Université de Montréal

Affine and Generalized Affine Models : Theory and Applications

par
Bruno Feunou Kamkui

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 Philosophiæ Doctor (Ph.D.)
en sciences économiques

Juin, 2009

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Affine and Generalized Affine Models : Theory and Applications

présentée par:

Bruno Feunou Kamkui

a été évaluée par un jury composé des personnes suivantes:

Benoit Perron,	président-rapporteur
Nour Meddahi,	directeur de recherche
René Garcia,	codirecteur
Peter Christoffersen,	membre du jury
Christian Gouriéroux,	examinateur externe
Brahim Boudarbat,	représentant du doyen de la FES

A

Thèse acceptée le: ....

RÉSUMÉ

Le thème principal de cette thèse est d'introduire plus de flexibilité dans les modèles d'évaluation des actifs financiers en temps discret tout en maintenant leur tractabilité . Nous appliquons nos modèles à différents domaines de la finance, incluant l'évaluation des produits dérivés, l'analyse de la structure à terme des taux d'intérêt, et l'évaluation du risque. D'un point de vue théorique, nous montrons comment évaluer les produits dérivés quand il y a non-normalité et hétéroscédasticité conditionnelle. L'approche de la modélisation des séries financières étudiée dans cette thèse est nouvelle et consiste en la spécification de la dynamique de la fonction caractéristique conditionnelle. Elle est motivée par le fait que de nombreux problèmes structurels rencontrés en analyse des risques peuvent naturellement s'écrire en terme de fonction caractéristique conditionnelle du processus d'intérêt.

Le premier chapitre construit un modèle dénommé "modèle affine généralisé". Les modèles affines sont très populaires dans la modélisation des séries financières parce qu'ils permettent un calcul analytique de la structure à terme des taux d'intérêt et des prix des produits dérivés. La principale propriété des modèles affines est que la fonction cumulant conditionnelle du processus d'intérêt, qui se définie comme le logarithme de la fonction caractéristique conditionnelle, est affine en cette variable d'intérêt. Par conséquent, un modèle affine est Markovien, comme les modèles autorégressifs, ce qui est une limite d'un point de vue empirique. Ce chapitre généralise les modèles affines en ajoutant dans l'expression de la fonction cumulant d'aujourd'hui le passé de la fonction cumulant. Par conséquent, les modèles affines sont non-Markoviens comme les modèles ARMA et GARCH, permettant de dissocier les dynamiques de long et de court terme du processus d'intérêt. Ce chapitre étudie les propriétés statistiques du nouveau modèle, dérive les équations des moments conditionnels, les formules analytiques des moments inconditionnels, et la prévision de la distribution pour une maturité donnée, ce qui est important dans l'analyse des problèmes de structure à terme. Dans ce chapitre, nous dérivons également les formules analytiques de la structure à terme du taux d'intérêt ainsi que des options Européennes. Différentes méthodes d'estimation sont proposées,

incluant le maximum de vraissemblance (MLE), le quasi-maximum de vraissemblance (QMLE), la méthode des moments généralisées (GMM) et la fonction caractéristique empirique (ECF).

Le deuxième chapitre étudie de façon plus spécifique le modèle de structure à terme VARMA avec absence d'opportunités d'arbitrage (AOA). Dans ce modèle nous montrons que les taux d'intérêt sont affines en la variable d'état et son espérance conditionnelle. Le facteur d'escompte stochastique est similaire à celui proposé par Ang et Piazzesi(2003), à la différence que le prix du risque est une fonction affine de la variable d'état et de l'espérance conditionnelle de sa réalisation future. Pour un choix particulier de la variable d'état (constituée de deux taux d'intérêt, court et long, et de deux variables macroéconomiques, inflation et mesure du niveau d'activité), nous investiguons les performances empiriques du modèle VARMA aussi bien en séries temporelles qu'en coupe transversale. Dans la dimension séries temporelles, nous trouvons des composantes MA significatives, en coupe transversale nous trouvons des prix du risque significatifs pour l'espérance conditionnelle de la variable d'état. Un exercice de prévision établit qu'un modèle VARMA d'AOA sur les trois facteurs traditionnels prédit mieux toute la courbe des taux et à tous les horizons qu'un VAR d'AOA sur les mêmes facteurs et une variable macroéconomique.

