation can be very costly. We propose below simple ways to assess the underestimation of both value at risk and expected shortfall.

3.2.1 A Simple Test of VaR Underestimation

When the model is well specified the proportion of violations is exactly p. Formally by denoting $\pi = E[I_t]$, the expected proportion of violations with VaR computed under the true data generating process, we should have $\pi = p$. However, if the value at risk is underestimated, the expected number of violations will be larger than p, i.e. $\pi > p$. Then the testing hypothesis is

$$H_0 : \pi = p$$

$$\text{against}$$

$$H_1 : \pi > p$$

The estimator of π is $\widehat{\pi}_T = \frac{1}{T}\sum_{t=1}^{T} I_t$. Under the null, we have $I_t \rightsquigarrow i.i.d.$ Bernoulli(p). Therefore, it is a simple and well-known unilateral test for the mean of a variable.

3.2.2 A Simple Test of ES Underestimation

If we are using the correct risk model the expected shortfall should be exactly the mean of the returns in the portfolio tail with probability p. Therefore the regression of $(-X_t - ES_t)\, I\left(X_t \leq -VaR_p\left(X_t\right)\right)$

on a constant term, will give zero coefficient. So, if we perform the regression

$$-X_t - ES_t = c + \varepsilon_t, \qquad \text{where } X_t \leq -VaR_p\left(X_t\right).$$

A correct model will produce zero for the estimate of c, while a model which underestimates the expected shortfall will produce $c > 0$. We then want to test the hypothesis

$$H_0 : c = 0$$

against

$$H_1 : c > 0$$

Once again, this is a simple unilateral test based on a Student-t statistic. The estimate of the parameter c is proportional to the difference between the sample mean and the model mean both computed for losses beyond the VaR.

4. Empirical Investigation

4.1 Data

We use equity and bond index data for US and Canada. The US equity returns are based on the SP 500 index, while the Canadian equity returns are computed with the Datastream index. The bond series are indices of five-year government bonds computed by Datastream. These bond indices are available daily and are chain linked allowing the addition and removal of bonds without affecting the value of the index. All returns are expressed in US dollar on a weekly basis from January 01, 1985 to December 21, 2004, which corresponds to a sample of 1044 observations

To perform VaR forecasting, we split the full sample in two part[7]. The in-sample period starts from January 01, 1985 to December 23, 2003 and the out-of-sample period from December 23, 2003 to December 21, 2004. Descriptive statistics are provided in tables 1&3.

[7]The models are estimated in-sample and the out-of-sample data allow us to assess the forecasting ability of models. We perform backtests in all sample due to the limited size of out of sample data, however this is enough to take into account the out-of-sample effects.

4.2 Dependence Structure

Using the AIC and BIC[8] criteria, we perform a comparison between two symmetric copulas (Gaussian and Student-t) and two asymmetric ones (Clayton and Gumbel[9]). Table 4 shows that the Student-t copula is a better fit for our data than the Gaussian one, while the Gumbel copula is the best of all our compared dependence functions. From the sample correlation coefficients and the left tail dependence coefficients (TDC) computed from the Gumbel copula, one can notice that the dependence between equity indices is characterized by a higher level of correlation and also a stronger level of TDC than the dependence between bond indices.

4.3 Testing Results

To backtest our risk model, we perform both unconditional and conditional coverage tests for VaR estimation. Underestimation tests are done for both VaR and ES. All these tests are performed on an equally weighted US-Canada equity and bond portfolios

4.3.1 Equity Portfolio

Given the large correlation and strong asymmetric dependence between equity indices, their portfolio results are more relevant in term of the asymmetric effect on risk estimation (see table 5). The results show that at a 5% level all three copula specification models provide good VaR estimates. For this level of violation probability, there is no need to go beyond Gaussian DCC to estimate VaR even if it is important to notice that a good risk model for VaR at a specific level does not insure good estimation for ES at the same level. In fact, the ES depends on the entire shape of the distribution beyond the VaR at a given level. In other words, to see if a risk model will give a good estimation for ES or not based on VaR, one should extend the estimation to more lower levels of violation probability. However, at a 1% violation level, the two symmetric models (Gaussian, and

[8] AIC and BIC refer to Aikaike and Schwarz's Bayesian information criteria respectively.

[9] Gumbel copula here refers to rotated (or survival) Gumbel copula. Since Gumbel copula has a left tail independence and right tail dependence, the rotated Gumbel copula exhibits a right tail independence and a left tail dependence that is the fact widely observed from data.

Student-t) fail the two tests at a 10% critical level, while the Gaussian model fails the tests at a 5% critical level. For lower violation level (0.5%) which represents in practice a very prudential risk management, the Gaussian does worse, while the Student-t improves very well. The Gumbel copula model passes both tests at all probability levels.

4.3.2 Bond Portfolio

The dependence between bond indices is relatively low and therefore the DCC model seems to work regardless of the copula specification. Even if as shown in table 6, the Gaussian copula risk model produces more violations at all probability levels, it passes all tests except only the unconditional coverage at a 10% critical level. The Student-t and Gumbel specifications perform statistically well.

4.3.3 Underestimation of VaR and ES

In the first panel of the tables 5 and 6, the nimber of violation T_1 is especially large for Gaussian dependence. Given this large frequency of violations observed with the sample estimation of coverage probability compared to the required level (5%, 1%, and 0.5%), we test the risk underestimation tendency of three models. Both VaR and ES measures are tested.

As shown in the p-value row of panel 3 table 5, the Gaussian model seriously underestimates the VaR for low levels (1% and 0.5%) for the equity portfolio, while the Student-t which underestimates the VaR at the 1% level gives a good estimation at 0.5%. Intuitively, this can be explained by the fact that the strong tail dependence of the Student-t corrects the effect of asymmetry for the lower tail distribution and reduces the magnitude of underestimation. Surprising although explicable, is the fact that the ES is more likely underestimated for a large level (5%) (see the p-value row of panel 4, table 5 and 6). In fact all models provide good estimate of VaR at 5% level, however when the shape in the tail are not fat enough, the ES will be underestimated.

The figure 1 clearly shows the ranking of the VaR level for the three dependence model specifications.[10] The Gaussian model presents lower estimates of the VaR, while the Gumbel copula

[10] We don't present the out-of-sample graphs for other violation levels (5% and .5%) because they have the same shape and same conclusion in terms of ranking than the 1% level. The ES ranking is the same like the VaR ranking.

presents upper estimates. This is essentially due to the residual downside risk that the Gaussian copula cannot capture, while the Student-t partially incorporates it and the Gumbel copula takes it into account in a more effective way.

5. Conclusion

We provide arguments to explain the fact that symmetric dependence specifications tend to underestimate extreme risk in the presence of asymmetric dependence. In the DCC framework, we find that the Gaussian and Student-t specifications perform relatively well when the correlation or the tail dependence is low. However, in the presence of strong asymmetry like it is the case for equity indices, these symmetric specifications tend to underestimate VaR and ES. Therefore for the accuracy of risk measures, it is important in presence of asymmetric dependence to use an asymmetric model such as the Gumbel copula which allows lower tail dependence and upper tail independence.

We use bivariate models in this work, what is enough to show the effects of the dependence structure on the risk measures accuracy. However in a practical context to estimate portfolio risk, one should need a large dimension of multivariate models to capture asymmetric dependence. It would be necessary to point out that the building of large dimension asymmetric copulas with more flexibility in the dependence structure among different couples remains a challenge in the statistical literature. The sensitivity to modeling asymmetry in the marginal distributions combine with asymmetric dependence would be another interesting issue.

6. Appendix A

Proof of Proposition 1.

To prove this result, we need the below lemma.

Lemma: *Let* $F \equiv (F_1, F_2, C)$ *and* $(X_1', X_2') \rightsquigarrow F' \equiv (F_1, F_2, C')$.

$C' \prec^{st} C$ *is equivalent to* $F' \prec^{st} F$.

Proof of lemma

Let us assume that $C' \prec^{st} C$ and let g be an increasing function such that $\int gdF'$ and $\int gdF$ exist. we want to show that

$F' \prec^{st} F$, i.e.$\int g(x_1, x_2)\, dF'(x_1, x_2) \leq \int g(x_1, x_2)\, dF(x_1, x_2)$

by defining $h(u, v) = g\left(F_1^{-1}(u), F_2^{-1}(v)\right)$, since F_i are increasing functions, h is also an increasing function. Therefore $C' \prec^{st} C$ implies

$\int h(u_1, u_2)\, dC'(u_1, u_2) \leq \int h(u_1, u_2)\, dC(u_1, u_2)$

i.e. $\int g\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right) dC'(u_1, u_2) \leq \int g\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right) dC(u_1, u_2)$

and therefore $F' \prec^{st} F$.

conversely by assuming $F' \prec^{st} F$ and define $g(x, y) = h(F_1(x), F_2(y))$ we have the above lemma.

Q.E.D.

with the above lemma, we have by assumption that $F' \prec^{st} F$, for $w \in [0, 1]$ by taking

i) $g(x_1, x_2) = -I(wx_1 + (1-w)x_2 \leq -VaR_p(X))$, with $I(P) = 1$ if P is true and 0 if not.

g is an increasing function and we have

$\int g(x_1, x_2)\, dF(x_1, x_2) = -p \geq \int g(x_1, x_2)\, dF'(x_1, x_2)$

then

$-\int I(wx_1 + (1-w)x_2 \leq -VaR_p(X'))\, dF'(x_1, x_2)$

$\geq -\int I(wx_1 + (1-w)x_2 \leq -VaR_p(X))\, dF'(x_1, x_2)$

i.e. $-VaR_p(X') \leq -VaR_p(X)$ or $VaR_p(X') \geq VaR_p(X)$

ii) $g(x_1, x_2) = [wx_1 + (1-w)x_2]\, I(wx_1 + (1-w)x_2 \leq -VaR_p(X))$

g is an increasing function since $VaR_p(X) \geq 0$, and we have

$$\begin{aligned}
ES_p(X) &= -E[X|X \leq -VaR_p(X)] \\
&= -p^{-1}E[X.I(X \leq -VaR_p(X))] \\
&= -p^{-1}\int g(x_1,x_2)\,dF(x_1,x_2) \\
&\leq -p^{-1}\int g(x_1,x_2)\,dF'(x_1,x_2) \\
&= -p^{-1}\int [wx_1+(1-w)x_2]\,I(wx_1+(1-w)x_2 \leq -VaR_p(X))\,dF'(x_1,x_2) \\
&\leq -p^{-1}\int [wx_1+(1-w)x_2]\,I(wx_1+(1-w)x_2 \leq -VaR_p(X'))\,dF'(x_1,x_2) \\
&= ES_p(X')
\end{aligned}$$

Q.E.D.

Proof of Proposition 2.

Let C_N be a normal copula and C_{rG} be a rotated Gumbel copula that are "stochastically ordered", we want to show that $C_{rG} \prec^{st} C_N$. Let us suppose the reverse, .i.e. $C_N \prec^{st} C_{rG}$.

