Université de Montréal

Asymmetry Risk, State Variables and Stochastic Discount Factor Specification in Asset Pricing Models

par

Fousseni Chabi-Yo

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

Octobre 2004

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Asymmetry Risk, State Variables and Stochastic Discount Factor Specification in Asset Pricing Models

présentée par

Fousseni Chabi-Yo

a été évaluée par un jury composé des personnes suivantes:

Silvia Goncalves
président -rapporteur
Eric Renault
directeur de recherche
René Garcia
codirecteur de recherche
Bryan Campbell
membre du jury (Concordia University)
Angelo Melino
examinateur externe (University of Toronto)
Yoshua Bengio
représentant du doyen de la FES

Thèse acceptée le 4 Novembre 2004

Sommaire

Ma thèse est centrée sur l'introduction de l'asymétrie dans les modèles d'évaluation d'actifs financiers et dans le choix de portefeuille. Dans le premier chapitre, nous examinons comment l'équilibre d'un marché financier révèle, à la fois par les quantités détenues à l'équilibre et par les prix, les préférences des investisseurs pour trois types de caractéristique des rendements : leur espérance, leur variance et leur asymétrie. Dans le deuxième chapitre, en prenant en compte l'asymétrie, nous déterminons une nouvelle borne sur la variance de tout facteur d'actualisation stochastique (SDF) qui valorise correctement les rendements d'actifs financiers et les gains de produits dérivés qui sont des fonctions quadratiques des gains d'actifs risqués. Dans le troisième chapitre, nous construisons une économie où les préférences des investisseurs et leur consommation dépendent d'une variable d'état qui suit un processus de type Markovien à deux états et montrons que ce modèle économique produit et explique les énigmes de l'aversion pour le risque et du SDF mises en évidence par Jackwerth (2000, RFS). Dans le quatrième chapitre, nous proposons une approche pour l'évaluation de produits dérivés par les arbres lorsqu'une variable d'état non observable affecte le processus de prix du sous-jacent.[1]

Dans le premier chapitre, nous examinons comment l'équilibre d'un marché financier révèle, à la fois par les quantités détenues à l'équilibre et par les prix, les préférences des investisseurs pour trois types de caractéristique des rendements : leur espérance, leur variance et leur asymétrie. Deux types d'approche sont utilisés pour cela. D'abord, en considérant une situation au voisinage de la non-incertitude (expansion en petit bruit), on calcule les demandes des agents pour différents types d'actifs risqués. L'idée est de considérer un actif en offre non nulle, représentatif du portefeuille de marché, et des actifs dérivés en offre nette nulle mais dont les gains sont des fonctions non linéaires du portefeuille de marché. On s'aperçoit alors que la demande d'actifs dérivés est précisément justifiée par le goût des investisseurs pour l'asymétrie. Au niveau des prix, la rémunération du risque dépend non seulement du beta de marché, comme dans un contexte moyenne-variance classique, mais aussi d'un coefficient de coasymétrie par rapport au marché. Les conclusions obtenues par

[1]Le premier chapitre de cette thèse a été écrit en collaboration avec Dietmar Leisen et Eric Renault. Le deuxième, troisième et le quatrième chapitre ont été écrits en collaboration avec René Garcia et Eric Renault.

l'expansion en petit bruit peuvent ensuite être retrouvées dans des contextes plus généraux grace à la définition d'un facteur d'actualisation stochastique adapté. Cette double approche peut être ensuite étendue à un marché à deux périodes où d'autres phénomènes d'asymétrie doivent être pris en compte dans la dépendance temporelle des rendements d'une période à l'autre.

L'objet du deuxième chapitre est l'extension de l'approche des bornes de variance proposée par Hansen et Jagannathan (1991, JPE). Alors que Hansen and Jagannathan (1991, JPE) caractérisent la variance minimale que doit avoir un facteur d'actualisation stochastique (SDF) susceptible de valoriser correctement un ensemble donné d'actifs primitifs, nous considérons l'effet sur cette borne de variance de l'ajout de contraintes imposées par l'évaluation correcte des fonctions quadratiques des gains de ces actifs primitifs. Nous approchons ainsi le problème de l'évaluation d'actifs dérivés dont les gains sont par définition des fonctions non linéaires des gains des actifs sous-jacents. Il est alors éclairant de décrire la nouvelle frontière de variance ainsi obtenue dans un espace à trois dimensions mettant en jeu non seulement les rendements espérés et leur variance mais aussi leur coefficient d'asymétrie. De même que la frontière de variance de Hansen and Jagannathan (1991, JPE) présente une relation de dualité avec la frontière efficiente moyenne-variance du choix optimal de portefeuille au sens de Markowitz (1952, JF), la frontière que nous proposons peut être interprétée en termes de choix de portefeuille par minimisation du risque sous contrainte non seulement de coût et de rendement espéré, mais aussi d'une contrainte qui dépend de l'asymétrie du portefeuille. Nous montrons que la solution du problème de minimisation du risque sous contrainte de coût, du rendement espéré et d'asymétrie du portefeuille proposée par de Athayde et Flores (2004, JEDC) est un cas particulier de notre problème de choix de portefeuille. En ce sens, notre travail donne un nouvel éclairage à la question du choix de portefeuille en présence de rendements asymétriques.

Dans le troisième chapitre, nous présentons un modèle économique avec changements de régime qui produit et explique les énigmes de l'aversion pour le risque et du SDF mises en évidence dans Jackwerth (2000, RFS). Nous construisons un modèle où les préférences des investisseurs et leur consommation dépendent d'une variable d'état qui suit un processus de type Markovien à deux états et simulons les prix d'options d'achat européennes. En utilisant la méthodologie proposée par

Jackwerth (2000, RFS), nous déduisons la fonction d'aversion absolue pour le risque et du SDF pour chaque valeur de la richesse. Ces fonctions présentent les mêmes énigmes que celles observées par Jackwerth (2000, RFS). Lorsque nous appliquons la même méthodologie dans chaque état de l'économie, l'énigme de l'aversion absolue pour le risque disparaît. Nos résultats suggèrent que ce modèle rationalise et explique l'énigme de l'aversion pour le risque et du SDF mises en évidence par Jackwerth (2000, RFS).

Dans le quatrième chapitre, nous présentons un modèle d'évaluation des produits dérivés par la méthode d'arbre lorsque le processus du prix du sous-jacent est affecté par une variable d'état non observable. Ce modèle généralise les modèles d'arbre existants : Cox, Ross et Rubinstein (1979) et Boyle (1988). Dans ce modèle, la variable d'état non observable capture les faits marquants mis en évidence par l'observation des prix d'options, en particulier l'asymétrie et la dynamique de l'asymétrie présentes dans les actifs dérivés.

Mots clés: variable d'état, modèle d'arbre, choix de portefeuille, énigme de l'aversion pour le risque, énigme du facteur d'actualisation stochastique, asymétrie, facteur d'actualisation stochastique.

Summary

My thesis focuses on the introduction of asymmetry in asset pricing models and portfolio selection. In the first chapter, we use a small noise expansion approach to investigate how the market equilibrium discloses, through quantities and prices, investors' preferences for three characteristics of asset returns: expected return, variance and skewness. In the second chapter, taking into account asset higher moments, we find a new bound on the volatility of any admissible stochastic discount factor (SDF) that prices correctly a set of primitive asset returns and derivatives which payoffs are a quadratic function of the same primitive assets. We further propose a method for portfolio selection which accounts for higher moments, in particular skewness. In the third chapter, we develop a utility-based economic model with state dependence in fundamentals and preferences which rationalizes and explains the risk aversion and pricing kernel puzzles put forward in Jackwerth (2000, RFS). Chapter four proposes a lattice-based model for valuing derivatives when the underlying process is affected by an unobservable state variable.

The first chapter examines how the market equilibrium discloses, through quantities and prices, investors' preferences for three characteristics of asset returns: expected return, variance and skewness. We use a small-noise expansion approach to compute heterogeneous agents' demands for several risky assets. The idea is to consider a risky asset in positive net supply which represents the market portfolio and derivatives assets in zero net supply which payoffs are nonlinear functions of the market return. We observe that the demand for derivative assets comes from the fact that investors have a preference for skewness. With regard to the equilibrium prices of derivative assets, we find that the risk is priced through the market beta like in the standard mean-variance analysis but also through an additional parameter which is the co-skewness of derivative assets with respect to the market. These findings obtained under a small noise expansion approach can be found in a more general context if we define an appropriate stochastic discount factor. This methodology can be extended in a two-period market to see how other skewness effects should be taken into account to explain temporal dependence between asset returns across time.

The second chapter extends the well-known Hansen and Jagannathan (1991, JPE) volatility bound. Hansen and Jagannathan characterize the volatility lower bound of any admissible SDF

that prices correctly a set of primitive asset returns. We characterize this lower bound for any admissible SDF that prices correctly both primitive asset returns and quadratic payoffs of the same primitive assets. In particular, we aim at pricing derivatives which payoffs are defined as nonlinear functions of the underlying asset payoffs. We put forward a new volatility surface frontier in a three-dimensional space by considering not only asset expected payoffs and variances, but also asset skewness. Since there exists a duality between the Hansen and Jagannathan (1991, JPE) mean-variance frontier and Markowitz (1952, JF) mean-variance portfolio frontier, our volatility surface frontier can be interpreted in terms of portfolio selection by minimizing the portfolio risk subject to portfolio cost and expected return as usual, but also to an additional constraint which depends on the portfolio skewness. This approach which consists in finding the lower risk portfolio subject to portfolio cost, expected return and skewness, embeds the mean-variance-skewness portfolio choice of de Athayde and Flores (2004, JEDC). In this sense, our paper sheds light on portfolio selection when asset returns exhibit skewness.

The third chapter examines the ability of economic models with regime shifts to rationalize and explain the risk aversion and pricing kernel puzzles put forward in Jackwerth (2000, RFS). We build an economy where state dependences are introduced either in investors' preferences or fundamentals and simulate European call option prices. Following Jackwerth's (2000, RFS) nonparametric methodology, we recover the risk aversion and pricing kernel functions across wealth states. These functions exhibit the same puzzle found in the data. However, when we apply the same methodology within each regime the puzzles disappear. Our findings suggest that state dependence in preferences or fundamentals potentially explains the risk aversion puzzle.

The last chapter presents a lattice-based method for valuing derivatives when the underlying process is affected by an unobservable state variable. This model generalizes the existing lattice models: Cox, Ross and Rubinstein (1979) and Boyle (1988) trinomial pricing model. In this model, an unobservable state variable captures the salient features of derivatives such as options, in particular skewness and the dynamic effect of asset skewness.

Key words: lattice, portfolio choice, pricing kernel puzzle, risk aversion puzzle, skewness, state variable, stochastic discount factor.

Contents

List of Tables

List of Figures

À ma femme, Bignon,

à mon père, Sarebou,

à ma mère, Assanatou,

et à mes frères et soeurs.

Remerciements

De nombreuses personnes et institutions sont à remercier dans l'accomplissement de cette thèse. Tout d'abord mes directeurs de recherche Éric Renault et René Garcia qui ont su orienter et guider mes recherches de manière fructueuse. Durant les années de préparation de cette thèse, ils ont su m'encourager et me prodiguer des conseils. Sans leur soutien, j'aurais sûrement abandonné.

Je voudrais remercier mon co-auteur Dietmar Leisen pour sa collaboration significative dans le deuxième essai 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), MITACS (The Mathematics of Information Technology and Complex Systems) et Hydro-Quebec. Je voudrais également remercier la Banque du Canada pour son support dans la rédaction finale de cette thèse.

Je voudrais également remercier mon père Sarebou, ma mère Assanatou, mes frères et sœurs, Boni, Hassan, Sani, Mansou, Rachi et mon épouse Bignon de m'avoir non seulement enduré, mais encouragé et supporté dans les périodes les plus éprouvantes de mes études.

Introduction générale

La validité des modèles d'évaluation d'actif financiers dépend de leur capacité à reproduire les caractéristiques des prix observés sur le marché. Sous certaines conditions dont celle d'absence d'opportunité d'arbitrage, Hansen et Richard (1987) montrent qu'il existe un facteur d'actualisation aléatoire qui sert à évaluer le prix de tous actif financier. De part sa nature aléatoire, ce facteur est appelé facteur d'actualisation stochastique (SDF). Hansen et Richard (1987) montrent que le prix d'un actif financier s'écrit comme la valeur espérée du produit du SDF et du gain de cet actif. La spécification de ce facteur dépend en général des hypothèses sur les préférences des investisseurs. La ligne directrice de cette thèse est l'étude parcimonieuse des différentes spécifications de ce SDF et leur implication en terme d'évaluation d'actifs financiers, de produits dérivés, de préférence et de choix de portefeuille.

Dans le premier essai de cette thèse, nous examinons comment l'équilibre d'un marché financier révèle, à la fois par les quantités détenues à l'équilibre et par les prix, les préférences des investisseurs pour trois types de caractéristique des rendements : leur espérance, leur variance et leur asymétrie. Deux types d'approche sont utilisés pour cela. D'abord, en considérant une situation au voisinage de la non-incertitude (expansion en petit bruit), on calcule les demandes des agents pour différents types d'actifs risqués. L'idée est de considérer un actif en offre non nulle, représentatif du portefeuille de marché, et des actifs dérivés en offre nette nulle mais dont les gains sont des fonctions non linéaires du portefeuille de marché. En faisant une expansion en petit bruit au premier ordre, on s'aperçoit que la demande d'actifs dérivés est déterminée uniquement par l'aversion des investisseurs pour la variance. Au niveau des prix, la rémunération du risque dépend du bêta de marché comme dans un contexte moyenne-variance (voir Markowitz (1952)). Le SDF impliqué par ce modèle est une fonction linéaire du rendement de marché. En d'autres termes une expansion en petit bruit au premier ordre produit le modèle d'évaluation d'actifs financiers CAPM. Toutefois, de nombreuses études empiriques ont souligné que ce modèle n'est pas pertinent en terme d'évaluation d'actifs financiers. Ce qui nous a conduit à faire une expansion en petit bruit au deuxième ordre. Dans ce cas, on s'aperçoit alors que la demande d'actifs dérivés est précisément justifiée par le goût des investisseurs pour l'asymétrie. Au niveau des prix, la rémunération du risque dépend non seulement du bêta de marché, comme dans un contexte moyenne-variance, mais aussi d'un coefficient de coasymétrie par rapport au marché. Le SDF impliqué par une expansion en petit bruit au deuxième ordre est une fonction quadratique du rendement de marché. Cette approche peut être étendue à un marché à deux périodes où d'autres phénomènes d'asymétrie doivent être pris en compte dans la dépendance temporelle des rendements d'une période à l'autre.

Une fois un facteur d'actualisation stochastique identifié, un des aspects importants est de le comparer à un SDF de référence pour s'assurer de sa pertinence ou de sa validité. Différentes méthodes sont proposées pour comparer les modèles d'évaluation d'actifs financiers, tester leur validité et s'assurer de leur pertinence. Bien souvent, ces méthodes utilisent comme référence le SDF proposé par Hansen et Jagannathan (1991). Étant donné une série de rendements observés, Hansen et Jagannathan (1991) déterminent la variance minimale que doit avoir un SDF pour évaluer correctement les rendements d'actifs financiers. Le SDF de Hansen et Jagannathan (1991) dépend des deux premiers moments des rendements d'actifs financiers, donc ne prend pas en compte l'asymétrie observée dans ces rendements. Pour les actifs fondamentaux comme les indices boursiers et les indices obligataires qui servent à l'évaluation des modèles, les moments d'ordre supérieur à deux ne jouent pas en général un rôle déterminant. Toutefois, dans plusieurs études empiriques, il est admis que ces deux premiers moments ne caractérisent pas entièrement la distribution des rendements. Un des faits stylisés est que la distribution des rendements est souvent asymétrique. Comme nous l'avons précisé dans le premier chapitre, l'asymétrie peut s'avérer importante pour la prise de décisions d'investissement. D'abord, un investissement avec une distribution de cash-flows fortement etalee a droite peut etre attractif meme si son ratio de Sharpe n'est pas tres eleve. Ensuite, les contrats d'option d'achat ou de vente exhibent evidemment des payoffs tres assymetriques. Pour toutes ces raisons, la frontiere proposee par Hansen et Jagannathan (1991) dans le plan moyeene-variance pour decrire les SDF admissibles ne met sans doute pas assez l'accent sur la remuneration de l'assymetrie.

Dans le deuxième essai de cette thèse, nous proposons un SDF de variance minimale parmi les

SDF qui évaluent correctement non seulement les rendements d'actifs financiers fondamentaux mais aussi ceux de produits dérivés et de stratégies financières complexes telles que celles utilisées par les fonds spéculatifs. Nous supposons que le gain de tout produit dérivé peut être approximé par une fonction quadratique des actifs primitifs. Intuitivement, nous augmentons l'ensemble des opportunités d'investissement des agents économiques en considérant non seulement les actifs financiers mais aussi les produits dérivés qui sont fonction des actifs primitifs. Tout comme le SDF de Hansen et Jagannathan (1991), ce SDF est simple et facile à utiliser, il peut être utilisé pour comparer les modèles d'évaluation d'actifs financiers et pour tester leur validité. Il peut être interprété comme une simple extension du SDF de Hansen et Jagannathan (1991). Toutefois, sa particularité est qu'il prend en compte les moments d'ordres supérieurs des rendements d'actifs, en particulier l'asymétrie observée dans ces rendements. Le SDF proposé dans cet essai est une fonction quadratique des rendements d'actifs primitifs. Récemment, de nombreux auteurs parmi lesquels Harvey et Siddique (2000) et Barone-Adesi et al. (2004) ont souligné l'importance d'utiliser un SDF qui est une fonction quadratique du rendement de marché pour étudier l'impact de l'asymétrie sur les rendements espérés d'actifs financiers. Par exemple, Harvey et Siddique (2000) montrent qu'un SDF fonction quadratique du rendement de marché permet d'expliquer les variations en coupe transversale des rendements espérés entre différents actifs. Tout comme le SDF de Hansen et Jagannathan (1991), nous montrons que le SDF proposé dans cet essai peut être interprété en terme de choix de portefeuille en proposant une simple approche de choix de portefeuille sous asymétrie. Cette approche est une simple extension de l'approche moyenne-variance (voir Markowitz (1952)) et de l'approche moyenne-variance-asymétrie (voir de Athayde et Flores (2004)). Cette dernière approche consiste à chercher le portefeuille le moins risqué (portefeuille ayant la plus petite variance) parmi tous les portefeuilles ayant un même coefficient d'asymétrie et une même valeur espérée. En terme de choix de portefeuille, cet essai apporte deux contributions. Premièrement, nous généralisons le problème de choix de portefeuille résolu par de Athayde et Flores (2004). Deuxièmement, nous proposons une approche simple qui permet de déduire facilement (sans une résolution numérique) la solution à ce problème.

Dans une première application empirique, nous illustrons la perte d'information sur le SDF qui résulte d'une utilisation du SDF de Hansen et Jagannathan (1991) lorsqu'il y a de forte présomptions que l'asymétrie est évaluée sur le marché. Dans une deuxième application empirique, nous utilisons le SDF proposé dans cet essai pour vérifier si les modèles basés sur la consommation expliquent ou non l'énigme de la prime de risque mise en évidence par Mehra et Prescott (1985). Le SDF proposé dans ce essai rend l'énigme de la prime de risque encore plus difficile à expliquer. Dans une troisième application empirique nous montrons que les investisseurs qui ont une préférence pour l'asymétrie choisissent un portefeuille autre que celui proposé dans de Athayde et Flores (2004). Ils ne choisissent le portefeuille proposé par de Athayde et Flores (2004) que sous des hypothèses plus restrictives qui ne sont en général pas vérifiées empiriquement.

Dans le troisième essai, nous présentons un modèle économique avec changements de régime qui produit et explique les énigmes de l'aversion pour le risque et du SDF mises en évidence dans Jackwerth (2000). En résolvant le problème de choix de portefeuille de l'agent économique on s'aperçoit que le SDF peut être interprété comme un taux marginal de substitution intertemporel. En admettant que la fonction d'utilité de l'agent économique est concave (pour un agent économique averse au risque), le taux marginal de substitution intertemporel doit être une fonction décroissante de la richesse de l'agent économique, tout comme d'ailleurs la fonction d'aversion absolue pour le risque de l'agent économique doit également être une fonction décroissante de sa richesse. Toutefois, les études empiriques (voir Jackwerth (2000) et Ait Sahalia et Lo (2000)) montrent que ni le SDF, ni la fonction d'aversion absolue pour le risque n'apparaissent comme des fonctions décroissantes de la richesse. Pour expliquer ce paradoxe, nous construisons un modèle où les préférences des investisseurs et leur consommation dépendent d'une variable d'état qui suit un processus de type Markovien à deux états et simulons les prix d'options d'achat européennes. En utilisant la méthodologie proposée par Jackwerth (2000), nous déduisons la fonction d'aversion absolue pour le risque et le SDF pour chaque valeur de la richesse. Ces fonctions présentent les mêmes énigmes que celles observées par Jackwerth (2000) et Ait Sahalia et Lo (2000). Lorsque nous appliquons la même méthodologie dans chaque état de l'économie, l'énigme de l'aversion absolue pour le risque disparaît. Nos résultats suggèrent que ce modèle rationalise et explique l'énigme de l'aversion pour le risque et du SDF mises en évidence par Jackwerth (2000).

Dans le quatrième chapitre, nous présentons un modèle d'évaluation des produits dérivés par la méthode d'arbre lorsque le processus du prix du sous-jacent est affecté par une variable d'état non observable. Dans un modèle de marché à une période, nous montrons que ce modèle est observationnellement équivalent au modèle proposé dans Boyle (1988). Sur deux périodes, nous montrons que ce modèle généralise le modèle de Boyle (1988). Dans ce modèle, la variable d'état non observable capture les faits marquants mis en évidence par l'observation des prix d'options, en particulier l'asymétrie et la dynamique de l'asymétrie présentes dans les actifs dérivés.

Chapter 1

Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing

1. Introduction

Asymmetric shocks are common on markets and will lead to payoffs that are not normally distributed and exhibit skewness. Moreover, even when the primitive assets have symmetric payoffs, typical derivative assets display a high degree of skewness. The risk-return trade off on such payoffs may not be captured well by mean-variance analysis. However, Samuelson (1970) argued that mean-variance analysis is still a valid approach to characterize the optimal portfolio problem in general, i.e. even in those cases when the decision maker has a general concave von Neuman-Morgenstern utility function and asset returns are not normally distributed. His result is based on the limit of portfolio holdings under infinitesimal risk. We argue, in the presence of "small" risks it is necessary to study also the slope of portfolio holdings in the neighborhood of zero risk, and thereby incorporate skewness risk into the analysis. This paper extends Samuelson's analysis of financial decision making to derive agents' portfolio holdings and the equilibrium allocation under mean-variance-skewness risk.

We characterize portfolio holdings using risk-tolerance and a term we call skew-tolerance which contains the third derivative of an agent's utility function. Risk-tolerance captures the mean-variance trade-off and skew-tolerance the mean-variance-skewness trade-off. Using appropriately defined "average" risk-tolerance and "average" skew-tolerance we show that such an "average" agent sets prices while each heterogeneous agent's holdings are proportional to the difference between the agent's skew-tolerance and that of the "average" agent. The proportionality factor is determined through co-skewness with the market; two-fund separation theorems typically do not hold under skewness risk. A related work is Judd and Guu (2001) where Samuelson's analysis is also extended to an asymptotically valid theory for the trade-off between one risky asset and the riskless asset in single period setups. However, while their approach is based on bifurcation theory, our results are based directly on limits of first order conditions.

Our paper makes the following contributions

First, we generalize Samuelson's analysis by not imposing that risk premia are locally proportional to variance. By relaxing this restriction, we are able to characterize the price of skewness in equilibrium. A significant result is that, although separation theorems do not hold under skewness risk, it remains true that any risk is compensated only through its relationship with the market, either through the standard market beta or through market co-skewness which is akin to a beta with respect to the squared market return. In this respect, one may say that neither idiosyncratic variance nor idiosyncratic skewness are compensated in equilibrium. We thereby provide a foundation for empirical studies that extend the CAPM model using in an ad-hoc way the squared market return as a second factor. Furthermore this paper provides a method to determine portfolio holdings under skewness risk.

Second, we study extensively the pricing implications of a Stochastic Discount Factor (SDF) specification that is quadratic with respect to market return. Although motivated by the above small risk analysis à la Samuelson (1970), this study is valid under very general settings and can be compared to previous literature on the pricing implications of skewness risks. We revisit beta pricing under skewness as already considered by Kraus and Litzenberger (1976), Barone-Adesi (1985), Harvey and Siddique (2000), Dittmar (2002), and Barone-Adesi, Urga and Gagliardini (2004) among others. For the purpose of derivative asset pricing, we also relate skewness pricing to risk neutral variance (Rosenberg(2000)) and price of volatility contracts (Bakshi and Madan (2000)). We shed more light on beta pricing relationships as proposed by Harvey and Siddique (2000) by showing that they correspond to a limit case which is strictly speaking at odds with a no-arbitrage requirement, namely the case of a zero risk-neutral variance of the market. We put forward a more general beta pricing relationship which explicitly depends on the price of the squared return on the market portfolio, or equivalently, on the market risk neutral variance.

Finally, while the statistical identification of a significantly positive skewness premium is generally considered to be a difficult task (Barone-Adesi, Urga and Gagliardini (2004)), we provide some empirical evidence which suggests that such premia show up in a more manifest way when they are considered from a conditional point of view. This evidence is documented from simulated data calibrated on the GARCH factor model with in mean effects recently estimated by Bekaert and Liu(2004). Moreover, this empirical evidence also shows that neglecting the market risk neutral variance as Harvey and Siddique (2000) beta pricing model does lead to a severe underestimation of the skewness premium which may go so far as to invert its sign.

The remainder of the paper is organized as follows: the next section discusses portfolio choice and asset pricing in the context of infinitesimal risks. Section 3 studies quadratic pricing kernels in the conditional setup of Hansen and Richard (1987). Section 4 makes an empirical assessment of the order of magnitude of the various effects put forward in section 3. All proofs are postponed to the appendix.

2. Static Portfolio Analysis in Terms of Mean, Variance and Skewness

Samuelson (1970) argues that, for risks that are infinitely small, optimal shares of wealth invested in each security coincide with those of a mean-variance optimizing agent. However Samuelson (1970) also derives a more general theorem about higher order approximations. To further characterize the way the optimal shares vary locally in the direction of any risk, that is their first derivatives at the limit point of zero risk, one needs to push one step further the Taylor expansion of the utility function; carrying this out will lead us to a mean-variance-skewness approach.

We start here from a slight generalization of Samuelson's approximation theorem. Following closely his exposition, let us denote respectively by R_i, $i = 1, ...n$, the return from investing \$1 in each of security i=1,...n. The random vector $R = (R_i)_{1 \leq i \leq n}$ defines the joint probability distribution of interest, which is specified by the following decomposition:

$$R_i(\sigma) = \mu + \sigma^2 a_i(\sigma) + \sigma Y_i. \tag{2.1}$$

Here, $a_i(\sigma)$, $i = 1, ...n$, are positive functions of σ and μ is the gross return on the riskless (safe) security. The σ parameter characterizes the scale of risk that is crucial for our analysis. We are typically interested in this section in local approximations in the neighborhood of $\sigma = 0$. The small noise expansion (2.1) provides a convenient framework to analyze portfolio holdings and resulting equilibrium allocations for a given random vector $Y = (Y_i)_{1 \leq i \leq n}$ with $E[Y] = 0$, and $Var(Y) = \Sigma$

a symmetric and positive definite matrix[1].

In equation (2.1), the term $\sigma^2 a_i(\sigma)$ has the interpretation of the risk premium. Samuelson (1970) restricts the function $a_i(\sigma)$ to constants; under this assumption risk premia are proportional to the squared scale of risk; we relax this restriction throughout since it would prevent us from analyzing the price of skewness in equilibrium. Throughout we refer to $a(\sigma) = (a_i(\sigma))_{i=1,\dots,n}$ as the vector of risk premia.

2.1 The individual investor problem

We consider an investor with Von Neumann-Morgenstern preferences, i.e. she derives utility from date 1 wealth according to the expectation over some increasing and concave function u evaluated over date 1 wealth; for given risk-level σ she then seeks to determine portfolio holdings $(\omega_i)_{1\leq i\leq n} \in \mathbb{R}^n$ of risky assets that maximize her expected utility.

$$\max_{(\omega_i)_{1\leq i\leq n}\in\mathbb{R}^n} Eu\left(\mu + \sum_{i=1}^{n} \omega_i\left(R_i(\sigma) - \mu\right)\right) \tag{2.2}$$

Note that for the sake of notational simplicity, the initial invested wealth is normalized to one. The solution of this program is denoted by $(\omega_i(\sigma))_{1\leq i\leq n}$ and depends on the given scale of the risk σ. The question we ask is then the following: to what extent does a Taylor approximation of u allow us to understand well the local behavior of the shares $\omega_i(\sigma)$, $i = 1,\dots n$, in the neighborhood of zero risk, $\sigma = 0$, that is to correctly characterize the two quantities:

$$\omega_i(0) = \lim_{\sigma\rightsquigarrow 0^+} \omega_i(\sigma) \text{ and } \omega_i'(0) = \lim_{\sigma\rightsquigarrow 0^+} \omega_i'(\sigma) \tag{2.3}$$

for i=1...n? Samuelson (1970) stresses that a third-order Taylor expansion of u is needed to do the job. We slightly extend his result by showing that its remains valid even though the function $a_i(\sigma)$ are not assumed to be constant.[2] Let us then consider a third order Taylor expansion of u in the

[1] Samuelson (1970) provides a heuristic explanation of (2.1) that is of interest for readers accustomed to continuous-time finance models; he couches this terms of Brownian motion of time and identifies σ with the square root of time.

[2] Let $W(\sigma) = \mu + \sum_{i=1}^{n} \omega_i(R_i(\sigma) - \mu)$ denote end of period wealth and note that $W(0) = \mu$. For the sake of simplicity, we denote $W(\sigma) = W$.

neighborhood of the safe return μ:

$$u^*(W) = u(\mu) + u'(\mu)(W-\mu) + \frac{u''(\mu)}{2!}(W-\mu)^2 + \frac{u'''(\mu)}{3!}(W-\mu)^3. \tag{2.4}$$

Let us denote by $(\omega_i^*(\sigma))_{1 \leq i \leq n}$ the solution of the approximated problem:

$$\max_{(\omega_i)_{1 \leq i \leq n}} Eu^*\left(\mu + \sum_{i=1}^{n} \omega_i (R_i(\sigma) - \mu)\right) \tag{2.5}$$

$\omega_i^*(0)$ and $\omega_i^{*'}(0)$, i=1...n are defined accordingly by continuity extension as in (2.3). We prove:

Theorem 2.1 *Under suitable smoothness and concavity assumptions, the solution to the general problem (2.2) is related asymptotically to that of the 3-moment problem (2.5) by the tangency equivalences:*

$$\begin{aligned} \omega_i(0) &= \omega_i^*(0), \\ \omega_i'(0) &= \omega_i^{*'}(0) \text{ for all } i=1,\ldots,n. \end{aligned}$$

This theorem states that third-order Taylor expansions give tangency equivalence. The intuition behind this result is:

1. The optimal shares of wealth invested $\omega_i(0)$, i=0,...n, in the limit case $\sigma \rightsquigarrow 0$ depend only on its first two derivatives $u'(\mu)$ and $u''(\mu)$. Thus a second order Taylor expansion of u, that is a mean-variance approach, provides a correct characterization of these shares.

2. The first derivatives with respect to σ, $\omega_i'(0)$ i=1,...,n of optimal shares, in the limit case $\sigma \rightsquigarrow 0$, depend on the utility function u only through its first three derivatives $u'(\mu)$, $u''(\mu)$ and $u'''(\mu)$. Thus a third order Taylor expansion of u, that is a mean-variance-skewness approach, does the job.

As far as optimal shares are concerned, theorem 2.2 below confirms that they are conformable to standard mean-variance formulas, that is formulas usually obtained with an assumption of joint normality of returns:

Theorem 2.2 *The vector $\omega(0) = (\omega_i(0))_{1 \leq i \leq n}$ of shares of wealth invested in the limit case $\sigma \rightsquigarrow 0$ is given by:*

$$\omega(0) = \tau \Sigma^{-1} a(0),$$

where $a(0) = (a_i(0))_{1 \leq i \leq n}$ is the vector of risk premia and $\tau = -\frac{u'(\mu)}{u''(\mu)}$ is the risk tolerance coefficient.

To see the equivalence with standard formulas commonly derived under an assumption of joint normality, two remarks are in order:

1. While joint normality with a general utility function would lead to introduce a kind of average risk tolerance coefficient $\left(-Eu'(W)/Eu''(W)\right)$ with $W = \mu + \sum_{i=1}^{n} \omega_i (R_i - \mu)$, this actually coincides with τ in the limit case $\sigma \rightsquigarrow 0$.

2. Joint normality would imply, in equilibrium, constant functions $a_i(\sigma)$ (see theorem 2.4 below). In such a case, the formula of theorem 2.2 can be rewritten:

$$\omega(0) = \tau \left(Var R(\sigma)\right)^{-1} \sigma^2 a,$$

where $\sigma^2 a$ defines the vector of risk premia.

Generally speaking, following theorem 2.2, if we see optimal shares of wealth invested $\omega(\sigma)$ as equivalent to $\tau \Sigma^{-1} a(\sigma)$ in the neighborhood of $\sigma = 0$, we get a Sharpe ratio for optimal portfolios equivalent to:

$$\frac{E\left[\omega^{\top}(\sigma)\left(R(\sigma) - \mu\right)\right]}{\left(Var\left[\omega^{\top}(\sigma) R(\sigma)\right]\right)^{\frac{1}{2}}} = \sigma P(0),$$

where

$$P(0) = \left[a^{\top}(0) \Sigma^{-1} a(0)\right]^{\frac{1}{2}} \tag{2.6}$$

denotes, by unit of scaling risk σ, the potential performance of the set R of returns as in traditional mean variance analysis [see e.g. Jobson and Korkie (1982)]. Of course, the above analysis neglects the variation in equilibrium of the risk premium functions $a(\sigma)$. We are going to see in theorem 2.4 below that these functions will not be constant, even locally in the neighborhood of $\sigma = 0$, as soon as asset return joint probability distribution features some asymmetries.

These asymmetries will actually play a double role in the local behavior of optimal shares of wealth invested. First, preferences for skewness would increase, *ceteris paribus*, asset demands in the direction of positive skewness. Second, market equilibrium induced variations in risk premium which potentially erase this effect. To see this, let us define the co-skewness of asset k in portfolio ω as:

Definition 2.3 *The co-skewness of asset k in portfolio ω is:*

$$c_k(\omega) = \frac{Cov\left(Y_k, (\omega^\top Y)^2\right)}{Var\left[\omega^\top Y\right]} = \frac{\omega^\top \Gamma_k \omega}{\omega^\top \Sigma \omega}, \tag{2.7}$$

where $\Gamma_k = E\left[YY^\top Y_k\right]$ is the matrix of covariances between Y_k and cross product $Y_i Y_j$, i,j=1,...,n.

We will see in section 3 below that this notion of co-skewness is tightly related to a measure put forward by Kraus and Litzenberger (1976) (see also Ingersoll (1987), p 100). For the optimal portfolio $\omega(0)$ characterized in theorem 2.2, we have $c_k(\omega(0)) = c_k$ defined as:

$$c_k = \frac{1}{P^2(0)} a^\top(0) \Sigma^{-1} \Gamma_k \Sigma^{-1} a(0). \tag{2.8}$$

Typically, asymmetry in the joint probability distribution of the vector R of returns means that at least some matrices Γ_k, k=1,...,n are non-zero. We get the following result:

Theorem 2.4 *The slope $\omega'(0)$ of the vector $\omega(0)$ of optimal shares of wealth invested in the neighborhood of $\sigma = 0$ is given by:*

$$\omega'(0) = \tau \Sigma^{-1}\left[a'(0) + \rho P^2(0)\, c\right],$$

where $a'(0) = \left(a_i'(0)\right)_{1 \le i \le n}$ is the vector of marginal risk premia, $c = (c_k)_{1 \le k \le n}$ defined by (2.8) is the vector of co-skewness coefficients and $\rho = \frac{\tau^2}{2}\frac{u'''(\mu)}{u'(\mu)}$ is the skew tolerance coefficient.

In other words, up to variations $a^{'}(0)$ of risk premiums in equilibrium, a positive co-skewness of asset k will have a positive effect on the demand of this asset. This positive effect will be all the more pronounced that the skew tolerance coefficient ρ is large. Of course, this interpretation is based on two implicitly maintained assumptions:

1. The skew tolerance coefficient is nonnegative ($\frac{u^{'''}(\mu)}{u^{'}(\mu)} > 0$). This assumption conforms to both the literature on prudence [Kimball (1990)] and the literature on preferences for high order moments [Dittmar (2002), Harvey and Siddique (2000), Ingersoll (1987)]

2. The vector $c(\omega) = (c_k(\omega))_{1<k<n}$ represents a multivariate notion of skewness that investors do like to get positive, componentwise. This assertion is justified by the fact that on average:
$$\sum_{k=1}^{n} \omega_k c_k(\omega) = \frac{E\left[(\omega^{\top} Y)^3\right]}{Var\left[\omega^{\top} Y\right]}$$
is positive if and only if the portfolio return is positively skewed. Of course, individual preferences for positive skewness will increase, ceteris paribus, the equilibrium price of assets with positively skewed returns. This will actually appear in the equilibrium value $a^{'}(0)$ of risk premium slopes in the neighborhood of $\sigma = 0$.

2.2 Equilibrium allocations and prices

Let us consider consider asset markets for risky assets i=1,2,...,n with agents s=1,...,S. Each agent is characterized by a Von Neumann-Morgenstern utility functions u_s and associated preference coefficients:
$$\tau_s = -\frac{u_s^{'}(\mu)}{u_s^{''}(\mu)} \text{ and } \rho_s = \frac{\tau_s^2}{2}\frac{u_s^{'''}(\mu)}{u_s^{'}(\mu)}.$$
Note that for the sake of notational simplicity, we assume that the net supply of each risky asset i=1,...,n is exogeneous and fixed to unity as a normalization. Then, in the limit case $\sigma \rightsquigarrow 0$, the market clearing conditions can be written:
$$\begin{aligned} \sum_{s=1}^{S} \omega^{(s)}(0) &= e_n, \qquad (2.9) \\ \sum_{s=1}^{S} \omega^{(s)'}(0) &= 0, \end{aligned}$$

where $\omega^{(s)}(0) = (\omega_{si}(0))_{1<i<n}$ and e_n denotes the n-dimensional column vector, the components of which are all equal to 1. Below, it will be convenient to consider an average investor characterized by average holdings $\overline{\omega}$, an average risk tolerance $\overline{\tau}$ and average skew tolerance $\overline{\rho}$, where

$$\overline{\omega} = \frac{1}{S}e_n, \quad \overline{\tau} = \frac{1}{S}\sum_{s=1}^{S}\tau_s, \quad \overline{\rho} = \frac{\sum_{s=1}^{S}\rho_s\tau_s}{\sum_{s=1}^{S}\tau_s}. \tag{2.10}$$

If all individual were identical, each one would buy the average portfolio $\overline{\omega}$. The link between the two average preference coefficients $\overline{\tau}$ and $\overline{c}$ and individual portfolios demands is characterized by theorem 2.5.

Theorem 2.5 *In equilibrium, in the limit case $\sigma \rightsquigarrow 0$, the optimal shares of wealth invested $\omega_s(\sigma)$ by agents $s = 1, ..., S$ is characterized by:*

$$\begin{aligned}
\omega^{(s)}(0) &= \frac{\tau_s}{\overline{\tau}}\overline{\omega}, \\
\omega^{(s)\prime}(0) &= \tau_s\left[\rho_s - \overline{\rho}\right]P^2(0)\,\Sigma^{-1}c(\overline{\omega}) \quad \textit{for } s=1,\ldots,S,
\end{aligned}$$

where $P^2(0) = \frac{1}{\overline{\tau}^2}\overline{\omega}^\top\Sigma\overline{\omega}$ is the (squared) market Sharpe ratio and $c(\overline{\omega})$ is the vector of the market co-skewness coefficients.

In other words, in the limit case $\sigma \rightsquigarrow 0$, the vector $\omega^{(s)}(\sigma)$ of optimal shares of wealth invested is as in a standard mean-variance separation theorem. All individuals buy a share of the market portfolio e_n, the size of this share being determined by the comparison of individual risk tolerance τ_s with respect to average one. Preferences for skewness only play a role for the slopes $\omega^{(s)\prime}(0)$ of the shares of wealth invested in the neighborhood of zero. A positive market co-skewness $c_k(\overline{\omega})$ will have a positive effect on the individual s demand of asset k if and only if his skew tolerance coefficient is more than the average one $\overline{\rho}$. On the contrary, if $\rho_s < \overline{\rho}$, the positive effect of asset k co-skewness on its market price is higher than required to compensate the investor's preference for skewness.

In order to characterize the asset pricing implications of risk tolerance and preference for skewness, we deduce the local behavior of the risk premium in equilibrium:

Theorem 2.6 *In the limit case $\sigma \rightsquigarrow 0$, the equilibrium risk premium vector $a(\sigma)$ is such that the average portfolio $\overline{\omega}$ is optimal for the average investor:*

$$\overline{\omega} = \overline{\tau}\Sigma^{-1} a(0)$$

and its slope in the neighborhood of zero is given by:

$$a^{'}(0) = -\overline{\rho} P^2(0)\, c(\overline{\omega}),$$

where $P^2(0) = \frac{1}{\overline{\tau}^2}\overline{\omega}^{\intercal}\Sigma\overline{\omega}$ is the (squared) market Sharpe ratio and $c(\overline{\omega})$ is the vector of the market co-skewness coefficients.