Dans le troisième chapitre, nous fournissons des résultats sur l'évaluation des options Européennes pour une large classe de dynamiques du sous-jacent. Notre principe d'évaluation utilise une mesure martingale équivalente (EMM), s'applique en temps discret, et dans un espace état à dimension infinie en s'appuyant sur le principe d'AOA. Notre approche s'accommode à toutes les formes d'hétéroscédasticité du sous-jacent, et les résultats d'évaluation dans un cadre homoscédastique en sont un cas particulier. La non-normalité conditionnelle est prise en compte, ce qui est important étant donné que l'hétéroscédasticité conditionnelle à elle seule ne suffit pas pour capter les "smiles" observés sur les prix d'options. La dynamique risque neutre des rendements est de la même famille que l'historique. Nous ne faisons aucune restriction sur la prime de risque, encore moins sur la forme de non normalité. Par conséquent notre cadre englobe les résultats de Duan (1995), Heston et Nandi (2000). Nous donnons des extensions dans le cadre des modèles discrets de volatilité stochastique, et nous analysons les relations entre les principes d'évaluations en temps discret et continu. Un exercice empirique démontre l'utilité de la non normalité conditionnelle dans la réplication des faits stylisés dénommés "smirk".

Le quatrième chapitre développe un modèle affine à facteurs multiples en temps discret et à composantes inobservables dans lequel la variance et l'asymétrie conditionnelle des rendements sont stochastiques. De façon cohérente, nous dissocions la dynamique de la variance conditionnelle de celle de l'asymétrie conditionnelle. Notre approche permet à la distribution des rendements journaliers courants d'être asymétrique conditionnellement aux facteurs courants. Dans notre modèle, l'asymétrie conditionnelle est la résultante, d'une part des effets de levier, et d'autre part de l'asymétrie de la distribution des rendements courants conditionnellement aux facteurs courants. Nous dérivons des formules analytiques pour différentes conditions de moments utiles pour l'inférence par la méthode des moments généralisée. En appliquant notre approche aux rendements journaliers de plusieurs indices boursiers, nous montrons que la distribution des rendements courants conditionnellement à la volatilité courante est positivement asymétrique, et nécessaire pour reproduire l'asymétrie inconditionnelle et les corrélations négatives entre rendements courants et carrés des rendements futurs. L'effet de levier est significatif et négatif tandis que l'asymétrie conditionnelle est positive, impliquant que l'asymétrie de la distribution des rendements courants conditionnellement à la volatilité courante domine l'effet de levier dans la détermination de l'asymétrie conditionnelle.

Mots clés: Modèles affines, fonction cumulant, structure à terme des taux d'intérêt, VARMA, prix du risque, GARCH, principe d'évaluation risque-neutre, absence d'arbitrage, innovations non-normales, volatilité stochastique, asymétrie stochastique, effet de levier, méthode des moments généralisée.

ABSTRACT

The main goal of this thesis is to introduce more flexibility in discrete time financial models while maintaining tractability . We apply our models in several domains in Finance including derivative pricing, term structure of interest rate and evaluation of risk. From a theoretical point of view, we show how we can still price derivative whenever non-normality, heteroskedasticity and time varying higher moments are taken into account. We introduce a different way of modeling financial time series, notably using conditional characteristic functions directly. The main motivation of this new approach is to extend affine model to non-Markovian ones.