Let $g(u_1,u_2) = -I(u_1 \leq u, u_2 \leq u)$ which is an increasing function. By assuming that $C_N \prec^{st} C_{rG}$ we have

$\int -I(u_1 \leq u, u_2 \leq u)\,dC_N(u_1,u_2) \leq \int -I(u_1 \leq u, u_2 \leq u)\,dC_{rG}(u_1,u_2)$

so then

$\int I(u_1 \leq u, u_2 \leq u)\,dC_N(u_1,u_2) \geq \int I(u_1 \leq u, u_2 \leq u)\,dC_{rG}(u_1,u_2)$

i.e.

$C_N(u,u) \geq C_{rG}(u,u)$.

However, since the tail dependence coefficient for normal copula $\tau_N^L = \lim_{u\to 0}\dfrac{C_N(u,u)}{u} = 0$ and for rotated Gumbel copula $\tau_{rG}^L = \lim_{u\to 0}\dfrac{C_{rG}(u,u)}{u} > 0$ then there exists $\eta \in (0,1)$ such that for any $u \in (0,\eta)$

$\dfrac{C_{rG}(u,u) - C_N(u,u)}{u} > 0$ what is equivalent to $C_{rG}(u,u) > C_N(u,u)$. This contradicts the assumption $C_N \prec^{st} C_{rG}$, so we necessary have $C_{rG} \prec^{st} C_N$

Q.E.D.

References

[1] Ang, Andrew and Joseph Chen (2002) "Asymmetric Correlations of Equity Portfolios," *Journal of Financial Economics*, 63, 443-494.

[2] Aas, Kjersti and Hobæk Haff, Ingrid (2006) "The Generalised Hyperbolic Skew Student's t-distribution ," *Journal of Financial Econometrics*, 4, .275-309

[3] Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 3, 203-228.

[4] Basak, S. and A. Shapiro (2001), "Value-at-risk-based risk management: Optimal policies and asset prices," *Review of Financial Studies*, 14, 371–405.

[5] Bollerslev, Tim (1990) "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498-505.

[6] Christoffersen, P. (1998), P "Evaluating interval forecasts," *International Economic Review*, 39 841–862

[7] Christoffersen, P (2003) *Elements of Financial Risk Management* Academic Press.

[8] Engle, Robert F. (1982) "Autoregressive Conditional Heteroskedasticity Models with estimation of Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

[9] Engle, Robert F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models", *Journal of Business and Economic Statistics*, 20, 339-350.

[10] Falk, Michael (1984) "Relative Deficiency of Kernel Type Estimators of Quantiles," *Annals of Statistics*, 12, 261-268.

[11] Fantazzini, D. (2006) "Dynamic copula modelling for value at risk," Frontiers in Finance and Economics

[12] Longin, François and Bruno Solnik (2001) "Extreme correlations in international Equity Markets," *Journal of Finance*, 56, 649-676.

[13] Garcia, Rene, Eric Renault, and David Veredas (2004) "Estimation of Stable Distributions by Indirect Inference," Working Paper, UNC at Chapel Hill.

[14] Garcia, Rene and Georges Tsafack (2007) "Dependence Structure and Extreme Comovements in International Equity and Bond Markets," Working Paper, Université de Montréal.

[15] Gourieroux, Christian, Jean Paul. Laurent, and Olivier Scaillet (2000) "Sensitivity Analysis of Values at Risk," *Journal of Empirical Finance*, 7, 225-245.

[16] Lee, T.-H. and X. Long (2006), "Copula-based Multivariate GARCH Model with Uncorrelated Dependent Errors," *Journal of Econometrics*, Forthcoming.

[17] Patton, A. (2004), "On the importance of skewness and asymmetric dependence in stock returns for asset allocation," *Journal of Financial Econometrics*, 2 130–168.

[18] Pelletier, D. (2002), "Regime switching for dynamic correlation," *Journal of Econometrics*, 131 445–473.

[19] Sheather, Simon; J. S. Marron "Kernel Quantile Estimators", *Journal of the American Statistical Association*, 85, 410-416.

Table 12: Estimates of GARCH (1, 1) parameters for all bond and equity returns. The figures between brackets represent standard deviations of the parameter estimates.

	μ	ω	α	β
US Equity	3.20e-03	5.02e-06	9.19e-01	8.07e-02
	(7.00e-04)	(2.39e-06)	(9.80e-03)	(1.28e-02)
US Bond	1.50e-03	2.06e-06	8.99e-01	5.26e-02
	(2.00e-04)	(1.02e-06)	(3.60e-02)	(1.66e-02)
CA Equity	3.24e-03	1.53e-04	4.21e-01	3.55e-01
	(6.23e-04)	(1.71e-05)	(3.27e-02)	(2.38e-02)
CA Bond	1.78e-03	1.26e-05	8.32e-01	6.65e-02
	(3.41e-04)	(5.58e-06)	(5.42e-02)	(1.68e-02)

Table 13: Sample correlation coefficient (Corr) and left tail dependence coefficient (TDC) computed with the Gumbel copula.

	US-CA Equities		US-CA Bonds	
	Corr	TDC	Corr	TDC
Full Sample	0.7182	0.5058	0.4706	0.3564
In Sample	0.7221	0.5076	0.4779	0.3647
Out of Sample	0.6531	0.4864	0.3817	0.2297

Table 14: Aikaike (AIC) and Schwarz's Bayesian (BIC) information criteria for two symmetric copulas (Gaussian and Student-t and two asymmetric copula (Clayton and Gumbel). Estimation is performed over the in-sample period.

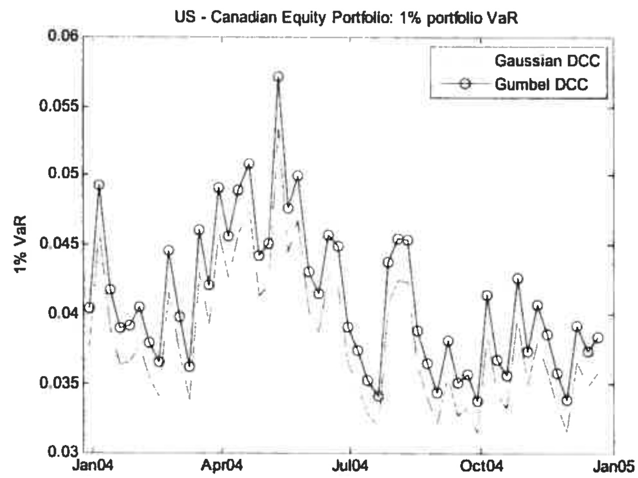
	Symmetric Copula		Asymmetric Copula	
	Gaussian	Student-t	Clayton	Gumbel
		US-CA Equities		
LogLikelihood	3.22e+02	3.29e+02	2.95e+02	3.31e+02
AIC	-320.7354	-326.5728	-293.7713	-329.5698
BIC	-318.2604	-321.6229	-291.2964	-327.0949
		US-CA Bonds		
LogLikelihood	1.29e+02	1.39e+02	1.00e+02	1.43e+02
AIC	-128.0589	-137.2090	-99.4721	-142.4054
BIC	-125.5840	-132.2591	-96.9971	-139.9304

Table 15: Test results over the full sample for the US and Canada equity 50/50 portfolio ($w = 0.5$) : the first panel presents the unconditional backtest. LR_{uc} is the likelihood ratio for unconditional coverage with a χ_1^2 distribution. In the second panel, LR_{cc} is the likelihood ratio for conditional coverage with a χ_2^2 distribution. π_T in the third panel is the sample estimation of violation probability which is used to test the underestimation. The last panel presents the test results for the ES. c is the difference between the sample mean of violations and the ES produced by the model. Three dependence models are tested. The Gaussian (Gauss.), the Student-t (t), and the Gumbel (Gumb.). Numbers in bold emphasize the statistical significance at the 5 or 10 percent level.

	5 %VaR			1 %VaR			0.5 %VaR		
	Gauss.	t	Gumb.	Gauss.	t	Gumb.	Gauss.	t	Gumb.
T_1	56	52	50	19	17	14	12	8	7
LR_{uc}	0.2925	0.0005	0.0945	5.7219	3.5118	1.1147	6.4753	1.2839	0.5543
p-value	0.5886	0.9822	0.7585	**0.0168**	**0.0609**	0.2911	**0.0109**	0.2572	0.4566
T_{00}	933	941	945	1005	1009	1015	1019	1026	1028
T_{11}	3	3	3	1	1	1	1	0	0
LR_{cc}	0.3853	0.1506	0.3277	6.5658	4.6692	2.8854	8.7773	1.3086	0.5328
p-value	0.8248	0.9275	0.8489	**0.0375**	**0.0968**	0.2362	**0.0124**	0.5198	0.7661
π_T	5.37%	4.99%	4.79%	1.82%	1.63%	1.34%	1.15%	0.77%	0.67%
$\pi_T - p$	0.37%	-0.01%	-0.21%	0.82%	0.63%	0.34%	0.65%	0.27%	0.17%
Std Err	0.0070	0.0067	0.0066	0.0041	0.0039	0.0036	0.0033	0.0027	0.0025
t Stat.	0.5286	-0.0213	-0.3115	1.9833	1.6058	0.9601	1.9691	0.9880	0.6766
p-value	0.2986	0.5085	0.6223	**0.0238**	**0.0543**	0.1686	**0.0246**	0.1617	0.2494
c	4.14e-4	4.14e-4	3.17e-4	2.58e-4	2.40e-4	2.16e-4	2.28e-4	2.26e-4	2.08e-4
Std Err	2.25e-4	2.22e-4	2.17e-4	1.97e-4	1.94e-4	1.90e-4	1.90e-4	1.86e-4	1.82e-4
t Stat.	1.8422	1.8622	1.4576	1.3098	1.2357	1.1343	1.2026	1.2117	1.1392
p-value	**0.0329**	**0.0314**	**0.0726**	**0.0952**	0.1084	0.1284	0.1147	0.1130	0.1274

Table 16: Test results over the full sample for the US and Canada bond 50/50 portfolio ($w = 0.5$) : the first panel presents the unconditional backtest. LR_{uc} is the likelihood ratio for unconditional coverage with a χ_1^2 distribution. In the second panel, LR_{cc} is the likelihood ratio for conditional coverage with a χ_2^2 distribution. π_T in the third panel is the sample estimation of violation probability which is used to test the underestimation. The last panel presents the test results for the ES. c is the difference between the sample mean of violations and the ES produced by the model. Three dependence models are tested. The Gaussian (Gauss.), the Student-t (t), and the Gumbel (Gumb.). Numbers in bold emphasize the statistical significance at the 5 or 10 percent level.