Note that, by comparison of theorems 2.4 and 2.6, the equilibrium slopes are precisely such that the average agent would have no motive to deviate from the market portfolio $\overline{\omega}$ ($\omega^{'}(0) = 0$ for the average investor).

Theorem 2.6 gives as a new asset pricing model. While approximating risk premia by their limit values $a_i(0)$ would clearly give the Sharpe-Lintner CAPM, approximating them by higher order expansions $a_i(0) + \sigma a_i^{'}(0)$ gives a new mean-variance-skewness asset pricing model. A convenient way to describe the implications of an asset pricing model is to characterize it through a Stochastic Discount Factor (henceforth SDF), see e.g Cochrane (2001). By definition, a SDF $m(\sigma)$ must be able to price correctly all available securities; here we therefore need: $Em(\sigma) = \frac{1}{\mu}$ and $E\left[m(\sigma)\left(\mu + \sigma^2 a_i(\sigma) + \sigma Y_i\right)\right] = 1$ for i=1,...,n. We denote $R_M(\sigma) = \sum_{i=1}^{n} R_i(\sigma)$ the market return. We are then able to translate theorem 2.6 in terms of a SDF:

Theorem 2.7 *The random variable:*

$$m(\sigma) = \frac{1}{\mu} - \frac{1}{\mu S \overline{\tau}}\left(R_M(\sigma) - ER_M(\sigma)\right) + \frac{\overline{\rho}}{\mu S^2 \overline{\tau}^2}\left[\left(R_M(\sigma) - ER_M(\sigma)\right)^2 - E\left(R_M(\sigma) - ER_M(\sigma)\right)^2\right]$$

is a SDF consistent with variance-skewness risk premium defined by $a(\sigma) = a(0) + \sigma a^{'}(0)$ where $a(0)$ and $a^{'}(0)$ are given by theorem 2.6.

The conjunction of theorems 2.6 and 2.7 summarizes what we have learnt so far about portfolio choice and asset pricing in the context of mean-variance-skewness preferences:

1. Due to heterogeneity in preferences for skewness, the common two-fund CAPM separation theorem is violated: different individuals may hold in equilibrium different risky portfolios.

2. However, the pricing implications of a common separation theorem remain true in some respect. Somewhat unexpectedly, the market return alone is still able to summarize the pricing of risk. Of course, since not only market betas but also market co-skewness must be taken into account, both the actual market return and its squared value enter linearly in the pricing kernel.

Following the seminal paper by Kraus and Litzenberger (1976), Harvey and Siddique (2000) and Dittmar (2002) among others have recently studied the empirical implications of a SDF which involves a quadratic function of market return. Theorem 2.7 above provides a theoretical basis for doing so. Section 3 will elaborate more on the pricing implications of such a SDF.

3. Nonlinear Pricing Kernels

The pricing implications of a SDF formula that is quadratic with respect to the market return are studied in this section, first with a linear beta pricing point of view and second in terms of derivative pricing.

3.1 Beta pricing

In their paper about conditional skewness in asset pricing tests, Harvey and Siddique (2000) start with the maintained assumption that the SDF is quadratic in the market return:

$$m_{t+1} = \nu_{0t} + \nu_{1t} R_{Mt+1} + \nu_{2t} R_{Mt+1}^2. \tag{3.11}$$

It actually suffices to revisit our section 2 above with a conditional viewpoint to see theorem 2.7 as a theoretical justification of (3.11). Then, the coefficients ν_{0t}, ν_{1t} and ν_{2t} are functions of the conditioning information I_t at time t.

From theorem 2.7, we interpret the factors coefficients as:

$$\nu_{2t} = \frac{\overline{\rho}}{\mu_t S^2 \overline{\tau}^2} > 0, \tag{3.12}$$

and

$$\nu_{1t} = -\frac{1}{\mu_t S \overline{\tau}} - \frac{2\overline{\rho}}{\mu_t S^2 \overline{\tau}^2} E_t \left(R_{Mt+1} \right) < 0. \tag{3.13}$$

It is worth characterizing the role of the two factors R_{Mt+1} and R^2_{Mt+1} in the SDF (3.11) in terms of beta pricing relationships. Assuming the existence of a conditionally risk-free asset (with return μ_t), we can write for the net excess return $r_{it+1} = R_{it+1} - \mu_t$ of any asset i:

$$E_t \left[r_{it+1} m_{t+1} \right] = 0,$$

that is:

$$\frac{1}{\mu_t} E_t \left[r_{it+1} \right] + \nu_{1t} Cov_t \left[r_{it+1}, R_{Mt+1} \right] + \nu_{2t} Cov_t \left[r_{it+1}, R^2_{Mt+1} \right] = 0$$

or, using the market net excess return, we get:

$$\frac{1}{\mu_t} E_t \left[r_{it+1} \right] + \left(\nu_{1t} + 2\mu_t \nu_{2t} \right) Cov_t \left[r_{it+1}, r_{Mt+1} \right] + \nu_{2t} Cov_t \left[r_{it+1}, r^2_{Mt+1} \right] = 0,$$

that is:

$$E_t \left[r_{it+1} \right] = \lambda_{1t} Cov_t \left[r_{it+1}, r_{Mt+1} \right] - \lambda_{2t} Cov_t \left[r_{it+1}, r^2_{Mt+1} \right],$$

with:

$$\lambda_{1t} = -\mu_t \left(\nu_{1t} + 2\mu_t \nu_{2t} \right) \text{ and } \lambda_{2t} = \mu_t \nu_{2t}.$$

If ν_{1t} and ν_{2t} are interpreted in terms of preferences of an average investor as in (3.12) and (3.13), we deduce:

$$\lambda_{1t} = \frac{1}{S \overline{\tau}} + \frac{2\overline{\rho}}{S^2 \overline{\tau}^2} \left(E_t R_{Mt+1} - \mu_t \right) \text{ and } \lambda_{2t} = \frac{\overline{\rho}}{S^2 \overline{\tau}^2}.$$

Note that λ_{2t} is something like a structural invariant, only time varying through the value of preference parameters computed from the derivatives of the utility function at μ_t. λ_{2t} should be non-negative and all the more positive that preference for skewness is high. Similarly, λ_{1t} is expected to be positive and time varying insofar as the market risk premium $(E_t R_{Mt+1} - \mu_t)$ is.

To summarize:

Theorem 3.1 *Under the maintained assumption (3.11) of a quadratic SDF, net expected returns are given by:*

$$E_t r_{it+1} = \lambda_{1t} Cov_t \left[r_{it+1}, r_{Mt+1}\right] - \lambda_{2t} Cov_t \left[r_{it+1}, r_{Mt+1}^2\right].$$

If in addition, theorem 2.7 applies, λ_{1t} and λ_{2t} are non negative. $\lambda_{2t} = \frac{\overline{\rho}}{S^2\overline{\tau}^2}$ is determined by average preferences for skewness while:

$$\lambda_{1t} = \frac{1}{S\overline{\tau}} + 2\lambda_{2t} E_t r_{Mt+1}.$$

Note that λ_{1t} has two components which are both increasing with the average risk aversion, first as $1/\overline{\tau}$ and second as the market risk premium $E_t r_{Mt+1}$. When applying theorem 3.1 to the market return itself $(r_{it+1} = r_{Mt+1})$, we get even more insight on what makes λ_{1t} large:

Corollary 3.2 *Under the assumptions of theorem 3.1*

$$\lambda_{1t} = \frac{E_t r_{Mt+1}}{Var_t r_{Mt+1}} + \lambda_{2t} \frac{Skew_t\left(r_{Mt+1}\right)}{Var_t r_{Mt+1}},$$

where $Skew_t\left(r_{Mt+1}\right) = Cov_t\left(r_{Mt+1}, r_{Mt+1}^2\right)$.

In particular, we can see that theorem 3.1 coincides with the standard Sharpe-Lintner CAPM formula when $\lambda_{2t} = 0$, that is the average preference for skewness is zero. By contrast, λ_{1t} is augmented in the general case by an additive term which is proportional to both λ_{2t} and

$$Skew_t\left(r_{Mt+1}\right) = Cov_t\left(r_{Mt+1}, r_{Mt+1}^2\right) = E_t r_{Mt+1}^3 - \left(E_t r_{Mt+1}\right)\left(E_t r_{Mt+1}^2\right).$$

This notion of market co-skewness has already been put forward by Harvey and Siddique (2000) and theorem 3.1 and corollary 3.2 correspond to their formulas (7). It is also worth rewriting the pricing relationship of theorem 3.1 and corollary 3.2 in term of betas:

$$E_t r_{it+1} = \left(\lambda_{1t} Var_t r_{Mt+1}\right)\beta_{imt} - \left(\lambda_{2t} Var_t r_{Mt+1}^2\right)\gamma_{imt}, \tag{3.14}$$

and

$$E_t r_{it+1} = (E_t r_{Mt+1})\,\beta_{imt} - \lambda_{2t} Var_t r^2_{Mt+1}\left(\gamma_{imt} - \gamma_{mmt}\beta_{imt}\right), \tag{3.15}$$

where $\beta_{imt} = \frac{Cov_t[r_{it+1}, r_{Mt+1}]}{Var_t r_{Mt+1}}$ is the standard market beta while the beta coefficient with respect to the squared market return: $\gamma_{imt} = \frac{Cov_t\left[r_{it+1}, r^2_{Mt+1}\right]}{Var_t r^2_{Mt+1}}$ is tightly related to the measure of co-skewness already introduced in section 2. More precisely, it is straightforward to see that the return decomposition of section 2 gives: $\gamma_{imt} = \frac{1}{\sigma^2} c_{imt}(\sigma)$ with $c_{imt}(0)$ as introduced in definition 2.3. While we had already seen in theorem 2.6 that risk premiums in equilibrium where influenced by skewness preferences in proportion of the vector $c(\overline{\omega})$ of market co-skewness coefficients, the same vector shows up in the beta pricing relationship (3.14) with $\gamma_{imt} = \frac{1}{\sigma^2} c_{imt}(\sigma)$. Note that what Harvey and Siddique (2000) call "market co-skewness" is actually $Skew_t(r_{Mt+1}) = \gamma_{mmt}\left(Var_t r^2_{Mt+1}\right)$.

The beta pricing model (3.14) with a second beta coefficient interpreted in terms of co-skewness with the market is observationally equivalent to a conditional version of the three-moments CAPM first proposed by Kraus and Litzenberger (1976) (see also Ingersoll (1987), p100). While they put forward a measure of co-skewness defined as:

$$\delta_{imt} = \frac{Cov_t\left(R_{it+1}, (R_{it+1} - ER_{it+1})^2\right)}{Cov_t\left(R_{Mt+1}, (R_{Mt+1} - ER_{Mt+1})^2\right)},$$

we have preferred to remain true to a genuine notion of beta coefficient as γ_{imt}. However, the difference between the two is just a matter of normalization and is immaterial in terms of asset pricing. In particular (3.15) enhances as formula (64) in Ingersoll (1987) that the beta pricing relationship differs from Sharpe-Lintner CAPM by a factor proportional to the difference between the two betas. It is however worth noticing that these authors derive this pricing relationship by using a utility function directly defined over mean, standard deviation and skewness. The small noise expansion approach of section 2 affords more theoretical underpinnings for doing so.

Normalization in terms of beta coefficient is usually convenient since it allows a direct interpretation of beta loadings in terms of factor risk premium. For instance, when $\lambda_{2t} = 0$, (3.14) applied to the market gives the usual formula: $\lambda_{1t} = P_{Mt}^{(1)}$ with

$$P_{Mt}^{(1)} = \frac{E_t r_{Mt+1}}{Var_t r_{Mt+1}}.$$

It is, however, worth noticing that in general case, λ_{1t} and λ_{2t} can not be read as simple risk premium associated respectively to the two payoffs r_{Mt+1} and r_{Mt+1}^2. Even if we assume that r_{Mt+1}^2 does correspond to a payoff of a portfolio available in the market with price η_t, the risk premium on such a payoff:

$$P_{Mt}^{(2)}(\eta_t) = \frac{E_t \frac{r_{Mt+1}^2}{\eta_t} - \mu_t}{Var_t \left(\frac{r_{Mt+1}^2}{\eta_t}\right)} \tag{3.16}$$

will not coincide with $(-\lambda_{2t}\eta_t)$. The difference comes from the fact that the two factors are not orthogonal. λ_{1t} does depend on λ_{2t} (see corollary 3.2) and the expression of λ_{2t} in function of the equilibrium prices is more involved:

Theorem 3.3 *If $\eta_t = E_t\left[m_{t+1} r_{Mt+1}^2\right]$ denotes the equilibrium price of a payoff r_{Mt+1}^2, we have:*

$$\lambda_{2t} = \frac{\gamma_{mmt} P_{Mt}^{(1)} - \frac{1}{\eta_t} P_{Mt}^{(2)}(\eta_t)}{1 - \rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)},$$

where according to (3.16), $P_{Mt}^{(2)}(\eta_t)$ is the risk premium on the asset with payoff r_{Mt+1}^2 and $\rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)$ denotes the square (conditional) linear correlation coefficient between r_{Mt+1} and r_{Mt+1}^2.

It is worth considering the limit case when r_{Mt+1}^2 is almost worthless. From (3.16):

$$\lim_{\eta_t \rightsquigarrow 0} \frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t r_{Mt+1}^2}{Var_t\left(r_{Mt+1}^2\right)}. \tag{3.17}$$

In this limit case, one gets:

$$\lambda_{2t} = \frac{\gamma_{mmt} P_{Mt}^{(1)} - \frac{E_t r_{Mt+1}^2}{Var_t\left(r_{Mt+1}^2\right)}}{1 - \rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)} \tag{3.18}$$

which actually coincides with the formula put forward by Harvey and Siddique (2000). However, this limit case appears to be at odds with a no-arbitrage condition since $\eta_t = E_t\left[m_{t+1} r_{mt+1}^2\right]$ should be positive. Indeed, as shown in subsection 3.2 below, $\sigma_{mt}^{*2} = \eta_t \mu_t$ may be interpreted as the risk neutral variance of the market return.

Besides the no-arbitrage condition, the fact that risk neutral variance is significantly positive is of course an empirical question. Since, from (3.16):

$$\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t r_{Mt+1}^2 - \sigma_{mt}^{*2}}{Var_t\left(r_{Mt+1}^2\right)}, \tag{3.19}$$

one may expect that considering the limit case (3.17), that is $\sigma_{mt}^{*2} = 0$ leads to overestimate $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t}$ and then to underestimate λ_{2t}. The relevant empirical issue (see section 4) is then to decide if considering only the limit case (3.18) leads to an economically significant underestimation of the weight λ_{2t} of coskewness in the two factors pricing relationship (3.15). If it is the case, we must realize that λ_{2t} actually depends on investors preferences for skewness as they show up either in the (market) price of squared market return or, equivalently, in the risk neutral variance of the market return.

3.2 Derivative pricing

The huge expansion of derivative asset markets, introducing asset payoffs which are nonlinear and often skewed functions of underlying primitive asset returns, has motivated the renewal of interest in asset payoffs skewness. For sake of notational simplicity, we consider in this subsection only options written on the market return. However, most of the results could be extended to other primitive assets. Let us then consider the pricing issue for a payoff $h_t(.)$ of the market return, the definition of which may depend on conditioning information. Maintaining, as we do in this section, the assumption that a valid SDF is quadratic means that the price of the payoff $h_t(R_{Mt+1})$ coincides with the price of its (conditional) affine regression $h_t^{(L2)}(R_{Mt+1}) = EL_t\left[h_t(R_{Mt+1})\,|R_{Mt+1}, R_{Mt+1}^2\right]$ on R_{Mt+1} and R_{Mt+1}^2 :

$$E_t\left[m_{t+1}h_t(R_{Mt+1})\right] = E_t\left[m_{t+1}h_t^{(L2)}(R_{Mt+1})\right].$$

Besides this, understanding why a CAPM Sharpe-Lintner pricing does not accommodate well the pricing of derivatives is akin to showing why the (conditional) affine regression

$$h_t^{(L1)}(R_{Mt+1}) = EL_t\left[h_t(R_{Mt+1})\,|R_{Mt+1}\right]$$

of $h_t(R_{Mt+1})$ on R_{Mt+1} does not summarize the risk which is compensated in equilibrium:

$$E_t[m_{t+1}h_t(R_{Mt+1})] \neq E_t\left[m_{t+1}h_t^{(L1)}(R_{Mt+1})\right].$$

Starting with the simplest nonlinear payoff $h_t(R_{Mt+1}) = R_{Mt+1}^2$, we are then led to study the difference between the price

$$\pi_t = E_t\left[m_{t+1}R_{Mt+1}^2\right]$$

of the so-called "volatility contract" (see Bakshi, Kapadia and Madan (2003)) and the price of its linear approximation:

$$E_t\left[m_{t+1}EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right]\right].$$

To enhance the role of skewness it is first worth noticing that:

Lemma 3.4 *The conditional linear regression of R_{Mt+1}^2 on R_{Mt+1} is:*

$$\begin{aligned} EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right] &= E_tR_{Mt+1}^2 + 2(E_tR_{Mt+1})(R_{Mt+1} - E_tR_{Mt+1}) + \\ &\quad \frac{E(R_{Mt+1} - E_tR_{Mt+1})^3}{VarR_{Mt+1}}(R_{Mt+1} - E_tR_{Mt+1}). \end{aligned}$$

Note that by contrast, the Taylor expansion of R_{Mt+1}^2 arround E_tR_{Mt+1}

$$R_{Mt+1}^2 \approx (E_tR_{Mt+1})^2 + 2(E_tR_{Mt+1})(R_{Mt+1} - E_tR_{Mt+1})$$

does not take into account the crucial role of the skewness term. This remark casts some doubts on theories of higher moments pricing which are based on Taylor expansions. In that respect, the small noise expansion appears to be more reliable.

In any case, it is worth relating the price of the volatility contract with risk neutral pricing popular for derivative pricing. We have:

$$\pi_t = E_t\left[m_{t+1}R_{Mt+1}^2\right] = \frac{1}{\mu_t}E_t^*\left[R_{Mt+1}^2\right],$$

where E_t^* denotes the conditional expectation with respect to a risk neutral probability measure. By definition, $E_t^*[R_{Mt+1}] = \mu_t$, so that

$$\sigma_{mt}^{*2} = E_t^*(R_{Mt+1} - \mu_t)^2 = \mu_t(\pi_t - \mu_t) = \mu_t\eta_t$$

can be interpreted as a risk neutral variance (see Rosenberg (2000)).

As already noticed from theorem 3.3, we expect that higher is the price η_t of r^2_{Mt+1}, the smaller will be the premium $P^{(2)}_{Mt}$ and the larger will be price λ_{2t} of co-skewness. Therefore, one way to assess the strength of preference for skewness is to describe the factors which tend to increase ceteris paribus the risk neutral variance $\sigma^{*2}_{mt} = \mu_t \eta_t$. For doing so, we first state a useful relation between risk neutral variance σ^{*2}_{mt} and historical one $\sigma^2_{mt} = Var_t\left(R_{Mt+1}\right)$:

Theorem 3.5

$$\sigma^{*2}_{mt} = \sigma^2_{mt}\left(1 - \sigma^2_{mt}\left(P^{(1)}_{Mt}\right)^2\right) - P^{(1)}_{Mt} E_t\left(R_{Mt+1} - E_t R_{Mt+1}\right)^3 + \mu_t Cov_t\left(m_{t+1}, \varepsilon_{t+1}\right)$$

where $\varepsilon_{t+1} = R^2_{Mt+1} - EL_t\left[R^2_{Mt+1} | R_{Mt+1}\right]$ *denotes the residual of the (conditional) affine regression of* R^2_{Mt+1} *on* R_{Mt+1}.

Note that theorem 3.5 is valid under the very general assumption that a positive SDF m_{t+1} is able to price the asset of interest and in particular to define risk neutral conditional expectations as $\frac{1}{\mu_t} E^*_t h_t\left(R_{Mt+1}\right) = E_t\left[m_{t+1} h_t\left(R_{Mt+1}\right)\right]$. It is then worth revisiting skewness pricing by studying the factors which may potentially increase the risk neutral variance. Theorem 3.5 basically puts forward two factors. One factor is model dependent, through $Cov_t\left(m_{t+1}, \varepsilon_{t+1}\right)$ while the other terms can be directly observed from the market return. Typically, in the case of a positive return skewness $\left(E_t\left(R_{Mt+1} - E_t R_{Mt+1}\right)^3 > 0\right)$, the risk neutral variance is inversely related to the risk premium $P^{(1)}_{Mt}$. Intuitively, high risk neutral variance, that is high compensation for skewness, may compensate a low risk premium $P^{(1)}_{Mt}$. By contrast, the effect encapsulated in $Cov_t\left(m_{t+1}, \varepsilon_{t+1}\right)$ depends in general explicitly on the SDF specification, that is on the investor preferences. There is however a case where the risk neutral variance is preference free, in the sense that it is completely determined by the observation of the risk-free interest rate and the market risk premium. This is the case of joint log normality which is an extension (see Garcia, Ghysels and Renault (2003)) of the risk neutral valuation relationships first introduced by Brennan (1979):

Theorem 3.6 *If* $(\log m_{t+1}, \log R_{Mt+1})$ *is jointly normal given the conditioning information,*

$$\sigma_{mt}^{*2} = \sigma_{mt}^2 \cdot \left[\frac{\mu_t}{E_t R_{Mt+1}}\right]^2 < \sigma_{mt}^2.$$

Theorem 3.6 confirms in a particular case the above discussion of the difference between risk neutral variance and historical one. While we expect the former to be smaller than the latter in case of positive skewness, the difference between the two is inversely related to the market risk premium.

In the general case, the role of investor preference for skewness in increasing the risk neutral variance can be characterized from the following result:

Theorem 3.7 *With a quadratic SDF,*

$$m_{t+1} = \nu_{0t} + \nu_{1t} R_{Mt+1} + \nu_{2t} R_{Mt+1}^2,$$

the term $Cov_t(m_{t+1}, \varepsilon_{t+1})$ *is given by:*

$$Cov_t(m_{t+1}, \varepsilon_{t+1}) = \nu_{2t}\left(Var_t R_{Mt+1}^2\right)\left(1 - \rho_t^2\left(R_{Mt+1}, R_{Mt+1}^2\right)\right).$$

Therefore, we do expect that this term increases the risk neutral variance, all the more that R_{Mt+1} and R_{Mt+1}^2 are weakly correlated and the average skewness tolerance $\overline{\rho} = \nu_{2t}\mu_t S^2 \overline{\tau}^2$ is large. The main message of this subsection is that empirical assessments of risk neutral variance as recently proposed by Rosenberg (2000) from derivative asset prices may also be seen as a way to characterize preferences for skewness.

4. Empirical Illustration

4.1 The general issue

The empirical relevance issue of the asset pricing model with coskewness as developed in previous sections is encapsulated in the asset pricing equation (3.15):

$$E_t(r_{it+1}) = (E_t r_{Mt+1})\,\beta_{imt} - \lambda_{2t} Var_t r_{Mt+1}^2 \left(\gamma_{imt} - \gamma_{mmt}\beta_{imt}\right). \quad (4.20)$$

The question is: does this asset pricing equation significantly deviate from standard CAPM?, that is should we maintain a significantly positive skewness premium λ_{2t}?

It turn out that the statistical identification of this hypothesis is difficult since, as well noticed by Barone-Adesi, Gagliardini and Urga (2004), covariance and coskewness with market tend to be almost collinear across common portfolios, leading to hardly significant coskewness factors $(\gamma_{imt} - \gamma_{mmt}\beta_{imt})$. To shed more light on this identification issue, let us consider the (conditional) affine regression of net return of asset i on market return:

$$r_{it+1} = \alpha_{it} + \beta_{imt} r_{Mt+1} + u_{it+1}. \tag{4.21}$$

It is clear that asset i coskewness can be interpreted as the covariance between the residual of this regression with squared market return:

$$\left(Var_t\left(r^2_{Mt+1}\right)\right)\left(\gamma_{imt} - \gamma_{mmt}\beta_{imt}\right) = Cov_t\left(u_{it+1}, r^2_{Mt+1}\right) = Cov_t\left(u_{it+1}, R^2_{Mt+1}\right). \tag{4.22}$$

Therefore, a positive sign for λ_{2t} can be identified only insofar as one can observe some asset returns r_{it+1} with positive (negative) coskewness $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ and check that they command a lower (higher) expected return than explained by standard CAPM. The problem is that $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ will be more often than not close to zero since u_{it+1} is by definition (conditionally) uncorrelated with r_{Mt+1}. Of course non correlation does not imply independence (except in linear market models like the Gaussian one) and one may hope that some portfolios i exhibit a significantly positive (or negative) covariance $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$. However, as long as a linear approximation is valid, $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ is almost zero leading to:

$$Cov_t\left(r_{it+1}, r^2_{Mt+1}\right) \approx \beta_{imt} Cov_t\left(r_{Mt+1}, r^2_{Mt+1}\right)$$

almost collinear with β_{imt} across portfolios.

To avoid such a perverse linearity effect, Barone-Adesi, Gagliardini and Urga (2004) focus on a quadratic market model first introduced by Barone-Adesi (1985). Thanks to this specification, they estimate a slightly significantly positive coefficient λ_{2t}, at least when the risk free rate is a free parameter, not assumed to be observed by the econometrician. However, their approach is unconditional and this may explain the difficulty to identify the sign of λ_{2t}, in particular with respect to the risk free rate issue. To remedy that, we propose here to consider instead an asymmetric GARCH in mean model recently estimated by Bekaert and Liu (2004). Since this model exhibits interesting time-varying non-linearities in the consumption process, it may allow an accurate identification of time varying conditional coskewness and in turn consumption-based preference for coskewness. The superior identification power of such a conditional approach will actually be confirmed below through a series of Monte Carlo simulations based on Bekaert and Liu (2004) parameters estimates.

4.2 The simulation set up:

Bekaert and Liu (2004) estimate a GARCH factor model with in mean effects for the trivariate process of logarithm of consumption growth X_{t+1}, logarithm of stock return $Log\left(R_{Mt+1}\right)$, and logarithm of bond return $Log\left(\mu_{t+1}\right)$:

$$Y_{t+1} = \left[Y_{1t+1}, Y_{2t+1}, Y_{3t+1}\right]' = \left[X_{t+1}, Log\left(R_{Mt+1}\right), Log\left(\mu_{t+1}\right)\right]',$$

that is a model of the form:

$$Y_{t+1} = c_t + AY_t + \Omega e_{t+1} \tag{4.23}$$

where the coefficient c_{it} of c_t, $i = 1, 2, 3$, is an affine function of $Var_t\left[Y_{it+1}\right]$ and all the time variation in volatility is driven by time varying uncertainty in consumption growth: the conditional probability distribution of e_{t+1} given information I_t is normal with zero mean and a diagonal covariance matrix , the coefficients of which are constant except the first one which follows an asymmetric GARCH(1,1):

$$Var_t\left[e_{1t+1}\right] = \delta_1 + \alpha\left(e_{1t}\right)^2 + \beta Var_{t-1}\left[e_{1t}\right] + \xi\left(Max\left[0, -e_{1t}\right]\right)^2. \tag{4.24}$$

To further limit parameter proliferation, they assume that all the off diagonal coefficients of the matrix Ω are zero except in the first column; in other words the consumption shock is the only factor. For sake of normalization, the diagonal coefficients of Ω are fixed to the value 1. Table 1.1 gives the parameters estimates obtained by Bekaert and Liu (2004) on monthly US data. These estimates will be considered below as true population values for simulating a sample path and we don't care about estimation errors.

A convenient feature of the above model for our purpose is that, since it maintains a conditional joint normality assumption for log-consumption and log-market return, it allows us to apply theorem 3.6 to assess the risk neutral variance without need of a preference specification. More precisely, insofar as the log-pricing kernel is, given I_t, a linear combination of the first two components of Y_{t+1}, as it is not only in the Lucas (1978) consumption based CAPM with isoelastic preferences but also more generally in the Epstein and Zin (1991) recursive utility model, we are sure that theorem 3.6 applies. Then, our simulation set up is as follows:

For a given simulated path of the process (Y_{t+1}), specifications (4.23) and (4.24) allow us to compute corresponding paths of:

1 Compute $\sigma_{mt}^{*2} = \sigma_{mt}^2 \left[\mu_t / E_t\left(R_{Mt+1}\right)\right]^2$,

2 Compute $\eta_t = \sigma_{mt}^{*2}/\mu_t$,

3 Compute $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t r_{Mt+1}^2 - \sigma_{mt}^{*2}}{Var_t\left(r_{Mt+1}^2\right)}$,

4 Compute $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t}\eta_t$,

5 Compute $\lambda_{2t} = \frac{\gamma_{mmt} P_{Mt}^{(1)} - \frac{1}{\eta_t} P_{Mt}^{(2)}(\eta_t)}{1 - \rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)}$.

By contrast, the limit case put forward by Harvey and Siddique (2000) corresponds to the alternative formula:

$$\lambda_{2t}^* = \frac{\gamma_{mmt} P_{Mt}^{(1)} - \frac{E_t r_{Mt+1}^2}{Var_t\left(r_{Mt+1}^2\right)}}{1 - \rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)},$$

the path of which is also easy to build from above simulations.

Of course, by introducing only one risky asset, this setting does not allow us to compare coskewness across portfolios. However, the focus of our interest here is to get time series of λ_{2t} and λ_{2t}^* in order to assess their sign and their differences both date by date and in average. Note moreover that return skewness in this market is not as trivial as log-normality may lead to think. Over two periods, temporally aggregated asset returns will feature some sophisticated skewness, first due to the asymmetric effect in the variance dynamics and second due to time varying risk premium. A detailed characterization of induced dynamic skewness pricing is beyond the scope of this paper.

4.3 Monte Carlo results

All the simulated paths considered correspond to 500 months. The main message conveyed by these simulated series is well summarized by figure 1.1 where we plot on the same graph both the path of λ_{2t} corresponding to our formula for the price of coskewness and of λ_{2t}^* corresponding to Harvey and Siddique (2000) limit case. The conclusions draw from this graph are twofold:

First, while the series of λ_{2t} does show a positive price for coskewness as expected (4.25 in average), the series λ_{2t}^* of displays some implausible huge negative price of coskewness (-67.82 in average). This tends to prove that neglecting the price η_t of squared net returns (or equivalently the risk neutral variance) leads to a severe understimation of coskewness price, so severe that it may reverse the direction of the effect of coskewness in asset prices. The time series of η_t (figure 1.2) and risk premium $P_{Mt}^{(2)}(\eta_t)$ (figure 1.3) as well confirm that they are positive. Note also that while λ_{2t} and λ_{2t}^* are stationary processes-in particular first order differences $(\lambda_{2t} - \lambda_{2t-1})$ and $(\lambda_{2t}^* - \lambda_{2t-1}^*)$ have a zero time average - the former is more stable than the latter: the standard error of the series $(\lambda_{2t} - \lambda_{2t-1})$ is only 4.93 while it is 8.75 for $(\lambda_{2t}^* - \lambda_{2t-1}^*)$. This give some support to our interpretation of λ_{2t} as a kind of preference-based structural invariant, which is time varying only through the value of utility derivatives at point μ_t.

Second, our simulations confirm that the positive sign of the price for coskewness should be hardly identifiable in an unconditional setting. While the series λ_{2t} does show a positive average price of 4.25 for coskewness, it comes with a standard error of 4.06. This may explain why Barone-Adesi, Gagliardini and Urga (2004) found it difficult to identify a positive price in an unconditional setting. They actually get a t-statistic of 1.01, which has the same order of magnitude as our informal assessment. Of course, a rigorous unconditional study should not be simply based on time averages. By contrast, figure 1.1 shows that spot values of the process series λ_{2t} may cover the full interval between 0 and 20, making them likely significant for a number of dates. This enhances the important contribution of Harvey and Siddique (2000) who stress that coskewness must be addressed in a conditional setting. However, even an unconditional approach would not make the simplified price series λ_{2t}^* meaningful since their standard error is only 7.45, which does not compensate their negative average of (-67.82).

Overall, we conclude that there should be a positive price for coskewness, but not so high and hardly identifiable in an unconditional setting. One way to interpret the limited level of this price is to realize that buying the squared net market return commands a positive risk premium (see figure 1.3) which, by theorem 3.3 leads to lower the price λ_{2t}. This does not mean that skewness is worthless but only that, by lemma 3.4, a part of its value is already captured by the linear pricing of squared return. In order word, a positive skewness implies a positive correlation between market return and squared market return, so that the two components of asset prices cannot be interpreted separately.

Finally, one ought to realize that quadratic pricing kernels cannot be more than local approximations of a true pricing kernel, for instance in the neighborhood of small risk as in section 2. In particular, while a representative agent with a convex utility function would imply that the pricing kernel is decreasing with respect to the market return, this cannot be the case on the full range of returns with a quadratic function. More precisely, a quadratic pricing kernel as characterized by (3.11), (3.12), and (3.13) with a positive coskewness price λ_{2t} will become increasing when the market returns exceeds its conditional expectation by more than $(S_T/2\rho)$. This kind of paradoxical increasing shape of pricing kernels for large levels of market return already showed up in the empirical evidence documented by Dittmar (2002). Of course, a negative λ_{2t} as in the case of the zero-price η_t approximation would produce an even weirder behavior with increasing pricing kernel for any value of the market return below its expectation.

As far as Dittmar's paradox is concerned, it does not mean that one should give up nonlinear polynomial pricing kernels because their decreasing shape cannot be enforced on the whole range of possible market returns. One must only remember that polynomial approximations are local and ought to be used cautiously. For instance, it is clear that market information about risk neutral variance or equivalently about the price η_t of squared net market return may be helpful for a better control of a quadratic pricing kernel on the range of interest. Since this information may be in practice backed out of observed derivative asset prices, it is worth checking how it works on simulated paths. Figure 1.4 displays the pricing kernel surface as well as its time average as a function of the net market return. This figure is obtained with our value of η_t (time average of 6.4 10^{-3}) which determines the coefficients λ_{1t} and λ_{2t} of the pricing kernel by application of corollary 3.2 and theorem 3.3. No paradoxical behavior of the pricing kernel is observed in this figure: on the range of interest for the net market return, the pricing kernel is always decreasing. If now one increases the value of η_t, by fixing somewhat arbitrarily the price of the squared market return at the level 1.02, which in turns implies a time-varying η_t (with a time average of 15.6 10^{-3}), one gets figure 1.5. Then, one may observe that, by contrast with figure 1.4, on the same range of values of the market return, the aforementioned increasing shape of the pricing kernel for large returns may show up.

5. Conclusion

This paper investigates the relevance of nonlinear pricing kernels both at the theoretical and empirical levels. We first show that considering pricing kernels that are quadratic functions of the market return is a well-founded approximation of actual expected utility behavior when one wants to characterize locally the demand for risky asset in the neighborhood of zero risk. Such quadratic pricing kernels disclose some pricing for skewness, but only through co-skewness with the market. In other words, while taking heterogeneity of skewness preferences into account yields a violation of common separation theorems in terms of asset demands, it remains true that idiosyncratic risk is not priced, both in terms of variance and skewness.

While statistical identification of positive skewness premium may be difficult since covariance and co-skewness tend to be almost collinear across common portfolios, we are able to show through simulated data calibrated on actual estimation of a factor GARCH model of returns with in mean effect that a conditional set up is much more informative to capture relevant nonlinearities in pricing kernels. Such non linearities imply some level of risk neutral variance for the market which cannot be neglected. This observation leads us to a generalization of the Harvey and Siddique (2000) beta pricing model for skewness; by contrast with theirs, our model considers the price of the squared market return as a free parameter whose actual value might be backed out from observed derivative asset prices.

Although conditional, our study is purely static in the sense that investors only maximize a one-period utility function. As an intertemporal extension of this study is still work in progress, it will point out the role of various kinds of asymmetries in a dynamic setting. Typically, while only conditional skewness of asset returns shows up in the current paper, a multiperiod setting will also enhance the role of dynamic asymmetry, that is some instantaneous correlation between asset returns and their volatility process. Such an effect has been dubbed the leverage effect by Black (1976) and specific leverage-based dynamic risk premia should be the result of non-myopic intertemporal optimization behavior of investors with preferences for skewness.