The first chapter builds a new class of model termed "generalized affine models". Affine models are very popular in modeling financial time series as they allow for analytical calculation of prices of financial derivatives like treasury bonds and options. The main property of affine models is that the conditional cumulant function, defined as the logarithmic of the conditional characteristic function, is affine in the state variable. Consequently, an affine model is Markovian, like an autoregressive process, which is an empirical limitation. The chapter generalizes affine models by adding in the current conditional cumulant function the past conditional cumulant function. Hence, generalized affine models are non-Markovian, such as ARMA and GARCH processes, allowing one to disentangle the short term and long-run dynamics of the process. Importantly, the new model keeps the tractability of prices of financial derivatives. This chapter studies the statistical properties of the new model, derives its conditional and unconditional moments, as well as the conditional cumulant function of future aggregated values of the state variable which is critical for pricing financial derivatives. It derives the analytical formulas of the term structure of interest rates and option prices. Different estimating methods are discussed (MLE, QML, GMM, and characteristic function based estimation methods).

The second chapter models joint dynamics of short term rate, term spread, inflation and economic growth factor in a Vector Autoregression and Moving Average (VARMA). We combine VARMA processes with the no-arbitrage restrictions and study the fore-

castability of yields and macroeconomic variables. The paper shows that adding a Moving Average [MA] component to a standard VAR process offers substantial improvements in forecasting future yields, inflation, real activity and future interest rate risk premia where our benchmarks are either a standard VAR model or a dynamic version of the Nelson-Siegel model. An important hindsight from our results is that using VARMA processes break the tight link between current value of the state variable and the current conditional expectation of the future realization of the state variable, implicit in VAR models. Moreover, we show that the state variable follows a VARMA process under the risk-neutral probability measure only if the price of risk is linear in the current value of the state variable and the current conditional expectation of the future value of the state variable.

In the third chapter, we provide results for the valuation of European style contingent claims for a large class of specifications of the underlying asset returns. Our valuation results obtain in a discrete time, infinite state-space setup using the no-arbitrage principle and an equivalent martingale measure. Our approach allows for general forms of heteroskedasticity in returns, and valuation results for homoskedastic processes can be obtained as a special case. It also allows for conditional non-normal return innovations, which is critically important because heteroskedasticity alone does not suffice to capture the option smirk. We analyze a class of equivalent martingale measures for which the resulting risk-neutral return dynamics are from the same family of distributions as the physical return dynamics. In this case, our framework nests the valuation results obtained by Duan (1995) and Heston and Nandi (2000) by allowing for a time-varying price of risk and non-normal innovations. We provide extensions of these results to more general equivalent martingale measures and to discrete time stochastic volatility models, and we analyze the relation between our results and those obtained for continuous time models.

Finally, the fourth chapter develops a conditional arbitrage pricing theory (APT) model where factors and idiosyncratic noises are both heteroscedastic and asymmetric. The model features both stochastic volatility and conditional skewness (SVS model), as well as conditional leverage effects. We explicitly allow asset prices to be asymmetric conditional on current factors and past information, termed contemporaneous asymmetry. Conditional skewness is driven by conditional leverage effects (through factor loadings) and contemporaneous asymmetry (through idiosyncratic skewness). We estimate and test three versions of the SVS model using several equity and index daily returns, as well as daily index option data. Results suggest that contemporaneous asymmetry is particularly important in several dimensions. It helps to match sample return skewness, negative and significant cross-correlations between returns and squared returns, as well as positive and significant cross-correlations between returns are cubed returns. Further diagnostics suggest that SVS models with contemporaneous asymmetry show a better option pricing performance compared to contemporaneous normality and existing affine GARCH models, especially, but not only, for in-the-money call options and short-maturity contracts.

Keywords: Affine models; cumulant function; option pricing; term structure of interest rates; VARMA; Price of risk; GARCH; risk-neutral valuation; no-arbitrage; non-normal innovations; stochastic volatility; stochastic skewness; leverage effect; GMM.

TABLE DES MATIÈRES

LISTE DES TABLES

LISTE DES FIGURES

LISTE DES APPENDICES

A ma mère, Jeanne Djoumessi, et mon père, Emmanuel Kamkui

ACKNOWLEDGMENTS

Je suis particulièrement heureux d'écrire ces mots pour exprimer toute ma reconnaissance à toutes les personnes qui m'ont soutenu de près ou de loin et sans condition pendant toutes ces années de thèse.