	5 %VaR			1 %VaR			0.5 %VaR		
	Gauss.	t	Gumb.	Gauss.	t	Gumb.	Gauss.	t	Gumb.
T_1	55	47	43	13	12	11	10	8	6
LR_{uc}	0.1612	0.5529	1.7928	0.5932	0.2277	0.0309	3.4730	1.2839	0.1132
p-value	0.6881	0.4571	0.1806	0.4412	0.6332	0.8605	**0.0624**	0.2572	0.7365
T_{00}	937	950	958	1016	1018	1020	1022	1026	1030
T_{11}	5	2	2	0	0	0	0	0	0
LR_{cc}	1.6695	0.6319	1.8841	0.8746	0.4524	0.2021	3.5933	1.3086	0.0451
p-value	0.4340	0.7291	0.3898	0.6458	0.7976	0.9039	0.1659	0.5198	0.9777
π_T	5.27%	4.51%	4.12%	1.25%	1.15%	1.05%	0.96%	0.77%	0.58%
$\pi_T - p$	0.27%	-0.49%	-0.88%	0.25%	0.15%	0.05%	0.46%	0.27%	0.08%
Std Err	0.0069	0.0064	0.0062	0.0034	0.0033	0.0032	0.0030	0.0027	0.0023
t Stat.	0.3947	-0.7684	-1.4244	0.7169	0.4556	0.1727	1.5197	0.9880	0.3212
p-value	0.3466	0.7788	0.9227	0.2368	0.3244	0.4315	**0.0644**	0.1617	0.3741
c	6.61e-5	7.33e-5	3.37e-5	4.37e-5	3.82e-5	2.29e-5	3.11e-5	2.77e-5	2.15e-5
Std Err	3.65e-5	3.51e-5	3.29e-5	2.35e-5	2.24e-5	2.00e-5	2.05e-5	1.93e-5	1.65e-5
t Stat.	1.8133	2.0848	1.0225	1.8585	1.7061	1.1444	1.5123	1.4348	1.3055
p-value	**0.0350**	**0.0187**	0.1534	**0.0317**	**0.0441**	0.1264	**0.0654**	**0.0758**	**0.0960**



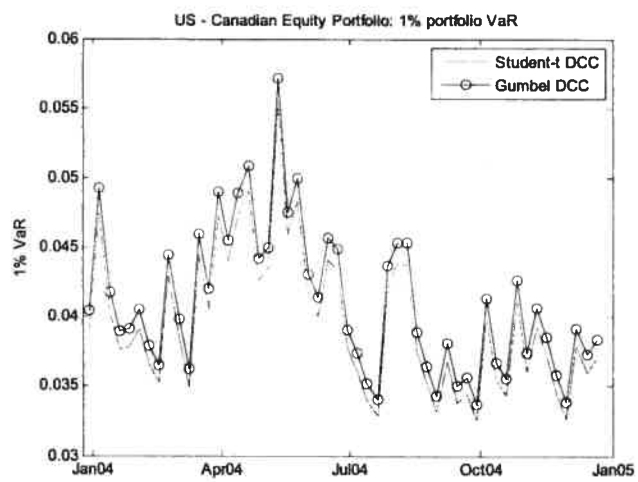


Figure 10: Out-of-sample forecasts for the equity portfolio 1% VaR using the three dependence specification models.

Chapter 3

Proper Conditioning for Coherent VaR in Portfolio Management

1. Introduction

Value at risk (VaR) - the amount of money such that there is typically a 95% or 99% probability of a portfolio losing less than that amount over a certain horizon, has become a central concept in risk management[1]. Financial institutions, regulators as well as nonfinancial corporations use this method to measure financial risk. Although the VaR risk measure seems to agree with a concept of maximum loss popular with practitioners, it is not a coherent measure of risk, as stressed by Artzner et al.(1999), since it is not subadditive.

In practice, VaR is also as a tool for risk to manage and control risk. Recently, several authors have pointed out that the use of VaR as a risk magement tool may entail perverse effects. Basak and Shapiro (2001) show that VaR risk managers choose a larger exposure to risky assets than non-risk managers and as a result incur larger losses in the worst states of the world. In general equilibrium, risk regulation may have the perverse effect of amplifying price fluctuations, as demonstrated by Basak and Shapiro (2001) and Danielsson, Shin and Zigrand (2004)).

In this paper, we focus on a decentralized portfolio management system, widespread among financial institutions, that relies on VaR as a risk measure and risk control tool. In such a system, subadditivity appears as a natural requirement. Typically, the supervisory unit in charge of portfolio management wants to decentralize the management of certain parts of the portfolio to specialists of market segments. If the unit wants to impose a global VaR amount on the whole portfolio, subadditivity will allow to decentralize its VaR constraint into several VaR constraints, one per specialist. The supervisory unit will then be assured that the VaR of the global risk will not surpass the sum of the individual VaRs. The problem is that, even if one may be ready to assume that individual risks are mutually independent, Artzner et al.(1999) have precisley given examples of the nonsubadditivity of VaR even in the case of aggregation of independent risks. A notable exception is the case where all risks are jointly normally (or more generally elliptically) distributed , since the quantiles satisfy subadditivity as long as probabilities of exceedance are smaller than a half.

[1] See, for example, Jorion (2001).

We provide an analysis of the feasibility of decentralized risk management through VaR objectives even when subaddivity is violated. The main idea is that, when tail thickness is responsible for violation of subadditivity, eliciting proper conditioning information may restore a rationale for a decentralized risk management system based on VaR. In such a Rent-a-Trader system, as it is often called, the specialists are hired because they have access to specific information on which they condition their portfolio decisions. Therefore, the central management unit possesses only a subset of the conditioning information which belongs to each specialist. Naturally, in such a context, a distribution appears always more leptokurtic to the central unit than to the specialist. Because of a lack of information, VaR may appear non-subadditive to the central management unit, but without bad consequences for the actual risk incurred. We are then able to show that decentralized portfolio management with a VaR allocation to each specialist will work despite evidence to the contrary. VaR is therefore decentralizable if specialists obey their VaR requirements.

Of course, central management will still want to assess that specialists meet the VaR requirements. We distinguish the case where central management has access to some information from the case where they have to rely on unconditional information only.We provide an illustration where traders have access to private information, which is unobservable to both central management and the other traders, but where they communicate their individual portfolio shares to central management. We discuss ways for central management to improve backtesting of VaR requirements in this context.

To provide a theoretical underpinning to the validity of such a Rent-a-Trader system for risk control, we proceed in two steps. First, we show that violations of VaR subadditivity in case of aggregation of independent risks are basically due to a perverse effect of fat tails. More precisely, we show that when tails are suficiently thin to ensure the existence of finite absolute expected returns, a rather realistic assumption, VaR subadditivity is guaranteed at least for sufficiently large confidence levels, or equivalently for sufficiently small levels of the fixed probability of VaR exceedance. The example of stable distributions is well suited to assess how small is small. In this setting, it is shown that with a reasonable level of skewness in asset returns, VaR at common confidence levels will be subadditive when absolute expected returns have a finite expectation. A contrario, we show that arbitrarily thick tails may produce arbitrarily large violations of the subaddivity property.

Second, we provide the key argument in favor of the rent-a-trader approach by linking fat tailness to a lack of conditioning information. Through the consideration of higher order moments, we extend the argument of Clark (1973) to note that in general, the more information we condition the returns of a portfolio upon, the thinner are the tails of the resulting distribution. This argument works in particular for scale mixing variables, like stochastic volatility. Moreover, by appealing to a scale mixture representation property of stable distributions, we show that our framework nests the family of stable distributions. In other words, we have a way of recovering VaR subadditivity through a proper conditionning which is valid in particular for stable distributions. This is a satisfactory result when remembering the early work by Mandelbrot (1963) and Fama (1965), showing that stable distributions accommodate well heavy-tailed financial series, with the consequence that it produces measures of risk based on the tails of the distribution, such as value at risk, which are more reliable (see in particular Mittnik and Rachev, 1993, and Mittnik, Paolella and Rachev, 2000).

The rest of the paper is organized as follows. Section 2 describes the Rent-a-Trader system whereby portfolio management is decentralized to specialists. Section 3 appeals to some elements of probability theory to put forward the logical relationships between fat tails, violation of VaR subadditivity and conditioning information. In section 4, we show that deconditioning by central management always increase tail fatness and spuriously make VaR look non-subadditive. In section 5, we provide a simple illustration with private signals to traders and show how the transmission of information in the form of portfolio shares helps to assess risk more accurately. Section 6 concludes. Proofs of the various propositions are collected in an Appendix.

2. VaR Decentralization with Differential Information

In this section we describe a decentralized portfolio management system also known that uses VaR as a tool for risk management. Senior management is interested in the value at risk $VaR_p(X)$

associated with the random value X of its portfolio:

$$P[X \leq -VaR_p(X)] = p. \tag{2.1}$$

This value X will be conformable to a VaR requirement VaR_p^* if and only if: $VaR_p(X) \leq VaR_p^*$, that is if:

$$P[X \leq -VaR_p^*] \leq p. \tag{2.2}$$

Suppose that senior management hires n traders to manage parts of its portfolio. Then $X = \Sigma_{j=1}^{n} X_j$ is the aggregation of the net results X_j of n traders $j = 1, \ldots, n$. Failure of subadditivity for senior management in this context means that $VaR_p(\sum_{j=1}^{n} X_j)$ may exceed $\sum_{j=1}^{n} VaR_p(X_j)$. We will show in the next section that violation of subadditivity is a perverse effect of fat tails in the distribution of X. We will then see that fat tails can be reduced by relying on some conditioning information.

Trader j has obtained result X_j by building a portfolio $\theta_j(I_j)$, which is a function of his private information I_j. A decentralized management system is used precisely to exploit this private information I_j since it is inaccessible to central management. Let us consider that trader j has received from senior management a target S_j in terms of VaR, that is:

$$VaR_p(X_j \,|I_j\,) \leq S_j \tag{2.3}$$

where:

$$P[X_j \leq -VaR_p(X_j \,|I_j\,) \,|I_j\,] = p. \tag{2.4}$$

Note that S_j is a given number chosen by senior management. Typically, the quantity $VaR_p(X_j \,|I_j\,)$, which depends on private information, cannot be observed at the central level. Therefore, senior management cannot check directly that the requested target (2.3) has been met or equivalently that:

$$P[X_j \leq -S_j \,|I_j] \leq \ p. \tag{2.5}$$

Therefore it must rely on backtesting, with two objectives. First, as usual, senior management must check on a time series of portfolios returns that $X = \Sigma_{j=1}^{n} X_j$ fulfills the VaR requirement (2.2). It is quite natural for central management to imagine that this requirement will be ensured by the enforcement of targets S_j if and only if these targets are chosen ex-ante in order to fulfill:

$$\Sigma_{j=1}^{n} S_j \leq VaR_p^*. \tag{2.6}$$

Second, even though (2.5) cannot be observed, senior management should be interested to seek valuable information about individual trader j behavior. Of course, historical observation allows him to test for a weak consequence of (2.5), that is:

$$P[X_j \leq -S_j] \leq p. \tag{2.7}$$

But, for the targets to appear credible, a tighter control should be performed. Often senior management will request that traders communicate their portfolio shares. We will see how this information can help exercise a better control, but even if this information on individual portfolio shares is transmitted, senior management will never recover fully in practice the information I_j of the individual traders.