6. Appendix: Proofs

PROOF. The solution $\omega(\sigma) = (\omega_i(\sigma))_{1<i<n}$ of problem (2.2) determines a terminal wealth

$$W(\sigma) = \mu + \sum_{i=1}^{n} \omega_i(\sigma)(R_i - \mu)$$

according to the first order conditions:

$$Eu'(W(\sigma))(R_i - \mu) = 0. \tag{6.1}$$

These conditions could be written:

$$Eh_i(\sigma) = 0 \tag{6.2}$$

with

$$h_i(\sigma) = u'(W(\sigma))(\sigma a_i(\sigma) + Y_i).$$

(6.2) implies

$$E\frac{dh_i}{d\sigma}(\sigma) = 0.$$

which also implies

$$\lim_{\sigma \rightsquigarrow 0^+} E\frac{dh_i}{d\sigma}(\sigma) = 0. \tag{6.3}$$

An easy calculation gives:

$$\text{(A3)} \Leftrightarrow \sum_{i=1}^{n} \omega_i(0)\, Cov(Y_i, Y_k) = -\frac{u'(\mu)}{u''(\mu)} a_k(0).$$
$$\Leftrightarrow \omega(0) = \Sigma^{-1}\tau a(0).$$

The Sharpe ratio for optimal portfolio is equivalent to

$$\frac{E[\omega^\intercal(\sigma)(R-\mu)]}{(Var[\omega^\intercal(\sigma)R])^{\frac{1}{2}}} = \sigma P(0).$$

Then,

$$\sigma^2 P^2(0) = \frac{(E[\omega^\intercal(\sigma)(R-\mu)])^2}{Var[\omega^\intercal(\sigma)R]}.$$

and

$$\sigma^2 P^2(0) = \frac{\left(\tau a^\top(0)\Sigma^{-1}\sigma^2 a(0)\right)^2}{\tau a^\top(0)\Sigma^{-1}(\sigma^2\Sigma)\Sigma^{-1}\tau a(0)} = \sigma^2\frac{\left(a^\top(0)\Sigma^{-1}a(0)\right)^2}{a^\top(0)\Sigma^{-1}a(0)}.$$

Then

$$P(0) = \left[a^\top(0)\Sigma^{-1}a(0)\right]^{\frac{1}{2}}.$$

Similarly, (6.2) implies

$$\lim_{\sigma\rightsquigarrow 0^+} E\frac{d^2 h_i}{d^2\sigma}(\sigma) = 0,$$

that is:

$$\sum_{i=1}^n \omega_i'(0)\, Cov(Y_i, Y_k) = \frac{\rho}{\tau}\sum_{i=1}^n \omega_i^2(0)\, EY_i^2 Y_k + 2\frac{\rho}{\tau}\sum_{i<j}^n \omega_i(0)\,\omega_j(0)\, EY_iY_jY_k + \tau a_k'(0) \quad (6.4)$$

$$\Leftrightarrow \sum_{i=1}^n \omega_i'(0)\, Cov(Y_i, Y_k) = \frac{\rho}{\tau}\frac{1}{\sigma^2}Cov\left[\left(\omega^\top(0)R\right)^2, Y_k\right] + \tau a_k'(0),$$

where $\rho = \frac{\tau^2}{2}\frac{u'''(\mu)}{u'(\mu)}$. We now define the co-skewness of asset k in portfolio ω as:

$$c_k(\omega) = \frac{Cov\left[\left(\omega^\top R\right)^2, Y_k\right]}{Var\left[\omega^\top R\right]}.$$

Then,

$$c_k(\omega(0)) = \frac{Cov\left[\left(\omega^\top(0)R\right)^2, Y_k\right]}{Var\left[\omega^\top(0)R\right]} = \frac{a^\top(0)\Sigma^{-1}\Gamma_k\Sigma^{-1}a(0)}{P^2(0)}.$$

(6.4) is equivalent to

$$\begin{aligned}\omega'(0) &= \tau\Sigma^{-1}\left[c(\omega(0))\frac{\rho}{\tau^2}\frac{1}{\sigma^2}Var\left[\omega^\top R\right] + a'(0)\right]\\ &= \tau\Sigma^{-1}\left[c(\omega(0))\frac{\rho}{\tau^2}\frac{1}{\sigma^2}\left[\tau a^\top(0)\Sigma^{-1}(\sigma^2\Sigma)\Sigma^{-1}\tau a(0)\right] + a'(0)\right]\\ &= \tau\Sigma^{-1}\left[c(\omega(0))\,\rho P^2(0) + a'(0)\right].\end{aligned}$$

Therefore

$$\omega'(0) = \tau\Sigma^{-1}\left[c(\omega(0))\,\rho P^2(0) + a'(0)\right] \tag{6.5}$$

Now, we consider asset markets for risky assets i=1,...,n with s agents s=1,...,S. Each agent is characterized by a Von-Neuman Morgenstern utility function u_s and associated preference coefficients:

$$\tau_s = -\frac{u_s'(\mu)}{u_s''(\mu)} \text{ and } \rho_s(\mu) = \tau_s^2 \frac{u_s'''(\mu)}{2u_s'(\mu)}.$$

We also assume that the net supply in each risky asset i=1,...,n is exogeneous and fixed to unity (normalization): Then in the limit case

$$\begin{aligned}\sum_{s=1}^{S} \omega^{(s)}(0) &= e,\\ \sum_{s=1}^{S} \omega^{(s)'}(0) &= 0.\end{aligned}$$

We deduce:

$$\omega^{(s)'}(0) = \Sigma^{-1}\tau_s a(0).$$

Then

$$\sum_{s=1}^{S} \Sigma^{-1}\tau_s a(0) = e$$

which implies

$$a(0) = \frac{1}{\overline{\tau}}\Sigma\overline{\omega}. \tag{6.6}$$

We rewritte the marginal shares of agent s

$$\omega^{(s)'}(0) = \tau_s\Sigma^{-1}\left[c(\overline{\omega})\,\rho_s P^2(0) + a'(0)\right]$$

Taking the sum of this equation for s=1,...,S, we get

$$\sum_{s=1}^{S} \omega^{(s)'}(0) = \sum_{s=1}^{S} \tau_s\Sigma^{-1}\left[c(\overline{\omega})\,\rho_s P^2(0) + a'(0)\right] = 0.$$

This expression is equivalent to

$$a^{'}(0) = -\overline{\rho}c(\overline{\omega})P^{2}(0).$$

Then

$$\begin{aligned} a_{k}^{'}(0) &= -\overline{\rho}c_{k}(\overline{\omega})P^{2}(0) \\ &= -\overline{\rho}a^{\top}(0)\Sigma^{-1}\Gamma_{k}\Sigma^{-1}a(0). \end{aligned}$$

Since

$$a(0) = \frac{1}{\overline{\tau}}\Sigma\overline{\omega}, \tag{6.7}$$

we then have

$$\begin{aligned} a_{k}^{'}(0) &= -\overline{\rho}\frac{1}{\left(\sum_{s=1}^{S}\tau_{s}\right)^{2}}e^{\top}\Gamma_{k}e \\ &= -\frac{\overline{\rho}}{\overline{\tau}^{2}}\left(\overline{\omega}^{\top}\Gamma_{k}\overline{\omega}\right). \end{aligned} \tag{6.8}$$

From (6.5), we rewritte the marginal shares of agent s

$$\begin{aligned} \omega^{(s)'}(0) &= \tau_{s}\Sigma^{-1}\left[c(\overline{\omega})\rho_{s}P^{2}(0) + a^{'}(0)\right] \\ &= \tau_{s}\left[\rho_{s} - \overline{\rho}\right]P^{2}(0)\Sigma^{-1}c(\overline{\omega}). \end{aligned}$$

Therefore,

$$\omega^{(s)'}(0) = \frac{\tau_{s}}{\overline{\tau}^{2}}\Sigma^{-1}\left[\rho_{s} - \overline{\rho}\right]\Lambda \text{ for } s = 1, ..., S.$$

A variable m is a valid SDF when risk premium of asset k is $\sigma^{2}\left[a_{k}(0) + \sigma a_{k}^{'}(0)\right]$ if and only if:

$$Em(R_{k} - \mu) = 0 \Leftrightarrow Cov(m, R_{k}) = -\frac{1}{\mu}\sigma^{2}\left[a_{k}(0) + \sigma a_{k}^{'}(0)\right].$$

Note from (6.7) that:

$$\sigma^{2}a_{k}(0) = \frac{1}{\overline{\tau}S}Cov\left(\sum_{i=1}^{n}(\sigma Y_{i}), R_{k}\right)$$

and from (6.8),

$$\sigma^3 a_k^{'}(0) = Cov\left(\frac{\overline{\rho}}{(\overline{\tau}S)^2}\left(\sum_{i=1}^{n}(\sigma Y_i)\right)^2, R_k\right)$$

Therefore,

$$-\frac{1}{\mu}\sigma^2\left[a_k(0)+\sigma a_k^{'}(0)\right] = Cov\left(-\frac{1}{\mu\sum\limits_{s=1}^{S}\tau_s}\sum_{i=1}^{n}(\sigma Y_i)+\frac{\overline{\rho}}{\mu\left(\sum\limits_{s=1}^{S}\tau_s\right)^2}\left(\sum_{i=1}^{n}(\sigma Y_i)\right)^2, R_k\right)$$

and to identify this covariance with $Cov(m, R_k)$ it is suffices to choose:

$$\begin{aligned} m &= a-\frac{1}{\mu\sum\limits_{s=1}^{S}\tau_s}\sum_{i=1}^{n}(\sigma Y_i)+\frac{\overline{\rho}}{\mu\left(\sum\limits_{s=1}^{S}\tau_s\right)^2}\left(\sum_{i=1}^{n}(\sigma Y_i)\right)^2 \\ &= a-\frac{1}{\mu\sum\limits_{s=1}^{S}\tau_s}\left(\sum_{i=1}^{n}(R_i-ER_i)\right)+\frac{\overline{\rho}}{\mu\left(\sum\limits_{s=1}^{S}\tau_s\right)^2}\left(\sum_{i=1}^{n}(R_i-ER_i)\right)^2 \end{aligned}$$

for some well chosen constant a.

Let us denote

$$R_M = \sum_{i=1}^{n} R_i.$$

Then, we rewrite m

$$\begin{aligned} m &= a-\frac{1}{\mu\sum\limits_{s=1}^{S}\tau_s}(R_M-ER_M)+\frac{\overline{\rho}}{\mu\left(\sum\limits_{s=1}^{S}\tau_s\right)^2}(R_M-ER_M)^2 \\ &= a-\frac{1}{S\mu}\frac{1}{\overline{\tau}}(R_M-ER_M)+\frac{1}{S^2\mu}\frac{\overline{\rho}}{\overline{\tau}^2}(R_M-ER_M)^2. \end{aligned}$$

Using the fact that

$$Em = \frac{1}{\mu},$$

we get

$$a = \frac{1}{\mu} - \frac{1}{S^2\mu}\frac{\overline{\rho}}{\overline{\tau}^2}E(R_M-ER_M)^2$$

and

$$m = \frac{1}{\mu} - \frac{1}{\mu S \overline{\tau}} (R_M - ER_M) + \frac{\overline{\rho}}{\mu S^2 \overline{\tau}^2} \left[(R_M - ER_M)^2 - E (R_M - ER_M)^2 \right].$$

■

PROOF OF THEOREM 3.3. We apply (3.15) to the asset net return:

$$r_{it+1} = \frac{r^2_{Mt+1}}{\eta_t} - \mu_t.$$

We get:

$$E_t r^2_{Mt+1} - \mu_t \eta_t = (E_t r_{Mt+1}) \gamma_{mmt} \frac{Var r^2_{Mt+1}}{Var r_{Mt+1}} - \lambda_{2t} \left(Var r^2_{Mt+1} \right) \left(1 - \rho^2_t \left(r_{Mt+1}, r^2_{Mt+1} \right) \right),$$

that is:

$$\frac{P^{(2)}_{Mt}}{\eta_t} = \gamma_{mmt} P^{(1)}_{Mt} - \lambda_{2t} \left(1 - \rho^2_t \left(r_{Mt+1}, r^2_{Mt+1} \right) \right).$$

This gives the announced value for λ_{2t}. ■

PROOF OF LEMMA 3.4. The conditional linear regression of R^2_{Mt+1} on R_{Mt+1} is of the form

$$EL_t \left[R^2_{Mt+1} | R_{Mt+1} \right] = E_t R^2_{Mt+1} + a F_{1t+1},$$

where

$$F_{1t+1} = R_{Mt+1} - E_t R_{Mt+1}.$$

The residual of the conditional linear regression of R^2_{Mt+1} on R_{Mt+1}, that is $R^2_{Mt+1} - EL_t \left[R^2_{Mt+1} | R_{Mt+1} \right]$, is orthogonal to F_{1t+1}. Consequently,

$$Cov_t \left(R^2_{Mt+1} - EL_t \left[R^2_{Mt+1} | R_{Mt+1} \right], F_{1t+1} \right) = 0.$$

Solving this equation gives

$$a = \frac{Cov_t \left(R^2_{Mt+1}, R_{Mt+1} \right)}{Var_t \left(R_{Mt+1} \right)}.$$

Then,

$$EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right]=\frac{Cov_t\left(R_{Mt+1}^2,R_{Mt+1}\right)}{Var_t\left(R_{Mt+1}\right)}\left(R_{Mt+1}-E_tR_{Mt+1}\right). \tag{6.9}$$

But,

$$\begin{aligned}Cov_t\left(R_{Mt+1}^2,R_{Mt+1}\right) &= Cov_t\left(\left(R_{Mt+1}-E_tR_{Mt+1}+E_tR_{Mt+1}\right)^2,R_{Mt+1}\right)\\ &= Cov_t\left(\left(R_{Mt+1}-E_tR_{Mt+1}\right)^2+2\left(R_{Mt+1}-E_tR_{Mt+1}\right)E_tR_{Mt+1},R_{Mt+1}\right)\\ &= Cov_t\left(\left(R_{Mt+1}-E_tR_{Mt+1}\right)^2,R_{Mt+1}\right)+2E_tR_{Mt+1}Var_t\left(R_{Mt+1}\right)\\ &= E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3+2E_tR_{Mt+1}Var_t\left(R_{Mt+1}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right] &= E_tR_{Mt+1}^2+\frac{\left(R_{Mt+1}-E_tR_{Mt+1}\right)E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3}{Var_t\left(R_{Mt+1}\right)}+\\ &\quad 2\left(R_{Mt+1}-E_tR_{Mt+1}\right)E_tR_{Mt+1}.\end{aligned}$$

This ends the proof. ■

PROOF OF THEOREM 3.5. Let us first note that

$$\sigma_{mt}^{*2}=\mu_tE_tm_{t+1}\left(R_{Mt+1}-\mu_t\right)^2$$

where,

$$\left(R_{Mt+1}-\mu_t\right)^2=R_{Mt+1}^2+\mu_t^2-2R_{Mt+1}\mu_t. \tag{6.10}$$

But the squared market return can be rewritten as

$$R_{Mt+1}^2=EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right]+\varepsilon_{t+1}$$

where,

$$E_t\varepsilon_{t+1}=0.$$

We replace this last expression into (6.10) and get

$$
\begin{aligned}
(R_{Mt+1}-\mu_t)^2 &= EL_t\left[R_{Mt+1}^2|R_{Mt+1}\right]+\varepsilon_{t+1}+\mu_t^2-2R_{Mt+1}\mu_t \\
&= E_tR_{Mt+1}^2+\frac{(R_{Mt+1}-E_tR_{Mt+1})\,E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3}{Var_t\left(R_{Mt+1}\right)}+ \\
&\quad 2\left(R_{Mt+1}-E_tR_{Mt+1}\right)E_tR_{Mt+1}+\varepsilon_{t+1}+\mu_t^2-2R_{Mt+1}\mu_t.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sigma_{mt}^{*2} &= E_tR_{Mt+1}^2+\frac{\mu_t\left(1-\frac{1}{\mu_t}E_tR_{Mt+1}\right)E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3}{Var_t\left(R_{Mt+1}\right)}+ \\
&\quad 2\mu_t\left(1-\frac{1}{\mu_t}E_tR_{Mt+1}\right)E_tR_{Mt+1}+\mu_tCov\left(m_{t+1},\varepsilon_{t+1}\right)-\mu_t^2 \\
&= E_tR_{Mt+1}^2-P_{Mt}^{(1)}E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3+2\mu_t\left(1-\frac{1}{\mu_t}E_tR_{Mt+1}\right)E_tR_{Mt+1}+ \\
&\quad \mu_tCov\left(m_{t+1},\varepsilon_{t+1}\right)-\mu_t^2 \\
&= \left(E_tr_{Mt+1}^2-2\left(E_tr_{Mt+1}\right)^2\right)-P_{Mt}^{(1)}E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3+\mu_tCov\left(m_{t+1},\varepsilon_{t+1}\right) \\
&= \sigma_{mt}^2\left(1-\left(P_{Mt}^{(1)}\right)^2\sigma_{mt}^2\right)-P_{Mt}^{(1)}E_t\left(R_{Mt+1}-E_tR_{Mt+1}\right)^3+\mu_tCov\left(m_{t+1},\varepsilon_{t+1}\right).
\end{aligned}
$$

This ends the proof. ■

PROOF OF THEOREM 3.6. Assume that the joint process $(m_{t+1},R_{Mt+1})^{\top}$ is conditionally lognormal. Then,

$$
\begin{bmatrix} Log\left(m_{t+1}\right) \\ LogR_{Mt+1} \end{bmatrix}/I_t \rightsquigarrow N\left[\begin{bmatrix}\mu_{mt}\\ \mu_{Mt}\end{bmatrix},\begin{bmatrix}\sigma_t^2 & \sigma_{mrt}\\ \sigma_{mrt} & \sigma_{Mt}^2\end{bmatrix}\right].
$$

Let us denote

$$
c_{mt}=E_tm_{t+1}R_{Mt+1}^2.
$$

The market return risk neutral variance σ_{mt}^{*2} is

$$
\sigma_{mt}^{*2}=E_t^*R_{Mt+1}^2-\mu_t^2.
$$

Where,

$$
E_t^*R_{Mt+1}^2=\mu_tE_tm_{t+1}R_{Mt+1}^2.
$$

We know that:

$$Log\left(m_{t+1}R_{Mt+1}^2\right) = Log\left(m_{t+1}\right) + 2Log\left(R_{Mt+1}\right).$$

Therefore,

$$\begin{aligned} E_t m_{t+1} R_{Mt+1}^2 &= \exp\left(\mu_{mt} + 2\mu_{Mt} + 0.5\sigma_t^2 + 2\sigma_{Mt}^2 + 2\sigma_{mrt}\right) \\ &= \exp\left(-\mu_{mt} - 0.5\sigma_t^2\right)\exp\left(2\mu_{Mt} + 2\sigma_{Mt}^2\right)\exp\left(-2\mu_{Mt} - \sigma_{Mt}^2\right) \times \\ &\quad \left[\exp\left(\mu_{mt} + \mu_{Mt} + 0.5\sigma_t^2 + 0.5\sigma_{Mt}^2 + \sigma_{mrt}\right)\right]^2. \end{aligned}$$

But

$$E_t m R_{Mt+1} = 1 \Leftrightarrow \exp\left(\mu_{mt} + \mu_{Mt} + 0.5\sigma_t^2 + 0.5\sigma_{Mt}^2 + \sigma_{mrt}\right)$$

Therefore,

$$\begin{aligned} E_t m_{t+1} R_{Mt+1}^2 &= \exp\left(-\mu_{mt} - 0.5\sigma_t^2\right)\exp\left(2\mu_{Mt} + 2\sigma_{Mt}^2\right)\exp\left(-2\mu_{Mt} - \sigma_{Mt}^2\right) \\ &= \mu_t \frac{ER_{Mt+1}^2}{\left(E_t R_{Mt+1}\right)^2}. \end{aligned}$$

Consequently,

$$\sigma_{mt}^{*2} = \mu_t^2 \frac{E_t R_{Mt+1}^2}{\left(E_t R_{Mt+1}\right)^2} - \mu_t^2 = \sigma_{mt}^2 \left(\frac{\mu_t}{E_t R_{Mt+1}}\right)^2 < \sigma_{mt}^2.$$

■

PROOF OF THEOREM 3.7. Assume that

$$m_{t+1} = \nu_{0t} + \nu_{1t} R_{Mt+1} + \nu_{2t} R_{Mt+1}^2.$$

Then,

$$Cov_t\left(m_{t+1}, \varepsilon_{t+1}\right) = \nu_{2t} Cov_t\left(R_{Mt+1}^2, \varepsilon_{t+1}\right).$$

But

$$\begin{aligned}
Cov_t\left(R_{Mt+1}^2, \varepsilon_{t+1}\right) &= Cov_t\left(R_{Mt+1}^2, R_{Mt+1}^2 - \frac{\left(R_{Mt+1} - E_t R_{Mt+1}\right) Cov_t\left(R_{Mt+1}^2, R_{Mt+1}\right)}{Var_t\left(R_{Mt+1}\right)}\right) \\
&= Var_t\left(R_{Mt+1}^2\right) - \frac{Cov_t^2\left(R_{Mt+1}^2, R_{Mt+1}\right)}{Var_t\left(R_{Mt+1}\right)} \\
&= Var_t\left(R_{Mt+1}^2\right)\left[1 - \frac{Cov_t^2\left(R_{Mt+1}^2, R_{Mt+1}\right)}{Var_t\left(R_{Mt+1}^2\right) Var_t\left(R_{Mt+1}\right)}\right] \\
&= Var_t\left(R_{Mt+1}^2\right)\left[1 - \rho_t^2\left(R_{Mt+1}^2, R_{Mt+1}\right)\right].
\end{aligned}$$

■

References

[1] Bakshi, G., N. Kapadia and D. Madan, 2003, "Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options," Review of Financial Studies (Winter 2003) 101-143.

[2] Bakshi, G., and D. Madan, 2000, "Spanning and Derivative Security Valuation," Journal of Financial Economics, 55(2), 205-238.

[3] Barone-Adesi, G., 1985, "Arbitrage Equilibrium with Skewed Asset Returns," Journal of Financial and Quantitative Analysis, 20, 299-313.

[4] Barone-Adesi, G., G. Urga and P. Gagliardini 2004, "Testing Asset Pricing Models with Coskewness," Forthcoming in Journal of Business and Economic Statistics.

[5] Bekaert, G., and J. Liu, 2004, "Conditioning Information and Variance Bounds on Pricing Kernels," The Review of Financial Studies, 17(2) 339-378.

[6] Black, F., 1976, "The Pricing of Commodity Contracts," Journal of Financial Economics, 3, 167-179.

[7] Brennan, M. J., 1979, "The Pricing of Contingent Claims in Discrete Time Models," Journal of Finance, 24, 53-68.

[8] Cochrane, J., Asset Pricing, 2001, Princeton NJ: Princeton University Press.

[9] Dittmar, R. F., 2002, "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from Cross section of Equity Returns," Journal of Finance, 57, 368-403.

[10] Epstein, L., and S. Zin, 1991, "Substitution, Risk Aversion and the Temporal Behaviour of Consumption and Asset Returns: An Empirical Analysis," Journal of Political Economy 99, 263-286.

[11] Garcia, R., E. Ghysels, and E. Renault, 2003, The Econometrics of Option Pricing, Handbook of Financial Econometrics, Y. Ait Sahalia and L. P. Hansen (eds), North Holland (forthcoming)

[12] Juud, G., and S. Guu, 2001, "Asymptotic Methods for Asset Market Equilibrium Analysis," Economic Theory, 18, 127-157.

[13] Kraus, A., and R. H. Litzenberger, 1976, "Skewness Preference and the Valuation of Risk Assets," Journal of Finance 31(4), 1085-1100.

[14] Lucas, R. Jr., 1978, "Asset Prices in an Exchange Economy", Econometrica 46, 1429-1445.

[15] Hansen, L., and S. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," Econometrica, 55, 587-613.

[16] Harvey, C. R., and A. Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," The Journal of Finance, LV, 3, 1263-1295.

[17] Ingersoll, J., 1987, Theory of Financial Decision Making, Rowman & Littlefield Studies in Financial Economics.

[18] Kimball, M., 1990, "Precautionary Saving in the Small and in the Large," Econometrica 58, 53-73.

[19] Jobson, J. D., and B. M. Korkie, 1982, "Potential Performance and Tests of Portfolio Efficiency," Journal of Financial Economics, 10, 433-466.

[20] Markowitz, H., 1952, "Portfolio Selection," Journal of Finance, 7 (1), 77-91.

[21] Rosenberg J., 2000, "Asset Pricing Puzzles: Evidence from Options Markets," Working Stern School of Business.

[22] Samuelson, P. A., 1970, "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Mean, Variance and Higher Moments,", Review of Economic Studies 37, 537-42.

Table 1.1: Estimated results of the Factor GARCH in mean (see Bekaert and Liu (2004))

Equations		Coefficients		
	c_t	Y_{1t}	Y_{2t}	Y_{3t}
Y_{1t+1}	0.0030	0.361	-0.029	0.008
	(0.0005)	(0.033)	(0.022)	(0.005)
Y_{2t+1}	0.0056-162.65 $(e_{1t})^2$	-0.198	0.738	-0.0002
	(0.0006) (0.0001)	0.031	(0.037)	(0.0043)
Y_{3t+1}	$0.0188 - 58.02\,(e_{1t})^2$	-1.734	1.029	0.077
	(0.0083) (0.0003)	0.005	(0.014)	(0.034)
	constant	α	β	ξ
$Var_t\,(e_{1t+1})$	0.000019	-0.0265	0.0008	0.2705
	0.000018	(0.0807)	(0.7898)	(0.0426)
δ_2	0.000014	0	0	0
	(0.000002)			
δ_3	0.006134	0	0	0
	(0.00103)			
	$\sigma_{13} = -0.0564$		$\sigma_{12} = 3.182$	
	(0.1425)		(0.003)	

Notes: In this table, we reproduce the results of the Factor GARCH in mean estimated by Bekaert and Liu (2004). $\delta_2 = Var_t\,(e_{2t+1})$ and $\delta_3 = Var_t\,(e_{3t+1})$.

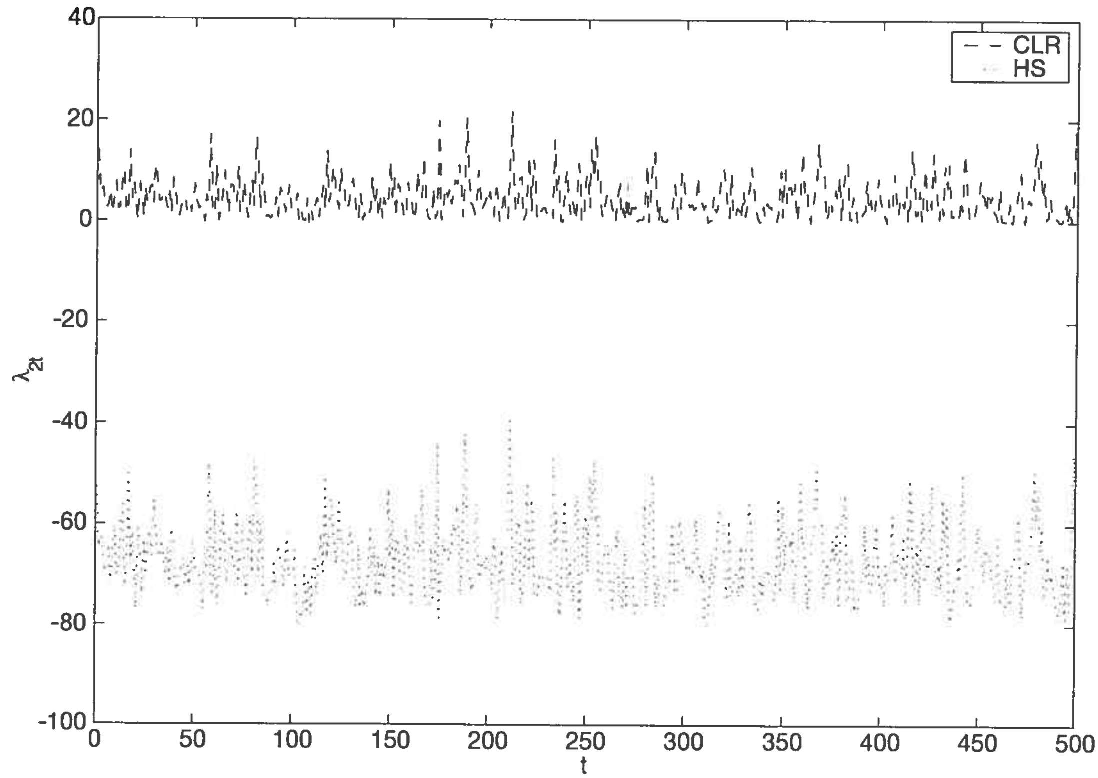


Figure 1.1: **Price of coskewness:** Price of coskewness inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004). HS indicates the price of coskewness corresponding to Harvey and Siddique (2000) limit case. CLR indicates the price of coskewness corresponding to our formula.

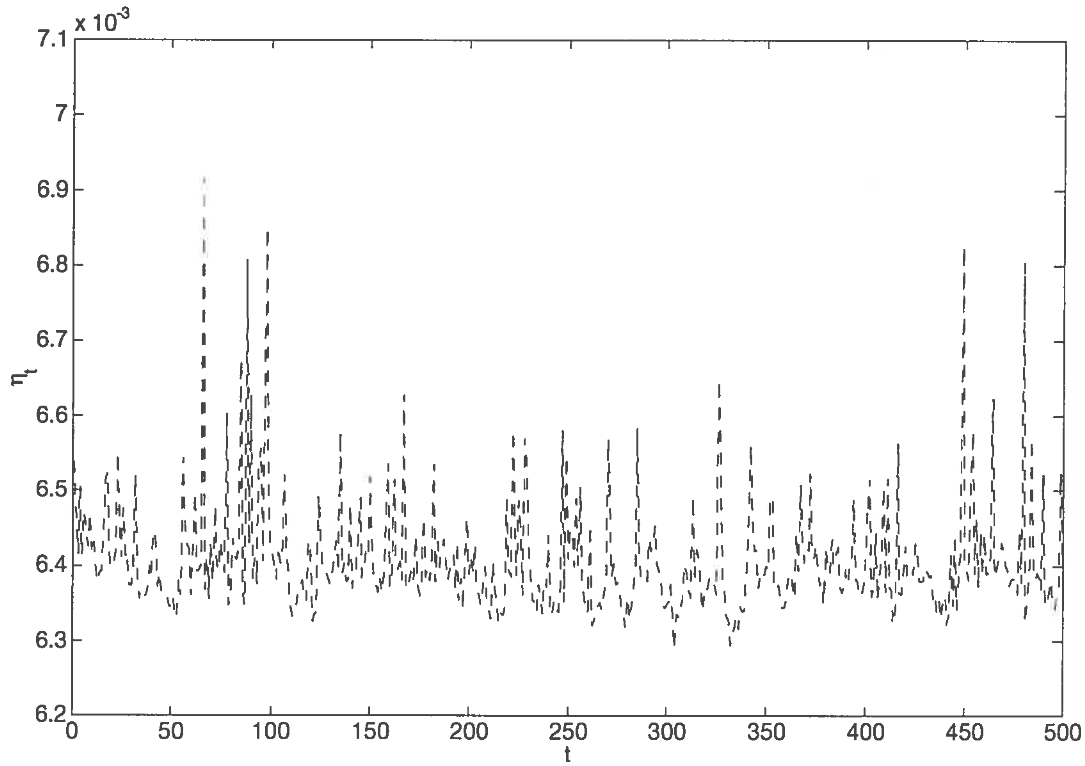


Figure 1.2: **Price of squared net return:** Price of squared net return inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004).

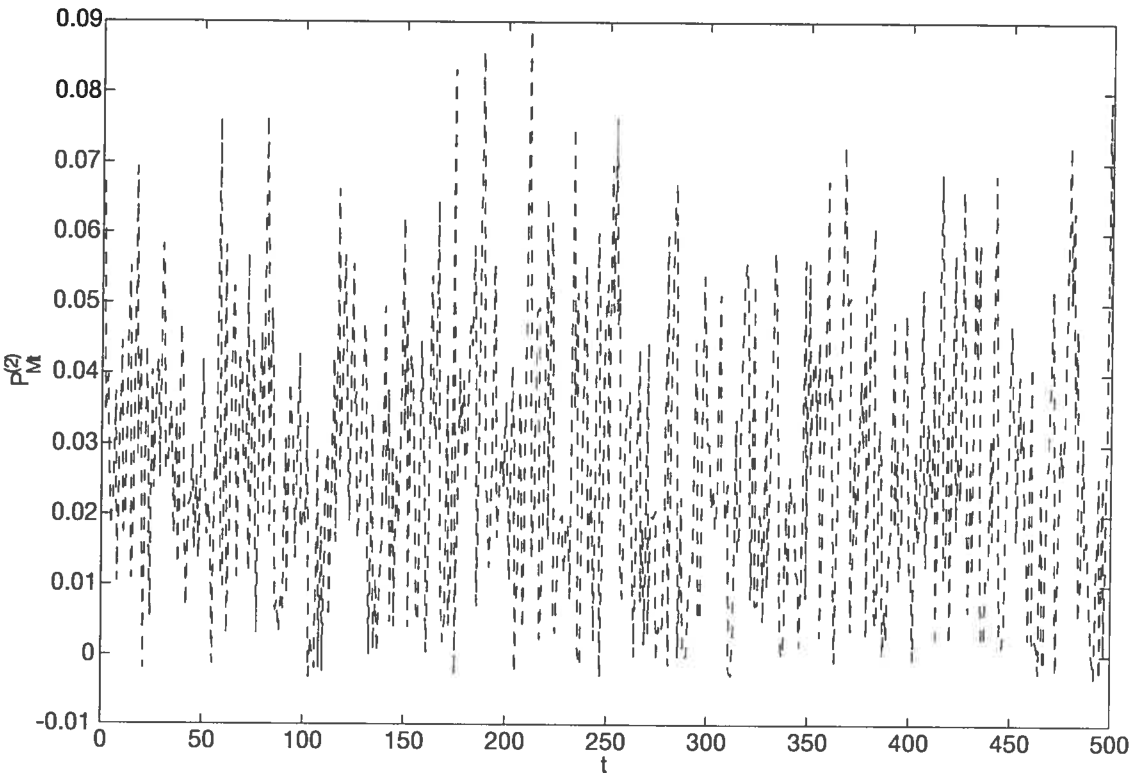


Figure 1.3: **Risk premium on the squared net return:** Risk premium on the squared net return inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004).

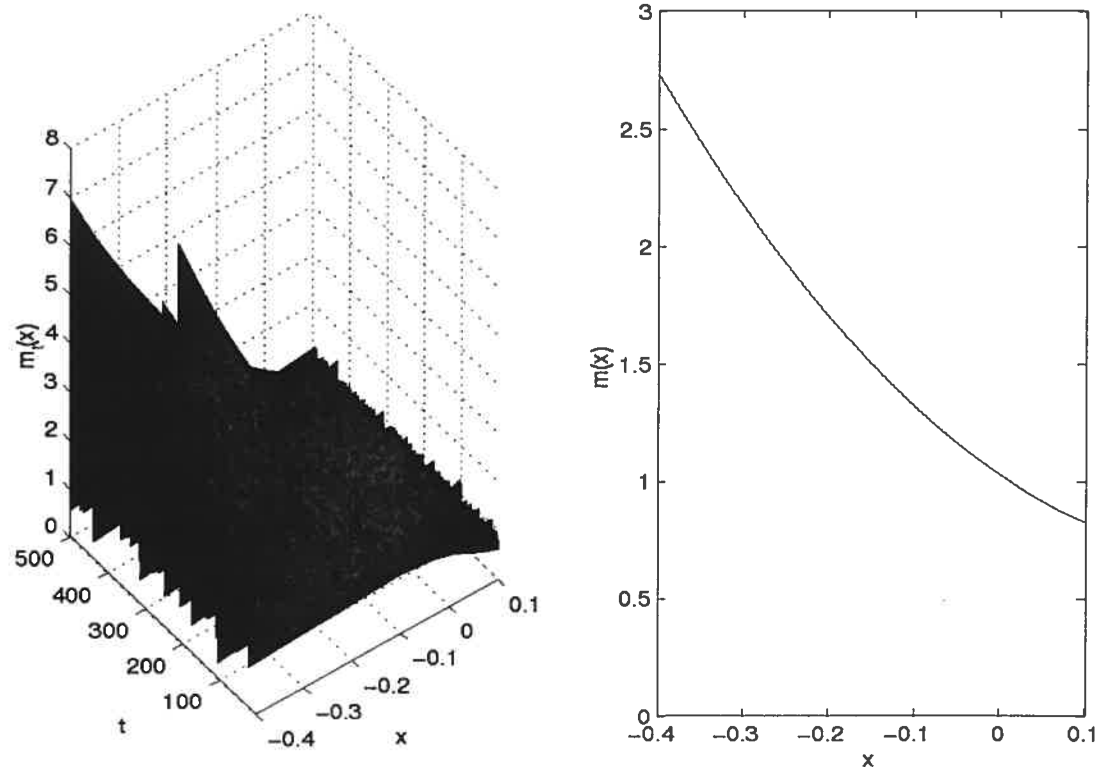
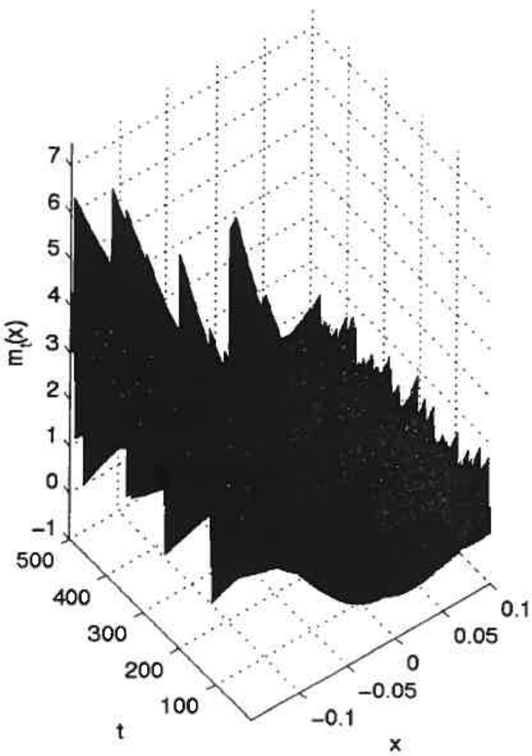


Figure 1.4: **Quadratic SDF:** Quadratic SDF inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004). In the left hand side graph, we plot the pricing kernel m_{t+1} as a function of $t+1$ and r_{Mt+1}. In the right hand side, we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$.



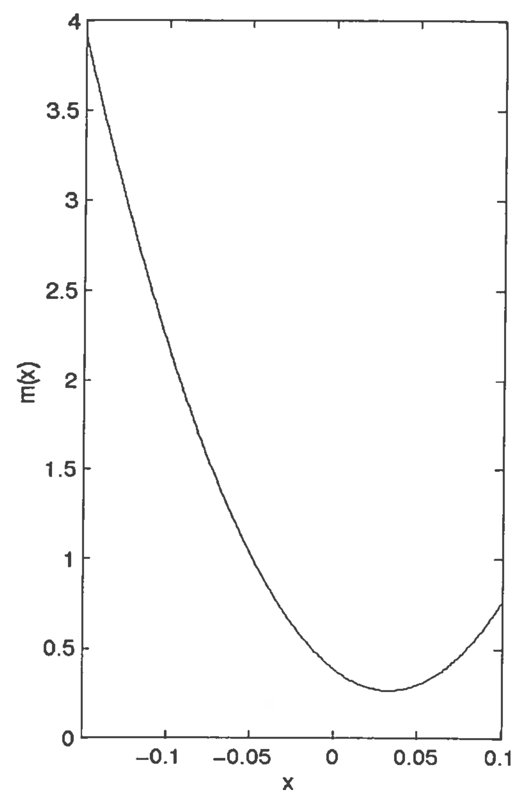


Figure 1.5: **Quadratic SDF**: Fixing the price of the squared market return at the level 1.02, which in turns implies a time varying η_t, we infer the quadratic pricing kernel for simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004). In the left hand side graph, we plot the pricing kernel m_{t+1} as a function of $t+1$ and $r_{Mt+1} = x$. In the right hand side, we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$.

Chapter 2

Stochastic Discount Factor Volatility Bound and Portfolio Selection Under Higher Moments

1. Introduction

Hansen and Richard (1987) introduced the concept of a stochastic discount factor (SDF) to the financial econometrics literature and defined a stochastic discount factor as a random variable that discounts payoffs differently in different states of the world. Since this seminal contribution, it has become evident that the empirical implications of asset pricing models can be characterized through their SDFs [Cochrane (1991)]. In this context, Hansen and Jagannathan (1991) address the question of what asset returns data may be able to tell us about the behavior of the SDF volatility. They found a lower bound on the volatility of any admissible SDF that prices correctly a set of asset returns. Their bound has been applied to a variety of financial issues. For example, the Hansen and Jagannathan (HJ) bound is used to test if a particular SDF implied by a model is valid or not. Recently, Barone-Adesi et al. (2004) assume a quadratic specification of the SDF in terms of the market return and test asset pricing models with co-skewness. They found evidence that asset skewness (co-skewness) is priced in the market through the cost of the squared market return even if the squared market return is not a traded asset. This line of thinking had been initiated by Ingersoll (1987) and pursued more recently by Harvey and Siddique (2000) and Dittmar (2002). They look at extensions of the CAPM framework by considering asset skewness. Assuming higher skewness is preferred, Ingersoll (1987) shows that a decrease in co-skewness requires an increase in expected return to induce the same holding of the asset at the margin. Furthermore if we use a Taylor series of derivatives' payoff functions as quadratic functions of the underlying asset return, we realize that the price of the derivatives is a function of the cost of the squared return and this cost is tightly related to return skewness. The cost of the squared portfolio return is, therefore, particularly relevant when pricing derivatives. Since the HJ volatility bound considers admissible SDFs that price correctly only a set of asset returns, it appears useful to construct a new variance bound for any admissible SDF that prices correctly not only a set of primitive assets but also the squared returns of the same primitive assets.

The first contribution of this paper is to find such a lower bound. While HJ minimizes the SDF variance for a given SDF mean under the assumption that the admissible SDFs price correctly a set of primitive asset returns, we minimize the SDF variance for a given SDF mean under the assumption that the admissible SDFs price correctly not only a set of primitive asset returns but also the squared return of the same primitive assets. Our variance bound tightens the HJ bound by an additional quantity which is a function of the assets' co-skewness and the cost of the squared primitive asset returns. We derive necessary and sufficient conditions to get the well-known HJ bound as a particular case. In this more general setting, our minimum variance SDF can be rewritten as a quadratic function of asset returns. By this, we mean a linear combination of two vectors: R and $R^{(2)}$ where R represents a vector of primitive asset returns and $R^{(2)}$ is a vector of the squared primitive asset returns whose components are of the form $R_i R_j$ with $i \leq j$. When R is the market return, we get a quadratic specification of the SDF in terms of the market return which is often used to underline the importance of skewness (co-skewness) in asset pricing models [Ingersoll (1987), Harvey and Siddique (2000), Dittmar (2002)]. We use the return on the Standard and Poors 500 stock index and the commercial paper index from 1889 to 1994 to illustrate our SDF volatility surface frontier. We also use the consumption on non-durables and services over the same period to relate the CRRA and Epstein and Zin (1989) preference models to our volatility bound for particular values of the relative risk aversion coefficient. We illustrate how our SDF variance frontier tightens the HJ variance frontier and makes the equity premium puzzle even more difficult to solve.

The second contribution of the paper is to offer a new approach for portfolio selection with higher moments. This approach is based on factors that span our minimum variance stochastic discount factor. The intuition behind our portfolio selection analysis is motivated by the duality between the HJ minimum variance SDF and Markowitz mean-variance analysis [Campbell Lo and Mc Kinlay (1997)]. Since we have found a minimum variance SDF that tightens the HJ minimum variance SDF, it is of interest to also give a portfolio selection approach which is based on our minimum variance SDF. Our approach consists in minimizing the portfolio risk subject to the portfolio expected return and an additional constraint (cost of the squared portfolio return) which depends on the portfolio skewness. The question we thereafter ask is: under which conditions is our portfolio choice observationally equivalent to the standard portfolio selection under skewness? We then generalize the standard portfolio selection approach under skewness which consists in minimizing the portfolio risk subject to the portfolio expected return and skewness [see Lai (1991), de Athayde and Flores (2004)]. Our more general approach is relevant since it first provides a formal bridge between the SDF variance bound and the portfolio selection under higher moments. Second, it shows that the standard approach of portfolio selection under skewness may overlook an important factor.

We also provide an empirical illustration for portfolio selection. We use daily asset returns for four individual firms. Our portfolio selection approach depends on the cost of the squared asset returns. To compute this cost, we assume that the joint process of the SDF and asset returns is lognormally distributed. The lognormal distribution is flexible and allows for skewness in asset returns. Many asset pricing tests assume that the joint process of SDF-asset returns is conditionally jointly lognormal. Moreover, diffusion models imply a locally lognormal distribution. Our results suggest that the cost of the squared portfolio return and portfolio higher moments have a significant impact on the portfolio mean-variance frontier.

The rest of the paper is organized as follows. Section 2 gives the theoretical background and an empirical illustration for the generalized SDF variance bound. In Section 3, we offer a portfolio selection approach based on factors that span our minimum variance SDF. Section 4 gives an empirical illustration for portfolio selection under higher moments. The last section concludes the paper.

2. The Minimum-Variance Stochastic Discount Factor

In this section, we first review the HJ bound and derive the SDF variance bound under higher moments. In section 2.2, we provide conditions under which the cost of the squared returns affects the variance bound and give some empirical implications of our new bound. Section 2.3 discusses the variance bound when we restrict admissible SDFs to be positive.