Je pense à Nour Meddahi, mon directeur de thèse. Nour m'a souvent dit que je suis le premier étudiant qu'il a encadré en temps que directeur, j'en suis très fier. Tout n'a pas été parfait, mais nous avons su surmonter quelques incompréhensions et développer de vrais liens solides. Je pense ensuite à René Garcia, mon co-directeur. René m'a toujours fait confiance et a beaucoup contribué à mon initiation à la recherche et à l'économie financière. Je vous remercie tous les deux d'avoir cru et de continuer à croire que je pourrai faire un bon chercheur.

J'ai ensuite une pensée à tous mes co-auteurs, en particulier Peter Christoffersen, Kris Jacobs, Redouane Elkamhi, Chistian Bontemps et mes amis Roméo Tedongap et Jean-Sébastien Fontaine. J'ai beaucoup appris et je continue de le faire à leurs côtés. Je suis particulière reconnaissant à Rym Ben Hamadi pour ses précieux commentaires et suggestions.

Je suis très flatté de l'attention que Christian Gouriéroux a porté à mes travaux, en m'accueillant au CREST et en acceptant d'être l'examinateur de cette thèse. Je le remercie également pour ses précieux commentaires. Je voudrais, par ailleurs, exprimer ma reconnaissance à Kris Jacobs et à Peter Christoffersen pour avoir accepté de faire partie du jury.

Je suis reconnaissant au département de sciences économiques de l'Université de Montréal, au Centre Interuniversitaire de Recherche en Économie Quantitative, au Centre Interuniversitaire de Recherche en Analyse des Organisations, à l'Institut de Finance Mathématique de Montréal, à la Banque Laurentienne et au Centre de Recherche en Économie et Statistique pour les fonds de recherche qui ont financé ce travail.

Je remercie tout le personnel administratif du département de sciences économiques de l'Université de Montréal. Je pense principalement à Lyne Racine, Josée Lafontaine, Jocelyne Demers et Suzanne Larouche. Je remercie également le groupe de finance et de comptabilité de l'Imperial College London ainsi que le groupe de finance et d'assurance du CREST de Paris pour l'hospitalité qui m'a été offerte de travailler avec leurs chercheurs.

Pendant ma visite au CREST, j'ai eu le privilège de connaître et de sympathiser avec des personnes très accueillantes que je tiens à remercier. Je pense à Fanda Traoré, Nadine et William Guedj, Imen Ghattassi, Nazim Regnard, Mathieu Rosembaum, Nicolas Dromel, Fulvio Pegoraro, Idriss Kharroubi, Guy Lalanne, Jean-François Chassagneux, Kenza Benhima, Sophie Osotimehin, Marie-Anne Valfort, Olivier Simon et Vincent Lapègue.

Pendant mes années doctorales, j'ai eu la chance de continuer à entretenir de bon rapports professionnels, personnels et fraternels avec des collègues et amis de longue date. Je remercie Paul Samuel Njiki, Didier et Carine Tatouchoup, Foutse Khomh, Eric Bahel, Bertrand Djembissi, Constant Lonkeng, Eric Tchouaket, Anicet Choupo et Hervé Momeya pour le chemin que nous avons parcouru ensemble et pour toute l'harmonie, l'estime et l'entente qui animent encore aujourd'hui notre collaboration et notre amitié. Les rapports entretenus avec tous mes collègues et amis de l'Université de Montréal et d'ailleurs m'ont été bénéfiques. Je remercie Selma Chaker, Abderrahim Taamouti, Prosper Dovonon et son épouse Olivia, Firmin Doko, Octave Keutiben, Fulbert Tchana, Bertrand Hounkannounon, Pavel Sevcik et Rachidi Kotchoni.