In section 4, we will show that, under a set of natural assumptions, both goals of backtesting may be met. In other words, it will be true that the enforcement of targets conformable to (2.6) will ensure (2.2). Moreover, senior management will have at its disposal some relevant information for a more powerful test of (2.5) than only through its weak consequence (2.7).

3. Conditioning Information, Tails and VaR Subadditivity

We show in a first subsection that violation of subadditivity in the case of aggregation of independent risks is basically a perverse effect of fat tails. In a second subsection, we study the logical relationships between conditioning information and fat tailness, in particular in the context of scale mixtures of distributions.

3.1 Tails and Subaddivity

Let us consider two stochastically independent real variables X and Y with cumulative distribution functions:

$$\begin{aligned} F_X(x) &= P[X \leq x] \\ F_Y(y) &= P[Y \leq y] \end{aligned}$$

A straightforward adaptation of the proof of Feller convolution theorem (Feller, 1971, p 278) allows us to claim that, when the variable x tends to $(-\infty)$, the distribution function $F_{X+Y}(x) = P[X+Y \leq x]$ of the sum $(X+Y)$ is equivalent to the sum of the distribution functions:

$$F_{X+Y}(x) \sim F_X(x) + F_Y(x) \tag{3.8}$$

Let $VaR_p(X)$ and $VaR_p(Y)$, as defined in (2.1), the values at risk respectively associated with the random values X and Y of some portfolio at a given horizon. A first implication of the convolution property (3.8) is that VaR subadditivity is not really an issue if one of the two risks has much thinner tails than the other. Suppose for instance that, when x tends to $(-\infty)$, $F_Y(x)$ is infinitely small with respect to $F_X(x)$. Then, large losses in the aggregate portfolio $(X+Y)$ will likely come from X and thus, for sufficiently small levels p of probability exceedence, $VaR_p(X)$ is the right measure of risk to control in order to control the aggregate $VaR_p(X+Y)$, irrespective of possible violations of subadditivity. In other words, violations of subadditivity may occur but they are negligible for sufficiently small levels p .

A more interesting case occurs when X and Y both have a distribution function with a left-tail behaviour of the following type. For some given function g increasing and unbounded on the positive real line, there are two positive real numbers a_X and a_Y such that:

$$\begin{aligned} a_X &= Lim_{x \longmapsto -\infty} g(-x) F_X(x) \\ a_Y &= Lim_{x \longmapsto -\infty} g(-x) F_Y(x) \end{aligned}$$

Hence, by application of (3.8), we have:

$$a_X + a_Y = Lim_{x \longmapsto -\infty} g(-x)[F_{X+Y}(x)]$$

Then, if we assume for notational simplicity that the cumulative distribution functions F_X and F_Y are continuous and strictly increasing in the neighborhood of $(-\infty)$ to allow a unambiguous definition of inverse functions, we conclude that for sufficiently small levels of probability p of VaR exceedance:

$$\begin{aligned} VaR_p(X) &\sim g^{-1}(\frac{a_X}{p}), \\ VaR_p(Y) &\sim g^{-1}(\frac{a_Y}{p}), \\ VaR_p(X+Y) &\sim g^{-1}(\frac{a_X + a_Y}{p}). \end{aligned}$$

The following proposition is easily deduced from these asymptotic equivalences:

Proposition 3.1

If X and Y are two independent random variables such that for some positive numbers a_X and *a_Y and a given function g strictly increasing and continuous on the positive real line:*

$a_X = Lim_{x \longmapsto -\infty} g(-x) F_X(x)$ *and* $a_Y = Lim_{x \longmapsto -\infty} g(-x) F_Y(x)$

we have:

(i) If for all positive u and *v: $g(u+v) > g(u) + g(v)$,*

There exists $p_0 \in]0,1]$ such that, for any $p \in]0, p_0[$:

$$VaR_p[X+Y] < VaR_p(X) + VaR_p(Y)$$

(ii) If for all positive u and *v: $g(u+v) < g(u) + g(v)$,*

There exists $p_o \in]0,1[$ such that, for any $p \in]0, p_0[$

$$VaR_p[X+Y] > VaR_p(X) + VaR_p(Y)$$

Proof: See the Appendix.

It is worth noticing that the subadditivity property, maintained above on all the real line either for the function g or for the function $(-g)$ is actually binding only for arbitrarily large u and v. Since the function g has been defined to characterize the tail behaviour of the distribution function, only its behaviour in the neighborhood of $(+\infty)$ really matters.

Distributions with such left-tail behavior are all distributions with Pareto-like left tails given in Feller (1971), $F_X(x) \sim_{x \longrightarrow -\infty} a_X[-x]^{-a}[Log(-x)]^{\gamma}$, with $\alpha > 0$ and γ any real number. Conditions *(i)* and *(ii)* about the concavity or convexity of g translate into conditions on α, i.e. $\alpha > 1$ for *(i)* and $\alpha < 1$ for *(ii)*. Therefore, up to the limit case $\alpha = 1$, subadditivity of VaR_p for sufficiently small levels p of probability exceedence is tantamount to finite expectation for absolute returns. In this respect, non-subadditive VaR remains quite an extreme situation[2]. However, in the case of very heavy- tail distributions (α close to zero), we show in the Appendix that violation of subadditivity may be arbitrarily extreme.

What is most important for our purpose is to be able to ensure subaddivity for sufficiently small levels p of probability of exceedence. We want to ensure that the commonly used small values of p like 1%, 5% or 10% are within the range of maintained subaddivity. To shed more light on the relevant order of magnitude, we propose to consider the case of stable distributions as a benchmark example of variables with Pareto-like tails.

A random variable X is said to follow a stable distribution[3] $S_\alpha(\sigma, \beta, \mu)$ for $2 > \alpha > 0, \alpha \neq 1, \sigma > 0, \mid \beta \mid < 1$ and μ any real number if its characteristic function is given by:

$$E\exp(i\theta X) = \exp\left\{-\sigma^\alpha \left|\theta\right|^\alpha \left[1 - i\beta\left(sign\ \theta\right)\tan\frac{\Pi\alpha}{2}\right] + i\mu\theta\right\} \tag{3.9}$$

The parameters σ, β and μ are uniquely defined. μ is a location parameter , σ is a scale parameter and β characterizes the skewness of the distribution: a positive (resp. negative) β implies a distribution skewed to the right (resp. to the left) while a zero β gives a symmetric distribution.

[2]Recently, Ibragimov (2004) obtained a similar result with a different approach based on the analysis of majorization properties of linear combinations of random variables.

[3]See Samorodnitsky and Taqqu (1994) for a thorough treatment of stable distributions.

In particular:

$$X \rightsquigarrow S_\alpha\left[\sigma,\beta,\mu\right] \Leftrightarrow \frac{X-\mu}{\sigma} \rightsquigarrow S_\alpha\left[1,\beta,0\right] \tag{3.10}$$

and:

$$X \rightsquigarrow S_\alpha\left[\sigma,\beta,0\right] \Leftrightarrow (-X) \rightsquigarrow S_\alpha\left[\sigma,-\beta,0\right]$$

In all cases, the support of the distribution is the whole real line. Note that for the sake of expositional simplicity, we have excluded the limit cases $\alpha = 2$ (normal distribution) $\alpha = 1$, and $|\beta| = 1$ (distribution concentrated on one half of the real line).

This parametric family of distributions has Pareto-like tails. If $X \rightsquigarrow S_\alpha\left[\sigma_X,\beta,\mu_X\right]$, we have:

$$Lim_{x\longmapsto -\infty}(-x)^\alpha F_X(x) = a_x,$$

with:

$$a_X = \frac{(1-\alpha)(1-\beta)\sigma_X^\alpha}{2\Gamma(2-\alpha)\cos(\frac{\Pi\alpha}{2})}.$$

The advantage is that in this case we can assess the level p_0 below which the subadditivity property is characterized. We can therefore state the following corollary to proposition 3.1.

Corollary 3.2

If X and Y are two independent stable variables with similar tails and the same degree of skewness:

$$X \rightsquigarrow S_\alpha\left[\sigma_X,\beta,\mu_X\right] \text{ and } Y \rightsquigarrow S_\alpha\left[\sigma_Y,\beta,\mu_Y\right],$$

and we consider a probability p of VaR exceedence such that:

$$p < P[S_\alpha(1,\beta,0) < 0]$$

Then:

(i) If $\alpha > 1$:

$$VaR_p[X+Y] < VaR_p(X) + VaR_p(Y)$$

(ii) If $\alpha < 1$:

$$VaR_p[X+Y] > VaR_p(X) + VaR_p(Y)$$

It is important to stress that this corollary can be applied for values of the probability p which are not excessively small. For instance, if $\beta = 0$, it applies for $p < 1/2$. Irrespective of the value of β, it applies exactly when $VaR_p(X) > \mu_X$ and $VaR_p(Y) > \mu_Y$. Note that when $\alpha > 1$, the shift parameter μ is equal to the mean. In other words, it is sufficient to have absolute returns with finite means and to consider possible amounts of losses $VaR_p(X)$ and $VaR_p(Y)$ beyond the opposite of the respective means μ_X and μ_Y to ensure subaddivity of the VaR[4].

3.2 Tails and Conditioning Information

Given the importance of tail thickness for VaR subaddivity, we want to argue in this section that, in general, the larger the conditioning information set is, the thinner the tails should be. Of course, this claim rests upon some measurement of tail thickness. Extending the common idea of kurtosis measurement, we characterize tail thickness through higher-order moments.

Let us consider some random variable Y such that $|Y|^n$has a finite expectation for some positive real number n. Let m be another real number such that $0 < m < n$. We argue that the larger the ratio $\frac{E\{|Y|^n\}}{\{E\{|Y|^m\}\}^{\frac{n}{m}}}$ is, the fatter the tails of the distribution of Y should be. According to Malevergne and Sornette (2006), the major contribution to the magnitude of the moment $E\{|Y|^n\}$ comes from the values of Y in the vicinity of the maximum of $|y|^n f_Y(y)$, where $f_Y(y)$ is the probability density function (pdf) of Y. The magnitude of this quantity increases fast with the order of the moment we consider. The faster it increases, the fatter are the tails of the pdf of Y. This ratio is the

[4] Not that we have assumed that X and Y have the same skewness parameter. This assumption, which was not needed to apply the convolution property, may appear overly restrictive to the point where only $\beta = 0$ has some practical content. Hopefully, the subadditivity should not be lost for not too different skewness parameters.

standard kurtosis measurement when $m = 2$ and $n = 4$, with Y measured in deviations from the mean. To accomodate the case of variables with possibly infinite variance and even infinite mean, we generalize the standard argument to moments of any order for variables not expressed in mean-deviation form. To characterize the effect of conditioning information, we extend the result previously derived by Clark (1973)[5] to the case of kurtosis.