2.1 The general framework

In this subsection we construct a new bound on the volatility of any admissible SDF which tightens the HJ volatility bound. By a SDF, we mean a random variable that can be used to compute the market price of an asset today by discounting payoffs differently in different states of the world in the future. HJ have proposed a way to find the lower bound on the volatility of any SDF that prices correctly a set of primitive asset returns. Their approach treats the unconditional mean of the stochastic discount factor as an unknown parameter $\overline{m}$. For each possible parameter $\overline{m}$, HJ form a stochastic discount factor $m_{HJ}(\overline{m})$ as a linear combination of asset returns and show that the variance of $m_{HJ}(\overline{m})$ represents a lower bound on the variance of any stochastic discount factor that has mean $\overline{m}$ and satisfies:

$$EmR = l,$$

where l represents an N-vector with components unity and R is a set of N primitive asset returns. Let $\mathcal{F}_1(\overline{m})$ denote the set of SDFs that have mean $\overline{m}$ and that price correctly R. Therefore,

$$\mathcal{F}_1(\overline{m}) = \left\{m \in L^2 : Em = \overline{m},\ EmR = l\right\}.$$

Thus, $m_{HJ}(\overline{m})$ is the solution to the problem:

$$\underset{m \in \mathcal{F}_1(\overline{m})}{Min}\ \sigma(m).$$

HJ show that

$$m_{HJ}(\overline{m}) = \overline{m} + (l - \overline{m}ER)' \Omega^{-1}(R - ER)$$

and

$$Var\left[m_{HJ}(\overline{m})\right] = (l - \overline{m}ER)' \Omega^{-1}(l - \overline{m}ER),$$

where Ω is the covariance matrix of the asset returns. Here, the N assets are risky and no linear combination of the returns in R is equal to one with probability one so that Ω is nonsingular. Using the HJ bound, it is then possible to derive an admissible region for mean and standard deviations of candidate SDFs using only asset returns data. By plotting these regions, the HJ approach provides an appealing graphical technique through which to gauge the specification of many asset pricing models. However it appears important for any admissible SDF to price correctly not only a set of primitive assets but also payoffs which are nonlinear functions of primitive assets' payoffs. For instance, a Taylor expansion series of derivatives' payoffs around a benchmark return will imply, in general, that the cost of squared portfolio returns is relevant when pricing derivatives.

Suppose $r_p = \omega' R$ represents a portfolio return, where $\omega = (\omega_1, \omega_2, ...\omega_N)'$ is a vector of portfolio weights which satisfies $\omega' l = 1$ with $l = (1, 1, ...1)'$. The squared return of the portfolio can be represented by:

$$r_p^2 = \left(\omega' R\right)^2 = (\omega \otimes \omega)' (R \otimes R),$$

where $\otimes$ stands for the Kronecker product. The cost of the squared portfolio return is therefore,

$$\begin{aligned} \widetilde{C}\left(r_p^2\right) &= Emr_p^2 \\ &= (\omega \otimes \omega)' Em (R \otimes R) \\ &= \omega^{(2)'} EmR^{(2)}, \end{aligned}$$

where $\omega^{(2)}$ represents a column vector whose components are of the form,

$$\omega_{ij} = \begin{cases} 2\omega_i \omega_j \text{ if } i < j \\ \omega_i^2 \text{ if } i = j \end{cases}$$

and $R^{(2)}$ represents a column vector, the components of which are of the form $R_i R_j$ with $i \leq j$. It can be observed that the cost of the squared portfolio return is a function of the cost of the "squared" asset returns, $R^{(2)}$.[1] The question we ask is whether we can we tighten significantly the HJ volatility bound by considering any admissible SDF that correctly prices payoffs that can be expressed as a quadratic function of the primitive assets. The idea is to consider a set of SDFs that correctly price the N asset returns, R, and the "squared" asset returns $R^{(2)}$. If $\mathcal{F}_2(\overline{m}, \eta)$ denotes

[1] For portfolio algebra using the inverse of covariance matrices, we prefer using $R^{(2)}$ than $R \bigcirc R$ since the latter has a singular covariance matrix.

a set of SDFs that correctly price R and $R^{(2)}$, we have,

$$\mathcal{F}_2(\overline{m},\eta) = \left\{m \in L^2 : Em = \overline{m},\ EmR = l,\ EmR^{(2)} = \eta\right\}.$$

where η denotes the vector of prices of squared returns. Notice that $\mathcal{F}_2(\overline{m},\eta) \subset \mathcal{F}_1(\overline{m})$. Intuitively, we exclude any admissible SDF that does not correctly price derivatives with payoffs that can be written as a quadratic function of a set of primitive assets. We then treat the unconditional mean $\overline{m}$ of the SDF and the cost η of the "squared" primitive asset, $R^{(2)}$, as unknown parameters. For each $\overline{m}$ and η, we form a candidate SDF, $m^{mvs}(\eta,\overline{m})$, as a quadratic function of asset returns:

$$m^{mvs}(\eta,\overline{m}) = \alpha(\eta,\overline{m}) + \beta(\eta,\overline{m})' R + \gamma(\eta,\overline{m})' R^{(2)} \tag{2.1}$$

with

$$\alpha(\eta,\overline{m}) = \overline{m} - \beta(\eta,\overline{m})' ER - \gamma(\eta,\overline{m})' ER^{(2)},$$

since $Em^{mvs}(\eta,\overline{m}) = \overline{m}$. Therefore, we exploit the pricing formulas $E(Rm) = l$ and $E\left(R^{(2)}m\right) = \eta$ to compute the parameters,

$$\begin{aligned}\beta(\eta,\overline{m}) &= \Omega^{-1}(l - \overline{m}ER) - \Omega^{-1}\Lambda\gamma(\eta,\overline{m}),\\ \gamma(\eta,\overline{m}) &= \left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right]^{-1}\left[\eta - \overline{m}ER^{(2)} - \Lambda'\Omega^{-1}(l - \overline{m}ER)\right],\end{aligned}$$

with,

$$\begin{aligned}\Sigma &= ER^{(2)}\left(R^{(2)} - ER^{(2)}\right)',\\ \Lambda' &= E\left(R^{(2)} - ER^{(2)}\right)R'.\end{aligned}$$

Note that Λ is related to the notion of co-skewness [see Ingersoll (1987), Harvey and Siddique (2000)]. The expansion $\Psi = \Sigma - \Lambda'\Omega^{-1}\Lambda$ denotes the residual covariance matrix in the regression of $R^{(2)}$ on R. We assume that the matrix Ψ is nonsingular; that is, no squared returns are redundant with respect to the primitive assets. This assumption will be maintained hereafter for the sake of notational simplicity. A simple application of the HJ argument to the vector $\left[R, (diag\ \eta)^{-1} R^{(2)}\right]$

of returns, (where $diag\ \eta$ denotes the diagonal matrix with coefficients defined by the components of η) ensures that $m^{mvs}(\eta, \overline{m})$ gives the volatility lower-bound in $\mathcal{F}_2(\overline{m}, \eta)$. That is, it solves:

$$\min_{m \in \mathcal{F}_2(\overline{m}, \eta)} \sigma(m).$$

To compare this minimum variance SDF to the HJ minimum variance stochastic discount factor associated with only the vector R of returns, we rewrite $m^{mvs}(\eta, \overline{m})$ as a function of the HJ minimum variance SDF.

Proposition 2.1 *The minimum variance stochastic discount factor among any admissible stochastic discount factors that correctly price not only a set of primitive assets but also derivatives the payoffs of which can be written as a quadratic function of the same primitive assets as follows:*

$$m^{mvs}(\eta, \overline{m}) = m_{HJ}(\overline{m}) + \gamma(\eta, \overline{m})' \left[R^{(2)} - ER^{(2)} - \Lambda'\Omega^{-1}(R - ER)\right]$$

where,

$$\gamma(\eta, \overline{m}) = \left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right]^{-1} \left[\eta - \overline{m}ER^{(2)} - \Lambda'\Omega^{-1}(l - \overline{m}ER)\right].$$

We are now going to discuss the necessary and sufficient conditions to get the HJ minimum variance SDF.

Proposition 2.2 *The minimum variance stochastic discount factor, $m^{mvs}(\eta, \overline{m})$, collapses to the Hansen and Jagannathan minimum variance stochastic discount factor, $m_{HJ}(\overline{m})$, if and only if*

$$\eta = \overline{m}ER^{(2)} + \Lambda'\Omega^{-1}(l - \overline{m}ER).$$

Proof. Of course, if $\gamma(\eta, \overline{m}) = 0$, we have $m^{mvs}(\eta, \overline{m}) = m_{HJ}(\overline{m})$. Conversely, assume that $m^{mvs}(\eta, \overline{m}) = m_{HJ}(\overline{m})$, thus it follows that

$$\gamma(\eta, \overline{m})' \left[R^{(2)} - ER^{(2)} - \Lambda'\Omega^{-1}(R - ER)\right] = 0.$$

If we premultiply this equality by $m^{mvs}(\eta, \overline{m})$, we get

$$\gamma(\eta, \overline{m})' m^{mvs}(\eta, \overline{m}) \left[R^{(2)} - ER^{(2)} - \Lambda'\Omega^{-1}(R - ER)\right] = 0.$$

Taking the expectation of this quantity, it is easy to show that

$$\eta - \overline{m}ER^{(2)} - \Lambda'\Omega^{-1}(l - \overline{m}ER) = 0.$$

This implies that $\gamma(\eta, \overline{m}) = 0$. ■

Note that propositions 2.1 and 2.2 have been derived under the maintained assumption that squared returns are not redundant assets, that is $R^{(2)}$ does not coincide with its affine regression R:

$$R^{(2)} - ER^{(2)} - \Lambda'\Omega^{-1}(R - ER).$$

However, when ever this residual has zero price, we see from proposition 2.2 that $m^{mvs}(\eta, \overline{m})$ and $m_{HJ}(\overline{m})$ coincide, that is when its product by the SDF has a zero expectation.

Our volatility bound can be used to assess the specification of a particular asset pricing model as is usually done with the HJ bound. But our bound is tighter:

$$\sigma[m^{mvs}(\eta, \overline{m})] \geq \sigma[m_{HJ}(\overline{m})] \text{ for all } \eta. \tag{2.2}$$

To see how our volatility bound can be used to check if a particular asset pricing model explains asset returns, let us consider a proposed SDF, $m(x)$, where x represents a set of relevant variables, for example the ratio of consumption, $x = \frac{C_{t+1}}{C_t}$, or the first difference of consumption $C_{t+1} - C_t$. To gauge if the proposed SDF passes our volatility bound, we need to first compute $\eta = Em(x)R^{(2)}$ and $Em(x)$ and then check if $\sigma[m(x)] \geq \sigma[m^{mvs}(\eta, Em(x))]$. If the proposed SDF passes the HJ bound but not our variance bound, it means that the proposed SDF variance is too low so that this SDF cannot correctly price derivatives, the payoffs of which are a quadratic function of the primitive assets. Since the price of such derivatives can be written as a function of the cost of $R^{(2)}$ and that this cost is a function of asset skewness, the failure of the proposed SDF is akin to a failure to price skewness correctly.

2.2 Why does the price of squared returns matter?

By the inequality (2.2), we realize that our variance bound is greater than the HJ bound. The first question we then ask is: is there pricing condition(s) under which our variance bound coincides with the HJ bound? Under these conditions, the squared return cost does not matter and we have failed to shed more light on the SDF variance bound. Proposition 2.3 summarizes this issue.

Proposition 2.3 *Consider the linear regression of the squared returns, $R^{(2)}$, on the return, R, that is:*

$$EL\left[R^{(2)}|R\right] = ER^{(2)} + \Lambda'\Omega^{-1}(R - ER).$$

and

$$\eta^* = \overline{m}ER^{(2)} + \Lambda'\Omega^{-1}(l - \overline{m}ER),$$

the price of this regression. Then, there exists $\eta > 0$ such that

$$\sigma\left[m^{mvs}(\eta, \overline{m})\right] = \sigma\left[m_{HJ}(\overline{m})\right]$$

if and only if $\eta^* > 0$.

PROOF. We have $\gamma(\eta, \overline{m}) = \left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right]^{-1}[\eta - \eta^*]$. Then if $\eta^* > 0$, $\eta = \eta^*$ implies $\gamma(\eta, \overline{m}) = 0$. We, therefore, have $\sigma\left[m^{mvs}(\eta^*, \overline{m})\right] = \sigma\left[m_{HJ}(\overline{m})\right]$. Conversely assume that there exists $\eta > 0$ such that $\sigma\left[m^{mvs}(\eta, \overline{m})\right] = \sigma\left[m_{HJ}(\overline{m})\right]$. This implies that $\gamma'(\eta, \overline{m})\left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right]\gamma(\eta, \overline{m}) = 0$. But

$$\gamma'(\eta, \overline{m})\left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right]\gamma(\eta, \overline{m}) = (\eta - \eta^*)'\left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right](\eta - \eta^*).$$

Therefore, $(\eta - \eta^*)'\left[\Sigma - \Lambda'\Omega^{-1}\Lambda\right](\eta - \eta^*) = 0$. Since in this paper we assume that the matrix $\Sigma - \Lambda'\Omega^{-1}\Lambda$ is nonsingular, we conclude by the Cauchy-Schwarz inequality that $\Sigma - \Lambda'\Omega^{-1}\Lambda$ is positive definite, then $\eta^* = \eta > 0$. ■

In other words, when $\eta^* < 0$, $\sigma\left[m^{mvs}(\eta, \overline{m})\right] > \sigma\left[m_{HJ}(\overline{m})\right]$ for all $\eta > 0$. Then taking into account the cost of squared returns will always have a significant impact on the volatility bound.

Recently, Kan and Zhou (2003) proposed an alternative way to tighten the HJ bound. They assume that they can find a vector x of state variables such that the conditional expectation of $m_{HJ}(\overline{m})$ given x coincides with its affine regression. Under this maintained assumption, they are able to show that any admissible SDF $m(x)$ which is a deterministic function of x has a larger volatility than a bound $\sigma^2[m_{KZ}]$ defined by:

$$\sigma^2[m(x)] \geq \sigma^2[m_{KZ}] = \frac{1}{\rho^2_{m_{HJ},x}}\sigma^2[m_{HJ}(\overline{m})],$$

where $\rho_{m_{HJ},x}$ is the multiple linear correlation coefficient between $m_{HJ}(\overline{m})$ and x. By considering $x = [R, R^{(2)}]$, we can then claim that:

$$\inf_{\eta} \sigma^2[m^{mvs}(\eta, \overline{m})] \geq \sigma^2[m_{KZ}].$$

Therefore, the Kan and Zhou volatility bound does not make our bound irrelevant. The cost of squared returns may matter significantly.

We now give empirical illustrations showing that squared return cost may be important. We first consider the annual excess simple return of the Standard and Poors 500 stock index over the commercial paper from 1889 to 1994. In this case, $q = 1$ and our SDF variance bound is easy to illustrate graphically. Figures 2.1 and 2.2 illustrate our variance bound surface and the HJ bound respectively. It can be seen from these figures that the cost of the squared asset excess return has a significant impact on the SDF mean-standard deviation frontier. For example, for a SDF mean in the neighborhood of 1, the HJ SDF standard-deviation is about 0.3 whereas our SDF standard deviation is greater than 0.6 for any positive value of the squared return cost. According to proposition 2.3, this should be a case where the cost η^* of the affine regression of $R^{(2)}$ on R is negative. Furthermore, when the SDF mean in the neighborhood of 1, our minimum variance SDF standard-deviation highly depends on the cost of the squared asset excess return. Thus the squared returns cost is relevant for determining the SDF variance bound. Similarly to the HJ volatility bound, our volatility bound can be used to illustrate if a particular asset pricing model fails to explain a set of asset returns. To give this illustration, we consider several consumption-based models. The first model assumes that there is a representative agent who maximizes a time-separable power utility function, so that:

$$u(C_{t+1}) = \frac{C_{t+1}^{1-\alpha} - 1}{1 - \alpha},$$

where α is the coefficient of relative risk aversion and C_{t+1} is the aggregate consumption. Therefore, it can be shown that the representative agent optimization problem yields a SDF of the form:

$$m_{t+1} = \beta \frac{u'(C_{t+1})}{u'(C_t)},$$

where $\beta \in (0,1)$ is a subjective discount parameter. For this CRRA preference model, we set $\beta = .95$. Using consumption on non durables and services over the same period, 1889 to 1994, Campbell, Lo and Mc Kinlay (1997) show that the variance of m_{t+1} enters in the HJ feasible region if the relative risk aversion coefficient α is greater than 25. This can be seen through Figure 2.3. Varying exogenously α from 0 to 27, the point $(Em_{t+1}, \sigma(m_{t+1}))$ does fall into the feasible region until the coefficient of the relative risk aversion reaches a value of 25.

Since our SDF variance bound is greater than the HJ variance bound, it is clear that for $\alpha \leq 24$, the point $(Em_{t+1}, \sigma(m_{t+1}), \eta)$ with $\eta = Em_{t+1}R^{(2)}$ does not enter into our feasible region. We need to check if any particular relative risk aversion $\alpha \geq 25$ produces a point $(Em_{t+1}, \sigma(m_{t+1}), \eta)$ which enters our feasible region. To proceed to our graphical illustration, for $\alpha = 25$ and $\alpha = 27$, we compute η and find the corresponding feasible region. We check, thereafter, if the point $(Em_{t+1}, \sigma(m_{t+1}))$ enters our feasible region. While Figure 2.4 shows that for various relative risk aversion coefficients, our variance bound never coincide with the HJ bound, the two bounds give nevertheless the same conclusion about the candidate SDFs produced by this model.

We repeat the same calibration exercise using the Epstein and Zin (1989) state-non-separable preferences. Following Epstein and Zin (1989), we assume that the state-non-separable preferences are given by $V_t = U[C_t, E_t V_{t+1}]$ where

$$U[C_t, V] = \frac{\left[(1-\beta) C_t^{1-\rho} + \beta \left[1 + (1-\beta)(1-\alpha) V\right]^{\frac{1-\rho}{1-\alpha}}\right]^{\frac{1-\alpha}{1-\rho}} - 1}{(1-\beta)(1-\alpha)}.$$

The elasticity of intertemporal substitution is $1/\rho$. The representative agent SDF is

$$m_{t+1} = \left[\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right]^{\frac{1-\alpha}{1-\rho}} R_{mt+1}^{\left(\frac{1-\alpha}{1-\rho}\right)-1},$$

where R_{mt+1} is the return on the market portfolio. Figure 2.5 plots the bound and the representative agent SDF volatilities for Epstein and Zin (1989) consumption based model. For this consumption-based model, the parameters used are $\beta = 0.96$. We use the same data set as in the CRRA case.

Figure 2.5 reveals that for $\beta = 0.96$, $(\rho, \alpha) = (3.05, 6.86)$ the point $(Em_{t+1}, \sigma(m_{t+1}))$ enters the IIJ feasible region, but this point does not enter our feasible region. This means that taking into account the cost of quadratic derivatives makes the equity premium puzzle even more difficult to solve. For reasonable value of the preference parameters, we also realize through Figure 2.5 that our variance bound never coincides with the IIJ bound. This underlines why squared returns cost should be taken into account in asset pricing models.

Next, we consider a model with state dependence in preferences. Several authors [see e.g., Gordon and St-Amour (2000), Melino and Yang (2003)] have pointed to countercyclical risk aversion as a potential source of misspecification that may account for the equity premium puzzle. It is then noteworthy to check if these models can explain this puzzle when using our variance bound on admissible SDF. We consider Gordon and St-Amour (2000) and Melino and Yang (2003) state-dependent preference model.

The Gordon and St. Amour (2000) stochastic discount factor is of the form

$$m_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\alpha(U_t)} \left(\frac{C_{t+1}}{\theta} \right)^{\alpha(U_t) - \alpha(U_{t+1})},$$

where the coefficient of relative risk aversion depends on a latent state variable U_t and $\frac{C_{t+1}}{\theta}$ is the ratio of next period's level of consumption to a scale parameter θ. For the state variable, we set the transition matrix to[2]

$$\Pi = \begin{bmatrix} 0.9909 & 0.0061 \\ 0.0091 & 0.9939 \end{bmatrix}.$$

Since the frontiers are very close under the two bounds, we find that when the implied Gordon and St-Amour SDF passes the IIJ bound it also passes our variance bound. We report here (see Figure 2.6) only the case: $\alpha = (3.7, 2.23)$, $\theta = 12, 18$. Melino and Yang (2003) generalize the model of Epstein and Zin (1989) by allowing the representative agent to display state-dependent preferences and show that these preferences can add to the explanation of the equity premium puzzle. They consider several state-dependent preference cases: state-dependent Constant Relative

[2]In this matrix, the probability of staying in state 1 is 0.9909 and the probability of staying in sate 2 is 0.9939.

Risk Aversion (CRRA), state-dependent Elasticity of Intertemporal Substitution (EIS) and state dependent subjective discount parameter β. Without loss of generality, we consider the Melino and Yang (2003) stochastic discount factor when the EIS and the subjective discount parameter β are constant:

$$m_{t+1} = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right]^{\frac{1-\alpha(U_t)}{1-\rho}} R_{mt+1}^{\left(\frac{1-\alpha(U_t)}{1-\rho}\right)-1}.$$

In Figure 2.7, we give two examples of SDF bounds. For fixed values of $(\beta, \rho) = (0.98, 3.58)$, the first example shows that for the state-dependent preferences parameter $\alpha = (7.8, 9.4)$, the state-dependent implied SDF passes the HJ bound but does not pass our bound whereas in the second example, $\alpha = (8.8, 9.35)$ produces a SDF which passes both bounds. In the next subsection, we provide insights on the SDF variance bound under a positivity constraint.

2.3 Positivity constraint on the SDF

So far we have ignored the arbitrage restriction that an admissible SDF must be nonnegative. HJ show that when an unconditionally riskless asset exists, it is straightforward to find the HJ minimum variance SDF with a nonnegativity constraint. But they show that this SDF may not be unique. In our case, when the unconditionally riskless asset exists, it can be shown that the minimum variance SDF with positivity constraint is:

$$m^{mvs}(\eta)^{+} = \left(\widetilde{\beta}(\eta)' R + \widetilde{\gamma}(\eta)' R^{(2)}\right)^{+},$$

where $x^{+} = \max(0, x)$ represents the nonnegative part of x. The parameters $\widetilde{\beta}(\eta)$ and $\widetilde{\gamma}(\eta)$ can be computed by solving the nonlinear equations:

$$\begin{aligned} ERm^{mvs}(\eta)^{+} &= l, \\ ER^{(2)}m^{mvs}(\eta)^{+} &= \eta. \end{aligned}$$

These two equations are nonlinear in the parameter vectors $\widetilde{\beta}(\eta)$ and $\widetilde{\gamma}(\eta)$ and the solution $(\widetilde{\beta}(\eta)$, $\widetilde{\gamma}(\eta))$ cannot be represented in terms of matrix manipulations. Similarly to HJ, it can be shown that this solution exists but may not be unique. Once this solution is found, however, it is easy to show that $m^{mvs}(\eta)^+$ has a minimum variance among any admissible SDF in $\mathcal{F}_2^+(\eta)$ where

$$\mathcal{F}_2^+(\eta) = \left\{m \in L^2 : m > 0,\ EmR = l,\ EmR^{(2)} = \eta\right\}.$$

To understand this more clearly, consider any other nonnegative admissible SDF in $\mathcal{F}_2^+(\eta)$ and note that

$$\begin{aligned} E\left[m^{mvs}(\eta)^+ m\right] &= E\left[m\left(\widetilde{\beta}(\eta)' R + \widetilde{\gamma}(\eta)' R^{(2)}\right)^+\right] \\ &\geq \widetilde{\beta}(\eta)' EmR + \widetilde{\gamma}(\eta)' EmR^{(2)} \\ &= \widetilde{\beta}(\eta)' Em^{mvs}(\eta)^+ R + \widetilde{\gamma}(\eta)' Em^{mvs}(\eta)^+ R^{(2)} \\ &= E\left[\left(m^{mvs}(\eta)^+\right)^2\right]. \end{aligned}$$

It follows that

$$Em^2 \geq E\left[\left(m^{mvs}(\eta)^+\right)^2\right]$$

and

$$\sigma(m) \geq \sigma\left(m^{mvs}(\eta)^+\right).$$

HJ find a similar inequality, but in their framework,

$$\sigma(m) \geq \sigma\left(m_{HJ}^+\right),$$

for any admissible SDF in $\mathcal{F}_1^+ = \left\{m \in L^2 : m > 0, EmR = l\right\}$, where $m_{HJ}^+ = \left(\beta'_{HJ}R\right)^+$. Since $\mathcal{F}_2^+(\eta) \subset \mathcal{F}_1^+$, it is straightforward to show that:

$$\sigma\left(m^{mvs}(\eta)^+\right) \geq \sigma\left(m_{HJ}^+\right).$$

Therefore, when the riskless asset exists and if we use a nonnegativity constraint on m, our variance bound also tightens the HJ variance bound. Following the same idea as Hansen and Jagannathan (1991), this result can be generalized to deal with the case in which there is no unconditionally riskless asset. In the rest of this paper, we work without a positivity constraint on admissible SDFs.

Motivated by the duality between the HJ frontier and the Markowitz mean-variance portfolio frontier, we offer, in the next section, a portfolio selection approach based on our minimum variance SDF surface frontier.

3. Portfolio Selection

Markowitz mean-variance analysis is the central tenet of portfolio selection in financial theory. Since any asset pricing model can be represented by a SDF model, a number of papers establish a connection between Markowitz mean-variance analysis and the HJ bound on the SDF volatility. See for example, Campbell, Lo and Mc Kinlay (1997), Nijman and de Roon (2001) and Penaranda and Sentana (2001). The leading assumption in Markowitz mean-variance analysis is that investors are interested in three characteristics of their portfolio: expected payoff, cost, and variance. Under these assumptions, it can be shown that the HJ minimum variance SDF is spanned by two factors and that the Markowitz optimization problem (which entails minimizing the (unit cost) portfolio variance subject to the portfolio expected return) yields an optimal mean-variance portfolio which can be written as a function of the same two factors.

In this section, we assume that investors are not only interested in these three characteristics of their portfolio but they are also interested in the cost of their squared portfolio return.

We first use these four characteristics to decompose the SDF as a function of factors which we use to provide a portfolio selection approach. Our main contribution is to show that the recent portfolio selection approach based on mean-variance-skewness may miss an important factor.

3.1 A SDF decomposition

Let $\mathcal{P}_N$ be the set of payoffs which is given by the linear span of primitive assets and $\mathcal{G}_N$ be the set of the payoffs which is given by the linear span of "squared" primitive assets $R^{(2)}$. The elements of $\mathcal{P}_N$ will be of the form

$$\sum_{i=1}^{N} \omega_i R_i.$$

Similarly the elements of $\mathcal{G}_N$ will be of the form

$$\sum_{i<j}^{N} \omega_{ij} R_i R_j.$$

$\mathcal{P}_N, \mathcal{G}_N$ are closed linear subspaces of L^2, where L^2 denotes the Hilbert space under the Mean-Square inner product defined as $\langle x, y \rangle = Exy$ and the associated mean-square norm $\langle x, x \rangle^{1/2}$ with x, y $\in L^2$. Assume that investors are interested at least in four characteristics of their portfolio $p = \omega^{'} R$: the (normalized) cost of their portfolio, their portfolio expected payoff value, the variance of their portfolio payoff and the cost of their squared portfolio returns which are given by $C(p) = \omega^{'} l$, $E(p) = \omega^{'} \nu$, $V(p) = \omega^{'} \Omega \omega$ and $\widetilde{C}(p^2)$ respectively.

For convenience, we denote

$$\begin{aligned} \Gamma &= ERR^{'}, \\ \Gamma^{(2)} &= ER^{(2)} R^{(2)'}. \end{aligned}$$

Under the law of one price, we can interpret both $C(.)$, $E(.)$ as linear functionals that map the elements of $\mathcal{P}_N$ into the real line. In this sense, the Riesz representation theorem says that there exists two unique elements of $\mathcal{P}_N$, p^+ and p^{++}, such that:

$$C(p) = E(p^+ p) \quad \forall p \in \mathcal{P}_N,$$

with

$$p^+ = a^{+'} R, \text{ with } a^{+'} = l^{'} \Gamma^{-1}, \tag{3.3}$$

and

$$E(p) = E(p^{++} p) \quad \forall p \in \mathcal{P}_N, \tag{3.4}$$

with

$$p^{++} = a^{++'} R, \text{ with } a^{++'} = \nu^{'} \Gamma^{-1}.$$

Similarly, $\widetilde{C}(.)$ can be viewed as a linear functional that maps the elements of $\mathcal{G}_N$ into the real line. The Riesz representation theorem aigain implies that there exists a unique element of $\mathcal{G}_N$ such that:

$$\widetilde{C}(p) = E(p^* p) \quad \forall p \in \mathcal{G}_N, \tag{3.5}$$

with

$$p^* = a^{*'} R^{(2)},$$

where $a^{*'} = \eta' \left[\Gamma^{(2)}\right]^{-1}$. The following theorem shows that these three vectors p^+, p^{++} and p^* are able to span the minimum variance SDFs.

Theorem 3.1 *For any $\eta \neq \eta^*$, the minimum variance stochastic discount factor $m^{mvs}(\eta, \overline{m})$ can be decomposed as:*

$$m^{mvs}(\eta, \overline{m}) = m_{HJ}(\overline{m}) + cF_3,$$

with $m_{HJ}(\overline{m}) = \overline{m} + aF_1 + bF_2$ where,

$$\begin{aligned} F_1 &= p^+ - Ep^+, \\ F_2 &= p^{++} - EL\left[p^{++} | F_1\right], \\ F_3 &= p^* - EL\left[p^* | F_1, F_2\right] \end{aligned}$$

and

$$\begin{aligned} a &= \frac{l'\Gamma^{-1}l - \overline{m}Ep^+}{Var(p^+)}, \\ b &= \frac{\left(\nu'\Gamma^{-1}l - \overline{m}Ep^{++}\right)}{Cov(F_2, p^{++})} - a\frac{Cov(F_1, p^{++})}{Cov(F_2, p^{++})} \\ c &= \frac{\left(\eta'\left[\Gamma^{(2)}\right]^{-1}\eta - \overline{m}Ep^*\right)}{Cov(F_3, p^*)} - a\frac{Cov(F_1, p^*)}{Cov(F_3, p^*)} - b\frac{Cov(F_2, p^*)}{Cov(F_3, p^*)} \end{aligned}$$

The notation $EL[.|\mathcal{F}]$ indicates the linear regression on $\mathcal{F}$.

We now use this SDF decomposition to provide a portfolio selection approach.

3.2 Application to portfolio choice

It can be shown that the Markowitz portfolio selection approach which consists in minimizing the (unit cost) portfolio risk subject to the portfolio expected return is based on factors p^+ and p^{++}. Markowitz minimizes the portfolio risk subject to the portfolio cost and expected payoff,

$$\begin{aligned} & \min_p \sigma(p). \\ \text{s.t } Ep &= \mu_p, C(p) = 1 \end{aligned}$$

If p^{mv} denotes the optimal solution to the problem above, p^{mv} is the only linear combination of p^+ and p^{++} satisfying the constraints. We consider now a portfolio selection approach based not only on p^+ and p^{++} but also on p^*.

Definition 3.2 *Given the portfolio expected return, the cost of the squared primitive asset η and the cost of squared portfolio return, c^*, the mean-variance-cost optimal portfolio is defined as the solution to the following program.*

$$\begin{aligned} & \min_p \quad \sigma(p) \\ s.t \ Ep &= \mu_p, C(p) = 1 \ , \widetilde{C}(p^2) = c^*, \end{aligned} \tag{3.6}$$

where $C(p)$ represents the cost of the portfolio p and $\widetilde{C}(p^2)$ the cost of the squared portfolio return.

The difference between our optimization problem and the Markowitz optimization problem is that we minimize portfolio risk subject to an additional constraint which takes into account the portfolio skewness. We first solve (3.6) and then show the relationship between our portfolio selection approach and the standard portfolio selection under skewness. If p^{mvs} denotes the optimal solution for problem (3.6), we have,

$$p^{mvs} = \alpha_1 p^+ + \alpha_2 F_2 + \alpha_3 F_3,$$

where, α_1, α_2 and α_3 are determined by equations below:

$$\begin{aligned} \alpha_1 E p^+ + \alpha_2 E F_2 + \alpha_3 E F_3 &= \mu_p, \\ \alpha_1 E\left(p^+ p^+\right) + \alpha_2 E\left(F_2 p^+\right) + \alpha_3 E\left(F_3 p^+\right) &= 1, \\ \alpha_1 E\left(p^+ p^*\right) + \alpha_2 E\left(F_2 p^*\right) + \alpha_3 E\left(F_3 p^*\right) &= c^*. \end{aligned}$$

The variance of p^{mvs} is

$$\sigma^2\left(c^*, \mu_p\right) = \alpha_1^2 Var\left(p^+\right) + \alpha_2^2 Var\left(F_2\right) + \alpha_3^2 Var\left(F_3\right). \tag{3.7}$$

To get the optimal portfolio weights, we have,

$$p^{mvs} = \omega_p^{'} R = R^{'} \omega_p.$$

Thus, premultiplying p^{mvs} by R and taking the expectation, we deduce

$$\omega_p = \alpha_1 \Gamma^{-1} E\left(R p^+\right) + \alpha_2 \Gamma^{-1} E\left(R F_2\right) + \alpha_3 \Gamma^{-1} E\left(R F_3\right). \tag{3.8}$$

We refer to $\mathcal{E}_1$ as being the set

$$\mathcal{E}_1 = \left\{\left(\mu_p, c^*, \sigma\left(p^{mvs}\right) : \left(\mu_p, c^*\right) \in \mathbb{R}^2\right)\right\},$$

where $\mathcal{E}_1$ represents the mean-variance-cost surface frontier. For each portfolio p^{mvs} in $\mathcal{E}_1$, we find the corresponding portfolio skewness $s_p = \frac{E\left(p^{mvs} - \mu_p\right)^3}{\sigma^3\left(c^*, \mu_p\right)}$. If we refer $\mathcal{E}_2$ as being the set

$$\mathcal{E}_2 = \left\{\left(\mu_p, s_p, \sigma\left(p^{mvs}\right) : \left(\mu_p, s_p\right) \in \mathbb{R}^2\right)\right\},$$

then $\mathcal{E}_2$ represents the mean-variance-skewness surface. Now, consider the payoff:

$$R^{mvs} = \frac{m^{mvs}\left(\eta, \overline{m}\right)}{C\left(m^{mvs}\left(\eta, \overline{m}\right)\right)}.$$

It follows that:

$$C\left(R^{mvs}\right) = 1. \tag{3.9}$$

If c^*_{mvs} denotes the cost of $\left(R^{mvs}\right)^2$, it can be shown that,

$$c^*_{mvs} = \left(\overline{m}^2 + \sigma^2\left[m^{mvs}\left(\overline{m}, \eta\right)\right]\right) E\left(R^{mvs}\right)^3.$$

Using (3.9), we show that

$$\frac{|1/\overline{m} - \mu_p|}{\sigma(p)} \leq \frac{|1/\overline{m} - ER^{mvs}|}{\sigma(R^{mvs})} = \frac{\sigma[m^{mvs}(\overline{m},\eta)]}{Em^{mvs}(\overline{m},\eta)} \leq \frac{\sigma(m)}{Em} \; \forall m \in \mathcal{F}_2(\overline{m},\eta). \tag{3.10}$$

Inequality (3.10) shows that no other portfolio with the same mean and same squared return cost has smaller variance than R^{mvs}. The return R^{mvs} belongs to the mean-variance-cost surface $\mathcal{E}_1$.

Theorem 3.3 *R^{mvs} is mean-variance-cost efficient. That is no other portfolio with the same squared portfolio cost and same mean has smaller variance.*

If we consider the return associated with the HJ minimum variance SDF, that is:

$$R^{mv} = \frac{m_{HJ}(\overline{m})}{C(m_{HJ}(\overline{m}))},$$

it can be shown that

$$\frac{|1/\overline{m} - ER^{mv}|}{\sigma(R^{mv})} = \frac{\sigma[m_{HJ}(\overline{m})]}{Em_{HJ}(\overline{m})}.$$

By proposition 2.1, we have:

$$\frac{\sigma[m_{HJ}(\overline{m})]}{Em_{HJ}(\overline{m})} < \frac{\sigma[m^{mvs}(\overline{m},\eta)]}{Em^{mvs}(\overline{m},\eta)}.$$

Therefore, the following inequality holds,

$$\frac{|1/\overline{m} - ER^{mv}|}{\sigma(R^{mv})} < \frac{|1/\overline{m} - ER^{mvs}|}{\sigma(R^{mvs})}.$$

The left hand side of (3.10) represents the portfolio Sharpe ratio under the assumption that the risk-free return exists. If the risk-free return (R_F) exists, that is $R_F = 1/\overline{m}$, R^{mvs} has a higher ratio than R^{mv}. In the light of this inequality and theorem 3.3, it is important to define in our setting the mean-variance-cost tangency portfolio.

Definition 3.4 *The mean-variance-cost tangency portfolio is the portfolio with the maximum Sharpe ratio of all possible portfolios with identical squared portfolio cost.*

We now investigate how the portfolio skewness affects the squared portfolio cost. To see how this cost is a function of the portfolio skewness, consider the linear regression of p^2 on p,

$$p^2 = Ep^2 + \frac{Cov\left(p, p^2\right)}{Var\left(p\right)}\left(p - Ep\right) + \upsilon.$$

The cost of the squared portfolio return can be written as:

$$\begin{aligned} c^* &= Em^{mvs}p^2 \\ &= \overline{m}Ep^2 + \left(1 - \overline{m}\mu_p\right)\left[2\mu_p + \sigma_p s_p\right] + Cov\left(\upsilon, m^{mvs}\right). \end{aligned}$$

with $s_p = \frac{E\left(p-\mu_p\right)^3}{\sigma_p^3}$. Through this expression, the cost of the squared portfolio return is a function of the portfolio skewness. Therefore it is reasonable to put forward the relationship between our portfolio selection approach and the standard portfolio selection under skewness. The later consists in minimizing the portfolio risk subject to the portfolio expected payoff and skewness. We formalize the standard portfolio selection approach as follows:

$$\begin{array}{cc} \min & \sigma\left(p\right), \\ C\left(p\right) = 1 & \\ Ep = \mu_p & \\ \frac{E\left(p-\mu_p\right)^3}{\sigma_p^3} = s_p & \end{array} \tag{3.11}$$

where s_p represents the portfolio skewness. Apart from the two portfolio constraints: expected return and portfolio cost, it can be observed that the difference between our optimization problem and standard portfolio selection under skewness comes from the third constraint. In standard portfolio selection under skewness approach, the third constraint is on the portfolio skewness whereas in our approach, the third constraint is on the cost of the squared portfolio return. De Athayde and Flores (2004) find a general solution to the problem (3.11). It is thus important to study the relationship between the two problems in (3.6) and (3.11). We will say that problems (3.6) and (3.11) are observationally equivalent if and only if any optimal solution to the problem (3.6) is also optimal to the problem (3.11) and vice versa.

To derive necessary and sufficient conditions that make our portfolio selection approach observationally equivalent to standard portfolio selection under skewness, we first show:

Proposition 3.5 *Consider a portfolio p and the linear regression of p^2 on p :*

$$p^2 = EL\left[p^2|p\right] + \upsilon.$$

Then, $Cov\left(\upsilon, m^{mvs}\right) = 0$ for all portfolios p if and only if the components of the price, $\eta = \overline{\eta}$, of $R^{(2)}$ are:

$$\overline{\eta}_{ii} = \overline{m}ER_i^2 + \left(1 - \overline{m}ER_i\right)\left[2ER_i + \frac{E\left(R_i - ER_i\right)^3}{Var\left(R_i\right)}\right] \quad \textit{for } i=1,\ldots,n.$$

and

$$\begin{aligned}\overline{\eta}_{ij} &= \frac{1}{2}\overline{m}\left(ER_i^2 + ER_j^2 + 2ER_iR_j\right) + \\ &\frac{\left[1 - \frac{1}{2}\left(ER_i + ER_j\right)\overline{m}\right] Cov\left(\left(R_i + R_j\right), \left(R_i + R_j\right)^2\right)}{\left[Var\left(R_i\right) + Var\left(R_j\right) + 2Cov\left(R_i, R_j\right)\right]} - \frac{1}{2}\left(\overline{\eta}_{ii} + \overline{\eta}_{jj}\right)\end{aligned}$$

for $i \neq j$.

Proposition 3.6 gives necessary and sufficient conditions to get standard portfolio selection under skewness, that is a maximum skewness portfolio [see de Athayde and Flores (2004)].

Proposition 3.6 *If $\mu_p \neq 1/\overline{m}$, consider a portfolio p and the linear regression of p^2 on p :*

$$p^2 = EL\left[p^2|p\right] + \upsilon.$$

Problem (3.11) and (3.6) are observationally equivalent if and only if $Cov\left(\upsilon, m^{mvs}\right) = 0$ for any portfolio p

PROOF. If $Cov\left(\upsilon, m^{mvs}\right) = 0$ we have,

$$c^* = \overline{m}\left(\sigma_p^2 + \mu_p^2\right) + \left(1 - \overline{m}\mu_p\right)\left[2\mu_p + \sigma_p s_p\right]. \tag{3.12}$$

This equation is equivalent to

$$\overline{m}\sigma_p^2 + \sigma_p s_p\left(1 - \overline{m}\mu_p\right) + \left(2\mu_p - \overline{m}\mu_p^2 - c^*\right) = 0.$$

From (3.12), it is obvious to show that

$$\left\{p : Ep = \mu_p \text{ and } \frac{E\left(p - \mu_p\right)^3}{\sigma_p^3} = s_p\right\}$$

and

$$\left\{p : Ep = \mu_p \text{ and } \widetilde{C}\left(p^2\right) = c^*\right\}$$

are equivalent. Therefore problem (3.11) and (3.6) are observationally equivalent.