Mes remerciements s'adressent également à tout mon réseau d'amis du Cameroun et d'ailleurs, ainsi qu'à toutes les personnes qui ont été plus ou moins proches de moi pendant toutes ces années. Je pense à Alex Chimé, Jean Silatchom, Eric Defo, Antoine Tambué, Hermine Gayum, Octavie Mbiakoup, Jean-Christophe Bobda, Justin Bem, Kouontchou Serge, Patrick Mzogang, Charlie Tchinda, Guy Chapda, Alain Job, Hugues Ibata et au Dr Raymond Todem.

Finalement, je remercie toute ma famille proche et lointaine, et en particulier mes parents, mon grand frère et ami Aubin, mes frères et soeurs Madeleine, Yves, Hervé, Laure, Carine, Lynda, Léa, Yannick, Cyrille, Estelle, mon oncle Jean-René et son épouse Yvonnes, ma fiancée Gwladys. Notre chemin a été tortueux et plein d'embûches, mais vous avez été présents tout le temps, m'encourageant et me soutenant dans toutes mes initiatives. Je vous en suis infiniment reconnaissant.

INTRODUCTION GÉNÉRALE

Cette thèse est constituée de quatre chapitres portant sur des modèles théoriques et empiriques d'évaluation des titres contingents. Tout d'abord elle construit un modèle plus flexible que le modèle affine, le modèle affine généralisé qui contient des modèles non Markoviens du type ARMA et GARCH, et garde l'atout majeur du modèle affine, à savoir les formules analytiques de la structure à terme des prix des titres contingents. Ensuite, une application est faite en structure à terme du taux d'intérêt où nous nous focalisons sur le modèle VARMA avec absence d'opportunités d'arbitrage (AOA). Par ailleurs, dans un autre registre nous nous intéressons à la problématique du changement de probabilité et à la caractérisation de la dynamique risque neutre des rendements quand l'historique est conditionnellement hétéroscédastique et non normale. Enfin une modélisation de l'asymétrie conditionnelle dans un cadre SV, par opposition au GARCH, est proposée.

Le premier chapitre introduit le modèle affine généralisé qui est une extension des modèles affines. Les modèles affines sont souvent utilisés dans la modélisation de la structure à terme des taux d'intérêt parce qu'ils permettent un calcul analytique du prix des bonds et des options, à toutes les maturités. En outre, les taux d'intérêt sont des fonctions linéaires de la variable d'état (taux court dans les modèles à un facteur), ce qui facilite l'inférence statistique. Cette approche a été introduite en temps continu par Vasicek (1977) où le taux court suit un processus autorégressif gaussien d'ordre 1. Une extension à plusieurs facteurs a ensuite été proposée par Duffie et Kan (1997). Une étude théorique des modèles affines en temps discret a été initiée par plusieurs travaux de recherche, parmi lesquels Darolles, Gourieroux et Jasiak (2006) et Gourieroux, Monfort et Polimenis (2002). D'autres études (Piazzesi (2005), Ang et Piazzesi (2003)) ont utilisé ces travaux théoriques pour caractériser les interactions entre la structure à terme des taux et la macroéconomie. Piazzesi (2003) a fait une synthèse des travaux sur la structure à terme affine des taux d'intérêt en temps continu. Dans le domaine de l'évaluation des produits dérivés tels que les options européennes, les processus affines sont utiles parce qu'ils permettent un calcul analytique du prix de ces produits. Dans cette littérature, on distingue comme en structure à terme des taux, les modèles en temps discret des modèles en temps continu. En temps continu Heston (1993) a décrit la dynamique jointe des rendements et de la volatilité à l'aide des modèles affines, et une généralisation avec sauts a été proposée par Duffie, Pan et Singleton (2000). En temps discret des modèles GARCH affine ont été proposés par Heston et Nandi (2001) et une généralisation avec asymétrie conditionnelle variable a été proposée par Christoffersen, Heston et Jacobs (2006).