Let Y and Z two random variables, where, for notational simplicity, we assume that Y is a positive real variable and : $E[(Y)^n] < +\infty$, $0 < m < n$. The tight relationship between conditioning and tail thickness , as measured by the comparison between higher and lower order moments, amounts to say that, more often than not:

$$E\left[\frac{E[(Y)^n \mid Z]}{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}}\right] < \frac{E[(Y)^n]}{\{E[(Y)^m]\}^{\frac{n}{m}}}. \tag{3.11}$$

In other words, conditioning on the variable Z reduces the distance between higher and lower order moments. We specialize the result to scale mixtures, with Z as a mixture variable, in the following proposition.

Proposition 3.3

If the distribution of the random variable Y is a scale mixture with Z as a mixture variable, that is if $Y = \sigma(Z)u$, with Z and u stochastically independent, and if, in addition: $E\{|u|^n\} < +\infty, E\{|\sigma(Z)|^n\} < +\infty, 0 < m < n$,

Then:

$$E\left[\frac{E[|Y|^n \mid Z]}{\{E[|Y|^m \mid Z]\}^{\frac{n}{m}}}\right] < \frac{E[|Y|^n]}{\{E[|Y|^m]\}^{\frac{n}{m}}}$$

Proof: See the Appendix.

The inequality of proposition 3.3 is very likely to hold in general[6] It does hold for a number of common models that are actually scale mixture models. A popular example is the stochastic

[5] See in particular corollary 4.1.

[6] For instance, in the classical case of a zero-mean variable Y with $m = 2$ and $n = 4$, inequality (6.22) indicates that it would take a perversely high positive correlation betwen conditional kurtosis and conditional variance to reverse the inequality.

volatility model without leverage effect, as first introduced by Taylor (1982) as a dynamic extension of Clark (1973). A less-known example is the case of symmetric stable distributions, which can always be seen as scale mixtures of stable distributions with less fat tails. Indeed, according to Samorodnitsky and Taqqu (1994)[7], if X is a random variable with a symmetric $\alpha-$ stable distribution, $X \rightsquigarrow S_\alpha(\sigma, 0, 0)$, $0 < \alpha < 2$, then X can be seen as a scale mixture of stable distributions: $X|A \rightsquigarrow S_{\alpha'}(\sigma A^{1/\alpha'}, 0, 0), 0 < \alpha < \alpha'$, where the probability distribution of the mixing variable A is defined by its Laplace transform: $E(\exp(-\gamma A)) = \exp(-\gamma^{\alpha/\alpha'})$. Therefore, a random variable with a symmetric stable distribution[8] can produce a stable distribution with less fat tails (higher α). This illustrates the general proposition above in terms of higher moments.

In this section we have shown that subadditivity of VaR is intimately related to fat tails and that fat tails are in turn very closely linked to conditioning information. In the next section, we want to use these two main principles to spell out conditions under which a decentralized portfolio management system will work in terms of risk control. These conditions will ensure that the VaR requirement is respected, that is $P[\Sigma_{j=1}^n X_j \leq -VaR_p^*] \leq p$.

4. Proper Conditioning for Subadditive VaR

We put forward in this section two crucial assumptions that will ensure VaR subadditivity in the decentralized management system described in section 2. We will assume that these assumptions are valid at a given probability level p, which will be seen as a relevant confidence level for VaR calculations such as 1% or 5%.

The first assumption amounts to consider that, even though subadditivity of VaR is not guaranteed at the senior management level, there exists a latent information I, nesting all individual information sets, such that the conditioning by this information would restore subadditivity of VaR.

[7] See Proposition 1.3.1 p. 20.

[8] This symmetry assumption is rather realistic for distributions produced by portfolio traders. Indeed, the central unit does not need to give a benchmark to the traders in the context of a decentralization portfolio management system. Therefore, the distribution of interest is not the deviations of the trader's portfolio returns from the benchmark, which ought to be skewed to the right, but simply the raw returns of the trader's strategy. The right skewness of the latter returns is less of a necessity.

Of course this conditioning will be unfeasible in practice, but it suffices that traders obey their VaR target for the risk control system to work. Moreover, it shows that eliciting some information from traders such as portfolio shares will be useful in terms of ex-post risk control or backtesting.

Assumption 1: *There exists* $I \supset \cup_{j=1}^{n} I_j$ *such that* $VaR_p(\Sigma_{j=1}^{n} X_j \,|I) \leq \Sigma_{j=1}^{n} VaR_p(X_j \,|I)$.

As seen in section 2, the larger the conditioning information set is, the thinner the tails are. In this case, VaR subadditivity is more likely to hold. In particular, Assumption 1 will be automatically fulfilled if the joint distribution of the vector $(X_j)_{1 \leq j \leq n}$ of returns is a multivariate scale mixture, that is for some n-dimensional variable $(u_j)_{1 \leq j \leq n}$ conformable to subadditivity (for instance a Gaussian vector with $p \leq \frac{1}{2}$) and independent from conditioning information, $X_j = \sigma_j(I) u_j$, for $j = 1, ..., n$.

The second assumption stated below will be fulfilled if in addition $\sigma_j(I)$ depends on information I only through trader's j information I_j. This appears as a rather natural assumption in such a delegated system where each trader is hired because he holds a specific information.

Assumption 2: *For all* $j = 1, \ldots, n$:$VaR_p(X_j \,|I) \leq VaR_p(X_j \,|I_j)$.

In other words, latent information other than I_j is irrelevant for forecasting the result X_j of trader's j investment. This latter condition, a kind of cross-sectional equivalent to a non-causality assumption (external information does not cause X_j given I_j), is fairly natural if one imagines trader j as an expert of his market segment. Trader j has at his disposal all the relevant information for his market segment[9].

However, assumption 2 is more general than this special case of cross-sectional non-causality. It only means that the part of latent information that trader j does not observe does not affect his perceived potential loss with probability p. In particular, we have:

[9]Note that given I_j, $VaR_p(X_j \,|I)$ is a random variable which can be constantly below $VaR_p(X_j \,|I_j)$ (with a common level of probability p) only if these two variables actually coincide almost surely. In other words, Assumption 2 is a non-causality property in terms of VaR_p . It is fulfilled in particular in case of global non-causality, that is if for all j, the conditional distributions of X_j given I or I_j coincide.

Proposition 4.1: *Assumption 2 implies that, I_j almost certainly:*

$$X_j \leq -VaR_p(X_j\,|I_j) \Leftrightarrow X_j \leq -VaR_p(X_j\,|I\,)$$

Under assumption 2, conditioning on the larger latent information set does not change the occurrence of VaR exceedance for j, I_j almost surely[10]. The most important result of this section is stated in the following proposition.

Proposition 4.2: *Under assumptions 1 and 2, $\Sigma_{j=1}^{n} S_j \leq VaR_p^*$ and $VaR_p(X_j\,|I_j) \leq S_j$ for all j implies that:*

$$P[\Sigma_{j=1}^{n} X_j \leq -VaR_p^*] \leq p.$$

Proofs for these two propositions are provided in the Appendix.

In other words, the VaR target $VaR_p(X_j\,|I_j) \leq S_j$ imposed to each specialist $j = 1, \ldots, n$ ensures that the VaR of the portfolio $\sum\limits_{j=1}^{n} X_j$ will not exceed $\sum\limits_{j=1}^{n} S_j$. It is worth emphasizing that this result has been obtained while VaR may not be subadditive for senior management, that is $VaR_p(\sum\limits_{j=1}^{n} X_j)$ may exceed $\sum\limits_{j=1}^{n} VaR_p(X_j)$. This convenient result has basically been obtained through an additional conditioning that has restored subadditivity without introducing additional perceived risk thanks to assumptions 1 and 2.

As already mentioned, assumption 2 may also be useful for the purpose of backtesting, or more precisely for ex-post control of the risk-taking behavior of the specialists. Senior management would like to check that the announced target S_j has been respected by specialist j, that is:

$$VaR_p(X_j\,|I_j) \leq S_j. \tag{4.12}$$

[10] Note that the converse is not true in general even though we have, by the law of iterated expectations: $I_j \subset I \Longrightarrow P[X_j \leq -VaR_p(X_j\,|I\,)|I_j] = p = P[X_j \leq -VaR_p(X_j\,|I_j\,)|I_j]$. But the equality of probabilities does not imply the equality of events.

Although senior management cannot observe the information set I_j, it has access to some partial information such as the specialists' actions. Let us assume, at it is often the case in practice, that the portfolio composition $\theta_j(I_j) \in I_j$, selected by each trader j is available to central management. Then, by the law of iterated expectations, (2.3)-(2.4) implies that:

$$P[X_j \leq -S_j \,|\theta_j(I_j)] \leq p \text{ for all } j = 1, \dots, n \tag{4.13}$$

If, as it is often the case, each specialist's information completely defines the return distribution of the fund in which the corresponding trader invests and the private information signals $I_j, j = 1, \cdots, n$,are mutually independent, then (4.13) actually means:

$$P[X_j \leq -S_j \,|\theta_k(I_k),\ k = 1, \dots, n] \leq p \text{ for all } j = 1, \dots, n \tag{4.14}$$

This condition can actually be tested by senior management from an econometric model of conditional probability distributions, including for instance ARCH effects (Engle, 1982). We discuss this issue in the next section. Note that, without maintaining a joint independence assumption, the non-causality assumption 2 actually implies even more since it ensures that:

$$P[X_j \leq -S_j \,|I] \leq p \text{ for all } j = 1, \dots, n. \tag{4.15}$$

Then, the control over trader j behavior appears a priori much more powerful than the solely unconditional control $P[X_j \leq -S_j] \leq p$ that could have been performed without taking advantage of the observation of specialists' actions and possibly resorting to assumption 2.

5. A Simple Illustration of a Rent-a-Trader System

We provide an illustration of the general propositions of the previous sections in a simple setting. The goal of the illustration is to provide a concrete yet basic example where conditioning on the private information of traders restore VaR subadditivity (assumption 1) and where only the information of trader j is relevant in forecasting the result X_j of trader's j investment (assumption

2). We also discuss in the framework of this example how eliciting information for traders can be helpful for backtesting VaR exceedances.

We assume that two traders can each choose a portfolio made up of one risk-free asset and two risky funds. The returns of the two risky funds depend on two state variables s_1 and s_2. State variable s_1 is observable to trader 1, but unobservable to trader 2 and to central management. Similarly, trader 2 is the only one to observe s_2. The two state variables are assumed to be *i.i.d. Bernoulli* (λ).

We can write the returns as $\widetilde{R}_1 = s_1 R_1^1 + (1 - s_1) R^0$, and $\widetilde{R}_2 = s_2 R_2^1 + (1 - s_2) R^0$, where R_1^1, R_2^1 and R^0 are mutually independent with R_1^1 and R_2^1 following the same probability distribution $N(\mu_1, \overline{\sigma}^2)$ and R^0 following $N(\mu_0, \overline{\sigma}^2)$. Moreover, the unconditional mean $[\lambda\mu_1 + (1 - \lambda\mu_0]$ of the two distributions is assumed to be equal to the risk-free rate[11]. These assumptions imply that, without any information on the state variables, a risk averse agent will only invest in the risk-free asset. Therefore, central management will have an incentive to hire traders 1 and 2, who have private information on state variables s_1 and s_2 respectively. In this context, if each trader forms his portfolio such that the VaR requirement imposed by central management is satisfied conditionally to any possible value of his private information, then the VaR requirement of the global portfolio will be satisfied and the apparent violation of subadditivity will not matter.