In other respects, assume that (3.6) and (3.11) are observationally equivalent. Thus problems (3.6) and (3.11) produce an identical solution. This implies that problem (3.11) can be used to compute the cost of the squared portfolio return. This is possible only if $Cov\left(v, m^{mvs}\right) = 0$. ■

This proposition shows how our portfolio selection approach generalizes standard portfolio selection under skewness and suggests that standard portfolio selection under skewness implicitly assume that the covariance of m^{mvs} with v is null for any portfolio p.

We assume that μ_p and s_p are known and give a simple methodology to get a maximum skewness portfolio solution to problem 3.11.

- First, under the assumption: $Cov\left(v, m^{mvs}\right) = 0$, compute the cost of $R^{(2)}$, $\eta = \overline{\eta}$, using proposition 3.5.

- Second, given the portfolio skewness and expected return, compute c^* as folows: The discriminant of equation (3.12) is

$$\Delta = s_p^2\left(1 - \overline{m}\mu_p\right)^2 - 4\overline{m}\left(2\mu_p - \overline{m}\mu_p^2 - c^*\right).$$

Assuming that $\Delta > 0$, this equation produces two solutions:

$$\sigma_p = \frac{-s_p\left(1 - \overline{m}\mu_p\right) - \sqrt{\Delta}}{2\overline{m}} \text{ or } \sigma_p = \frac{-s_p\left(1 - \overline{m}\mu_p\right) + \sqrt{\Delta}}{2\overline{m}}. \tag{3.13}$$

From (3.13),

$$\sigma_p^2\left(c^*, \mu_p\right) = \frac{\left[s_p\left(1 - \overline{m}\mu_p\right) + \sqrt{\Delta}\right]^2}{\left(2\overline{m}\right)^2} \text{ or } \sigma_p^2\left(c^*, \mu_p\right) = \frac{\left[s_p\left(1 - \overline{m}\mu_p\right) - \sqrt{\Delta}\right]^2}{\left(2\overline{m}\right)^2}. \tag{3.14}$$

If ($s_p > 0$ and $1 - \overline{m}\mu_p > 0$) or ($s_p < 0$ and $1 - \overline{m}\mu_p < 0$), then according to (3.7), the minimum variance portfolio is:

$$\sigma_p^2\left(c^*, \mu_p\right) = \alpha_1^2 Var\left(p^+\right) + \alpha_2^2 Var\left(F_2\right) + \alpha_3^2 Var\left(F_3\right) = \frac{\left[s_p\left(1 - \overline{m}\mu_p\right) - \sqrt{\Delta}\right]^2}{\left(2\overline{m}\right)^2} \tag{3.15}$$

with,

$$\begin{aligned} \alpha_1 &= 1/C\left(p^+\right), \\ \alpha_2 &= A_1 - A_2 c^*, \\ \alpha_3 &= A_3 c^* - A_4, \end{aligned}$$

where A_1, A_2, A_3 and A_4 are known parameters. Equation (3.15) is equivalent to

$$\begin{aligned} \frac{\left[s_p\left(1-\overline{m}\mu_p\right)-\sqrt{\Delta}\right]^2}{\left(2\overline{m}\right)^2} &= \alpha_1^2 Var\left(p^+\right) + \left[A_1^2 + A_2^2 c^{*2} - 2A_1A_2c^*\right] Var\left(F_3\right) + \quad (3.16) \\ & \left[A_3^2 + A_4^2 c^{*2} - 2A_4A_3c^*\right] Var\left(F_3\right). \end{aligned}$$

This equation can be rewritten in terms of Δ. There might be more than one solution to this equation. Choose the solution Δ that yields a smaller variance and use this Δ to compute c^*. The same methodology can be repeated if ($s_p < 0$ and $1 - \overline{m}\mu_p > 0$) or ($s_p > 0$ and $1 - \overline{m}\mu_p < 0$).

- Once c^* is computed, (3.16) gives the minimum variance to problem (3.11).

In the next section, we illustrate the portfolio selection and investigate empirically whether $Cov\left(v, m^{mvs}\right) \neq 0$.

4. Portfolio Selection: Empirical Illustration

To give an empirical illustration of our portfolio selection approach, we need to know the squared primitive assets cost. To compute this cost, we assume that the joint process of the SDF and asset returns is lognormal. This distribution is flexible and allows for skewness. It is often used to characterize asset probability models. For example, many asset pricing tests assume that the process SDF-asset returns is conditionally jointly lognormal. Diffusion models imply a locally lognormal distribution. The next proposition gives the squared primitive asset cost when the joint process of the SDF and asset returns is lognormal.[3]

[3] See the proof of proposition 4.1 in the Appendix.

Proposition 4.1 *Given a stochastic discount factor m, consider a set of N primitive assets. Assume that the joint process $(Log(m), Log(R))$ follows a multivariate normal distribution. thus the components of η are of the form:*

$$\begin{aligned} \eta_{ij} &= E(mR_iR_j) \\ &= \frac{1}{\overline{m}}\frac{ER_iR_j}{ER_iER_j}\ \forall i,j. \end{aligned}$$

To gauge the empirical importance of squared portfolio cost in portfolio selection, we collect daily asset returns from the Datastream data base for the sample period from January 2002 to June 2002. This data set consists of the daily returns of four highly liquid stocks: General Motors, Cisco Systems, Boeing and Ford Motors. Over the same period, we extract the U.S. 3 month Treasury-Bill rate (risk-free rate). The estimated U.S 3 month Treasury-Bill expected return is 1.0495. Table 2.1 reveals that Boeing has the lowest expected return and highest positive skewness, while Cisco Systems had a negative skewness. We use (3.6) to find the optimal portfolio. Figure 2.8 illustrates the Mean-Variance-Cost surface $\mathcal{E}_1$ and the associated Mean-Variance-Skewness surface $\mathcal{E}_2$. Slicing the surface at any level of squared portfolio cost, we get the familiar positively sloping portion of the mean-variance frontier. In the standard mean-variance analysis there is a single efficient risky-asset portfolio, but in our setting, there are multiple efficient portfolios. The mean-variance-skewness surfaces reveals that the squared portfolio cost and the portfolio skewness have a significant impact on the portfolio mean-variance frontier (this can be seen more clearly in Figure 2.10). Figure 2.10 shows how small changes in the cost of the squared portfolio return have a great impact on the portfolio mean-variance frontier. This indicates that the cost of the squared portfolio will significantly impact the tangency portfolio. Notice that at any level of the squared portfolio cost, we get the positively sloping portion of the mean-variance frontier. But at any level of the portfolio skewness, see the M-V-S surface, we do not have the usual positively sloping portion of the mean-variance frontier. This intuitively shows that our approach is not observationally equivalent to standard portfolio selection under skewness. However, from proposition 3.6 our approach is observationally equivalent to standard portfolio selection under skewness. Figure 2.9 illustrates the mean-variance-skewness surface when $Cov(m^{mvs}, v) = 0$. Through this Figure, at any level of the portfolio skewness, varying the portfolio mean produces the usual positively sloping portion of the mean variance frontier.[4] Figure 2.11 illustrates the implied covariance of the SDF with the residuals obtained when regressing the squared portfolio on the portfolio itself. Figure 2.11 provides empirical evidence that this covariance is different from zero and negative.

5. Conclusions and Extensions

In this paper, we derive a new variance bound on any admissible SDF that prices correctly a set of primitive assets and quadratic payoffs of the same primitive assets. Our bound tightens the HJ bound by an additional component which is a function of the squared primitive asset cost and asset co-skewness. We give necessary and sufficient conditions to get the well-known HJ bound. Using the Standard and Poors 500 stock index and commercial paper index from 1889 to 1994, we illustrate our volatility bound and show empirically that when the SDF mean is in the neighborhood of 1, our variance bound is twice as large as the HJ bound. We also found that the SDF implied from the consumption based models such as Epstein and Zin (1989) state non-separable preferences model passes the HJ bound for a particular values of the relative risk aversion coefficient but does not pass our variance bound making the equity premium puzzle even more difficult to solve.

Motivated by the duality between HJ bound and the Markowitz mean-variance analysis, we offer a portfolio selection approach based on factors that span our minimum variance SDF. We show that our portfolio selection approach generalizes standard portfolio selection under skewness which consists in minimizing the portfolio risk subject to the portfolio expected payoff and skewness. We use daily asset returns to illustrate empirically our findings. To proceed to our illustration, we assume that the process SDF-asset returns is jointly lognormal. This allows us to compute the squared primitive asset cost and then illustrate our portfolio selection approach. Empirical results suggest that the cost of the squared portfolio return and the portfolio skewness have a significant impact on the portfolio mean-variance frontier.

Since Bekaert and Liu (2004) and others use conditional information to tighten the HJ bound, it

[4] de Athayde and Flores (2004) use problem (3.11) to illustrate the Mean-Variance-Skewness surface.

would be of interest to examine how conditional information might be used to tighten our variance bound. In light of Hansen and Jagannathan (1997), it appears natural to develop a SDF-based distance measure for asset pricing models under this higher-moments framework. We leave these issues for future research.

6. Appendix: Proofs

Proof of Theorem 3.1. If $\eta = \eta^*$, $m^{mvs}(\eta, \overline{m})$ collapses to the HJ stochastic discount factor. Let us assume that $\eta \neq \eta^*$ and assume that the minimum variance stochastic discount factor $m^{mvs}(\eta, \overline{m})$ can be decomposed as:

$$m^{mvs}(\eta, \overline{m}) = \overline{m} + aF_1 + bF_2 + cF_3$$

where,

$$\begin{aligned} F_1 &= p^+ - Ep^+, \\ F_2 &= p^{++} - EL\left[p^{++}|F_1\right], \\ F_3 &= p^* - EL\left[p^*|F_1, F_2\right]. \end{aligned}$$

First,

$$Cov\left(m^{mvs}(\eta, \overline{m}), p^+\right) = aCov\left(F_1, p^+\right) = aVar\left(p^+\right).$$

Replacing p^+ by its expression (see (3.3)), we get:

$$\begin{aligned} Cov\left(m^{mvs}(\eta, \overline{m}), p^+\right) &= Em^{mvs}(\eta, \overline{m})\, p^+ - Em^{mvs}(\eta, \overline{m})\, Ep^+ \\ &= l'\Gamma^{-1}l - \overline{m}Ep^+. \end{aligned}$$

Therefore,

$$a = \frac{l'\Gamma^{-1}l - \overline{m}Ep^+}{Var\left(p^+\right)}.$$

Second,

$$Cov\left(m^{mvs}(\eta, \overline{m}), p^{++}\right) = aCov\left(F_1, p^{++}\right) + bCov\left(F_2, p^{++}\right).$$

Replacing p^{++} by its expression (see (3.4)), we get

$$Cov\left(m^{mvs}(\eta, \overline{m}), p^{++}\right) = Em^{mvs}(\eta, \overline{m})\, p^{++} - Em^{mvs}(\eta, \overline{m})\, Ep^{++} = \nu'\Gamma^{-1}l - \overline{m}Ep^{++}.$$

Consequently,

$$aCov\left(F_1,p^{++}\right)+bCov\left(F_2,p^{++}\right)=\nu'\Gamma^{-1}l-\overline{m}Ep^{++},$$

which implies

$$b=\frac{\left(\nu'\Gamma^{-1}l-\overline{m}Ep^{++}\right)-aCov\left(F_1,p^{++}\right)}{Cov\left(F_2,p^{++}\right)}.$$

Third,

$$Cov\left(m^{mvs}\left(\eta,\overline{m}\right),p^*\right)=aCov\left(F_1,p^*\right)+bCov\left(F_2,p^*\right)+cCov\left(F_3,p^*\right)$$

Replacing p^* by its expression (see (3.5)), we get:

$$Cov\left(m^{mvs}\left(\eta,\overline{m}\right),p^*\right)=Em^{mvs}\left(\eta,\overline{m}\right)p^*-\overline{m}Ep^*=\eta'\left[\Gamma^{(2)}\right]^{-1}\eta-\overline{m}Ep^*$$

Consequently,

$$c=\frac{\left(\eta'\left[\Gamma^{(2)}\right]^{-1}\eta-\overline{m}Ep^*\right)-aCov\left(F_1,p^*\right)-bCov\left(F_2,p^*\right)}{Cov\left(F_3,p^*\right)}.$$

It is obvious to show that the HJ stochastic discount factor can be written as:

$$m_{HJ}=\overline{m}+aF_1+bF_2.$$

This ends the proof. ■

PROOF OF PROPOSITION 3.5. The linear regression of r_p^2 on r_p gives

$$r_p^2=Er_p^2+\frac{Cov\left(r_p,r_p^2\right)}{Var\left(r_p\right)}\left(r_p-Er_p\right)+v$$

for any portfolio $r_p=\omega'R$. Therefore, the cost of the squared portfolio return is

$$c^*=\overline{m}Er_p^2+\frac{Cov\left(r_p,r_p^2\right)}{Var\left(r_p\right)}\left(1-\overline{m}Er_p\right)+Cov\left(v,m^{mvs}\right).$$

If $Cov\left(v,m^{mvs}\right)=0$, we have:

$$c^*=\overline{m}Er_p^2+\frac{Cov\left(r_p,r_p^2\right)}{Var\left(r_p\right)}\left(1-\overline{m}Er_p\right) \tag{A1}$$

Then if $\omega_i = 1$ and $\omega_j = 0$ for $j \neq i$, equation (A1) implies

$$\overline{\eta}_{ii} = \overline{m}ER_i^2 + (1 - \overline{m}ER_i)\frac{Cov\left(R_i, R_i^2\right)}{Var\left(R_i\right)} \text{ for i=1,...,n.}$$

For $\omega_i = \frac{1}{2}$, $\omega_j = \frac{1}{2}$ and $\omega_k = 0$ for $k \neq i$ and $k \neq j$. If we decompose the left hand side of (A1), we have

$$\begin{aligned} c^* &= Em\left(\frac{1}{2}R_i + \frac{1}{2}R_j\right)^2 \\ &= \frac{1}{4}Em\left(R_i^2 + R_j^2 + 2R_iR_j\right) \\ &= \frac{1}{4}\left[EmR_i^2 + EmR_j^2 + 2EmR_iR_j\right] \\ &= \frac{1}{4}\overline{\eta}_{ii} + \frac{1}{4}\overline{\eta}_{jj} + \frac{1}{2}\overline{\eta}_{ij} \end{aligned} \tag{A2}$$

where

$$\overline{\eta}_{ij} = EmR_iR_j$$

We also decompose the right hand side of (A1) and equate the resulting expression with (A2) to get

$$\begin{aligned} \overline{\eta}_{ij} &= \frac{1}{2}\overline{m}\left(ER_i^2 + ER_j^2 + 2ER_iR_j\right) + \\ &\quad \frac{\left[1 - \frac{1}{2}\left(ER_i + ER_j\right)\overline{m}\right] Cov\left(\left(R_i + R_j\right), \left(R_i + R_j\right)^2\right)}{\left[Var\left(R_i\right) + Var\left(R_j\right) + 2Cov\left(R_i, R_j\right)\right]} - \\ &\quad \frac{1}{2}\left(\overline{\eta}_{ii} + \overline{\eta}_{jj}\right) \end{aligned}$$

for $i \neq j$. But,

$$\begin{aligned} c^* &= Em^{mvs}r_p^2 \\ &= Em^{mvs}\left(\omega' R\right)^2 \\ &= Em^{mvs}\left(\omega^{(2)'}R^{(2)}\right) \\ &= \omega^{(2)'}Em^{mvs}R^{(2)} \\ &= \omega^{(2)'}\eta. \end{aligned}$$

Then,

$$c^* = \omega^{(2)'}\overline{\eta} = \overline{m}Er_p^2 + \frac{Cov\left(r_p, r_p^2\right)}{Var\left(r_p\right)}\left(1 - \overline{m}Er_p\right).$$

Conversely, let us assume that $\eta = \overline{\eta}$. We have,

$$\begin{aligned} c^* &= Em^{mvs}r_p^2 \\ &= Em^{mvs}\omega^{(2)'}R^{(2)} \\ &= \omega^{(2)'}Em^{mvs}R^{(2)} \\ &= \omega^{(2)'}\overline{\eta}. \end{aligned}$$

But, we know that

$$\omega^{(2)'}\overline{\eta} = \overline{m}Er_p^2 + \frac{Cov\left(r_p, r_p^2\right)}{Var\left(r_p\right)}\left(1 - \overline{m}Er_p\right).$$

Therefore,

$$c^* = \overline{m}Er_p^2 + \frac{Cov\left(r_p, r_p^2\right)}{Var\left(r_p\right)}\left(1 - \overline{m}Er_p\right).$$

But we know that:

$$c^* = \overline{m}Er_p^2 + \frac{Cov\left(r_p, r_p^2\right)}{Var\left(r_p\right)}\left(1 - \overline{m}Er_p\right) + Cov\left(v, m^{mvs}\right).$$

Consequently,

$$Cov\left(m^{mvs}, v\right) = c^* - \overline{m}Er_p^2 - \frac{Cov\left(r_p, r_p^2\right)}{Var\left(r_p\right)}\left(1 - \overline{m}Er_p\right) = 0.$$

■

Proof of Proposition 4.1. Assume that the joint process (m, R) is lognormal. This means that

$$\begin{bmatrix} Log\left(m\right) \\ Log\left(R\right) \end{bmatrix} \rightsquigarrow N\left[\begin{bmatrix} \mu_m \\ \mu_r \end{bmatrix}, \begin{bmatrix} \sigma_m^2 & \Sigma_{mr} \\ \Sigma_{mr} & \Sigma_r \end{bmatrix}\right].$$

We know that

$$EmR_i = 1 \; \forall i.$$

Let compute

$$\eta_{ij} = EmR_iR_j.$$

Therefore,

$$Log\,(mR_iR_j) = Log\,(m) + Log\,(R_i) + Log\,(R_j)\,.$$

Let μ_m and σ_m^2 denote the first two moments of $Log\,(m)$ and μ_i and σ_i^2 denote the first two moments of $Log\,(R_i)$. As result,

$$\begin{aligned}
\eta_{ij} &= \exp\left[\mu_m + \mu_i + \mu_j + \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2 + \sigma_m^2 + 2\sigma_{ij} + 2\sigma_{im} + 2\sigma_{jm}\right)\right] \\
&= \exp\left[\mu_m + \frac{1}{2}\sigma_m^2 + \mu_i + \frac{1}{2}\sigma_i^2 + \sigma_{im}\right]\exp\left[\mu_j + \frac{1}{2}\left(\sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm}\right)\right] \\
&= (EmR_i)\exp\left[\mu_j + \frac{1}{2}\left(\sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm}\right)\right] \\
&= \exp\left[\mu_j + \frac{1}{2}\left(\sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm}\right) + \mu_m + \frac{1}{2}\sigma_m^2\right]\exp\left[-\mu_m - \frac{1}{2}\sigma_m^2\right] \\
&= \exp\left[\mu_m + \frac{1}{2}\sigma_m^2 + \mu_j + \frac{1}{2}\sigma_j^2 + \sigma_{jm}\right]\exp\left[\frac{1}{2}\left(2\sigma_{ij}\right)\right]\exp\left[-\mu_m - \frac{1}{2}\sigma_m^2\right] \\
&= E\,(mR_j)\exp\left[\frac{1}{2}\left(2\sigma_{ij}\right)\right]\exp\left[-\mu_m - \frac{1}{2}\sigma_m^2\right]
\end{aligned}$$

But $E\,(mR_j) = 1$. Consequently,

$$\begin{aligned}
\eta_{ij} &= \exp\left[\frac{1}{2}\left(2\sigma_{ij}\right)\right]\exp\left[-\mu_m - \frac{1}{2}\sigma_m^2\right] \\
&= \frac{1}{m}\exp\left[\frac{1}{2}\left(2\sigma_{ij}\right)\right] \\
&= \frac{1}{m}\exp\left[\mu_i + \mu_j + \frac{1}{2}\left(\sigma_i^2 + \sigma_j^2 + 2\sigma_{ij}\right)\right]\exp\left[-\mu_i - \frac{1}{2}\sigma_i^2\right]\exp\left[-\mu_j - \frac{1}{2}\sigma_j^2\right] \\
&= \left[\frac{1}{m}E\,(R_iR_j)\right]\left[\frac{1}{ER_iER_j}\right] \\
&= \frac{1}{m}\frac{E\,(R_iR_j)}{ER_iER_j}.
\end{aligned}$$

■

References

[1] Athayde, G M. de., and R. G. Flores J.r, 2004, "Finding a Maximum Skewness Portfolio-a General Solution to Three Moments Portfolio Choice," Journal of Economic Dynamics & Control, 28, 1335-1352.

[2] Barone-Adesi, G., P. Gagliardini, and G. Urga, 2004, "Testing Asset Pricing Models with Co-skewness, Journal of Business Economics and Statistic," Forthcoming.

[3] Bekaert, G., and J. Liu, 2004, "Conditioning Information and Variance Bounds on Pricing Kernels," The Review of Financial Studies, 17(2) 339-378.

[4] Campbell, J., and J. H. Cochrane, 1999, "By Force of Habit: A Consumption Based Explanation of Aggregate Stock Market Behavior," Journal of Political Economy, 107 (2), 205-251.

[5] Campbell, J., A. Lo, and A. C. MacKinlay, 1997, The Econometrics of Financial Market, Princeton University Press, Princeton, N.J.

[6] Chabi-Yo F., D. P. J. Leisen, and E. Renault, 2003, "Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing," Working Paper, Departement de Sciences Economiques, Universite de Montreal.

[7] Cochrane, J. H., 2001, Asset Pricing. Princeton University Press.

[8] Dittmar, R. F., 2002, "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from Cross section of Equity Returns," Journal of Finance, 57, 368-403.

[9] Epstein, L., and Zin S. 1989, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theorical Framework," Econometrica, 57, 937-969.

[10] Gamba, A., and F. Rossi, 1998a, "A Three-Moment Based Portfolio Selection Model," WP, Universita di Venezia.

[11] Gamba, A., and F. Rossi, 1998b, "Mean-Variance-Skewness Analysis in Portfolio Choice and Capital Markets," Working Paper, Universita di Venezia.

[12] Gordon, S., and P. St-Amour, 2000, "A Preference Regime Model of Bull and Bear Markets," The American Economic Review, 90(4), 1019-1033.

[13] Hansen, L., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," Journal of Political Economy, 99, 225-262.

[14] Hansen, L., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," Journal of Finance, 52(2), 557-590.

[15] Hansen, L., and S. Richard, 1987, "The role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," Econometrica, 55, 587-613.

[16] Harvey, C. R., and A. Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," The Journal of Finance, LV, 3, 1263-1295.

[17] Harvey, C. R., C. J. Liechty, M. W. Liechty, and P. Muller, 2002, "Portfolio Selection with Higher Moments," Working Paper, Duke University.

[18] Ingersoll, J., 1987, Theory of Financial Decision Making, Rowman & Littlefield Studies in Financial Economics.

[19] Kan, R., and G. Zhou, 2003, "A New Variance Bound on the Stochastic Discount Factor," Working Paper, University of Toronto.

[20] Lai, T. Y., 1991, "Portfolio Selection with Skewness: A Multiple-Objective Approach," Review of Quantitative Finance and Accounting, 1, 293-305.

[21] Markowitz, H., 1952, "Portfolio Selection," Journal of Finance, 7 (1), 77-91.

[22] Mehra, R., and E. C. Prescott, 1985, "The Equity Premium: A puzzle," Journal of Monetary Economics, 15, 145-161.

[23] Melino, A., and A. X. Yang, 2003, "State Dependent Preferences Can Explain the Equity Premium Puzzle," Review of Economic Dynamics, 6(2), 806-830.

[24] Nijman, R.T., and F. A. de Roon, 2001, "Testing for Mean-Variance Spanning: A Survey," Working Paper, Department of Finance, Tilburg University.

[25] Penaranda, F., and F. Sentana, 2001, "Spanning Tests in Return and Stochastic Discount Factor Mean-Variance Frontiers: A Unifying Approach," Working Paper, CEMFI, Spain.

[26] Weil, P., 1989, "The Equity Premium Puzzle and The Risk-Free Rate Puzzle," Journal of Monetary Economics, 24, 401-421.

Table 2.1: Asset company return moments

Asset Company	Portfolio Variable ω	Expected Return	Variance	Skewness
General Motors	ω_1	1.0011	$0.2853e^{-3}$	0.2835
Cisco System	ω_2	1.0044	$0.3938e^{-3}$	−0.1244
Boeing	ω_3	0.9999	$0.3621e^{-3}$	0.6637
Ford Motors	ω_4	1.0049	$0.3777e^{-3}$	0.5045

Note: The skewness is measured by the third central moment divided by the cube of the standard deviation.

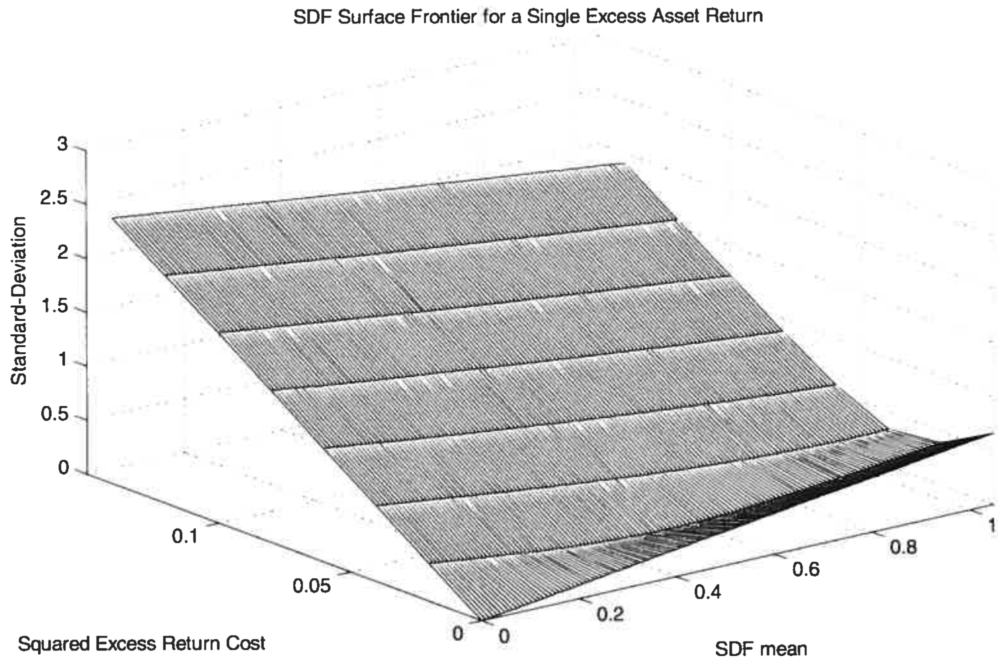


Figure 2.1: **SDF volatility surface frontier with a single excess return:** We use our approach to imply a Mean-Standard Deviation-Cost Surface for Stochastic Discount Factors using the excess simple return of the Standard and Poors 500 stock index over the commercial paper. Annual US data, from 1889 to 1994, are used to compute the SDF variance bound. The SDF feasible region is above this surface.

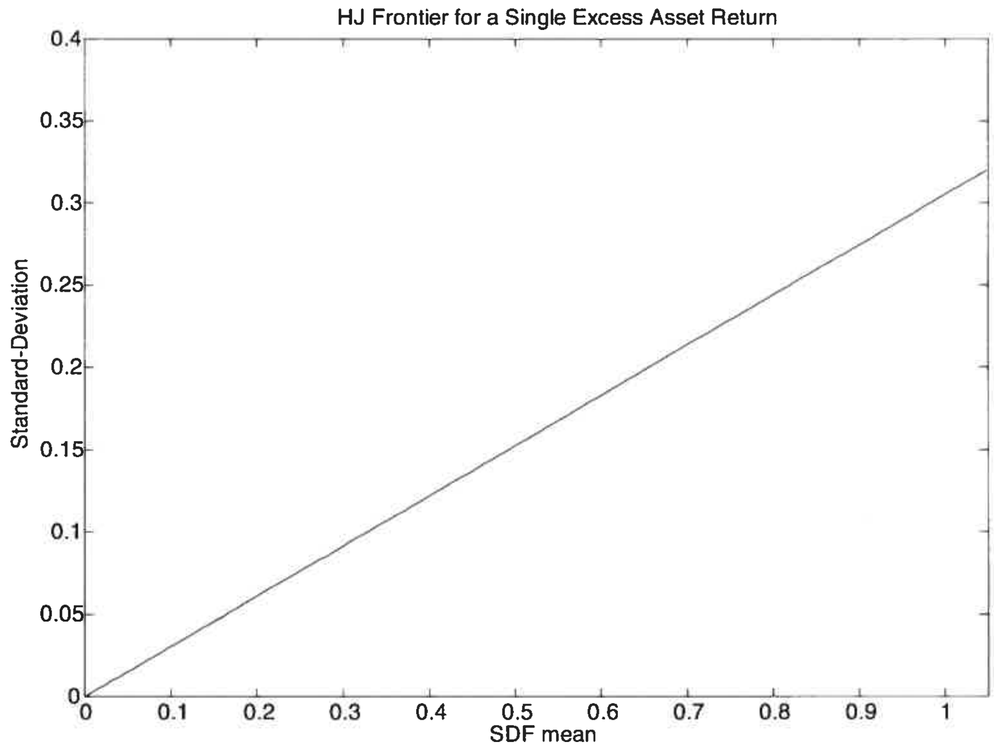


Figure 2.2: **HJ frontier with a single excess return**: We use the HJ approach to imply a Standard Deviation-Mean frontier for Stochastic Discount Factors using the excess simple return of the Standard and Poors 500 stock index over the commercial paper. Annual data from 1889 to 1994 are used to plot this frontier. The SDF feasible region is above this frontier.

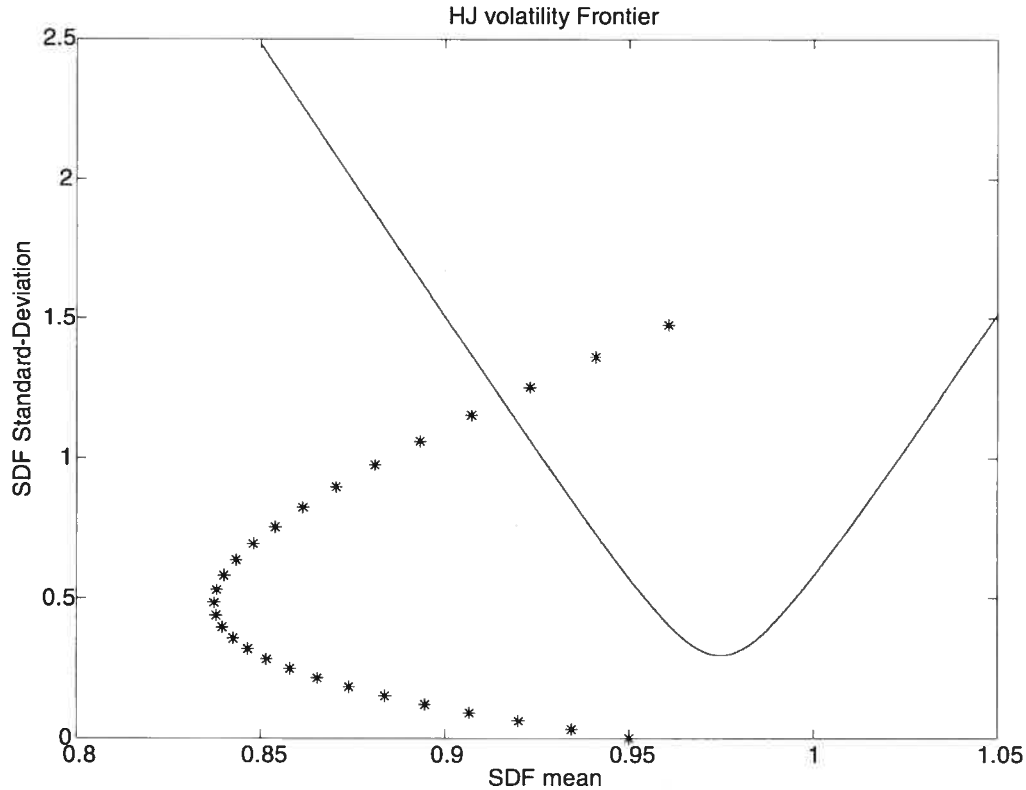


Figure 2.3: **HJ Volatility Frontier:** We imply a Mean-Standard Deviation frontier for Stochastic Discount Factors using the return of the Standard and Poors 500 stock index and the commercial paper . Annual US data, from 1889 to 1994 are used to compute the HJ variance bound. The SDF feasible region is above this frontier. With CRRA preferences we vary exogenously the relative risk aversion coefficient and trace out the resulting pricing kernels in this two-dimensional space. These pricing kernels are represented by the points *.

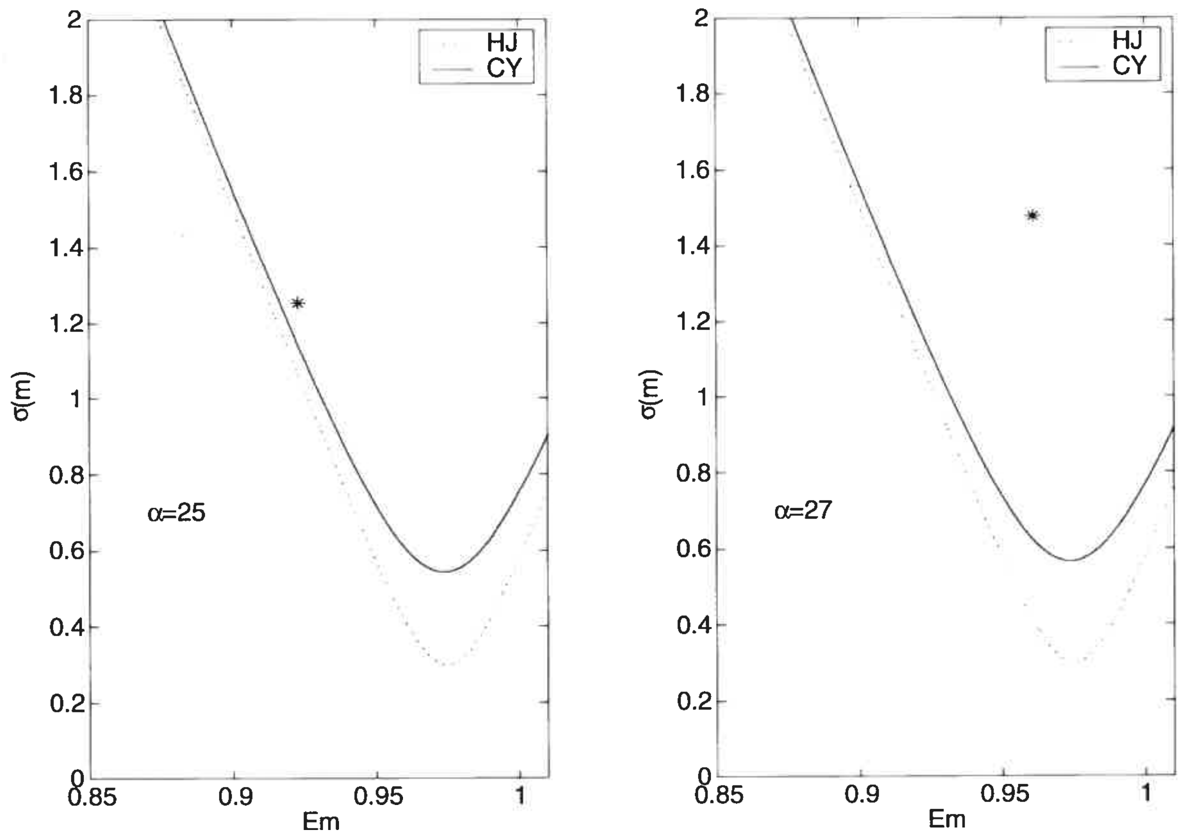


Figure 2.4: **SDF volatility frontier:** HJ represents the Hansen and Jagannathan volatility frontier and CY represents our volatility frontier. For each α, we find $\eta = EmR^{(2)}$ and trace out the point $(\overline{m}, \sigma(m^{mvs}(\overline{m}, \eta)))$ in a two-dimensional space. We also plot the point $(Em_{t+1}, \sigma(m_{t+1}))$ where m_{t+1} represents the SDF obtained in the investor optimization problem with CRRA preferences. We use the return of the Standard and Poors 500 stock index over the commercial paper. Annual US data, from 1889 to 1994, are used to compute the SDF variance bound.

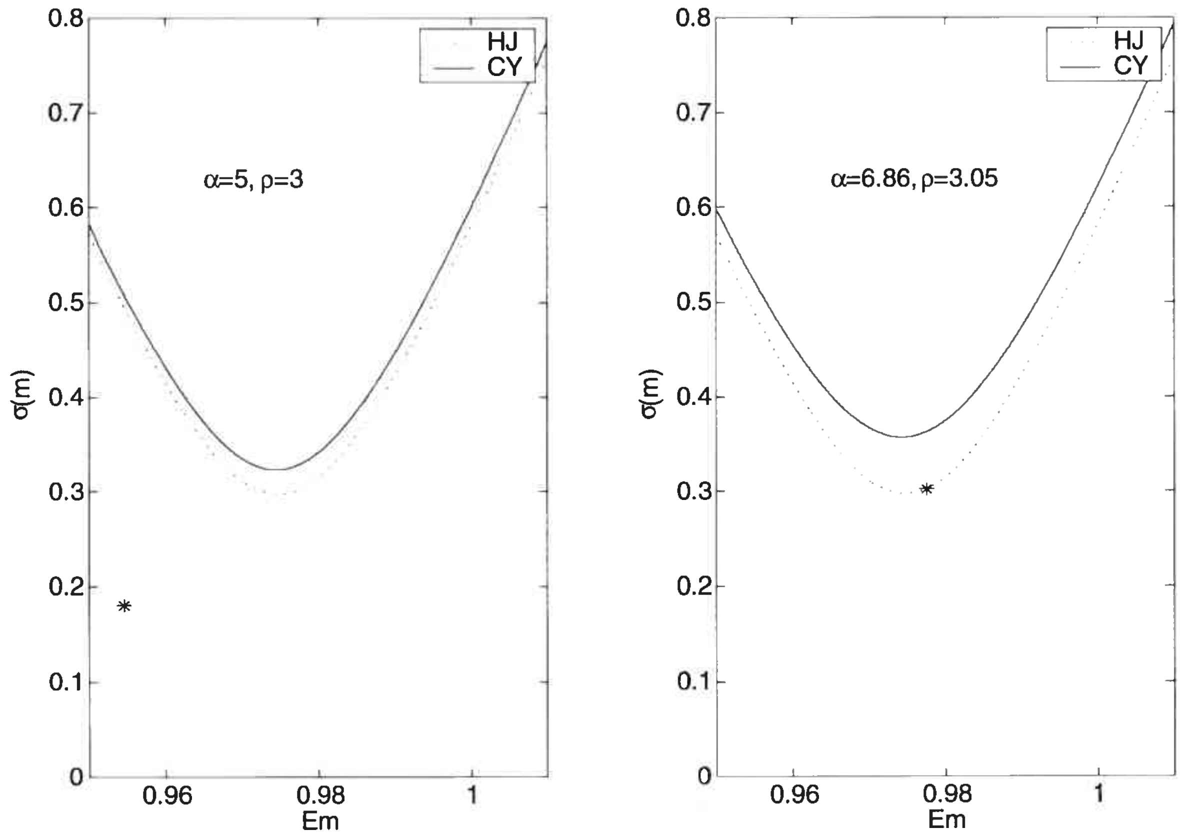


Figure 2.5: **SDF volatility frontier:** HJ represents the Hansen and Jagannathan volatility frontier and CY represents our volatility frontier. For each α, we find $\eta = EmR^{(2)}$ and trace out the point $(\overline{m}, \sigma(m^{mvs}(\overline{m}, \eta)))$ in a two-dimensional space. We also plot the point $(Em_{t+1}, \sigma(m_{t+1}))$ where m_{t+1} represents the SDF obtained with Epstein and Zin (1989) state non-separable preferences. We use the return of the Standard and Poors 500 stock index over the commercial paper. Annual US data, from 1889 to 1994, are used to compute the SDF variance bound.