Un processus en temps discret x_t est dit affine si sa fonction cumulant conditionnelle, notée $\psi_t(u)$, et définie comme le logarithme de sa fonction génératrice des moments conditionnels, c'est à dire,

$$\psi_t(u) \equiv \log[E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t]],$$

est donnée par

$$\psi_t(u) = \omega(u) + \alpha(u)x_t. \tag{1}$$

Tout processus autorégressif d'ordre un, AR(1), avec innovation *i.i.d.* est affine. Tout processus défini par (1) est Markovien, ce qui pourrait être une limitation dans la modélisation de certaines variables financières. Une des caractéristiques des séries financières, comme la volatilité des rendements des actifs, est la limitation des modèles ARCH Markovien de Engle (1982) qui ne repliquent pas les autocorrélations observées, ce qui a conduit à l'introduction des modèles du type GARCH par Bollerslev (1986). Par ailleurs, l'introduction des composantes non markoviennes dans un modèle, comme des composantes du type moyenne mobile (MA), permet de dissocier les dynamiques de court et de long terme du processus d'intérêt, ce qui pourrait être important pour la modélisation de la volatilité des rendements d'actifs et des taux courts (Andersen et Lund (1997)).

Des généralisations du modèle affine de base (1) ayant plus de mémoire et préservant les formules analytiques ont été proposées dans la littérature. Dai et Singleton (2003), Dai, Singleton et Yang (2006) ont supposé que les coefficients du modèle affine suivent un processus de Markov avec changement de régime. La pertinence empirique de cette approche a été démontrée, quoi que les méthodes de filtrage soient requises pour l'estimation et l'évaluation du modèle. Darolles, Gourieroux et Jasiak (2006) ont ajouté des retard de x_t dans (1), en proposant de définir des modèles affines d'ordre $p > 1$. Monfort et Pegoraro (2007a) ont appliqué cette approche avec succès à la structure à terme des taux, même si plusieurs retards sont requis entraînant ainsi un nombre élevé de paramètres à estimer (voir aussi Ang et Piazzesi (2003) où un VAR(12) est utilisé pour modéliser les variables macroéconomiques). Dans un autre article, Monfort et Pegoraro (2007b) ont combiné les deux approches susmentionnées, c'est à dire ajouter à la fois des retards et supposer que certains paramètres suivent un processus Markovien avec changement de régime. Cette approche nécessite des techniques de filtrage pour être implémentée en pratique.

Dans ce chapitre, nous adoptons une approche plus traditionnelle en incluant une composante du type MA dans le modèle affine. Les exemples suivant illustrent notre approche. Supposons que le processus x_t suive un modèle ARMA(1,1)

$$x_t = a + bx_{t-1} + \varepsilon_t - c\varepsilon_{t-1}, \ \ \varepsilon_t \text{ i.i.d.}, \ |b| < 1, \ |c| < 1,$$

où la fonction cumulant de ε est notée $\psi_\varepsilon(\cdot)$. On peut montrer (voir Section 2 du chapitre 1) que

$$\psi_t(u) = (ua + (1-c)\psi_\varepsilon(u)) + u(b-c)x_t + c\psi_{t-1}(u),$$

ce qui suggère d'étendre le modèle affine (1) comme suit

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(u). \tag{2}$$

Cette nouvelle généralisation du modèle affine est similaire à l'extension des modèles AR aux modèles ARMA. Mais l'interprétation dépasse le cadre de la modélisation AR. En effet l'équation (2) implique que toute puissance du processus x_t suit un modèle ARMA. Ceci est du au fait que la fonction cumulant conditionnelle de x_t est autorégressive.

Supposons que le processus x_t suit un modèle $VARMA(1,1)$ donné par

$$x_t = A + Bx_{t-1} + \varepsilon_t - C\varepsilon_{t-1}, \ \ \varepsilon_t \text{ i.i.d.}, \ |B| < 1, \ |C| < 1,$$

où la fonction cumulant de ε est notée $\psi_\varepsilon(\cdot)$. On peut montrer (voir la Section 2 du chapitre 1) que

$$\psi_t(u) = (u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u)) + u'(B-C)x_t + \psi_{t-1}(C'u),$$

ceci suggère la généralisation suivante du modèle (1)

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \psi_{t-1}(\theta u). \tag{3}$$