We further assume that each trader communicates his portfolio shares to central management. We show how, based on this information, central management can recover statistically the parameters of the conditional distributions of the traders' portfolio returns and assess whether traders have respected the VaR requirement or not. It is important to realize that this is just an example while in the general setting considered above, it has never been assumed that the knowledge of these individual portfolio shares was a sufficient statistic to recover fully the conditioning information of traders and, by the same token, to restore subadditivity.

[11]It is important to realize that funds 1 and 2 have the same conditional means μ_1 in state 1 and μ_0 in state 0 and the same conditional variance $\overline{\sigma}^2$ in any state. They differ only by the realization of the states, which do not necessarily coincide. For example, fund 1 could be in state 0 when fund 2 is in state 1.

5.1 Traders' Behavior

We start by computing the optimal mean-variance portfolio of traders 1 and 2 given their private information on s_1 and s_2 respectively. We assume for simplicity that the VaR of their optimal portfolio is always below S_j[12], the target set by central management.

We can write the portfolio return of trader 1 as $\widetilde{R}^1 = R_f + \theta_{11}\left(\widetilde{R}_1 - R_f\right) + \theta_{12}\left(\widetilde{R}_2 - R_f\right)$. The expectation and the variance conditional on the state are:

$$\begin{aligned} E\left(\widetilde{R}^1 \left| s_1 = i\right.\right) &= R_f + \theta_{11}\left(\mu_i - R_f\right) \\ Var\left(\widetilde{R}^1 \left| s_1 = i\right.\right) &= \theta_{11}^2\overline{\sigma}^2 + \theta_{12}^2\sigma^2 \end{aligned} \tag{5.16}$$

Normalizing initial wealth to one, and denoting the risk aversion coefficient of trader 1 by γ_1, the optimal portfolio is solution of:

$$\underset{\theta_1}{Max}\left\{\left(R_f + \theta_{11}\left(\mu_i - R_f\right)\right) - \tfrac{\gamma_1}{2}\left(\theta_{11}^2\overline{\sigma}^2 + \theta_{12}^2\sigma^2\right)\right\}$$

with $\theta_1 = (\theta_{11}, \theta_{12})$. The solution $\widehat{\theta}_1 = \left(\widehat{\theta}_{11}, \widehat{\theta}_{12}\right)$ is given by

$$\begin{aligned} \widehat{\theta}_{11} &= \left(\mu_i - R_f\right)/\gamma_1\overline{\sigma}^2 \\ \widehat{\theta}_{12} &= 0 \end{aligned} \tag{5.17}$$

The proportion invested in the risk-free asset is $1 - \widehat{\theta}_{11}$. Trader 1 never invests in fund 2 for which he has no information[13]. Moreover, traders will always include a non-zero share of their respective risky fund in their optimal portfolio along with the risk-free asset. In the good state,

[12] When the VaR constraint of trader 1 (resp. 2) binds, it can be shown that he may have to invest a nonzero part in asset 2 (resp. 1). Therefore, the distribution given the portfolio shares will be a mixture of normals and not a normal. Conditioning will still make the tails less fat as discussed in the earlier sections, but we prefer to keep things simple and recover normality and hence restore subadditivity.

[13] In a general framework, Merton (1987) assumes this result and justifies his assumption by the fact that the portfolios held by actual investors contain only a small fraction of the thousand of traded securities available. In our setting, the result follows directly from the private information held by the traders.

they will hold a long position, in the bad state they will sell the risky fund short[14].

Overall, we are typically in a situation where each specialist's information completely defines the return distribution of the fund in which the corresponding trader invests and the private information signals I_j, $j = 1, \dots n$, are mutually independent. Therefore, assumption 2 is fulfilled and the condition to test for backtesting is just (4.13). Note moreover that assumption 1 is trivially fulfilled for any p smaller than 1/2 since, given the private signals, the joint conditional probability distribution of traders' portfolio returns is normal.

5.2 Subadditivity Issues

Since assumptions 1 and 2 are fulfilled, we know from our general analysis above that the Rent-a-Trader system ensures a coherent risk management. However, in the type of setting described in the previous subsection, VaR may typically appear to violate subadditivity from the central management point of view, even for very small levels of confidence probability p. To see this, let us assume for simplicity that both traders have the same risk aversion γ and that their initial wealth is normalized to one. We further assume that both information variables s_1 and s_2 are in the good state ($s_1 = s_2 = 0$). Therefore, it follows that $\widehat{\theta}_{11} = \widehat{\theta}_{22} = \theta > 0$, traders' portfolios are $\widetilde{R}^1 = (1-\theta) R_f + \theta \widetilde{R}_1$ and $\widetilde{R}^2 = (1-\theta) R_f + \theta \widetilde{R}_2$ and the aggregate portfolio at central management level is $\widetilde{R}^1 + \widetilde{R}^2 = 2(1-\theta) R_f + \theta \left(\widetilde{R}_1 + \widetilde{R}_2 \right)$.

In such a context, since the central management does not observe the private signals, it is confronted with a mixture of normals for which subadditivity of VaR may be violated even at small probability levels p :

Proposition 5.1: *For any mixture probability values $\lambda < 1/2$, at the level $p = p(\lambda, \mu_0, \mu_1) = P\left(\widetilde{R}_1 + \widetilde{R}_2 \leq \mu_1 + \mu_0 \right)$ we have*

$$VaR_p\left(\widetilde{R}^1 + \widetilde{R}^2 \right) > VaR_p\left(\widetilde{R}^1 \right) + VaR_p\left(\widetilde{R}^2 \right)$$

[14] In a framework with only two assets (a risk-free asset and a risky asset with the same unconditional expected return), Sentana (2005) assumes and rationalizes the fact that the wealth invested in the risky asset is a linear function of the information that the investor has on this asset.

Moreover

$$\begin{cases} p(\lambda,\mu_0,\mu_1) = \lambda^2\Phi\left(-\overline{\mu}/\sqrt{2\sigma}\right) + \lambda(1-\lambda) + (1-\lambda)^2\,\Phi\left(\overline{\mu}/2\overline{\sigma}\right) \underset{\substack{\lambda\to 0\\ \overline{\mu}\to-\infty}}{\longrightarrow} 0 \\ p(\lambda,\mu_0,\mu_1) \underset{\overline{\mu}\to 0}{\longrightarrow} 1 \end{cases}$$

where Φ *denotes the standard normal distribution function and* $\overline{\mu} \equiv \mu_1 - \mu_0 < 0$.

Proof : see the Appendix.

In other words, the level p where violation of subadditivity occurs can be arbitrary close to λ when the expected return spread between the good and the bad states is arbitrarily large. Therefore, a small probability λ of occurrence of the bad state will produce a violation of subadditivity of VaR, even for small levels of the confidence probability level p. We then get an example of the surprising situation where subadditvity is violated even though decentralized risk management works, insofar as traders remain true to the VaR requirements sent to them by central management.

5.3 Backtesting VaR requirements

In this simple model, a central unit can test condition (4.13), which can be written $P[X_j \leq -S_j\,|\theta_k(I_k), k=1,2] \leq p$ for all j, when the shares of the portfolios held by traders 1 and 2 are known. It is important to realize that in this setting, knowing the portfolio composition of traders 1 and 2 is equivalent to knowing their private information s_1 and s_2. Indeed, if trader 1 takes a long position in risky fund 1 it means that s_1 is in the good state and inversely if he short-sells fund 1. Similarly, the position of trader 2 will be fully revealing. We can write $s_j = 1_{\{\theta_{jj}<0\}}$. Since each private information completely defines the return distribution of the fund in which the corresponding trader invests and given that the two informations s_1 and s_2 are independent, the condition to test is exactly $P[X_j \leq -S_j\,|\theta_j(s_j)] \leq p$ for $j = 1, 2$.

By the conditional normality of the return distributions, central management needs only to infer the means and variances in the good and bad states for each trader in order to test if each trader obeys his VaR limit with probability p. In section 4, we assumed that central management knew the underlying model. In practice, central management must estimate the model based on time series of returns and compute VaR conditionally on past returns. In this dynamic setting, assumption A2 can be rewritten as:

$$VaR_p\left[X_j|I_t, I_\tau, \tau \leq t\right] \leq VaR_p\left[X_j|I_{jt}, I_\tau, \tau \leq t\right]. \tag{5.18}$$

In other words, we will assume that all past information becomes common knowledge. Propositions equivalent to Propositions 4.1 and 4.2 can be derived in a dynamic context. Engle and Manganelli (2004) provide a useful framework to estimate value at risk in a dynamic context. They remark that VaR is simply a particular quantile of future portfolio values, conditional on current information. They provide a specification (CAViaR, Conditional Autoregressive Value at Risk) for calculating VaR_t as a function of variables known at time $t-1$ (which in our case could be the portfolio shares of the traders) and a set of parameters that are estimated using Koenker and Bassett's (1978) regression quantile framework.

6. Conclusion

In this paper, we have argued that, in the context of decentralized portfolio management, central management possesses only a fraction of information which belongs to each specialist. In such a context, a distribution appears always thicker to the central unit than to the specialist. Therefore, because of a lack of information, VaR may appear fallaciously non subadditive to the central management unit. We were then able to show that decentralized portfolio management with a VaR allocation to each specialist will work despite evidence to the contrary. Furthermore, we have shown that value at risk remains subadditive in many situations of practical interest. In the case of heavy-tail distributions, we have shown, at least for sufficiently small probabilities, that VaR remains subadditive when the possible loss has finite expectation. In this respect, non-subadditive VaR remains quite an extreme situation.

Even though we show that for all practical purposes VaR is not as incoherent a measure of risk as it is often argued, it remains that portfolio optimization practices using VaR as a simple substitute to variance (i.e.maximization of expected return under a VaR constraint) may generate perverse effects. In particular, there is a risk that a manager who is controlled only through a maximal loss level with probability $(1-p)$ will be enticed to expose himself to huge possible losses with probability p, as demonstrated by Basak and Shapiro (2001). To control for such a risk, one can add to VaR the expected loss beyond the VaR or consider a parameterized family of more limited possible risks. Alexander and Baptista (2004) compare VaR and conditional VaR constraints on portfolio selection with a mean-variance model. Ortobelli, Rachev and Schwartz (2000) provide a thorough analysis of optimal portfolio allocation with stable distributed returns, including with a safety-first optimal allocation problem, whereby investors maximize their expected wealth while minimizing at the same time the risk of loss. Efficient frontiers in the latter case are a function of the threshold VaR.

References

[1] Alexander, G. J. and A. M. Baptista (2004), "A Comparison of VaR and CVaR Constraints on Portfolio Selection with the Mean-Variance Model," *Management Science,* 50, 9, 1261–1273.

[2] Artzner, P., F. Delbaen, J.-M. Eber and D. Heath (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 3, 203-228.