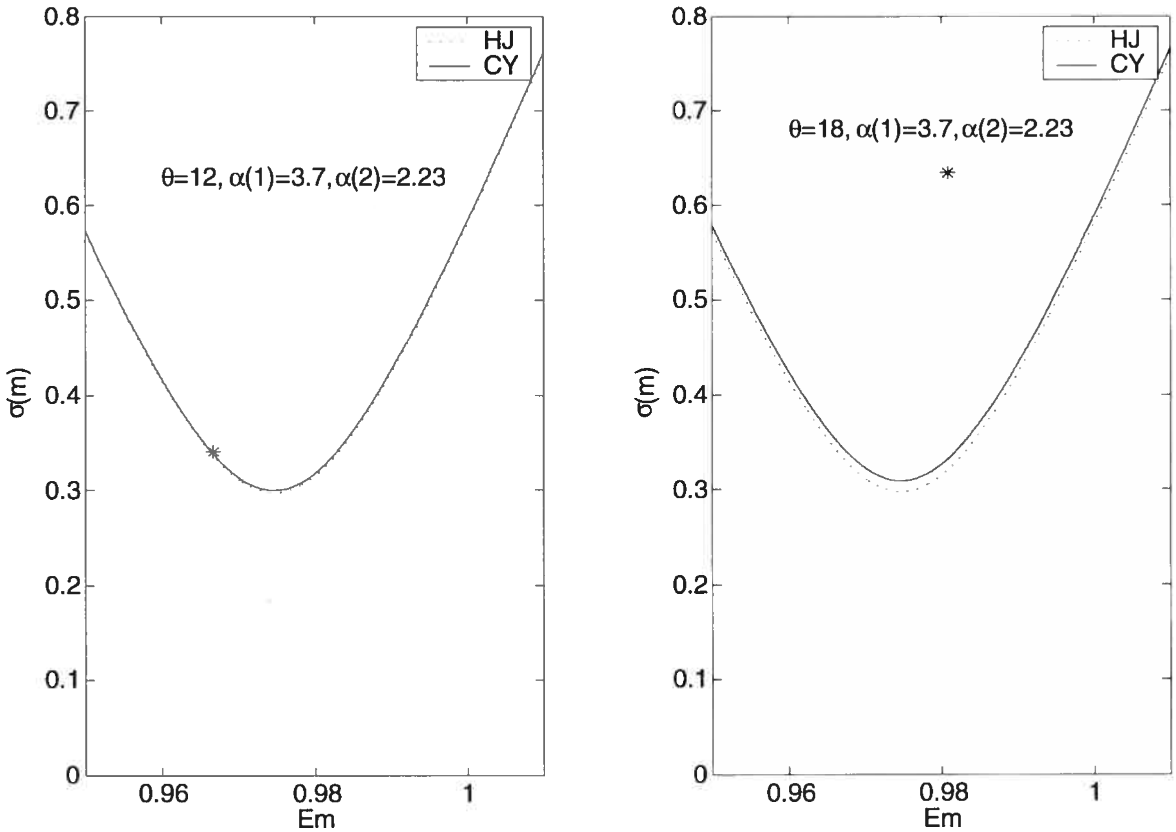


Figure 2.6: **SDF volatility frontier:** HJ represents the Hansen and Jagannathan volatility frontier and CY represents our volatility frontier. For each α, we find $\eta = EmR^{(2)}$ and trace out the point $(\overline{m}, \sigma(m^{mvs}(\overline{m}, \eta)))$ in a two-dimensional space. We also plot the point $(Em_{t+1}, \sigma(m_{t+1}))$ where m_{t+1} represents the SDF obtained with Gordon and St-Amour (2000) state dependent preferences. We use the return on the Standard and Poors 500 stock index over the commercial paper. Annual US data, from 1889 to 1994, are used to compute the SDF variance bound.

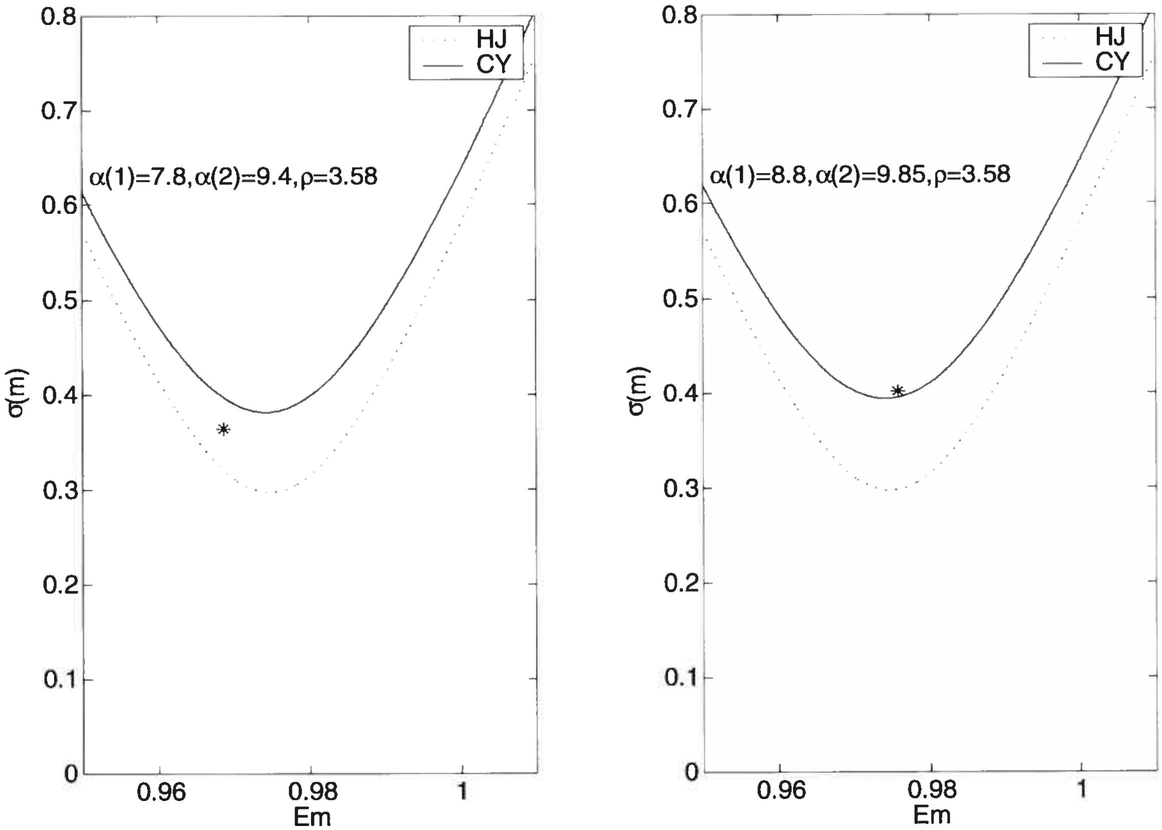


Figure 2.7: **SDF volatility frontier:** HJ represents the Hansen and Jagannathan volatility frontier and CY represents our volatility frontier. For each α, we find $\eta = EmR^{(2)}$ and trace out the point $(\overline{m}, \sigma(m^{mvs}(\overline{m}, \eta)))$ in a two-dimensional space. We also plot the point $(Em_{t+1}, \sigma(m_{t+1}))$ where m_{t+1} represents the SDF obtained with Melino and Yang (2003) state dependent preferences with constant EIS, constant β.and state dependent CRRA. We use the return of the Standard and Poors 500 stock index over the commercial paper. Annual US data, from 1889 to 1994, are used to compute the SDF variance bound.

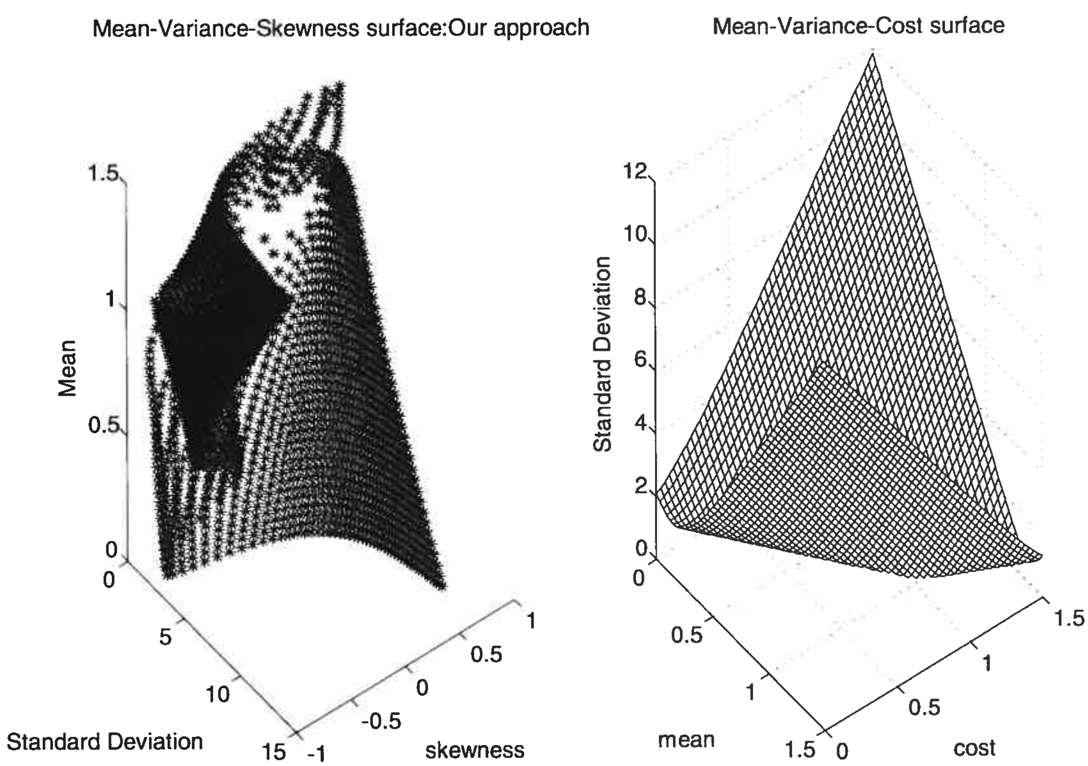


Figure 2.8: **Mean-Variance-Cost (M-V-C) and Mean-Variance-Skewness (M-V-S) surfaces:** Given the portfolio mean, μ_p, and squared portfolio cost, c^*, we solve problem (3.6) and plot in a three-dimensional space the optimal portfolio $\left(\mu_p, c^*, \sigma\left(p^{mvs}\right)\right)$. Then we vary exogenously μ_p and c^* and get the M-V-C surface. We then plot each optimal portfolio in a three-dimensional space: mean-standard deviation-skewness (see the M-V-S surface).

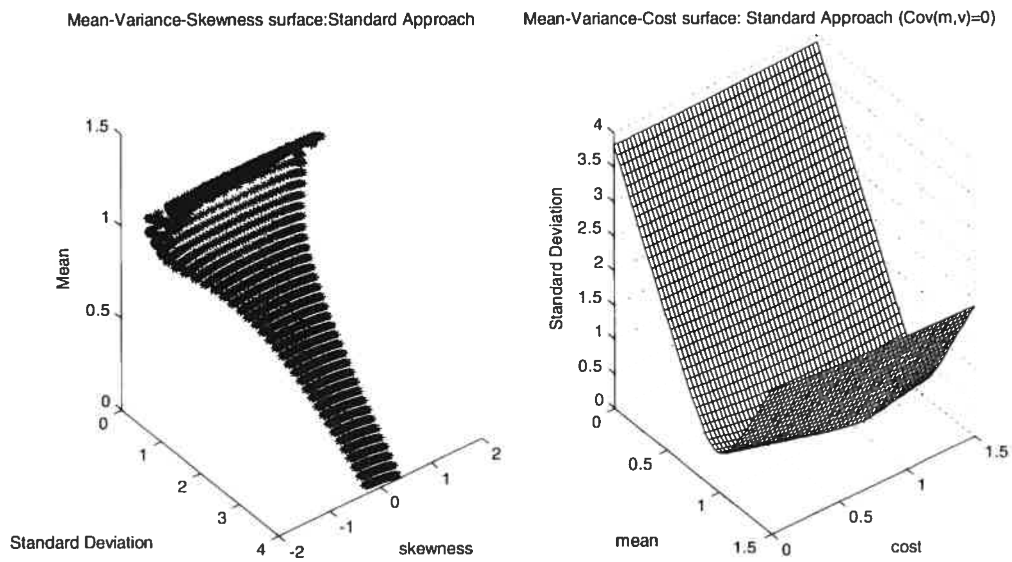


Figure 2.9: **Mean-Variance-Cost (M-V-C) and Mean-Variance-Skewness (M-V-S) surfaces:** We assume $Cov\left(m^{mvs}, v\right) = 0$. Given the portfolio mean, μ_p, and squared portfolio cost, c^*, we solve problem (3.6) and plot in a three-dimensional space the optimal portfolio $\left(\mu_p, c^*, \sigma\left(p^{mvs}\right)\right)$. Then we vary exogenously μ_p and c^* and get the M-V-C surface. We thereafter plot each optimal portfolio in a three-dimensional space: mean-standard deviation-skewness (see the M-V-S surface).

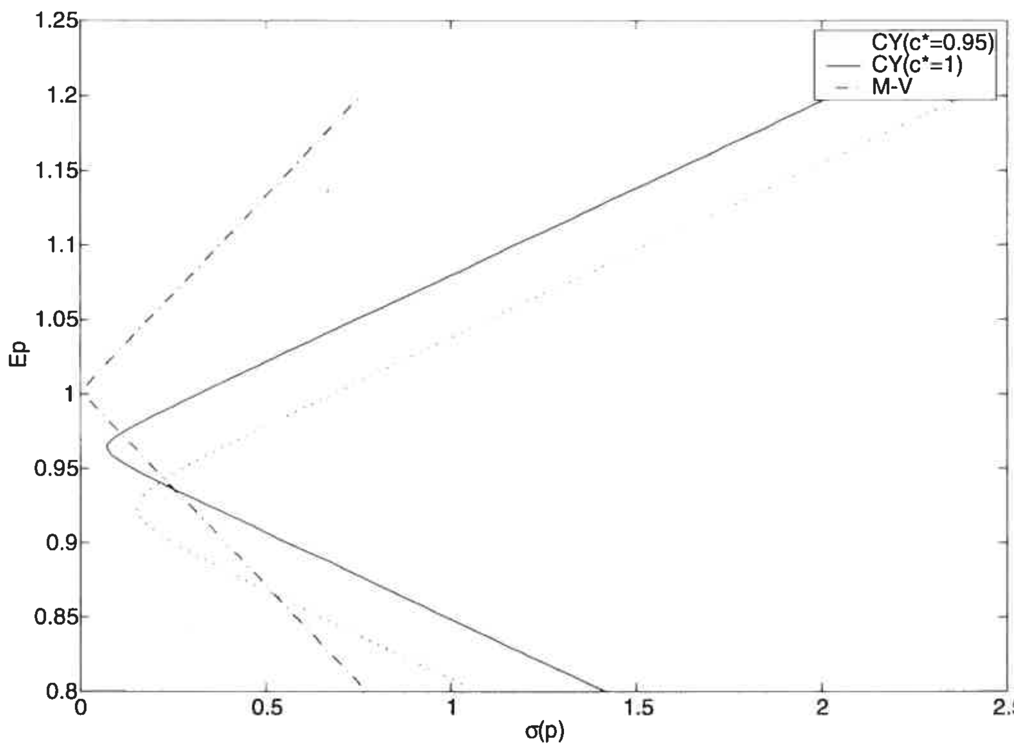


Figure 2.10: **Mean-Variance frontier:** We first plot Markowitz Mean-Variance (M-V) portfolio frontier, then our Mean-Variance portfolio frontier (CY) for $c^* = 0.95$ and 1.

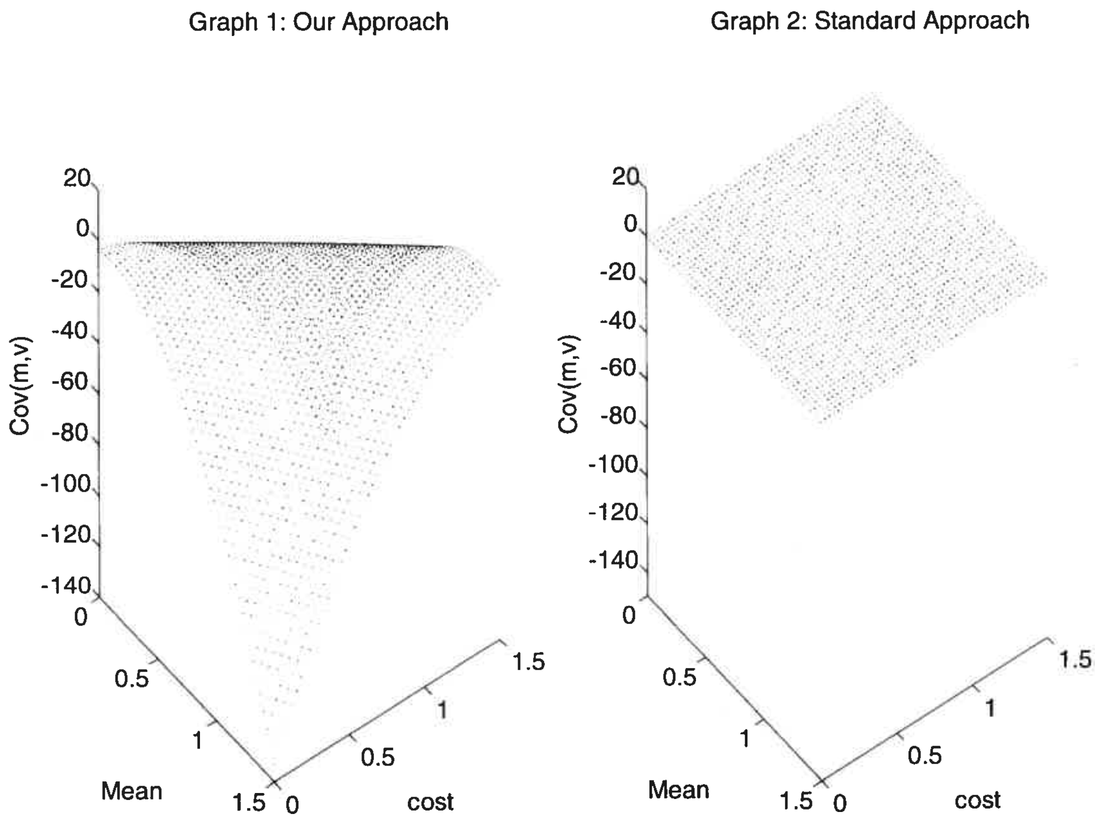


Figure 2.11: **The residual price:** For each portfolio p which belongs to the Mean-Variance-Cost surface, $\mathcal{E}_1$, see Figures 2.8 and 2.9, we plot within Graph 1 the point $(\mu_p, Cov\,(m^{mvs}, v)\,, c^*)$ where μ_p represents the portfolio mean, c^* is the cost of the squared portfolio return and $Cov\,(m^{mvs}, v)$ is the covariance of the SDF with the residuals obtained when regressing the squared portfolio on the portfolio itself. Graph 2 represents this covariance when the standard portfolio selection under skewness is used.

Chapter 3

State Dependence in Fundamentals and Preferences Explains the Risk Aversion Puzzle

1. Introduction

Recently, Jackwerth (2000) and Aït-Sahalia and Lo (2000) have proposed nonparametric approaches to recover risk aversion functions across wealth states from observed stock and option prices. In a complete market economy, which implies the existence of a representative investor, absolute risk aversion can be evaluated for any state of wealth by comparing the historical and risk neutral distributions.

To obtain the historical distribution, Jackwerth (2000) applied a nonparametric kernel density approach to a time series of returns on the S&P 500 index. The risk neutral distribution is recovered from prices on European call options written on the S&P500 index by applying a variation of the nonparametric method introduced in Jackwerth and Rubinstein (1996). The basic idea of this method is to search for the smoothest risk-neutral distribution, which at the same time explains the option prices.

Using simultaneously option prices and realized returns, Jackwerth (2000) and Jackwerth and Rubinstein (2001) find estimated values for absolute risk aversion that are nearly consistent with economic theory before the 1987 crash. However, for the post crash-period, Jackwerth (2000) finds that the implied absolute risk aversion function is negative around the mean wealth level and increasing for larger wealth levels. This empirical feature, called the risk aversion puzzle by Jackwerth (2000), has also been documented by Aït Sahalia and Lo (2000). Another way to express this puzzling result is through the pricing kernel across wealth states. A pricing kernel puzzle is observed when the ratio of the state price density to the historical density increases with wealth [see Brown and Jackwerth (2000)]. After looking at several potential explanations, Jackwerth (2000) concludes that these puzzling results are most probably due to the mispricing of some options by the market.

In this paper, we propose another explanation based on the existence of state dependence in preferences or in economic fundamentals. Garcia, Luger and Renault (2001) proposed a general pricing model where the pricing kernel depends on some latent state variables, observed only by the investor. This phenomenon can be understood in two possible ways. Either as in Melino and Yang

(2003), investors' preferences are state dependent. Or, as in Garcia, Luger and Renault (2003) the joint process of consumption and dividends follows a Markov switching regime distribution such that the current regime is only known to the investors. In this paper, we use the models developed in Garcia, Luger and Renault (2003) and Melino and Yang (2003) to generate artificial prices for stocks and options. To recover the risk neutral distribution, we develop a simple simulation method to create a bid-ask spread around option prices and apply the same nonparametric methodology as Jackwerth and Rubinstein (1996). The historical distribution is estimated based on a mixture of lognormals. In our model, by construction, the risk aversion functions are consistent with economic theory within each regime if we pool the data across regimes. However, as in Jackwerth (2000), we obtain negative estimates of the risk aversion function in some states of wealth. The pricing kernel function across wealth states, calculated from data pooled across regimes, also exhibits a puzzle even though this function is decreasing within each regime. We therefore provide another potential explanation for the puzzles put forward by Jackwerth (2000).

The remainder of this paper is organized as follows. In section 2, we present Jackwerth's (2000) approach for recovering the absolute risk aversion function across wealth states. In section 3, we build a utility-based economic model with state dependence in preferences and endowments and describe how to simulate artificial option and stock prices in this economy. In section 4, we recover the risk aversion and pricing kernel functions across wealth states.

2. The Pricing Kernel and Risk Aversion Puzzles

In this section, we recall the puzzles put forward by Jackwerth (2000) as well as the methodology used to exhibit these puzzles.

2.1 Theoretical underpinnings

Under very general non arbitrage conditions (Hansen and Richard (1987)), the time t price of an asset which delivers a payoff g_{t+1} at time $(t+1)$ is given by:

$$p_t = E_t\left[m_{t+1}g_{t+1}\right], \tag{2.1}$$

where $E_t[.]$ denotes the conditional expectation operator given investors' information at time t. Any random variable m_{t+1} conformable to (2.1) is called an admissible Stochastic Discount Factor (SDF). Among the admissible SDFs, only one denoted by m^*_{t+1} is a function of available payoffs. It is the orthogonal projection of any admissible SDF on the set of payoffs. If some rational investor is able to separate her utility over current and future values of consumption:

$$U[C_t, C_{t+1}] = u(C_t) + \beta u(C_{t+1}), \tag{2.2}$$

The first-order condition for an optimal consumption and portfolio choice will imply that m^*_{t+1} coincides with the projection of $\beta\frac{u\prime(C_{t+1})}{u\prime(C_t)}$ on the set of payoffs. Therefore, through a convenient aggregation argument, concavity of utility functions should imply that m^*_{t+1} is decreasing in current wealth.

Moreover, as shown by Hansen and Richard (1987), no arbitrage implies almost sure positivity of m^*_{t+1}. Therefore, $m^*_{t+1}/E_t m^*_{t+1}$ can be interpreted as the density function of the risk neutral probability distribution with respect to the historical one. In case of a representative investor with preferences conformable to (2.2), we deduce:

$$\frac{m^*_{t+1}}{E_t m^*_{t+1}} = \frac{u'(C_{t+1})}{E_t u'(C_{t+1})}.$$

Therefore:

$$\frac{\partial Log m^*_{t+1}}{\partial C_{t+1}} = \frac{u''(C_{t+1})}{u'(C_{t+1})} \tag{2.3}$$

is the negative of the Arrow-Pratt index of absolute risk aversion (ARA) of the investor.

2.2 The puzzles

For sake of simplicity, it is convenient to analyze these puzzles in a finite state space framework. If $j = 1, \cdots, n$ denote the possible states of nature, we get the density function of the risk neutral distribution probability with respect to the historical one as:

$$\frac{m^*_{t+1}}{E_t m^*_{t+1}} = \frac{p^*_j}{p_j} \text{ in state } j, \tag{2.4}$$

where p_j^* is the risk neutral probability across wealth states $j = 1, \cdots, n$ and p_j is the corresponding historical probability. Brown and Jackwerth (2000) use formula (2.4) to empirically derive the pricing kernel function from realized returns on the SNP 500 index and option prices on the index over a post-1987 period. For the center wealth states (over the range of 0.97 to 1.03 with wealth normalized to one), they found a pricing kernel function which is increasing in wealth. This is the so-called pricing kernel puzzle.

As explained in section 2.1 above, the increasing nature of function (2.1) in wealth is puzzling because it is akin to a convex utility function for a representative investor, which is obviously inconsistent with the general assumption of risk aversion. From (2.3), the ARA coefficient can actually be computed through a log-derivative of the pricing kernel. By using (2.4) we deduce:

$$ARA = -\frac{u''(C_{t+1})}{u'(C_{t+1})} = \frac{p_j'}{p_j} - \frac{p_j^{*\prime}}{p_j^*} \tag{2.5}$$

where p_j' and $p_j^{*\prime}$ are of the derivatives of p_j and p_j^* with respect to aggregate wealth in state j.

Jackwerth (2000) observes that the absolute risk aversion functions as computed from (2.5) dramatically change shape around the 1987 crash. Prior to the crash, they are positive and decreasing in wealth which is consistent with standard assumptions made in economic theory about investors' preferences. After the crash, they are partially negative and increasing (see figure 3 in Jackwerth (2000)). This result is called the risk aversion puzzle. One component of it is tantamount to the pricing kernel puzzle: ARA should be positive as the pricing kernel should be decreasing in aggregate wealth. Moreover, even when there is no pricing kernel puzzle (positive ARA), there remains a risk aversion puzzle when ARA is increasing in wealth. While the pricing kernel puzzle is only observed for the center of wealth states, the risk aversion puzzle (increasing ARA) remains for larger levels of wealth. Without any discretization of wealth states, Ait-Sahalia and Lo (2000) documented similar empirical puzzles for implied risk aversion.

2.3 Statistical methodology

Several statistical methodologies are possible to recover the historical distribution of future returns (on the underlying index) given current ones. As emphasized by Jackwerth (2000), the choice of a particular estimation strategy should not have any impact on the documented puzzles. For instance, a kernel estimation will be valid under very general stationarity and mixing conditions.

While historical probabilities p_j are recovered from a time series of underlying index returns, risk neutral probabilities p_j^* will be backed out in cross section from a set of observed option prices written on the same index. Concerning this issue, a pioneering work was Rubinstein (1994) who recommends to solve the following quadratic program:

$$\begin{aligned} &\min_{p^*} \sum_{j=1}^{n} \left(p_j^* - \bar{p}_j\right)^2 \\ &\sum_{j=1}^{n} p_j^* = 1, \quad p_j^* \geq 0, \\ C_i^* &= \frac{1}{R_f} \sum_{j=1}^{n} p_j^* \max\left[0, S_j - K_i\right], \\ &\frac{1}{R_f} \sum_{j=1}^{n} p_j^* S_j = S_0, \\ C_{ib}^* \leq C_i^* &\leq C_{ia}^* \text{ for i=1,...,m and } S_b \leq S_0 \leq S_a, \end{aligned} \tag{2.6}$$

where C_{ib}^* (C_{ia}^*) represents the call option bid (ask) price with strike price K_i. The bid and ask stock prices are respectively S_b and S_a. In other words, the implied risk neutral probabilities p_j^* are the closest to the prior $\bar{p}_j$ that result in option and underlying asset values that fall between the respective bid and ask prices. As stressed by Jackwerth and Rubinstein (1996), this methodology has the virtue that general arbitrage opportunities do not exist if and only if there is a solution. This remark is still valid when considering alternative quadratic programs based on other distances. For instance, Jackwerth and Rubinstein (1996) put forward the goodness of fit approach:

$$\min_{p^*} \sum_{j=1}^{n} \frac{\left(p_j^* - \bar{p}_j\right)^2}{\bar{p}_j} \tag{2.7}$$

while, following Hansen and Jagannathan (1997), one may prefer:

$$\min_{p^*} \sum_{j=1}^{n} \frac{\left(p_j^* - \bar{p}_j\right)^2}{p_j} \tag{2.8}$$

since, with obvious notations, the objective function (2.8) can be seen as $E_t(m^*_{t+1} - \overline{m}_{t+1})^2$.

However, Jackwerth and Rubinstein (1996) observe that the implied distributions are rather independent of the choice of the objective function when a sufficiently high number of options is available.[1]

Since we are to going to focus in this paper on a simulation exercise, we will choose 50 options in cross section in order to be sure that the solution is determined by the constraints (options and underlying asset values between bid and ask prices) and not by the objective function. In particular, the choice of the prior is immaterial and, as noticed by Jackwerth and Rubinstein (1996), even a pure smoothness criterion independent of any prior would do the job. They consider in particular:

$$\min_{p^*} \sum_{j=1}^{n} \left(p^*_{j-1} + p^*_{j+1} - 2p^*_j\right)^2 \tag{2.9}$$

when the states $j = 1, 2 \cdots, n$ are ranked in order of increasing wealth. However, to remain true to the traditional approach, we are going to use in the simulation section 4 the goodness of fit criterion. Prior risk neutral probabilities $\bar{p}_j$ will be computed, according to the Breeden and Litzenberger (1978) methodology, from second order derivatives of option prices with respect to the strike price. Note that a necessary source of difference between p^*_j and $\bar{p}_j$ is the discretization of the state space performed to define p^*_j.

3. Economies with regime shifts

In this section, we construct economies with regime shifts in endowments or preferences to simulate artificial stock and option prices.

[1] They notice that "as few as 8 option prices seem to contain enough information to determine the general shape of the implied distribution" and that "at the extreme, the constraints themselves will completely determine the solution"

3.1 The general framework

Consider an European call option with maturity T and strike price K. A straightforward multiperiod extension of (2.1) gives it's time t price as:

$$\pi_t = E_t\left[m_{t+1}m_{t+2}\cdots m_T\ (S_T - K)^+\right]. \tag{3.1}$$

Garcia, Luger and Renault (2001) provide a convenient set of general assumptions about the bivariate process $\left(m_{t+1}, \dfrac{S_{t+1}}{S_t}\right)$ which allow them to give closed-form formulas for the expectations (3.1) while encompassing the most usual option pricing models (see also Garcia, Ghysels and Renault(2003)). The maintained assumptions are:

Assumption A1

The variables $\left(m_{\tau+1}, \frac{S_{\tau+1}}{S_\tau}\right)_{1\leq\tau\leq T-1}$ are condionally independent given the path $U_1^T = (U_t)_{1\leq t\leq T-1}$ *of a vector* U_t *of state variables.*

Assumption A1 expresses that the dynamics of the returns is driven by the state variables. It is similar in spirit to common stochastic volatility models (the stochastic volatility process being the state variable) when standardized returns are assumed to be independent.

Assumption A2

The process $\left(m_t, \frac{S_t}{S_{t-1}}\right)$ *does not Granger-cause the state variables process* (U_t).

This assumption states that the state variables are exogenous. For common stochastic volatility or hidden Marlov processes, such an exogeneity assumption is usually maintained to make the standard filtering strategies valid. It should be noted that this exogeneity assumption does not preclude instantaneous causality relationships such as a leverage effect.

Assumption A3 *The conditional probability distribution of* $\left(\log m_{t+1}, \log\frac{S_{+1}}{S_t}\right)$ given U_1^{t+1} is a bivariate normal

$$\begin{bmatrix} \log\ m_{t+1} \\ \log \dfrac{S_{t+1}}{S_t} \end{bmatrix} \Big| U_1^{t+1} \rightsquigarrow \mathcal{N}\left[\begin{pmatrix} \mu_{mt} \\ \mu_{st} \end{pmatrix}, \begin{pmatrix} \sigma_{mt}^2 & \sigma_{mst} \\ \sigma_{mst} & \sigma_{st}^2 \end{pmatrix}\right].$$

Assumption A3 is a very general version of the mixture of normals model. A maintained assumption will be that investors observe U_t at time t, so that the conditioning information in the expectation operator (3.1) is:

$$I_t = \sigma\left[m_\tau, S_\tau, U_\tau, \tau \leq t\right]. \tag{3.2}$$

In our simulation exercise, the mixing variable U_{t+1} will be a two-state Markov chain with a transition matrix:

$$P = \begin{bmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{bmatrix}. \tag{3.3}$$

Indeed, following Garcia, Luger and Renault (2001), a general option pricing formula can be stated for any Markov process (U_t) conformable to A1, A2 and A3.

Proposition 3.1 *Under assumptions A1, A2 and A3*

$$\frac{\pi_t}{S_t} = \pi_t(x_t) = E_t\left\{Q_{ms}(t,T)\,\Phi\left(d_1(x_t)\right) - \frac{\tilde{B}(t,T)}{B(t,T)}e^{-x_t}\Phi\left(d_2(x_t)\right)\right\}$$

where $x_t = \log\frac{S_t}{KB(t,T)}$, $B(t,T) = E_t\left(\prod_{\tau=t}^{T-1} m_{\tau+1}\right)$ *is the time t price of a bond maturing at time T, and*

$$\begin{aligned}
d_1(x) &= \frac{x}{\bar{\sigma}_{t,T}} + \frac{\bar{\sigma}_{t,T}}{2} + \frac{1}{\bar{\sigma}_{t,T}}\log\left[Q_{ms}(t,T)\frac{B(t,T)}{\tilde{B}(t,T)}\right], \\
d_2(x) &= d_1(x) - \bar{\sigma}_{t,T}, \\
\bar{\sigma}^2_{t,T} &= \sum_{\tau=t}^{T-1}\sigma^2_{s\tau}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{B}(t,T) &= \exp\left(\sum_{\tau=t}^{T-1}\mu_{m\tau+1} + \frac{1}{2}\sum_{\tau=t}^{T-1}\sigma^2_{m\tau},\right) \\
Q_{ms}(t,T) &= \tilde{B}(t,T)\exp\left(\sum_{\tau=t}^{T-1}\sigma_{ms\tau+1}\right)E\left[\frac{S_T}{S_t}\middle| U_1^T\right].
\end{aligned}$$

As explicitly analyzed in Garcia, Ghysels and Renault (2003), this general option pricing formula encompasses most of the common pricing formulas for European options on equity.

In order to consider economically meaningful regime shifts in the SDF, it is convenient to start from a two-factor model as produced by Epstein and Zin (1989). Their recursive utility framework leads them to the following SDF:

$$m_{t+1} = \left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right]^{\gamma}\left[\frac{1}{R_{mt+1}}\right]^{1-\gamma} \tag{3.4}$$

where $\rho = 1 - \dfrac{1}{\sigma}$, σ is the elasticity of intertemporal substitution and $\gamma = \dfrac{\alpha}{\rho}$ with $(1-\alpha)$ the index of relative risk aversion. With a two-states mixing variable U_{t+1}, $\log m_{t+1}$ appears as a mixture of two normal distributions in two cases. In the first case of state dependent preferences, preference parameters are functions of U_{t+1} while in the second case, there are regime shifts in fundamentals and the joint probability distribution of $\left(\log \frac{C_{t+1}}{C_t}, \log R_{mt+1}\right)$ is a mixture of normals.

The case of state dependent-preferences has been analyzed recently by Melino and Yang (2003) while Garcia, Luger and Renault (2001, 2003) focus on shifts in fundamentals.[2]

3.2 State-dependent preferences or fundamentals

Let us first assume as Melino and Yang (2003) that the three preference parameters β, α, ρ are all state-dependent and then denoted as $\beta(U_t), \alpha(U_t)$ and $\rho(U_t)$. While these values, known by the investor at time t, define her time t utility level, she does not know at this date the next coming values $\beta(U_{t+1}), \alpha(U_{t+1})$ and $\rho(U_{t+1})$. Therefore, the resulting SDF will be more complicated than just replacing α, β and ρ in (3.4) by their state dependent value. Melino and Yang (2003) show that the SDF is:

$$m_{t+1} = \left[\beta(U_t)\left(\frac{C_{t+1}}{C_t}\right)^{\rho(U_t) - \frac{\rho(U_t)}{\rho(U_{t+1})}}\right]^{\gamma(U_t)} R_{mt+1}^{\frac{\alpha(U_t)}{\rho(U_{t+1})}-1} P_t^{\frac{\alpha(U_t)}{\rho(U_{t+1})}-\frac{\alpha(U_t)}{\rho(U_t)}} \tag{3.5}$$

where $\gamma(U_t) = \frac{\alpha(U_t)}{\rho(U_t)}$ and P_t is the time t price of the market portfolio. When $\beta(U_t)$, $\alpha(U_t)$, $\rho(U_t) = \rho(U_{t+1})$ are constants, this pricing kernel reduces to the Epstein and Zin SDF (3.4). By definition:

$$R_{mt+1} = \frac{P_{t+1} + C_{t+1}}{P_t},$$

[2]See also Gordon and St Amour (2000) for an alternative way to introduce state dependence in preferences in a CCAPM framework.

while the underlying asset return is $\frac{S_{t+1}+D_{t+1}}{S_t}$. Asset prices P_t and S_t are then determined as discounted values of future dividend flows by iteration of the following pricing formulas:

$$P_t = E_t\left[m_{t+1}\left(P_{t+1} + C_{t+1}\right)\right] \text{ and } S_t = E_t\left[m_{t+1}\left(S_{t+1} + D_{t+1}\right)\right]. \quad (3.6)$$

Garcia, Luger and Renault (2001) show that assumptions A1 and A2 are implied by similar assumptions stated for the joint process $\left(\frac{C_{t+1}}{C_t}, \frac{D_{t+1}}{D_t}\right)$. Assumption A3 will then also be implied by a similar assumption about fundamentals:

Assumption A3$'$: *The conditional probability distribution of* $\left(\log\frac{C_{t+1}}{C_t}, \log\frac{D_{+1}}{D_t}\right)$ given U_1^{t+1} is a bivariate normal

$$\begin{bmatrix} \log\frac{C_{t+1}}{C_t} \\ \log\frac{D_{+1}}{D_t} \end{bmatrix} |U_1^{t+1} \rightsquigarrow N\left[\begin{pmatrix} \mu_{X_{t+1}} \\ \mu_{Y_{t+1}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_{t+1}} & \sigma_{XY,t+1} \\ \sigma_{XY,t+1} & \sigma^2_{Y_{t+1}} \end{pmatrix}\right]$$

Proposition 3.2 bellow nests the results of Melino and Yang (2003) and Garcia, Luger, Renault (2001) in a common setting.

Proposition 3.2 : *Under assumptions A1, A2 and A3, with m_{t+1} given by (3.5), the conditional probability distribution of $\left(Logm_{t+1}, \log\frac{S_{t+1}}{S_t}\right)$ given U_1^{t+1} is jointly normal with mean and variances defined in the Appendix.*

In the simulation exercises conducted in section 4 we consider first regime changes in fundamentals and then regime changes in several configurations of the preference parameters in order to disentangle the respective roles of fundamentals and preferences. The general option pricing formula, which can also accomodate the case where both fundamentals and preferences change with the regime, is given in proposition 3.3 below:

First, it is worth noticing that the equilibrium model characterizes the asset prices P_t and S_t as:

$$\begin{aligned} \frac{P_t}{C_t} &= \lambda\left(U_1^t\right) = E_t\left[m_{t+1}\left(1 + \lambda\left(U_1^{t+1}\right)\right)\frac{C_{t+1}}{C_t}\right]. \\ \frac{S_t}{D_t} &= \varphi\left(U_1^t\right) = E_t\left[m_{t+1}\left(1 + \varphi\left(U_1^{t+1}\right)\right)\frac{D_{t+1}}{D_t}\right]. \end{aligned}$$

Then proposition 3.3 summarizes the option pricing implications of propositions 3.1 and 3.2 in the simplest case of a unit time to maturity (T=t+1):

Proposition 3.3 *Under (A1),(A2) and (A3) the European option price is given by:*

$$\pi_t = E_t\left[S_t Q_{XY}(t,t+1)\Phi(d_1) - K\widetilde{B}(t,t+1)\Phi(d_2)\right],$$

where,

$$d_1 = \frac{Log\left[\frac{S_t Q_{XY}(t,t+1)}{K\widetilde{B}(t,t+1)}\right]}{\left(\sigma^2_{Yt+1}\right)^{\frac{1}{2}}} + \frac{1}{2}\left(\sigma^2_{Yt+1}\right)^{\frac{1}{2}},\ d_2 = d_1 - \left(\sigma^2_{Yt+1}\right)^{\frac{1}{2}}$$

with,

$$\widetilde{B}(t,t+1) = a\left(U_1^{t+1}\right)\exp\left(\left[\alpha\left(U_{t+1}\right)-1\right]\mu_{X_{t+1}} + \frac{1}{2}\left[\alpha\left(U_{t+1}\right)-1\right]^2\sigma^2_{X_{t+1}}\right)$$

$$Q_{XY}(t,t+1) = \widetilde{B}(t,t+1)\, b_t^{t+1}\frac{\varphi\left(U_1^t\right)}{\varphi\left(U_1^{t+1}\right)}\exp\left(\left[\alpha\left(U_{t+1}\right)-1\right]\rho_{XY}\sigma_{XT}\sigma_{YT}\right)E_t\left[\frac{S_{t+1}}{S_t}|U_1^{t+1}\right]$$

and,

$$a\left(U_1^{t+1}\right) = \beta\left(U_t\right)^{\frac{\alpha(U_t)}{\rho(U_t)}}\left(1-\beta\left(U_t\right)\right)^{\frac{\alpha(U_t)}{\rho(U_{t+1})}-\frac{\alpha(U_t)}{\rho(U_t)}}\lambda\left(U_1^t\right)^{1-\frac{\alpha(U_t)}{\rho(U_t)}}\left[1+\lambda\left(U_1^{t+1}\right)\right]$$

$$b_t^{t+1} = \frac{1+\varphi\left(U_1^{t+1}\right)}{\varphi\left(U_1^t\right)}$$

PROOF. The proof is similar to the proof where $\alpha(U_t)$, $\beta(U_t)$, and $\rho(U_t)$ are constants, which can be found in Garcia, Luger and Renault (2003). ■

If the preference parameters α, β, and ρ are constants, proposition 3.3 collapses to the Garcia, Luger and Renault (2001) option pricing formula. Note that the definition of $\lambda\left(U_1^{t+1}\right)$ and $\varphi\left(U_1^{t+1}\right)$ is akin to

$$E_t Q_{XY}(t,t+1) = 1,\ \text{and}\ E_t\widetilde{B}(t,t+1) = B(t,t+1).$$

3.3 Simulating option and stock prices

First, we calibrate our economic models with regime shifts in the parameters describing preferences or economic fundamentals. In the case of state dependent fundamentals, we choose values that are close to those estimated in Garcia, Luger and Renault (2003) where preference parameters are not state-dependent. For state-dependent preferences, we disturb these particular values. All values are explicitly given in the figures. We then use proposition 3.3 to compute option prices with different strike prices. To use the methodology described in (2.7), we need to develop a simple technique to create bid-ask spreads around the simulated prices. This is done in three steps:

- **Step 1:** Given the stock price, S_t, we find a bid-ask spread sp by drawing in a lognormal distribution:

$$\log(sp) \rightarrow N\left(\mu_{sp}, \sigma_{sp}^2\right),$$

  where the parameter μ_{sp} and σ_{sp}^2 are chosen exogenously.

- **Step 2:** Given sp, we draw a real number x in the censored normal probability distribution $N(\mu_x, \sigma_x^2)$ given $0 \leq x \leq sp$.

- **Step 3:** We then compute the stock bid and ask prices:

$$\begin{aligned} ask\ price &= S_t + (sp - e^x), \\ bid\ price &= S_t - e^x. \end{aligned}$$

We apply a similar simulation methodology to create bid and ask prices for options. Based on these bid and ask option prices and stock prices, we recover the risk neutral probabilities using the nonparametric methodology described in section 2. It is important to note that our Monte-Carlo approach gives us the historical return distribution. Therefore, we do not need to use any nonparametric estimation technique to recover the historical distribution.

The whole procedure must be applied for each state $U_t \in \{0,1\}$ of the economy. At date t, given the state variable value, $U_t \in \{0,1\}$, we compute the call option prices:

$$\pi_t(U_t) = E\left[S_t Q_{XY}(t,t+1)\Phi(d_1) - K\widetilde{B}(t,t+1)\Phi(d_2)\,|U_t\right]$$

and perform steps 1, 2 and 3. We then use equations (2.4) and (2.5) to infer the conditional absolute risk aversion and pricing kernel functions across states (given the state variable U_t).

By construction, these quantities are computed from probabilities $p_j(U_t)$ and $p_j^*(U_t)$ which explicitly depend on the actual value of the latent state U_t. By constrast, a statistician who does not observe the state and performs a nonparametric estimation of the stationary historical distribution which does not account for unobserved heterogeneity, will estimate marginal probabilities p_j that are averaged across states:

$$p_j = P(U_t = 0)\,p_j(0) + P(U_t = 1)\,p_j(1)\,. \tag{3.7}$$

As far as risk neutral probability p_j^* are concerned, the issue is less clear. If we could be sure that not only the agents have observed the states U_t but also that the statistical observation of asset prices is in synchronized with observations, then the p_j^* computed from (2.6) and the real data should be $p_j^*(U_t)$. However, any synchronization problem may push the implied p_j^* towards their averaged values

$$p_j^* = P(U_t = 0)\,p_j^*(0) + P(U_t = 1)\,p_j^*(1)\,. \tag{3.8}$$

For reasons made explicit below, we choose to compare the implied risk aversion and pricing kernel computed state by state from $\left(p_j(U_t), p_j^*(U_t)\right)$ with the fully marginalized ones, that is computed from marginalized values $\left(p_j, p_j^*\right)$ given by (3.7) and (3.8) rather than using the possible mixed approach $\left(p_j, p_j^*(U_t)\right)$.

4. Empirical Results

Without loss of generality, we treat the cases of state-dependence in fundamentals and in preferences separately to illustrate our results.

4.1 Regime shifts in fundamentals

We first assume that only the fundamentals are affected by the latent state variables. Based on the prices generated with the procedures described in the previous sections, we follow the methodology described in section 2 and recover the risk aversion and pricing kernel functions across wealth states. The first graph in Figure 3.1 reveals that the unconditional pricing kernel increases in the center wealth states (over the range of 0.9 to 1.1). This feature is highlighted in Jackwerth and Brown (2001) as the kernel pricing puzzle. We use the term unconditional to emphasize that the pricing kernel function across wealth states is computed using marginalized probabilities given by (3.7) and (3.8). Also around the center wealth states, the unconditional absolute risk aversion function is negative as in Jackwerth (2000). However, within each regime, the conditional pricing kernel and absolute risk aversion function across wealth states are perfectly decreasing functions of the aggregate wealth: the puzzles disappear. When regimes are (or not) observed, we confirm that the results do not depend on the particular distance measure used. Figure 3.2 confirms the results when regimes are not observed. The same features are exhibited with the alternative Hansen and Jagannathan (1997) distance measure (2.8).

4.2 Regime shifts in preferences

We also consider state dependence in the investor's preference parameters and investigate several state-dependent preference cases. First, we assume a constant relative risk aversion (CRRA) and a state-dependent elasticity of intertemporal substitution (EIS). Second, we assume a state-dependent risk aversion and a constant EIS. Third, we assume cyclical CRRA and EIS and finally we assume state-dependent time preferences. For all combinations of state-dependent preference parameters, we get very similar results: both the unconditional pricing kernel and absolute risk aversion function exhibit the aforementioned puzzles while the puzzles disappear within each regime. Therefore, we only report the results for state-dependent relative risk aversion and constant EIS in Figure 3.3. Around the center wealth states, we observe an increasing marginal utility on the left panel while risk aversion shown in the right panel falls into negative values. Figure 3.4 confirms these results with the alternative distance measure (2.8).

4.3 General comments

The two above examples of regime shifts in fundamentals or in preferences lead us to the same general conclusion. While implied risk aversion and implied pricing kernel computed from marginalized probabilities $\left(p_j, p_j^*\right)$ display the same paradoxical features as in Ait-Sahalia and Lo (2000) and Jackwerth (2000), it turns out that taking into account unobserved heterogeneity through the state dependent probabilities $\left(p_j(U_t), p_j^*(U_t)\right)$ solve the puzzle. In other words, our results lead us to think that may be investors utility functions are not at odds with traditional economic theory, but investors observe a latent state variable which artificially creates a paradox when it is forgotten in the statistical procedure. As already mentioned, full observation (of states) by agents, with perfect synchronization with our observation of option prices may lead to use instead the probabilities $\left(p_j, p_j^*(U_t)\right)$. The implied risk aversion and pricing kernel observed with such mixed probabilities appear, according to a complementary simulation study available upon request, even wilder than the ones produced by marginal probabilities $\left(p_j, p_j^*\right)$. Since the latter look more conformable to the empirical evidence put forward by Jackwerth (2000), we have chosen to focus on them in this paper.

5. Conclusion

This paper investigates the ability of economic models with regime shifts to produce and solve the risk aversion and the pricing kernel puzzles put forward in Aït-Sahalia and Lo (2000) and Jackwerth (2000). We show that models with regime shifts in fundamentals or investor's preferences can explain and rationalize these puzzles. The absolute risk aversion and pricing kernel functions extracted from the simulated prices in these economies exhibit the same puzzling features as in the original papers and are inconsistent with the usual assumptions of decreasing marginal utility and positive risk aversion. However, within each regime, the absolute risk aversion and pricing kernel functions are consistent with economic theory: the investor utility is concave and her risk aversion remains positive. In other words, investors' behavior is not at odds with economic theory but depends on some factors that the statistician does not observe. We have also shown that this conclusion is robust to the choice of the statistical estimation procedure.

6. Appendix: Proofs

PROOF OF PROPOSITION 3.2. Rearranging equation (6.9) for the pricing kernel in Melino and Yang (2003)), one obtain:

$$m_{t+1} = \left[\beta(U_t)\left(\frac{C_{t+1}}{C_t}\right)^{\rho(U_t)-\frac{\rho(U_t)}{\rho(U_{t+1})}}\right]^{\gamma(U_t)} R_{mt+1}^{\frac{\alpha(U_t)}{\rho(U_{t+1})}-1} P_t^{\frac{\alpha(U_t)}{\rho(U_{t+1})}-\frac{\alpha(U_t)}{\rho(U_t)}},$$

where $\gamma(U_t) = \frac{\alpha(U_t)}{\rho(U_t)}$ and P_t is the equilibrium price of the market portfolio at time t. If $\rho(U_t) = \rho(U_{t+1})$ and $\beta(U_t)$, $\alpha(U_t)$, $\rho(U_t)$ are constants, this pricing kernel reduces to the Epstein and Zin (1989) pricing kernel. Let $\varphi(U_t) = \frac{S_t}{D_t}$ denote the price-dividend ratio and $\lambda(U_t) = \frac{P_t}{C_t}$ the price-earning ratio. The return on the market portfolio can be written as

$$R_{mt+1} = \frac{P_{t+1}+C_{t+1}}{P_t} = \left(\frac{\lambda(U_1^{t+1})+1}{\lambda(U_t)}\right)\left(\frac{C_{t+1}}{C_t}\right),$$

and the stock return:

$$\frac{S_{t+1}}{S_t} = \frac{\varphi(U_1^{t+1})}{\varphi(U_t)}\frac{D_{t+1}}{D_t}.$$

Let us assume that the conditional probability distribution of $\left(\log\frac{C_{t+1}}{C_t}, \log\frac{D_{t+1}}{D_t}\right)$ given U_1^{t+1} is a bivariate normal:

$$\begin{bmatrix} \log\frac{C_{t+1}}{C_t} \\ \log\frac{D_{t+1}}{D_t} \end{bmatrix} / U_1^T \rightsquigarrow N\left[\begin{pmatrix} \mu_{X_{t+1}} \\ \mu_{Y_{t+1}} \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_{t+1}} & \sigma_{XY,t+1} \\ \sigma_{XY,t+1} & \sigma^2_{Y_{t+1}} \end{pmatrix}\right], \tag{6.1}$$

with $U_1^{t+1} = (U_\tau)_{1\leq\tau\leq t+1}$. Taking the logarithm of m_{t+1}, we get

$$\begin{aligned}\log m_{t+1} &= \gamma(U_t)\log\beta(U_t) + \left(\frac{\alpha(U_t)}{\rho(U_{t+1})}-1\right)\log\left(\frac{\lambda(U_1^{t+1})+1}{\lambda(U_t)}\right) + \\ &\left(\frac{\alpha(U_t)}{\rho(U_{t+1})}-\frac{\alpha(U_t)}{\rho(U_t)}\right)\log(\lambda(U_t)C_t) + \\ &\left[\gamma(U_t)\left(\rho(U_t)-\frac{\rho(U_t)}{\rho(U_{t+1})}\right)+\left(\frac{\alpha(U_t)}{\rho(U_{t+1})}-1\right)\right]\log\left(\frac{C_{t+1}}{C_t}\right).\end{aligned}$$

The logarithm of the stock return is

$$\log\frac{S_{t+1}}{S_t} = \log\frac{\varphi(U_1^{t+1})}{\varphi(U_t)} + \log\frac{D_{t+1}}{D_t}.$$

Consequently,

$$\left[\begin{array}{c}\log m_{t+1}\\ \log \frac{S_{t+1}}{S_t}\end{array}\right] = A + B\left[\begin{array}{c}\log \frac{C_{t+1}}{C_t}\\ \log \frac{D_{t+1}}{D_t}\end{array}\right],$$

where $A = (a_1, a_2)'$ with

$$\begin{aligned} a_1 &= \gamma(U_t)\log\beta(U_t) + \left(\rho(U_t) - \frac{\rho(U_t)}{\rho(U_{t+1})}\right)\log\left(\frac{\lambda(U_1^{t+1})+1}{\lambda(U_t)}\right) + \\ &\quad \left(\frac{\alpha(U_t)}{\rho(U_{t+1})} - \frac{\alpha(U_t)}{\rho(U_t)}\right)\log\left(\lambda(U_t)C_t\right), \\ a_2 &= \log\frac{\varphi(U_1^{t+1})}{\varphi(U_t)}, \end{aligned}$$

and B is a diagonal matrix with diagonal coefficients:

$$\begin{aligned} b_{11} &= \left[\gamma(U_t)\left(\rho(U_t) - \frac{\rho(U_t)}{\rho(U_{t+1})}\right) + \left(\frac{\alpha(U_t)}{\rho(U_{t+1})} - 1\right)\right], \\ b_{22} &= 1. \end{aligned}$$

Using (6.1), it is straightforward to show:

$$\left[\begin{array}{c}\log m_{t+1}\\ \log \frac{S_{t+1}}{S_t}\end{array}\right] / U_1^{t+1} \rightsquigarrow N\left[\mu, \Sigma_{ms}\right]$$

with

$$\begin{aligned} \mu &= A + B\begin{pmatrix}\mu_{X_{t+1}}\\ \mu_{Y_{t+1}}\end{pmatrix}, \\ \Sigma_{ms} &= B\begin{pmatrix}\sigma^2_{X_{t+1}} & \sigma_{X_{t+1}Y_{t+1}}\\ \sigma_{X_{t+1}Y_{t+1}} & \sigma^2_{Y_{t+1}}\end{pmatrix}B. \end{aligned}$$

This completes the proof. ■

References

[1] Ait-Sahalia, Y., and A. W. Lo, 1998, "Nonparamtertric Estimation of State-Price Densities Implicit in Financial Asset Prices," Journal of Finance, 53, 499-547.

[2] Ait-Sahalia, Y., and A.W. Lo, 2000, "Nonparametric Risk Management and Implied Risk Aversion," Journal of Econometrics, 94, 9-51.

[3] Bates, D. S., 2000, "Post-87 Crash Fears in the S&P 500 Futures Option Market," Journal of Econometrics, 94,181-238.

[4] Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," Journal of Political Economy, 81, 637-659.

[5] Breeden, D., and R. Litzenberger, 1978, "Prices of State-Contingent Claims Implicit in Option Prices," Journal of Business, 51, 621-651.

[6] Brown , P., and J.C. Jackwerth, 2000, "The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory," Working Paper, University of Wisconsin at Madison, Finance Department, School of Business.

[7] Constantinides, G.M, 1982, "Intertemporal Asset Pricing with Heterogeneous Consumers and without Demand Aggregation," Journal of Business, 55, 253-267.

[8] Epstein, L., and S. Zin, 1989, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theorical Framework," Econometrica 57, 937-969.

[9] Garcia, R., E. Ghysels, and E. Renault, 2003, "The Econometrics of Option Pricing," in Handbook of Financial Econometrics, Y. Ait-Sahalia and L.P. Hansen (eds) North Holland (forthcoming).

[10] Garcia, R., R. Luger, and E. Renault, 2001, "Asymmetric Smiles, Leverage Effects and Structural Parameters," Working Paper Cirano 2001s-01.

[11] Garcia, R., R. Luger, and E. Renault, 2003, "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables," Journal of Econometrics, 116, 49-83.

[12] Gordon, S., and P. St-Amour, 2000, "A Preference Regime Model of Bull and Bear Markets," The American Economic Review, 90(4), 1019-1033.

[13] Hansen, L., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," Journal of Finance, 52(2), 557-590.

[14] Hansen, L., and S. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," Econometrica 55, 587-613.

[15] Jackwerth, J. C., 2000, "Recovering Risk Aversion From Option Prices and Realized Returns," Review of Financial Studies, 13(2), 433-451.

[16] Jackwerth, J. C., and M. Rubinstein, 2004, "Recovering Probabilities and Risk Aversion from Option Prices and Realized Returns," Essays in Honor of Fisher Black, editor: Bruce Lehmann, Oxford University Press.

[17] Jackwerth, J. C., and M. Rubinstein, 1996, "Recovering Probability Distributions from Contemporary Security Prices," Journal of Finance, 51, 1611-1631.

[18] Mehra, R., and E. C. Prescott, 1985, "The Equity Premium: A Puzzle," Journal of Monetary Economics, 15(2), 145-161.

[19] Melino, A., and A. X. Yang, 2003, "State Dependent Preferences Can Explain the Equity Premium Puzzle," Review of Economic Dynamics, 6(2), 806-830.

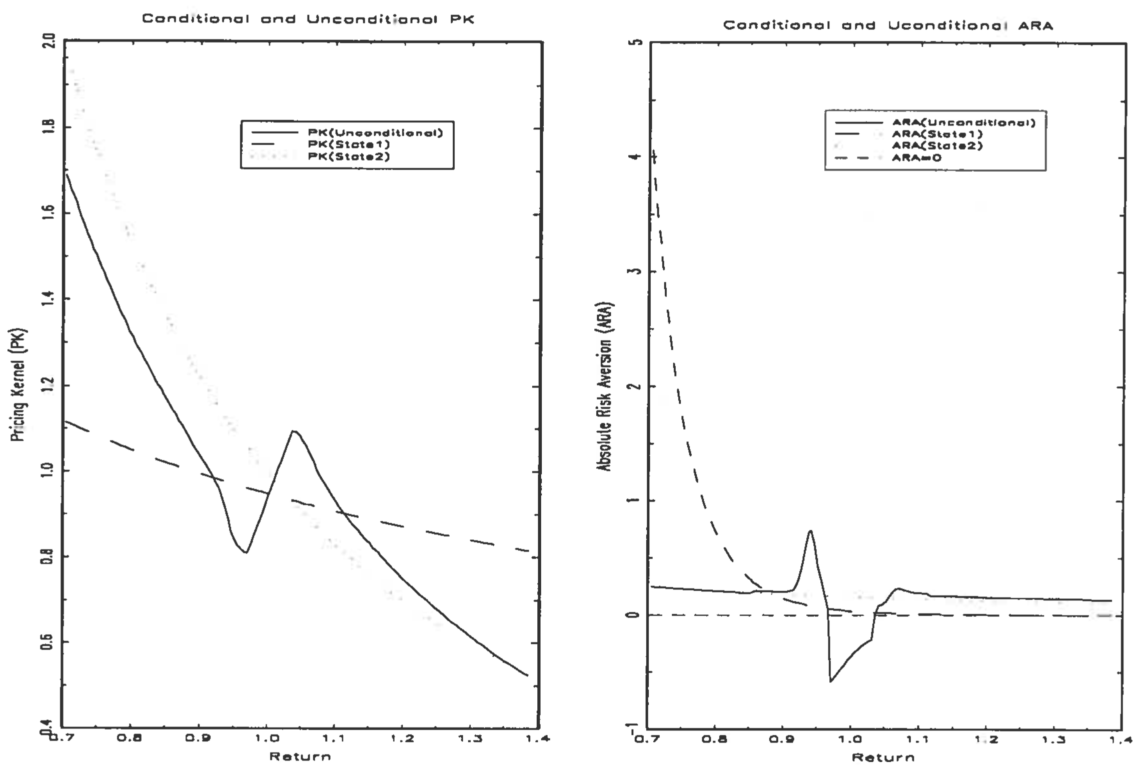


Figure 3.1: **Absolute Risk Aversion (ARA) and Pricing Kernel (PK) functions with state dependence in fundamentals.** The preference parameters are: $\beta = 0.95$, $\alpha = -5$, $\rho = -11$. The regime probabilities are: $p_{11} = 0.9$, $p_{00} = 0.6$. For the economic fundamentals, the means of the consumption growth rate are $\mu_{X_{t+1}} = (0.0015, -0.0009)$, and the corresponding standard deviations $\sigma_{X_{t+1}} = (0.0159, 0.0341)$. For the dividend rate, the parameters are $\mu_{Y_{t+1}} = (0, 0)$, $\sigma_{Y_{t+1}} = (0.02, 0.12)$. The correlation coefficient between consumption and dividends is 0.6. The number of options used is 50. The number of wealth states is $n = 170$. The left-hand panel contains the conditional and unconditional PK functions across wealth states. The right-hand panel contains the conditional and unconditional ARA functions across wealth states. The conditional ARA (PK) function is the ARA (PK) function computed within each regime. The unconditional ARA (PK) function is the ARA (PK) function computed when regimes are not observed.

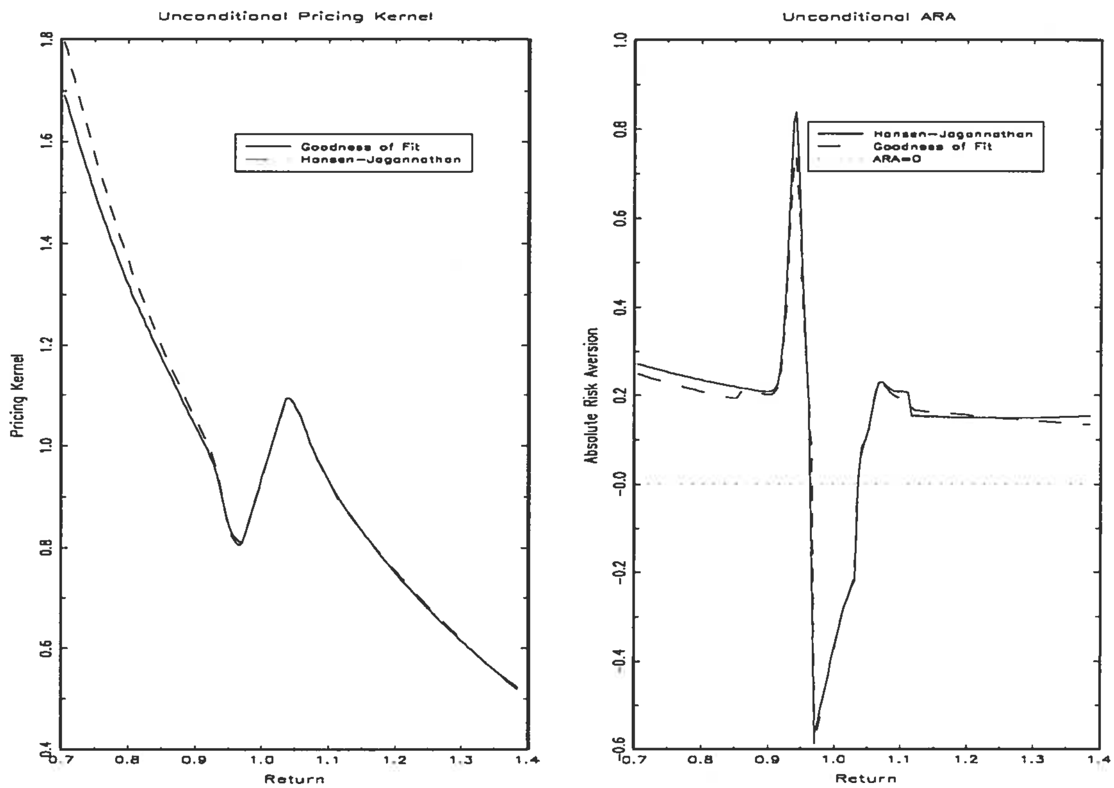


Figure 3.2: **Absolute Risk Aversion (ARA) and Pricing Kernel (PK) functions with state dependence in fundamentals:** The preference parameters are: $\beta = 0.95$, $\alpha = -5$, $\rho = -11$. The regime probabilities are: $p_{11} = 0.9$, $p_{00} = 0.6$. For the economic fundamentals, the means of the consumption growth rate are $\mu_{X_{t+1}} = (0.0015, -0.0009)$, and the corresponding standard deviations $\sigma_{X_{t+1}} = (0.0159, 0.0341)$. For the dividend rate, the parameters are $\mu_{Y_{t+1}} = (0, 0)$, $\sigma_{Y_{t+1}} = (0.02, 0.12)$. The correlation coefficient between consumption and dividends is 0.6. The number of options used is 50. The number of wealth states is $n = 170$. The left-hand panel contains the unconditional PK function across wealth states for the Goodness-of-fit and the Hansen and Jagannathan (1997) distance measures. The right-hand panel contains the unconditional ARA function across wealth states for the Goodness of Fit and the Hansen and Jagannathan (1997) distance measures. The unconditional ARA (PK) function is the ARA (PK) function computed when regimes are not observed.

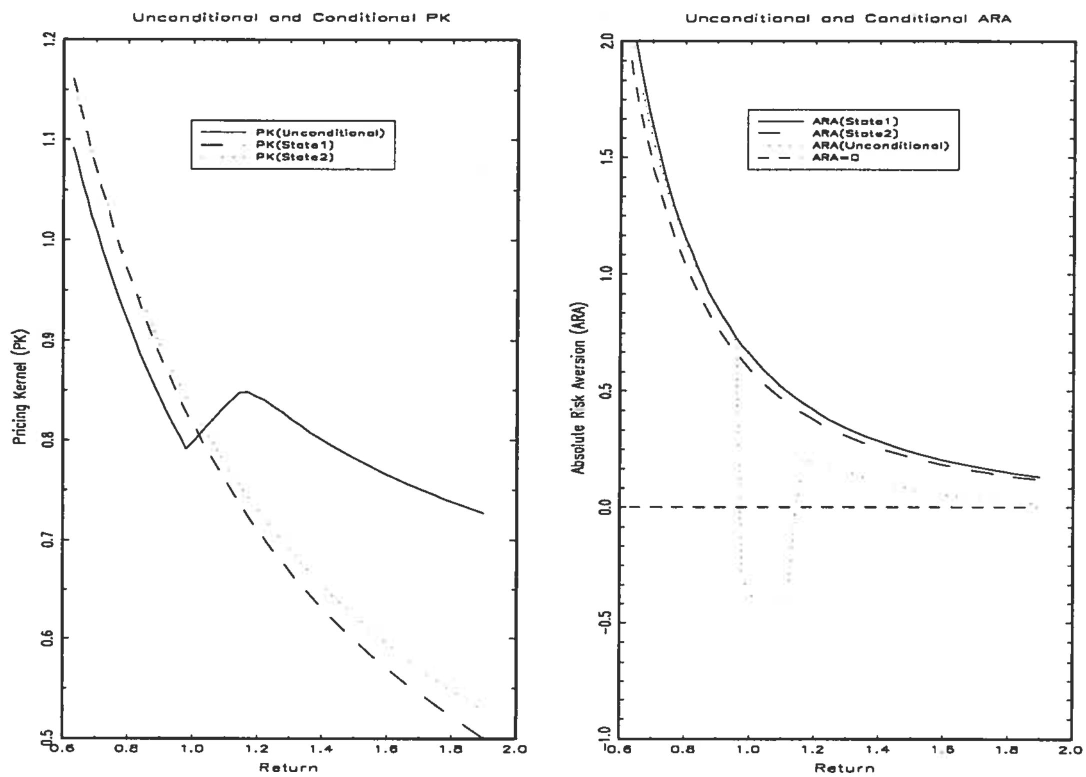


Figure 3.3: **Absolute Risk Aversion (ARA) and Pricing Kernel (PK) functions with state dependence in preferences.** The preference parameters are $\beta = 0.97$, $\alpha = (-7, -4.8)$, $\rho = -10$. The regime probabilities are $p_{11} = 0.9$, $p_{00} = 0.6$. For the economic fundamentals, the means of the consumption growth rate is $\mu_{X_{t+1}} = 0.018$ and the standard deviations $\sigma_{X_{t+1}} = 0.037$. For the dividend rate Y_{t+1}, the parameters are $\mu_{Y_{t+1}} = -0.0018$, $\sigma_{Y_{t+1}} = 0.12$. The correlation coefficient between consumption and dividend is 0.6. The number of options used is 50. The number of wealth states is $n = 170$. The left-hand panel contains the conditional and unconditional PK functions across wealth states. The right-hand panel contains the conditional and unconditional ARA functions across wealth states. The conditional ARA (PK) function is the ARA (PK) function computed within each regime. The unconditional ARA (PK) function is the ARA (PK) function computed when regimes are not observed.

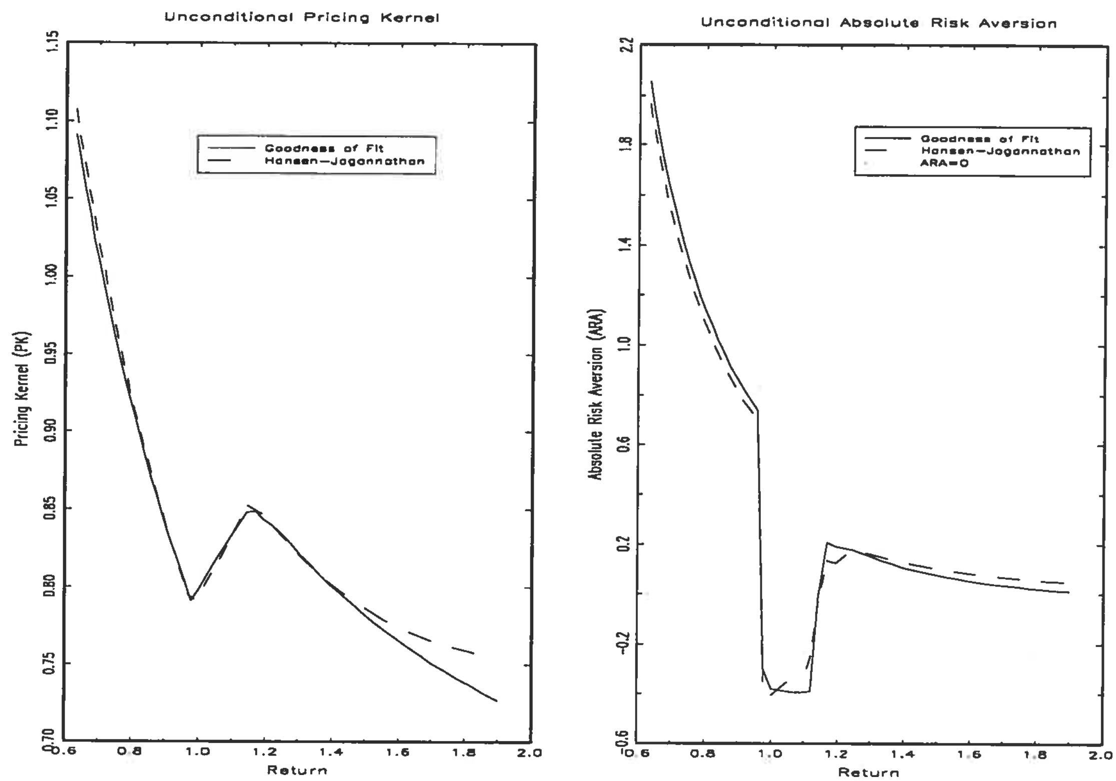


Figure 3.4: **Absolute Risk Aversion (ARA) and Pricing Kernel (PK) functions with state dependence in preferences.** The preference parameters are $\beta = 0.97$, $\alpha = (-7, -4.8)$, $\rho = -10$. The regime probabilities are $p_{11} = 0.9$, $p_{00} = 0.6$. For the economic fundamentals, the means of the consumption growth rate is $\mu_{X_{t+1}} = 0.018$ and the standard deviations $\sigma_{X_{t+1}} = 0.037$. For the dividend rate Y_{t+1}, the parameters are $\mu_{Y_{t+1}} = -0.0018$, $\sigma_{Y_{t+1}} = 0.12$. The correlation coefficient between consumption and dividend is 0.6. The number of options used is 50. The number of wealth states is $n = 170$. The left-hand panel contains the unconditional ARA function across wealth states for the Goodness-of-fit and the Hansen and Jagannathan (1997) distance measures. The right-hand panel contains the unconditional ARA function across wealth states for the Goodness of Fit and the Hansen and Jagannathan (1997) distance measures. The unconditional ARA (PK) function is the ARA (PK) function computed when regimes are not observed.

Chapter 4

Valuing Derivatives with a Trinomial Tree: A State Variable Approach

1. Introduction

Since Black and Scholes (1973), the empirical option pricing literature about European options on equity affords at least two kinds of extensions. First, while keeping the Black -Scholes and Merton continuous time paradigm, some authors have focused on problem surrounding the volatility parameter. While its is assumed to be constant in Black and Scholes initial geometric Brownian motion model, Hull and White (1987) considers that it actually follows a stationary process the current value of which belongs to the investors' information set. In this setting, the option price can be written as an expectation of the Black-Scholes price where the volatility parameter is replaced by the square root of a time average of the squared volatility process over the lifetime of the option and the expectation is conditional to the current value of the volatility process. Renault and Touzi(1996) and Renault (1997) have shown that this setting implies a symmetric volatility smile when implied Black and Scholes (BS) volatilities are backed out from observed option prices. Several extensions including leverage effect, multi-factors volatility process, long memory or jumps have addressed the issue of fitting better the observed volatility surfaces along the two dimensions of strike price and maturity. Irrespective of the detailed specification, a common feature of all these option pricing models is that there are some latent volatility factors the current value of which is assumed to be known to investors while not observed by the econometrician.

A second strand of literature, following the Cox, Ross and Rubinstein binomial reinterpretation of Black and Scholes, replaces the continuous time setting by a binomial tree. Generally speaking, the lattice kind of approach is more flexible than the diffusion model to accommodate complicated payoffs schedules. Of course, the simplest binomial tree is nothing but a discrete time approximation of the geometric Brownian motion and the states can be calibrated to ensure that the lattice option pricing model converges towards Black and Scholes when the time interval between two binomial draws goes to zero. Moreover, Boyle (1988) and Kamrad and Ritchken (1991) have shown that by considering multinomial trees, one can accommodate higher dimensions of uncertainty; market incompleteness as captured by a latent volatility risk that precludes the perfect hedge of an option contract by replicating it with a portfolio on the underlying asset and the risk-free asset can also be captured in a lattice framework with trinomial trees. However, the trinomial approach differs from the continuous time stochastic volatility model a la Hull and White (1987) by the fact that it erases the informational role of option prices as revealing the market assessment of unobserved current volatility.

The main goal of this paper is to bridge this gap. We will show that this is an important issue, not only for the econometric use of volatility assessments backed out from option prices, but also for option pricing and hedging itself. Actually, while our lattice based option pricing model will not differ from the trinomial one over one period, the difference will matter as soon as one is interested in pricing longer term options. The basic intuition is the following: if investors know something about the latent state that the econometrician does not know, they take advantage of this knowledge to forecast the future state better than the econometrician does and the resulting option prices are influenced by this knowledge.

To formalize this idea, we refer to the Markov switching literature do describe the latent uncertainty by a two-state Markov chain. When this binomial state interacts with a binomial tree, it will produce a lattice which is at first sight observationally equivalent to a trinomial or quadrinomial lattice. But over to periods, there is an important difference: when the investor knows the current realization of the state, he knows which binomial tree among two possible ones will be drawn. Over one period, this always defines a world with three or four possible sates. But, over two periods, the persistence of the Markov chain can be exploited to make better assessments of the probabilities of the various branches of the tree. This is the reason why the maintained assumption that the investor can eventually observe the realization of the state is important. Note that we are not the first to assume that the state is known to investors but not to econometricians. This is of course the case with state dependent preferences (Melino and Yang (2003)) and this is also the setting of option pricing models considered by Garcia , Luger and Renault (2001), (2003). Generally speaking, the motivation behind Markov switching regimes is that when the economic environment changes, the data generating process of the related financial variables also changes. One example is that the period of high volatility of the US short term interest rate coincides with change in the economic and political environments due to the October 1987 stock market crash, see for instance Hamilton (1990) and Gray (1996). Latent state variables models seek to capture such discrete shifts in the behavior of the financial variables by allowing the parameters of the underlying data generating process to take on different values in different time periods. While several papers have already addressed the issue of option pricing with Markov switching regimes, see Bollen (1998) among others, it turns out that the focus on the information content of option prices has still not be sufficiently put in the general framework of lattice pricing as it has been done for continuous time models. This paper bridges this gap.

The paper makes three contributions. First, we revisit Boyle's (1988) option pricing approach with the stochastic discount factor (hereafter SDF). By a stochastic discount factor, we mean a random variable that can be used to compute the market price of an asset today by discounting payoffs differently in the futures states of the world. Without building any replicating strategies, we show how the SDF can be used to derive the underlying risk neutral probabilities across wealth states. To do this, we use a fundamental valuation equation

$$Emg = \pi,$$

where g is the payoff of a traded derivative, π is the price of g and m is known as a stochastic discount factor [see e.g., Hansen and Richard (1987)]. The underlying asset risk neutral probabilities derived with a SDF coincides with the Boyle (1988) risk-neutral probabilities when the difference between the underlying historical and risk-neutral variance is null. We term this difference the risk neutral variance premium. This premiun can be intuitively interpreted as the market price of variance risk. Guo (1998) provides an empirical investigation of the risk-neutral variance process and the market price of variance risk implied in the foreign-currency options market. Assuming that the joint process of the asset return-SDF is lognormally distributed, this premium is not null, we show that the risk neutral probabilities are sensitive to small changes in the risk neutral variance premium.

Second, we develop a lattice trinomial tree to handle the situation in which the payoff from derivatives is affected by one latent state variable through the underlying asset (it is possible to extend this work to situation involving a higher number of state variables). Conditionally on this variable, we also assume (for simplicity) that the underlying asset follows a two-point jump process. We then provide the way to price derivatives and give under what condition (s), our pricing approach is observationally equivalent to Boyle (1988) option pricing approach. We, thereafter, extend our pricing approach in two-period (we have three dates, 0, 1 and 2).

The plan of the rest of the paper is as follows. Section 2 of the paper describes Boyle's (1988) approach to getting the underlying risk-neutral probabilities across wealth states and shows how these probabilities can be obtained in a stochastic discount factor framework. We, thereafter, propose a trinomial tree with state variable and show (in one-period) that this tree is equivalent to Boyle's (1988) trinomial tree. Section 3 of the paper extends the result of Section 2 in two-period and investigate under what conditions the TTSV is equivalent to Boyle (1988) trinomial tree. The last section concludes the paper.

2. One-Period Tree

In this section, the underlying asset follows a three-point jump process. Using this assumption, we characterize the structure of the SDF and derive the underlying-asset risk-neutral probabilities across wealth states. We, thereafter, propose an alternative trinomial tree where an unobservable state variable affects the underlying process and show that this tree is observationally equivalent to Boyle's (1988) trinomial tree (hereafter BTT).

2.1 Revisiting the Boyle trinomial model with a SDF

Over a small time interval, Boyle (1988) approximates the underlying return process by a three-point jump process:

$$\frac{S_{t+1}}{S_t} = u 1_{\frac{S_{t+1}}{S_t}=u} + 1_{\frac{S_{t+1}}{S_t}=1} + d 1_{\frac{S_{t+1}}{S_t}=d}, \tag{2.1}$$

with $u > d$. Without loss of generality, we assume that $t = 0$. The underlying probabilities across wealth states are represented by:

$\underline{Moves}$	$\underline{Asset\ Values}$	$\underline{Probabilities\ of\ Events}$
Up	$u.S_t$	p_1
$Horizontal$	S_t	p_2
$Down$	$d.S_t$	p_3

The probabilities of events p_1, p_2 and p_3 can be computed if the underlying expected return and variance are known. Let p_1^*, p_2^* and p_3^* represent the probabilities of events in a risk neutral world. Boyle (1988) imposes three conditions to solve for these probabilities in terms of u and other variables:

(i) the risk neutral probabilities sum to one,

(ii) under the risk neutral world, the mean of the discrete distribution, μ, is equal to the mean of the continuous lognormal distribution:

$$E_t^*\left[\frac{S_{t+1}}{S_t}\right] = \mu, \text{ and}$$

(iii) under the risk neutral world, the variance of the discrete distribution, $\sigma_t^{*2} = Var_t^*\left[\frac{S_{t+1}}{S_t}\right]$, is equal to the historical variance, σ_t^2, of the continuous lognormal distribution

$$\sigma_t^{*2} = \sigma_t^2.$$

So that the first two moments of the variable's return implied by the lattice match the first two moments implied by the underlying distribution. Boyle (1988) uses these equalities to compute the risk-neutral probabilities across wealth states:

$$\begin{aligned} p_1^* &= \frac{(\sigma_t^2+\mu^2-\mu)u-(\mu-1)}{(u-1)(u^2-1)}, \\ p_3^* &= \frac{(\sigma_t^2+\mu^2-\mu)u^2-u^3(\mu-1)}{(u-1)(u^2-1)}, \\ p_2^* &= 1-p_1^*-p_3^*. \end{aligned} \tag{2.2}$$

Therefore, to compute the price of derivatives in BTT, we need the underlying historical variance (σ_t^2). Before we stress the implications of this restriction when valuing derivatives, we give an alternative way to price derivatives.