[3] Basak, S. and A. Shapiro (2001), "Value-at-risk-based risk management: Optimal policies and asset prices," *Review of Financial Studies,* 14, 371–405.

[4] Clark, P. K. (1973), "A subordinated stochastic process model with finite variance for speculative prices," Econometrica, 41,135-156.

[5] Danielsson, J., H. S. Shin and J.-P. Zigrand (2004), "The impact of risk regulation on price dynamics," *Journal of Banking and Finance, 28, 1069–1087.*

[6] Engle, R. (1982) "Autoregressive Conditional Heteroskedasticity Models with estimation of Variance of United Kingdom Inflation," *Econometrica*, 50, 987-1007.

[7] Engle, R. and S. Manganelli (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics,* Volume 22, Number 4, 367-381

[8] Fama, E. (1965), "Behavior of Stock Market Prices," Journal of Business, 38, 34-105.

[9] Feller, W. (1971), An introduction to Probability Theory and Its Applications, Vol. II, 2nd edition. Wiley, New York.

[10] Gray, S. F. (1996) "Modeling the conditional distribution of interest rates as a regime-switching process", *Journal of Financial Economics,* 42, 27-62.

[11] Haas, M., S. Mittnik, and M. Paolella (2004), "A New Approach to Markov-Switching GARCH Models", *Journal of Financial Econometrics, 2*(4), 493-530.

[12] Hamilton, J.D. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

[13] Ibragimov, R. (2004), "On the Robustness of Economic Models to Heavy-Tailedness Assumptions", Working Paper, Yale University.

[14] Jorion, P. (2001) *Value at Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, New York.

[15] Koenker, R. and G. Bassett (1978), "Regression Quantiles," Econometrica, 46, 33-50.

[16] Malevergne, Y. and D. Sornette (2005), *Extreme Financial Risk: From dependence to risk management*, Springer Verlag.

[17] Mandelbrot, B. (1963), "The Variation of Speculative Prices", *Journal of Business*, 36, 394-419.

[18] Merton, R.C., (1987), "A Simple Model of Capital Market Equilibrium with Incomplete Information", *Journal of Finance*, 42, 483-510.

[19] Mittnik, S. and S. T. Rachev (1993), "Modeling Asset Returns with Alternative Stable Distributions", *Econometric Reviews*, 12(3), 261-330.

[20] Mittnik, S., M. S. Paolella,and S. T. Rachev (2000), "Diagnosis and treating the fat tails in financial returns data", *Journal of Empirical Finance*, 7, 389-416.

[21] Ortobelli, S., S. T. Rachev and E. Schwartz (2000), "The Problem of Optimal Asset Allocation with Stable Distributed Returns", Working Paper, Anderson Graduate School of Management.

[22] Samorodnitsky and M. S. Taqqu (1994), "Stable Non-Gaussian Random Processes - Stochastic Models with Infinite Variance," Chapman &Hall, New York.

[23] Sentana, E. (2005), "Least Squares Predictions and Mean-Variance Analysis", *Journal of Financial Econometrics*, 3(1), 56-78.

[24] Taylor, S. J. (1982), "Financial returns modelled by the product of two stochastic processes: a study of daily sugar prices 1961-79," in O. D. Anderson (Ed.), Time Series Analysis: Theory and Practice, 1, 203-226. Amsterdam: North-Holland.

Appendix

Proof of Proposition 3.1

To prove this proposition we need the following Lemma.

Lemma: *For an increasing function g, statements (a) and (b) below are equivalent*

(a) $g(u+v) > g(u) + g(v)$ *for all positive u and v,*

(b) $g^{-1}(u+v) < g^{-1}(u) + g^{-1}(v)$, *for all positive u and v.*

Proof of Lemma:

Suppose that (a) is true but (b) is not. So there exists positive u_0 and v_0 such that $g^{-1}(u_0+v_0) \geq g^{-1}(u_0) + g^{-1}(v_0)$. Since g is increasing, we have: $g\left(g^{-1}(u_0+v_0)\right) \geq g\left(g^{-1}(u_0) + g^{-1}(v_0)\right) \overset{\text{by (a)}}{>} g\left(g^{-1}(u_0)\right) + g\left(g^{-1}(v_0)\right)$, that is $u_0 + v_0 > u_0 + v_0$, an impossibility. Therefor, by contradiction (a) $\Rightarrow$ (b) and similarly (b) $\Rightarrow$ (a), and the lemma follows.

For the proposition, we have:

$$\begin{aligned} g^{-1}(\frac{a_X}{p}) &= Lim_{p\longmapsto 0} VaR_p(X), \\ g^{-1}(\frac{a_Y}{p}) &= Lim_{p\longmapsto 0} VaR_p(Y), \\ g^{-1}(\frac{a_X+a_Y}{p}) &= Lim_{p\longmapsto 0} VaR_p(X+Y) \end{aligned}$$

so, $Lim_{p\longmapsto 0}\left[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)\right] = g^{-1}(\frac{a_X+a_Y}{p}) - g^{-1}(\frac{a_X}{p}) - g^{-1}(\frac{a_Y}{p})$.

(i) If for all positive u and v: $g(u+v) > g(u)+g(v)$, then $g(\frac{a_X+a_Y}{p}) > g(\frac{a_X}{p}) + g(\frac{a_Y}{p})$. This implies

$g^{-1}(\frac{a_X+a_Y}{p}) < g^{-1}(\frac{a_X}{p}) + g^{-1}(\frac{a_Y}{p})$. Therefore:

$Lim_{p\longmapsto 0}\left[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)\right] < 0$. So, there exists $p_0 \in \left]0,1\right]$ such that, for any $p \in \left]0, p_0\right[$:

$$VaR_p\left[X+Y\right] < VaR_p(X) + VaR_p(Y).$$

(ii) If for all positive u and v: $g(u+v) < g(u)+g(v)$, then $g(\frac{a_X+a_Y}{p}) < g(\frac{a_X}{p}) + g(\frac{a_Y}{p})$. This implies

$g^{-1}(\frac{a_X + a_Y}{p}) > g^{-1}(\frac{a_X}{p}) + g^{-1}(\frac{a_Y}{p})$. Therefore:

$Lim_{p \longmapsto 0}\,[VaR_p(X+Y) - VaR_p(X) - VaR_p(Y)] > 0$.So, there exists $p_0 \in\,]0,1]$ such that, for any $p \in\,]0, p_0[$:

$$VaR_p\,[X+Y] < VaR_p(X) + VaR_p(Y), \mathbf{Q.E.D.}$$

Measuring the Degree of Violation of Subadditivity:

We propose below a way to measure the degree of violation of subadditivity. While nonsubadditivity means:

$$VaR_p\,(X+Y) > VaR_p(X) + VaR_p(Y), \tag{6.19}$$

that is the loss of the portfolio $(X+Y)$ can exceed $VaR_p(X) + VaR_p(Y)$ with probability p, the question is with what probability $kp, k \geq 1$, the loss of the portfolio $(X+Y)$ can exceed $VaR_p(X) + VaR_p(Y)$ with probability p, that is:

$$VaR_{kp}(X+Y) > (VaR_p(X) + VaR_p(Y). \tag{6.20}$$

While violation (6.19) of subadditivity means that (6.20) is fulfilled with $k = 1$, it cannot be fulfilled with $k = 2$ since, for any random variables X and Y:

$$VaR_{2p}(X+Y) \leq VaR_p(X) + VaR_p(Y) \tag{6.21}$$

Indeed:

$$P\,[X+Y \leq -VaR_p(X) - VaR_p(Y)] \leq 2p$$

$$\begin{aligned} & P\,[X+Y \leq -VaR_p(X) - VaR_p(Y)] \\ \leq\ & P\,[X \leq -VaR_p(X)] + P\,[Y \leq -VaR_p(Y)] = 2p. \end{aligned}$$

However, we can show for any k in $]0,2[$, there exists α_0 in $]0,1[$ such that for any α in $]0,\alpha_0[$, for $F_X(x) \sim_{x \longrightarrow -\infty} a_X[-x]^{-\alpha}$, there exists p_0 in $]0,1[$ such that for any p in $]0,p_0[$:

$$VaR_{kp}(X+Y) > VaR_p(X) + VaR_p(Y)$$

As a function of $\alpha \in]0,1[$, the function $\dfrac{[\sigma_X+\sigma_Y]^\alpha}{\sigma_X^\alpha+\sigma_Y^\alpha}$ is continuous and increasing, from the value 1/2 for $\alpha = 0$ to the value 1 for $\alpha = 1$. Then, for k given in $]0,2[$, there exists α_0 such that, for any α in $]0,\alpha_0[$:

$$\frac{[\sigma_X+\sigma_Y]^\alpha}{\sigma_X^\alpha+\sigma_Y^\alpha} < \frac{1}{k},$$

that is:

$$\sigma_X+\sigma_Y < k^{-1/\alpha}\,[\sigma_X^\alpha+\sigma_Y^\alpha]^{1/\alpha} = k^{-1/\alpha}\sigma_Z.$$

Since $\sigma_X = \lim\limits_{p\to 0} p^{1/\alpha} VaR_p(X)$, $\sigma_Y = \lim\limits_{p\to 0} p^{1/\alpha} VaR_p(Y)$
and $\sigma_Z = \lim\limits_{p\to 0}(kp)^{1/\alpha} VaR_{kp}(z)$, we have the existence of p_0 such that for any p in $]0,p_0[$:

$$p^{1/\alpha}VaR_p(X) + p^{1/\alpha}VaR_p(Y) < k^{-1/\alpha}(kp)^{1/\alpha}VaR_{kp}(Z)$$

that is $VaR_p(X) + VaR_p(Y) < VaR_{kp}(Z)$.

Proof of Proposition 3.3

Proposition 3.3. is a direct consequence of equation (3.11). Indeed, one can write:

$$\begin{aligned} E[(Y)^n] &= E\{E[(Y)^n \mid Z]\} = E\{k_{m,n}(Z)\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \\ &= E\{k_{m,n}(Z)\}E\{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} + Cov\{k_{m,n}(Z), \{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \end{aligned}$$

with $k_{m,n}(Z) = \frac{E[(Y)^n|Z]}{\{E[(Y)^m|Z]\}^{\frac{n}{m}}}$. Note that, since $0 < m < n$, Jensen's inequality gives:

$$E\{\{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\} \geq \{E\{E[(Y)^m \mid Z]\}\}^{\frac{n}{m}} = \{E[(Y)^m]\}^{\frac{n}{m}}.$$

The inequality becomes strict as soon as the conditioning information Z is not independent from Y. The spread widens when the information content of Z about Y increases. Therefore, we conclude that, as soon as Z and Y are not independent:

$$\frac{E[(Y)^n]}{\{E[(Y)^m]\}^{\frac{n}{m}}} > E\{k_{m,n}(Z)\} + \frac{Cov\{k_{m,n}(Z), \{E[(Y)^m \mid Z]\}^{\frac{n}{m}}\}}{E\{k_{m,n}(Z)\}} \tag{6.22}$$

To prove Proposition 3.3, we simply note that, in the case of a scale mixture, $k_{m,n}(Z)$ is a constant equal to $\frac{E[|u|^n]}{\{E[|u|^m]\}^{\frac{n}{m}}}$.

Proof of Proposition 4.1:

For A random event, we define the random variable:

$$\mathbf{1}_A = \begin{cases} 1 \text{ if } A \text{ occurs} \\ 0 \text{ otherwise.} \end{cases}$$

Assumption A2 implies that:

$$\mathbf{1}_{[X_j \leq -VaR_p(X_j|I_j)]} \leq \mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]}.$$

But these two random variables have the same conditional expectation given I_j since by definition:

$$\begin{aligned} E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I_j)]} \,|I_j\right] &= \Pr\{X_j \leq -VaR_p(X_j\,|I_j)\,|I_j\} \\ &= p \end{aligned}$$

and by the law of iterated expectations:

$$\begin{aligned} E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]} \,|I_j\right] &= E\left[E\left[\mathbf{1}_{[X_j \leq -VaR_p(X_j|I)]} \,|I\right] |I_j\right] \\ &= E\left[\Pr\{X_j \leq -VaR_p(X_j\,|I)\,|I\}\,|I_j\right] \\ &= E\left[p\,|I_j\right] \\ &= p \end{aligned}$$

Therefore, these two random variables must coincide I_j almost surely. This achieves the proof of proposition 2.1.

Proof of Proposition 4.2:

Since for all $j, VaR_p\,(X_j\,|I_j) \leq S_j$ and, by assumption A2:

$$VaR_p\,(X_j\,|I) \leq VaR_p\,(X_j\,|I_j)$$

we have:

$$p(I) = P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} S_j\,|I\right] \leq P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} VaR_p(X_j\,|I)\,|I\right]$$

But, by assumption A1:

$$VaR_p\left(\sum_{j=1}^{n} X_j\,|I\right) \leq \sum_{j=1}^{n} VaR_p(X_j\,|I).$$

Thus:

$$p(I) \leq P\left[\sum_{j=1}^{n} X_j \leq -VaR_p(\sum_{j=1}^{n} X_j\,|I)\,|I\right] = p.$$

Since this inequality is identically true, for all possible values of the random information set I, we conclude by considering unconditional expectations that:

$$P\left[\sum_{j=1}^{n} X_j \leq -\sum_{j=1}^{n} S_j\right] \leq p.$$

A fortiori:

$$P\left[\sum_{j=1}^{n} X_j \leq -VaR_p^*\right] \leq p$$

Proof of Proposition 5.1:

Let us define the following variables

$$X_1 \equiv R_1^1 - \mu_0 \sim N\left(\overline{\mu}, \overline{\sigma}^2\right)$$
$$X_2 \equiv R_2^1 - \mu_0 \sim N\left(\overline{\mu}, \overline{\sigma}^2\right)$$
$$Z \equiv R^0 - \mu_0 \sim N\left(0, \overline{\sigma}^2\right)$$

we have

$$\widetilde{R}_1 = s_1 X_1 + (1 - s_1)\,Z + \mu_0$$
$$\widetilde{R}_2 = s_2 X_2 + (1 - s_2)\,Z + \mu_0$$

and

$$
\begin{aligned}
p &= P\left[\widetilde{R}_1 + \widetilde{R}_2 \leq \mu_1 + \mu_0\right] \\
&= P\left[s_1 X_1 + (1-s_1) Z + s_2 X_2 + (1-s_2) Z \leq \mu_1 - \mu_0\right] \\
&= P\left[s_1 s_2 (X_1 + X_2) + s_1 (1-s_2)(X_1 + Z) + \right. \\
&\qquad \left. + s_2 (1-s_1)(X_2 + Z) + 2(1-s_1)(1-s_2) Z \leq \overline{\mu}\right] \\
&= \lambda^2 F_{X_1+X_2}(\overline{\mu}) + \lambda(1-\lambda) F_{X_1+Z}(\overline{\mu}) + \lambda(1-\lambda) F_{X_2+Z}(\overline{\mu}) + (1-\lambda)^2 F_{2Z}(\overline{\mu}) \\
&= \lambda^2 F_{X_1+X_2}(\overline{\mu}) + \lambda(1-\lambda) + (1-\lambda)^2 F_{2Z}(\overline{\mu})
\end{aligned}
$$

i.e. $p = \lambda^2 \Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) + \lambda(1-\lambda) + (1-\lambda)^2 \Phi(\overline{\mu}/2\overline{\sigma})$

Let us define

$$
\begin{aligned}
U_1 &\equiv s_1 X_1 + (1-s_1) Z \\
U_2 &\equiv s_2 X_2 + (1-s_2) Z
\end{aligned}
$$

We can see that $U_j = \widetilde{R}_j - \mu_0$, $j = 1, 2$. Therefore the proposition is equivalent to $VaR_p(U_1 + U_2) > VaR_p(U_1) + VaR_p(U_2)$. By definition of p, we have $VaR_p(U_1 + U_2) = -\overline{\mu}$, and since $U_1 \stackrel{d}{=} U_2$ then $VaR_p(U_1) = VaR_p(U_2)$ and the problem becomes $VaR_p(U_1 + U_2) > 2VaR_p(U_1)$ i.e. $VaR_p(U_1) < -\overline{\mu}/2$ or equivalently $\Pr(U_1 \leq \overline{\mu}/2) < p$.

Therefore, we have:

$$
\begin{aligned}
P(U_1 \leq \overline{\mu}/2) &= P\left[s_1 X_1 + (1-s_1) Z \leq \overline{\mu}/2\right] \\
&= \lambda F_{X_1}(\overline{\mu}/2) + (1-\lambda) F_Z(\overline{\mu}/2)
\end{aligned}
$$

i.e. $P(U_1 \leq \overline{\mu}/2) = \lambda \Phi(-\overline{\mu}/2\overline{\sigma}) + (1-\lambda) \Phi(\overline{\mu}/2\overline{\sigma})$

so

$$
\begin{aligned}
p - P(U_1 \leq \overline{\mu}/2) &= \lambda^2 \left[\Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) + \Phi(\overline{\mu}/2\overline{\sigma}) - 1\right] \\
&> 0
\end{aligned}
$$

since $\Phi\left(-\overline{\mu}/\sqrt{2}\overline{\sigma}\right) > \Phi(-\overline{\mu}/2\overline{\sigma})$ and $\Phi(-\overline{\mu}/2\overline{\sigma}) + \Phi(\overline{\mu}/2\overline{\sigma}) = 1$

i.e. $P(U_1 \leq \overline{\mu}/2) < p$, and the proposition follows. **Q.E.D**

Conclusion générale

Dans le premier chapitre de cette thèse, en établissant un lien entre la corrélation extrême et la dépendance des queues nous avons montré, que les modèles de changement de régime et GARCH classiques reproduisent partiellement ou ne reproduisent pas du tout l'asymétrie observée dans la dépendance. Nous avons alors proposé un modèle basé sur les copules qui permet de prendre en compte cette asymétrie. Au moyen de ces outils, nous avons pu construire un modèle à quatre variables, qui nous a permis non seulement d'analyser le comportement de la corrélation mais aussi la dépendance dans les queues. Ce modèle nous a permis d'analyser le comportement des rendements des actions et des obligations sur les marchés internationaux, notamment entre le Canada et les Etats-Unis d'une part, la France et l'Allemagne d'autre part. Les résultats empiriques ont révélé une plus forte dépendance internationale entre les actifs du même type par rapport à la dépendance entre les obligations et les actions même lorsqu'on considère le même pays. D'autre part, nous avons mis en exergue une relation entre la volatilité du taux d'échange et l'asymétrie dans la dépendance. Ainsi, la dépendance entre la France et l'Allemagne s'est révélée très asymétrique avant l'introduction de la monnaie unique européenne (Euro), alors que cette asymétrie a considérablement diminué après l'introduction de cette monnaie.

Ce modèle nous a permis d'apporter un élément d'explication additionnelle aux phénomènes de la faible diversification internationale et la tendance des investisseurs à se rabattre sur les actifs moins risqués comme les obligations au détriment des actions. L'intuition derrière cette explication est qu'en présence d'une forte dépendance dans les moments de baisse sur les marchés, les bénéfices liés à la diversification diminuent du fait que l'effet de la diversification s'estompe au moment ou les investisseurs en ont le plus besoin.

Dans le second chapitre, nous avons montrer que lorsque cette asymétrie n'est pas prise en compte dans un modèle et que la loi Gaussienne est utilisée comme c'est souvent le cas dans la pratique, les mesures de risques extrêmes telles que la VaR et la ES sont sous-estimées. Nous nous sommes rendu compte que bien que le cadre général du DCC estime assez bien la VaR au niveau 5% pour toutes les structures de dépendance, il sous-estime cette mesure pour des niveau beaucoup plus prudentiels (1%, 0.5%) lorsque l'asymétrie présente dans les données n'est pas prise en compte. Le fait que les modèles de dépendance asymétrique comme la Gaussienne et la t de Student sous-estiment l'ES au niveau 5%, bien qu'ils aient donné plutôt une bonne estimation de la VaR à ce même niveau de couverture est apparu comme un résultat inattendu bien que explicable. Une explication peu être liée au fait que si ces distributions génèrent des portefeuilles avec des queues moins épaisses qu'elles devraient l'être, alors elles auraient tendance à sous-estimer la moyenne dans les queues. La copule de Gumbel qui prend en compte l'asymétrie observée dans la structure de dépendance s'est révélée supérieure dans la précision de l'estimation des mesures de risques extrêmes en situation de dépendance asymétrique.

Dans le troisième chapitre, le problème de la cohérence de la VaR comme mesure de risque a été abordé. L'idée clef de notre démarche a été que si l'épaisseur des queues de distributions est responsable de la violation de la sous-additivité, une utilisation appropriée de l'information conditionnelle pourrait rendre la VaR plus rationnelle pour la gestion décentralisée du risque. Trois arguments soutiennent cette demarche. Premièrement, partant du fait que les traders possèdent sur leur segment de marché une information dite privée plus riche que le gestionnaire central, Ils doivent simplement respecter les contraintes prudentielles imposées par celui-ci pour que le contrôle de la VaR décentralisée fonctionne de façon cohérente. Nous avons montré par la suite, que dans ce contexte de décentralisation, si le gestionnaire central a accès ex-post à la composition du portefeuille des traders individuels, il pourra récupérer une bonne part de leur information privée. Finalement, en utilisant les distributions à queues épaisses telles que les lois stables et les lois de type Pareto, nous avons montré que l'épaisseur exigée des queues pour violer la sous-additivité même pour les petits niveaux de probabilités, induit une situation tellement extrême qu'elle correspond à une information tellement faible que la perte espérée est infinie. Nous concluons dont que l'incohérence de la gestion décentralisée par la VaR caractérisée par l'absence de sous-additivité avec une information assez riche, est une exception et non une règle, d'autant plus que dans la pratique, la moyenne conditionnelle ou inconditionnelle est en général supposée finie.