Consider a derivative, of which payoff is $g(S_{t+1})$. This payoff is quadratic in the underlying-asset because the underlying-asset follows a three-point jump process as specified in (2.1). We rewrite this payoff as:

$$g(S_{t+1}) = g_0 + g_1 S_{t+1} + g_2 S_{t+1}^2,$$

where, g_0, g_1, and g_2 are functions of the information set that investors use in buying or selling derivatives at date t. Of course, the coefficients g_0, g_1, and g_2 are solution to:

$$\begin{bmatrix} 1 & u.S_t & u^2 S_t^2 \\ 1 & S_t & S_t^2 \\ 1 & d.S_t & d^2 S_t^2 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} g(u.S_t) \\ g(S_t) \\ g(d.S_t) \end{bmatrix}.$$

According to the Hansen and Richard (1987) framework, the price of $g(S_{t+1})$ is:

$$\pi_t = E_t \left[m_{t+1} g(S_{t+1}) \right],$$

where m_{t+1} represents the SDF. We, therefore, rewrite this expression as:

$$\pi_t = E_t \left[m_{t+1}^* g(S_{t+1}) \right], \tag{2.3}$$

with $m_{t+1}^* = E_t [m_{t+1} | S_{t+1}]$. Using either m_{t+1}^* or m_{t+1} leads to the same pricing formula, since

$$E_t \left[\left(m_{t+1} - m_{t+1}^* \right) g(S_{t+1}) \right] = 0.$$

The random variable m_{t+1}^* can be interpreted as a SDF. This SDF follows a three-point jump because the underlying-asset also follows a three-point jump process. Thus, it seems reasonable to write m_{t+1}^* as a quadratic function in S_{t+1}:

$$m_{t+1}^* = a^* + b^* S_{t+1} + c^* S_{t+1}^2. \tag{2.4}$$

From equation (2.4), the coefficient c^* could be interpreted as a skewness parameter. Harvey and Siddique (2000), Dittmar (2002) among others demonstrate that a quadratic SDF introduces skewness in asset pricing. They show that asset skewness (co-skewness) is an important factor that explains asset expected returns. For instance, Harvey and Siddique (2000) show that investors are willing to accept negative expected return in presence of high positive skewness. Expression (2.4) shows how the skewness premium enters the trinomial tree. However, expression (2.4) can be used to price only derivatives of S_{t+1}. Substituting $g(S_{t+1})$ in (2.3) gives

$$\pi_t = g_0 (1 + r_f)^{-1} + g_1 S_t + g_2 E\left[m^*_{t+1} S^2_{t+1}\right]. \tag{2.5}$$

with $\mu = 1 + r_f$, r_f is the risk free rate. For sake of notational convenience, we use the following notation

$$E_t\left[m^*_{t+1} S^2_{t+1}\right] = \eta S^2_t,$$

where the parameter η can be rewritten as:

$$\eta = \left(Em^*_{t+1}\right) E_t\left[\frac{m^*_{t+1}}{Em^*_{t+1}} \frac{S^2_{t+1}}{S^2_t}\right] = \frac{\sigma^{*2}_t}{(1 + r_f)} + (1 + r_f),$$

where the specification of the risk-neutral measure dQ through the SDF change of measure is $\frac{m^*_{t+1}}{E_t m^*_{t+1}} dP$, dP represents the historical measure. If (p^*_1, p^*_3) denotes the solution to equations

$$S_t = \frac{1}{1 + r_f}\left(p^*_1 u.S_t + p^*_3 d.S_t + (1 - p^*_1 - p^*_3) S_t\right)$$

and

$$\eta S^2_t = \frac{1}{1 + r_f}\left(p^*_1 (u.S_t)^2 + p^*_3 (d.S_t)^2 + (1 - p^*_1 - p^*_3)(S_t)^2\right),$$

equation (2.5) can be rewritten as

$$\pi_t = g_0 (1 + r_f)^{-1} + g_1 S_t + g_2 \eta S^2_t = (1 + r_f)^{-1} E^*_t\left[g(S_{t+1})\right], \tag{2.6}$$

where $E^*_t(g)$ represents the expectation of g with respect to the underlying-asset risk-neutral probabilities (p^*_1, p^*_2, p^*_3):

$$\begin{aligned} p^*_1 &= \frac{(\sigma^2_t + \mu^2 - \mu - \delta)u - (\mu - 1)}{(u - 1)(u^2 - 1)}, \\ p^*_3 &= \frac{(\sigma^2_t + \mu^2 - \mu - \delta)u^2 - u^3(\mu - 1)}{(u - 1)(u^2 - 1)}, \\ p^*_2 &= 1 - p^*_1 - p^*_3, \end{aligned} \tag{2.7}$$

where $\delta = \sigma^2_t - \sigma^{*2}_t$ represents the risk-neutral variance premium. These risk-neutral probabilities are obtained without building any replicating portfolio strategy. The same technique can be applied to the well-known Cox, Ross and Rubinstein (1979) binomial model by assuming that the underlying asset follows a two-point jump process. Comparing our risk-neutral probabilities to the Boyle risk-neutral probabilities, we conclude that both the probabilities across wealth states are equal if and only if $\delta = 0$. This will certainly introduce errors in the computation of the true risk-neutral probabilities if $\delta \neq 0$. One example is that if the joint process underlying return-stochastic discount factor is conditionally lognormally distributed, the price of the squared underlying return is:[1]

$$\eta = E_t\left[m^*_{t+1}R^2_{t+1}\right] = \frac{1}{Em^*_{t+1}}\frac{E_t\left(R^2_{t+1}\right)}{\left(E_tR_{t+1}\right)^2},$$

with $R_{t+1} = \frac{S_{t+1}}{S_t}$. This last equality can be used to compute the underlying-asset risk-neutral variance:

$$\sigma^{*2}_t = \frac{\sigma^2_t}{\lambda^2}, \tag{2.8}$$

where $\lambda = 1 + \frac{E_t(R_{t+1}) - (1+r_f)}{(1+r_f)}$. In that case,

$$\delta = \sigma^2_t\left[1 - \frac{1}{\lambda^2}\right].$$

If $\lambda \neq 1$, the risk-neutral variance premium δ is not null. When the underlying return-stochastic discount factor is conditionally lognormally distributed, it follows that the underlying marginal distribution is also lognormally distributed. Over a small time interval, this return can be approximated by a three-point jump process [see Boyle (1988)]. Therefore, over a small time interval, equation (2.8) is still valid when the underlying return is approximated by a three-point jump process.

To compute the risk-neutral probabilities, recall that Cox, Ross and Rubinstein (1979) use

$$u = \exp\left[\sigma^*_t\sqrt{h}\right].$$

When this last expression is used to compute (2.2), the risk-neutral probabilities are sometime negative [see e.g., Boyle (1988)]. Instead of using this last equation to compute the risk-neutral probabilities, Boyle (1988) assumes

$$u = \exp\left[\lambda\sigma^*_t\sqrt{h}\right],$$

[1]See Chabi-Yo, Garcia and Renault (2003).

where h is the length of one time step (in this article, without loss of generality, we assume $h = 1$), and $\lambda > 1$. Boyle's intuition can be argued as follows: under the historical underlying-asset measure, approximating the lognormal distribution by a three-point jump process over a small time interval involves considering

$$u = \exp\left[\sigma_t \sqrt{h}\right].$$

From this last expression, if we replace the underlying historical variance, σ_t, by its expression given in (2.8), we get $u = \exp\left[\sigma_t \sqrt{h}\right] = \exp\left[\lambda \sigma_t^* \sqrt{h}\right]$. However in Boyle (1988), the parameter λ is exogenous. From the above example, explicit formula for λ is given when the joint process of underlying asset return-stochastic discount factor is lognormally distributed. In that case, the parameter λ depends on the underlying asset risk premium, $E_t\left(R_{t+1}\right) - (1 + r_f)$. High underlying risk premium implies high value for λ. For example, suppose $\sigma^* = 0.2$, $r_f = 0.1$, $h = 1$, Table 4.1 displays the values of u and the corresponding probabilities for a range of values of the underlying risk premium. Zero risk premium is not realistic when the joint process underlying asset return is lognormally distributed. This explains why the probability p_2^* is negative. It can be observed through Table 4.1 that the risk neutral probabilities highly depend on the underlying risk premium. In the next subsection, we propose a lattice model with an unobservable state variable and give conditions under which this model is equivalent to BTT.

2.2 Trinomial tree with state variable

We assume the underlying-asset process is affected by an unobservable variable namely U_{t+1}. This variable is not observed by investors at date $t = 0$ but is disclosed to them at date $t + 1$. Without loss of generality, we assume that this unobservable variable follows a two-point jump process, that is, $U_{t+1} = 0$ or 1. We also assume $U_0 = 1$. The random process (U_t) which affects the underlying process is a discrete first order homogeneous Markov chain with a transition matrix of the form:

$$P = \begin{bmatrix} \alpha_{11} & 1 - \alpha_{11} \\ 1 - \alpha_{00} & \alpha_{00} \end{bmatrix}.$$

Given U_{t+1}, we also assume that the underlying process follows a two-point jump process:

Regime 1 $(U_1 = 1|U_0 = 1)$ *Regime* 0 $(U_1 = 0|U_0 = 1)$

Moves	*Asset Values*	*Probabilities*	*Asset Values*	*Probabilities*
Up	$u(1).S_t$	p_{11}	$u(0).S_t$	p_{10}
Down	$d(1).S_t$	$1-p_{11}$	$d(0).S_t$	$1-p_{10}$

Assuming $u(1) = u$, $d(1) = 1 = u(0)$ and $u(0) = d(0)$, the underlying process follows a three-point jump process and the underlying probabilities across wealth states are characterized by equations below:

$$\begin{aligned} p_{11}P(U_1 = 1|U_0 = 1) &= p_1, \\ (1-p_{10})\,P(U_1 = 0|U_0 = 1) &= p_3, \\ (1-p_{11})\,P(U_1 = 1|U_0 = 1) + p_{10}P(U_1 = 0|U_0 = 1) &= p_2. \end{aligned}$$

The tree implied by this unobservable state variable will be referred to as a Trinomial Tree with State Variable (TTSV). The difference between this tree and the BTT is that, at date 1, investors are informed about the value of U_1 whereas in the BTT, investors are not informed. The TTSV and BTT produce identical wealth states. Although, it is important to investigate under what conditions the TTSV is observationally equivalent to BTT. To address this issue, we first compute the price of $g(S_1)$ using the TTSV.

Theorem 2.1 *In one-period, the price of a derivative with payoff* $g(S_1)$ *is:*

$$\pi_0 = E^*\left[C(U_1)\right],$$

with

$$C(U_1) = \left[\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)}\right]^{-1} E^*_{U_1}\left[g(S_1)\right],$$

where $\widetilde{B}_{(1,U_1)}(0,1) = E_0\left[m_1|U_1\right]$ *and* $Q_{(1,U_1)}(0,1) = E_0\left[m_1\frac{S_1}{S_0}|U_1\right]$,

- $E^*(.)$ *represents the expectation operator under the risk-neutral transition probability:*

$$P^*(U_1 = i|U_0 = 1) = Q_{(1,i)}(0,1)\, P(U_1 = i|U_0 = 1) \;\; for\ i = 0, 1.$$

- $E^*_{U_1}(.)$ *represents the expectation operator under the pseudo-risk-neutral probability* $\left(p^*_{1U_1}, 1 - p^*_{1U_1}\right)$:

$$p^*_{1U_1} = \frac{\frac{Q_{(1,U_1)}(0,1)}{\bar{B}_{(1,U_1)}(0,1)} - d(U_1)}{u(U_1) - d(U_1)} \;\; for\ i = 0, 1,$$

with $\frac{Q_{(1,U_1)}(0,1)}{\bar{B}_{(1,U_1)}(0,1)} > d(U_1)$.

From Theorem 2.1, the ratio $Q_{(1,i)}(0,1)$ shows that there exits a risk premium associated to the state variable U_1. This premium is measured through changes in this ratio. The next proposition shows that the TTSV is observationally equivalent to BTT.

Proposition 2.2 *In a one-period, the TTSV is observationally equivalent to BTT.*

In section 3, we extend the results of section 2 in two-period and derive conditions under which this model is observationally equivalent to BTT.

3. Two-Period Extension

This section extends the TTSV into two-period and provides a procedure for valuing derivatives. In this two-period, the state variable process (U_t) captures the salient features of derivatives such as options, in particular skewness and the dynamic effect of asset skewness. We describe conditions under which the TTSV is observationally equivalent to BTT.

3.1 The trinomial lattice description

We assume there is 3 dates, 0, 1 and 2. In BTT, at date t+1, the underlying asset return $\frac{S_{t+1}}{S_t}$ takes on 3 values: u, 1 and d whereas in the TTSV, given the conditioning variable U_{t+1}, the underlying asset return takes on two values $u(U_{t+1})$ and $d(U_{t+1})$. If J_t denotes the information set that investors use in buying or selling derivatives at date t in BTT,

$$J_t = \sigma\left\{(m_\tau, \frac{S_\tau}{S_{\tau-1}}), \tau \leq t\right\}.$$

Within the TTSV, the relevant information set is

$$I_t = \sigma\left\{(m_\tau, \frac{S_\tau}{S_{\tau-1}}, U_\tau), \tau \leq t\right\}.$$

It is straightforward to see that $J_t \subset I_t$ which means that investors are more informed in the TTSV than in BTT. In this sense, the TTSV allows investors to be fully informed about the trajectory of the state variable process (U_t) whereas in BTT, investors are not informed about this trajectory. Figure 4.1 describes the TTSV in two-period. To value derivatives, we need two assumptions:

- **Assumption A1**: The variables $\left(m_{\tau+1}, \frac{S_{\tau+1}}{S_\tau}\right)_{t\leq\tau\leq t+1}$ are conditionally serially independent given the path $U_1^{t+2} = (U_\tau)_{t+1\leq\tau\leq t+2}$ of a vector U_t of state variables,
- **Assumption A2**: The process $\left(m_{\tau+1}, \frac{S_{\tau+1}}{S_\tau}\right)_{t\leq\tau\leq t+1}$ does not cause the process (U_t).

Under assumptions **A1** and **A2**, the state variables summarize the dynamic effect of the underlying-asset skewness. To see this, let r_{t+1} denotes $Log\left(R_{t+1}\right)$. It follows that

$$Cov_0\left(r_1, r_2^2\right) = E_0\left[Cov_0\left(r_1, r_2^2|U_1^2\right)\right] + Cov_0\left(E\left(r_1|U_1^2\right), E\left(r_2^2|U_1^2\right)\right).$$

Under assumptions **A1** and **A2**, it follows that $Cov_0\left(r_1, r_2^2\right) = Cov_0\left(E\left(r_1|U_1\right), E\left(r_2^2|U_1^2\right)\right)$: the TTSV captures the salient features of derivatives, in particular skewness and the dynamic effect of asset skewness. In section 3.2, we provide a procedure for valuing derivatives in the TTSV and give under what conditions the TTSV is observationally equivalent to Boyle's trinomial tree.

3.2 Valuation of derivatives

Consider a derivative with payoff $g\left(S_2\right)$ at date 2. Theorem 3.1 gives the price of $g\left(S_2\right)$ at date 0.

Theorem 3.1 *In two-period, under assumptions* ***A1*** *and* ***A2****, the price of a derivative with payoff* $g\left(S_2\right)$ *is:*

$$\pi_t = E_{tr}^*\left(C\left[g\left(S_2\right)\right]\right),$$

with

$$C\left[g\left(S_2\right)\right]=\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1}\left[\frac{Q_{(1,U_1)}\left(0,1\right)}{\widetilde{B}_{(1,U_1)}\left(0,1\right)}\right]^{-1}E^*_{U_1}\left[E^*_{U_1^2}\left(g\left(S_2\right)\right)\right],$$

where $\widetilde{B}_{U_1^2}\left(1,2\right)=E_0\left[m_2|U_1^2\right]$, $Q_{U_1^2}\left(1,2\right)=E_0\left[m_2\frac{S_2}{S_1}|U_1^2\right]$ *and* $E^*_{tr}\left(.\right)$ *is the expectation under the risk-neutral probabilities:*

$$P^*\left[U_2=i,U_1=j|U_0=1\right]=P^*\left[U_2=i|U_1=j\right]\times P^*\left[U_1=j|U_0=1\right]\quad\forall\; i,j=0,1,$$

with

$$Q_{U_1^2}\left(1,2\right)=\frac{P^*\left[U_2=i|U_1=j\right]}{P\left[U_2=i|U_1=j\right]},$$

where $E^*_{U_1^2}\left(.\right)$ *is the expectation under the pseudo-risk-neutral probabilities:*

$$P^*\left[\frac{S_2}{S_1}=u(i)|U_1^2=(j,i)\right]=\frac{\frac{Q_{(j,i)}(1,2)}{B_{(j,i)}(1,2)}-d(i)}{u\left(i\right)-d\left(i\right)}.$$

$E^*_{U_1}\left(.\right)$ *is defined in theorem 2.1.*

Now, we state under what conditions the TTSV is observationally equivalent to BTT.

Proposition 3.2 *Under assumptions A1 and A2, if the the conditioning distribution of* (m_2,S_2) *given the set* (m_1,S_1,U_1) *equals the conditioning distribution of* (m_2,S_2) *given* S_1*, the TTSV is observationally equivalent to BTT.*

4. Conclusion

This paper develops a lattice for valuing derivatives when the underlying process is affected by an unobservable state variable. This model generalizes the existing lattice models by Cox, Ross and Rubinstein (1979) and Boyle's (1988) trinomial pricing model. In a future research, we intend to establish price convergence from this discrete framework to the case of continuous underlying asset.

5. Appendix: Proofs

PROOF OF THEOREM 2.1. At date 0, the price of $g(S_1)$ is

$$\pi_0 = E_0[m_1 g(S_1)] = E_0[E_0[m_1 g(S_1)|U_1]]$$

But,

$$E_0[m_1 g(S_1)|U_1] = E_0[E_0[m_1 g(S_1)|U_1, S_1]|U_1] = E_0[g(S_1) E_0[m_1|U_1, S_1]|U_1]$$

Consequently,

$$\pi_0 = E_0[E_0(m_1^*(U_1, S_1) g(S_1)|U_1)], \tag{5.9}$$

with

$$m_1^*(U_1, S_1) = E_0[m_1|U_1.S_1] = a(U_1) + b(U_1)\frac{S_1}{S_0},$$

where,

$$\begin{aligned} b(U_1) &= \widetilde{B}_{(1,U_1)}(0,1)\frac{\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)} - E_0\left(\frac{S_1}{S_0}|U_1\right)}{Var_0\left(\frac{S_1}{S_0}|U_1\right)}, \\ a(U_1) &= \widetilde{B}_{(1,U_1)}(0,1) - b(U_1)E_0\left(\frac{S_1}{S_0}|U_1\right) \end{aligned}$$

and

$$\begin{aligned} Q_{(1,U_1)}(0,1) &= E_0\left[m_1^*(U_1, S_1)\frac{S_1}{S_0}|U_1\right], \\ \widetilde{B}_{(1,U_1)}(0,1) &= E_0[m_1^*(U_1, S_1)|U_1]. \end{aligned}$$

Given the conditioning variable U_1, the underlying S_1 follows a two-point jump process. In that case

$$g(S_1) = g_0(U_1) + g_1(U_1) S_1.$$

This last equality is plugged in $E_0(m_1^*(U_1, S_1) g(S_1)|U_1)$, we then get:

$$E_0[m_1^*(U_1, S_1) g(S_1)|U_1] = \widetilde{B}_{(1,U_1)}(0,1)\left[g_0(U_1) + g_1(U_1) S_0\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)}\right].$$

If we denote $p^*_{1U_1} = \frac{\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)} - d(U_1)}{u(U_1) - d(U_1)}$ with $\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)} > d(U_1)$. This last expression can be rewritten as

$$E_0 \left[m_1^* (U_1, S_1) \, g(S_1) \, | U_1 \right] = \widetilde{B}_{(1,U_1)}(0,1) \left[E^*_{0,(U_1)} (g(S_1)) \right]$$

where $E^*_{U_1}(g)$ represents the expectation of g under the pseudo risk-neutral probability $\left(p^*_{1U_1}, 1 - p^*_{1U_1}\right)$. We plug this last expression in (5.9) and get:

$$\pi_0 = E^* \left[C(U_1) \right] \tag{5.10}$$

with

$$C(U_1) = \left(\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)} \right)^{-1} E^*_{U_1} \left[g(S_1) \right]$$

$E^*(g)$ represents the expectation of g under the risk-neutral transition probabilities

$$Q_{(1,i)}(0,1) = \frac{P^* \left[U_1 = i | U_0 = 1 \right]}{P \left[U_1 = i | U_0 = 1 \right]} \text{ for } i = 0, 1.$$

■

Proof of Proposition 2.2. Let $\pi_{BTT}[g(S_1)]$ be the price at date 0 of $g(S_1)$ in the BTT and $\pi_{TTSV}[g(S_1)]$ be the price at date 0 of $g(S_1)$ in the TTSV. For sake of notational convenience, we denote $P[U_1 = i | U_0 = j] = \alpha_{ji}$ and $P^*[U_1 = i | U_0 = j] = \alpha^*_{ji}$. BTT is equivalent to the TTSV if and only if:

$$\pi_{BTT}[g(S_1)] = \pi_{TTSV}[g(S_1)] \tag{5.11}$$

To investigate if the TTSV and BTT are equivalent, we equate (5.10) and (2.7). Let assume that the right hand side of (5.11) is known, if $g(S_1) = S_1$ and $g(S_1) = S_1^2$, equation (5.11) can be used to compute the underlying risk-neutral probabilities across wealth states in BTT. To see this, notice that

$$\pi_{BTT}[g(S_1)] = B(0,1) \left[p_1^* g(u.S_0) + p_2^* g(1.S_0) + p_3^* g(d.S_0) \right].$$

This last expression in plugged in (5.11),

$$\pi_{TTSV}\left[g\left(S_1\right)\right] = B\left(0,1\right)\left[p_1^* g\left(u.S_0\right) + p_2^* g\left(1.S_0\right) + p_3^* g\left(d.S_0\right)\right].$$

Using $g\left(S_1\right) = S_1$, this last equation reduces to:

$$S_0 = \pi_{TTSV}\left[S_1\right] = B\left(0,1\right)\left[p_1^* u.S_0 + p_2^* S_0 + p_3^* d.S_0\right].$$

If $g\left(S_1\right) = S_1^2$, we also have:

$$\eta S_0^2 = \pi_{TTSV}\left[S_1^2\right] = B\left(0,1\right)\left[p_1^*\left(u.S_0\right)^2 + p_2^*\left(1.S_0\right)^2 + p_3^*\left(d.S_0\right)^2\right]$$

with $p_3^* = 1 - p_1^* - p_2^*$. Solving the last two equations produces the underlying risk-neutral probabilities across wealth states (p_1^*, p_2^*, p_3^*).

Conversely if the left hand side of (5.11) is known, using $g\left(S_1\right) = S_1^2$, $g\left(S_1\right) = S_1$ and expanding the right hand side of (5.11) gives:

$$\begin{aligned}
\eta S_0^2 &= \pi_{BTT}\left[g\left(S_1\right)\right] = \pi_{TTSV}\left[S_1^2\right] = \alpha_{11}^* \frac{p_{11}^*\left(u.S_0\right)^2 + \left(1-p_{11}^*\right)\left(S_0\right)^2}{p_{11}^*\left(u-1\right)+1} + \\
&\quad \left(1-\alpha_{11}^*\right)\frac{p_{10}^*\left(S_0\right)^2 + \left(1-p_{10}^*\right)\left(d.S_0\right)^2}{p_{10}^*\left(1-d\right)+d} \\
S_0 &= \pi_{BTT}\left[g\left(S_1\right)\right] = \pi_{TTSV}\left[S_1\right] = \alpha_{11}^* \frac{p_{11}^* u.S_0 + \left(1-p_{11}^*\right)S_0}{p_{11}^*\left(u-1\right)+1} + \\
&\quad \left(1-\alpha_{11}^*\right)\frac{p_{10}^* S_0 + \left(1-p_{10}^*\right)d.S_0}{p_{10}^*\left(1-d\right)+d}.
\end{aligned}$$

Fixing α_{1i}^*, the above two equations can be solved for p_{10}^* and p_{11}^*. For this particular value α_{1i}^*, the BTT and TTSV are equivalent. This ends the proof. ■

Proof of Theorem 3.1. At date 0, the price of $g\left(S_2\right)$ is

$$\pi_0 = E_0\left[m_1 m_2 g\left(S_2\right)\right].$$

where $m_1 m_2$ is a two-period SDF. This last expression can be decomposed to:

$$E_0\left[m_1 m_2 g\left(S_2\right)\right] = E_0\left[m_1 E_0\left[m_2 g\left(S_2\right) | U_1^2, S_1, m_1\right]\right].$$

with

$$E_0\left[m_2 g\left(S_2\right)|U_1^2, S_1, m_1\right] = E_0\left[m_2 g\left(S_2\right)|U_1^2, S_1, m_1\right],$$

But,

$$E_0\left[m_2 g\left(S_2\right)|U_1^2, S_1, m_1\right] = E_0\left[E_0\left[m_2^*\left(U_1^2, S_2\right) g\left(S_2\right)|U_1^2, S_2, S_1, m_1\right]|U_1^2\right],$$

where

$$m_2^*\left(U_1^2, S_2\right) = E_0\left[m_2|U_1^2, S_2, S_1, m_1\right],$$

Under Assumption **A1**,

$$m_2^*\left(U_1^2, S_2\right) = a\left(U_1^2\right) + b\left(U_1^2\right)\frac{S_2}{S_1},$$

and

$$E_0\left[m_2^*\left(U_1^2, S_2\right) g\left(S_2\right)|U_1^2, S_2, S_1, m_1\right] = E_0\left[m_2^*\left(U_1^2, S_2\right) g\left(S_2\right)|U_1^2, S_2\right]$$

However, it can be shown that:

$$E_0\left[E_0\left[m_2^*\left(U_1^2, S_2\right) g\left(S_2\right)|U_1^2, S_2\right]|U_1^2\right] = \widetilde{B}_{U_1^2}\left(1,2\right) E_{U_1^2}^*\left[g\left(S_2\right)\right],$$

with,

$$\begin{aligned} b\left(U_1^2\right) &= \widetilde{B}_{U_1^2}\left(1,2\right)\frac{\frac{Q_{U_1^2}(1,2)}{\widetilde{B}_{U_1^2}(1,2)} - E_0\left(\frac{S_2}{S_1}|U_1^2\right)}{Var_0\left(\frac{S_2}{S_1}|U_1^2\right)}, \\ a\left(U_1^2\right) &= \widetilde{B}_{U_1^2}\left(1,2\right) - b\left(U_1^2\right) E_0\left(\frac{S_2}{S_1}|U_1^2\right) \end{aligned}$$

where

$$\widetilde{B}_{U_1^2}\left(1,2\right) = E_0\left[m_2^*\left(U_1^2, S_2\right)|U_1^2\right] \text{ and } Q_{U_1^2}\left(1,2\right) = E_0\left[m_2^*\left(U_1^2, S_2\right)\frac{S_2}{S_1}|U_1^2\right]$$

$E_{U_1^2}^*(.)$ represents the expectation under the pseudo risk-neutral probability:

$$P^*\left[\frac{S_2}{S_1} = u(U_2)|U_1^2\right] = \frac{\frac{Q_{U_1^2}(1,2)}{\widetilde{B}_{U_1^2}(1,2)} - d(U_2)}{u\left(U_2\right) - d\left(U_2\right)},$$

Therefore,

$$\begin{aligned}\pi_0 &= E_0\left[m_1\widetilde{B}_{U_1^2}(1,2)\,E^*_{U_1^2}\left[g(S_2)\right]\right]\\ &= E_0\left[E\left[m_1\widetilde{B}_{U_1^2}(1,2)\,E^*_{U_1^2}\left[g(S_2)\right]|U_1^2\right]\right]\\ &= E_0\left[\widetilde{B}_{U_1^2}(1,2)\,E\left[m_1E^*_{U_1^2}\left[g(S_2)\right]|U_1^2\right]\right]\end{aligned}$$

However,

$$\begin{aligned}E^*_{U_1^2=(i,j)}\left[g(S_2)\right] &= g(S_1u(i))\,P^*\left[\frac{S_2}{S_1}=u(i)|U_1^2=(i,j)\right]+\\ &\quad g(S_1d(i))\,P^*\left[\frac{S_2}{S_1}=u(i)|U_1^2=(i,j)\right].\end{aligned}$$

This last quantity only depends on S_1, similarly to theorem 2.1, we have:

$$E_0\left[m_1E^*_{U_1^2}\left[g(S_2)\right]|U_1^2\right]=E_0\left[m_1|U_1^2\right]E^*_{U_1}\left[E^*_{U_1^2}\left[g(S_2)\right]\right].$$

Under assumption **A2**, $E_0\left[m_1|U_1^2\right]=E_0\left[m_1^*|U_1,S_1\right]$. Consequently,

$$\begin{aligned}\pi_0 &= E_0\left[\widetilde{B}_{U_1^2}(1,2)\,\widetilde{B}_{(1,U_1)}(0,1)\,E^*_{U_1}\left[E^*_{U_1^2}\left[g(S_2)\right]\right]\right]\\ &= E^*_{tr}\left[\left(\frac{Q_{U_1^2}(1,2)}{\widetilde{B}_{U_1^2}(1,2)}\right)^{-1}\left(\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)}\right)^{-1}E^*_{U_1}\left[E^*_{U_1^2}\left(g(S_2)\right)\right]\right]\end{aligned}$$

$E^*_{tr}(.)$ represents the expectation under the risk-neutral probabilities:

$$P^*\left[U_2=i,U_1=j|U_0=1\right]=P^*\left[U_2=i|U_1=j\right]P^*\left[U_1=j|U_0=1\right]\quad\forall\, i,j=0,1$$

with:

$$Q_{U_1^2}(1,2)=\frac{P^*\left[U_2=i|U_1=j\right]}{P\left[U_2=i|U_1=j\right]}.$$

and $E^*_{U_1}(.)$ is defined in theorem 2.1. This ends the proof.

PROOF OF PROPOSITION 3.2. Let $\pi_{BTT,\tau}(g(S_{t+2}))$ be the price of $g(S_2)$ in BTT at date τ and $\pi_{TTSV,\tau}(g(S_{t+2}))$ the price of $g(S_2)$ in the TTSV. In a two-period, the BTT and TTSV are equivalent if and only if:

$$\pi_{BTT,0}\left[g(S_2)\right]=\pi_{TTSV,0}\left[g(S_2)\right].$$

Under **A1** and **A2**, Theorem 3.1 gives:

$$\pi_{TTSV,0}\left[g\left(S_2\right)\right] = E^*_{tr}\left[\left[\frac{Q_{(1,U_1)}\left(0,1\right)}{\widetilde{B}_{(1,U_1)}\left(0,1\right)}\right]^{-1} E^*_{U_1}\left[\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1} E^*_{U_1^2}\left(g\left(S_2\right)\right)\right]\right].$$

This last equation can be rewritten as:

$$\pi_{TTSV,0}\left[g\left(S_2\right)\right] = E^*\left[\left[\frac{Q_{(1,U_1)}\left(0,1\right)}{\widetilde{B}_{(1,U_1)}\left(0,1\right)}\right]^{-1} E^*_{U_2|U_1}\left[\left(E^*_{U_1}\left[\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1} E^*_{U_1^2}\left(g\left(S_2\right)\right)\right]\right)\right]\right]$$

where, E^* is the expectation under the risk neutral transition probabilities:

$$P^*\left[U_1 = i|U_0 = j\right].$$

and $E^*_{U_2|U_1}$ is the expectation under the risk neutral transition probabilities:

$$P^*\left[U_2 = i|U_1 = j\right].$$

It is obvious to see that:

$$\begin{aligned}
E^*_{U_2|U_1}\left[\left(E^*_{U_1}\left[\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1} E^*_{U_1^2}\left(g\left(S_2\right)\right)\right]\right)\right] &= E^*_{U_2|U_1}\left[\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1} E^*_{U_1^2}\left(g\left(S_2\right)\right)\right] \\
&= E^*_{U_2|U_1}\left[\left[\frac{Q_{U_1^2}\left(1,2\right)}{\widetilde{B}_{U_1^2}\left(1,2\right)}\right]^{-1} g\left(S_2\right)\right] \\
&= \pi_{TTSV,1}\left[g\left(S_2\right)\right]
\end{aligned}$$

Therefore,

$$\pi_{TTSV,0}\left[g\left(S_2\right)\right] = E^*\left[\left[\frac{Q_{(1,U_1)}\left(0,1\right)}{\widetilde{B}_{(1,U_1)}\left(0,1\right)}\right]^{-1} E^*_{U_1}\left(\pi_{TTSV,1}\left[g\left(S_2\right)\right]\right)\right],$$

But,

$$\pi_{TTSV,1}\left[g\left(S_2\right)\right] = E\left[m_2 g\left(S_2\right)|I_1\right]$$

Assuming the conditioning distribution of (m_2, S_2) given the the set (m_1, S_1, U_1) equals the conditioning distribution of (m_2, S_2) given S_1, we have

$$E\left[m_2 g\left(S_2\right)|I_1\right] = E\left[m_2 g\left(S_2\right)|S_1\right] = h\left(S_1\right),$$

where h is a positive function. This last expression can be viewed as the payoff of a traded derivative. In BTT, we recall that the price of $g(S_2)$ at date 1 can written as

$$\pi_{BTT,1}[g(S_2)] = E[m_2 g(S_2)|S_1] = h(S_1)$$

In that case,

$$\pi_{TTSV,0}[g(S_2)] = E^* \left[\left[\frac{Q_{(1,U_1)}(0,1)}{\widetilde{B}_{(1,U_1)}(0,1)} \right]^{-1} E^*_{U_1}(h(S_1)) \right]. \tag{5.12}$$

Applying proposition 2.2 to (5.12),

$$\begin{aligned} \pi_{TTSV,0}[g(S_2)] &= E[m_1 h(S_1)] \\ &= E[m_1 \pi_{BTT,1}[g(S_2)]]. \end{aligned}$$

The proof is completed. ■ ■

References

[1] Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," Journal of Political Economy, 81, 637-659.

[2] Bollen, N. P., 1998, "Valuing Options in Regime-Switching Models," Journal of Derivatives, 6, 38-49.

[3] Boyle, P., 1988, "A Lattice Framework for Option Pricing with Two State Variables," Journal of Financial and Quantitative Analysis, 23, 1-12.

[4] Chabi-Yo, F., R. Garcia, and E. Renault, 2004, "Stochastic Discount Factor Volatility Bound and Portfolio Selection Under Higher Moments," Working Paper, Department de Sciences Economiques, Université de Montréal.

[5] Cochrane, J. H., 2001, Asset Pricing, Princeton NJ: Princeton University Press.

[6] Cox, J., S. Ross, and M. Rubinstein, 1979, "Option Pricing: A simplified Approach," Journal of Financial Economics, 7, 229-264.

[7] Dai, Q., K. Singleton, and W. Yang, 2003, "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bound Yields," Working Paper, Stern School of Business, New York University.

[8] Garcia, R., R. Luger, and E. Renault, 2001, "Asymmetric Smiles, Leverage Effects and Structural Parameters," Working Paper Cirano 2001s-01.

[9] Garcia, R., R. Luger, and E. Renault, 2003, "Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables," Journal of Econometrics, 116, 49-83.

[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] Guo, D., 1998, "The Risk Premium of Volatility Implicit in Currency Options," Journal of Business and Economic Statistics, 16, 498-507.

[12] Hamilton, D.J., 1990, "Analysis of Time Series Subject to Changes in Regime," Journal of Econometrics, 45, 39-70.

[13] Hansen, L., and S. Richard, 1987, "The role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," Econometrica, 55, 587-613.

[14] Harvey, C. R., and A. Siddique, 2000, "Conditional Skewness in Asset Pricing Tests," The Journal of Finance, 55, 1263-1295.

[15] Hull, J. and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities," Journal of Finance, 42, 281-300.

[16] Kamrad, B., and P. Ritchken, "Multinomial Approximating Models for Options with k States Variables," Management Science 37 (12), 1640-1652.

[17] Melino, A., and A. X. Yang, 2003, "State Dependent Preferences Can Explain the Equity Premium Puzzle," Review of Economic Dynamics, 6(2), 806-830.

[18] Renault, E., 1997, "Econometrics Models of Option Pricing Errors," in Advances in Economics and Econometrics: Theory and Applications, 3, Kreps, D. and Wallis, K. eds., Cambridge University Press.

[19] Renault, E. and N. Touzi, 1996, "Option Hedging and Implied Volatilities in a Stochastic Volatility Model, Mathematical Finance," 6, 279-302.

Table 4.1: Jump amplitudes and jump probabilities

$\sigma^* = 0.2$, $r_f = 0.05$, $h = 1$				
Underlying Risk Premium	u	p_1^*	p_3^*	p_2^*
0	1.2214	0.5784	0.4306	**-0.0090**
0.03	1.2284	0.5473	0.4034	0.0492
0.05	1.2331	0.5280	0.3866	0.0854
0.06	1.2354	0.5187	0.3785	0.1028
0.07	1.2378	0.5097	0.3706	0.1197
0.08	1.2402	0.5009	0.3630	0.1362

Note: In this table, we compute the risk neutral jump probabilities for different values of the underlying risk premium.

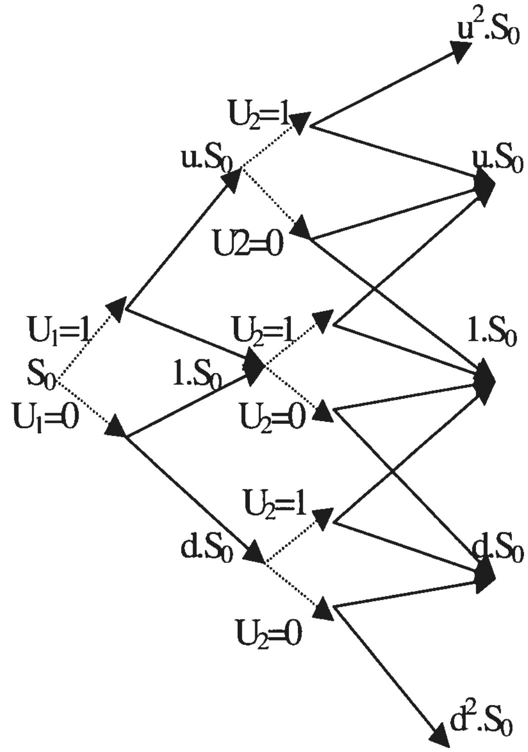


Figure 4.1: **Trinomial Tree with a State Variable**

Conclusion générale

Dans cette thèse, nous analysons différentes spécifications du SDF et ses implications en finance. Les différents sujets abordés ont pour thème commun la spécification du SDF.

Dans le premier essai nous analysons comment la préférence des agents économiques pour l'asymétrie affecte la demande et les prix des produits observés sur le marché. En considérant une situation au voisinage de la non-incertitude (expansion en petit bruit), on calcule les demandes des agents pour différents types d'actifs risqués. L'idée est de considérer un actif en offre non nulle, représentatif du portefeuille de marché, et des actifs dérivés en offre nette nulle mais dont les gains sont des fonctions non linéaires du portefeuille de marché. On s'aperçoit alors que la demande d'actifs dérivés est précisément justifiée par le goût des investisseurs pour l'asymétrie. Au niveau des prix, la rémunération du risque dépend non seulement du bêta de marché, comme dans un contexte moyenne-variance classique, mais aussi d'un coefficient de coasymétrie par rapport au marché. Les conclusions obtenues par l'expansion en petit bruit peuvent ensuite être retrouvées dans des contextes plus généraux grâce à la définition d'un facteur d'actualisation stochastique adapté. Cette double approche peut être étendue à un marché à deux périodes où d'autres phénomènes d'asymétrie doivent être pris en compte dans la dépendance temporelle des rendements d'une période à l'autre.

Le deuxième essai propose un SDF de référence qui à de nombreuses applications en finance notamment. Il peut servir notamment à comparer les modèles d'évaluation des actifs financiers ou à tester leur validité. Notre but est de présenter un SDF de référence qui prend en compte l'asymétrie observée dans les rendements des actifs financiers. Notre contribution est double. Premièrement nous présentons un SDF de référence simple et facile à utiliser. Deuxièmement, nous interprétons ce SDF en terme de choix de portefeuille sous asymétrie. Nous démontrons que notre approche de choix de portefeuille est une simple extension de l'approche moyenne variance (voir Markowitz (1952)) et de l'approche moyenne-variance-asymétrie (voir de Athayde et al. (2004)).

Dans une première application empirique, nous illustrons la perte d'information qui résulte de l'utilisation du SDF de Hansen et Jagannathan (1991). Dans une deuxième application, en utilisant le SDF proposé dans cet essai, on s'aperçoit que l'énigme de la prime de risque mis en évidence par Mehra et Prescott (1985) est encore plus difficile à expliquer. Dans une troisième application, nous illustrons le choix de portefeuille sous asymétrie et montrons qu'on perd de l'information sur le portefeuille choisi lorsque l'approche moyenne-variance-asymétrie proposée dans de Athayde et al. (2004) est utilisée.

Le troisième essai présente un modèle économique avec changement de régimes qui produit et explique les énigmes de l'aversion pour le risque et du SDF mises en évidence dans Jackwerth (2000) et Ait Sahalia et Lo (2000). Nous construisons un simple modèle où les préférences des investisseurs et leur consommation dépendent d'une variable d'état qui suit un processus de type Markovien à deux états et simulons les prix d'options d'achat européennes. En utilisant la méthodologie proposée par Jackwerth (2000), nous déduisons la fonction d'aversion absolue pour le risque et le SDF pour chaque valeur de la richesse. Ces fonctions présentent les mêmes énigmes que celles observées par Jackwerth. (2000) Lorsque nous appliquons la même méthodologie dans chaque état de l'économie, l'énigme de l'aversion absolue pour le risque disparaît. Nos résultats suggèrent que ce modèle rationalise et explique l'énigme de l'aversion pour le risque et du SDF mises en évidence par Jackwerth et Ait Sahalia et Lo (2000).

Le quatrième essai présente un modèle d'évaluation des produits dérivés par la méthode d'arbre lorsque le processus du prix du sous-jacent est affecté par une variable d'état non observable. Ce modèle généralise les modèles d'arbre existants: Cox, Ross et Rubinstein (1979) et Boyle (1988). Dans ce modèle, la variable d'état non observable capture les faits marquants mis en évidence par l'observation des prix d'options, en particulier l'asymétrie et la dynamique de l'asymétrie présentes dans les actifs dérivés.