Notre approche comporte plusieurs avantages. Elle nécessite moins de paramètres que l'approche de Darolles, Gourieroux et Jasiak (2006). L'évaluation des produits dérivés et les procédures d'estimation du modèle sont plus simples que celles des modèles avec changement de régime comme Dai, Singleton et Yang (2006). Un autre avantage important de notre approche est qu'elle permet de dissocier les dynamiques de court et de long terme du processus x_t. En effet la fonction $\alpha(u)$ définie dans (1) assume une double fonction, ce qui pourrait être contraignant. On sait par exemple dans la littérature sur la volatilité que les modèles GARCH permettent plus de persistance que les modèles ARCH et ceci est important empiriquement. Nos différents exercices empiriques attestent ce fait.

Certains modèles dynamiques de structure à terme des taux d'intérêt avec variables macroéconomiques introduisent des variables latentes dans le vecteur d'état ; voir Ang et Piazzesi (2003). Cette approche est justifiée par le fait que les variables macroéconomiques à elles seules n'expliquent que partiellement la structure à terme des taux. Cependant, c'est toujours un exercice difficile de comprendre et d'interpréter exactement ces variables latentes. Il est bien connu dans la littérature de la modélisation des séries temporelles que les modèles AR avec variables latentes, appelés modèles structurels, impliquent des formes réduites avec représentations ARMA sur les variables observables.

Par conséquent, on pourrait interpréter nos nouveaux modèles comme des formes réduites des modèles affines avec facteurs latents.

Nous introduisons une version plus générale que celle proposée dans l'équation (2) en permettant au coefficient qui pré-multiplie ψ_{t-1} d'être une fonction de u, c'est à dire, nous étudions le modèle défini comme suit :

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u). \tag{4}$$

Nous dénotons ce modèle par le modèle affine généralisé de type I. Pareillement au type I, nous introduisons une version légèrement plus générale que celle proposée dans l'équation (3) en permettant à l'argument de la fonction $\psi_{t-1}(.)$ d'être une fonction pas nécessairement linéaire de u et aussi en pré-multipliant $\psi_{t-1}(\theta(u))$ par un coefficient, autrement dit, nous étudions le modèle défini comme suit :

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(\theta(u)). \tag{5}$$

Ces extensions sont théoriquement importantes parce que l'équation (2) implique que le vecteur (x_t, m_t), où $m_t = E[x_{t+1} \mid x_\tau, \tau \leq t]$, suit un modèle bivarié affine, ce qui n'est pas le cas pour les modèles définis par les équations (4) et (5). En outre, nous permettons différents retards de x_t et $\psi_{t-1}(u)$ dans les équations (4) et (5), ce qui revient à considérer les modèles du type ARMA(p,q).

Ce chapitre a plusieurs contributions. Premièrement, nous étudions les propriétés statistiques des nouveaux modèles et nous dérivons les équations des moments conditionnels ainsi que les formules analytiques des moments inconditionnels. Nous caractérisons aussi, via la fonction cumulant conditionnelle, la distribution conditionnelle du vecteur $(x_{t+1}, x_{t+2}, \ldots, x_{t+h})$ impliquée par nos modèles. Cette fonction sera l'ingrédient clé du calcul analytique de la structure à terme taux d'intérêt et des prix des options. Nous étudions alors la structure à terme des taux d'intérêt en supposant que la dynamique du taux court est décrite par (4) ou (5) sous les probabilité historique et risque neutre (ceci nécessite une spécification du taux d'escompte stochastique). En outre, nous étudions l'évaluation des options européennes quand la dynamique jointe des log-rendements et de la volatilité est donnée par (4).

Nous étudions également les différentes méthodes d'estimation des modèles proposés. Dans certains cas nous pouvons calculer analytiquement la vraisemblance du modèle (ce qui est le cas des exemples empiriques proposés). Autrement, on pourrait utiliser les procédures ECF (fonction caractéristique empirique) proposées par Singleton (2001) ou les GMM (méthodes de moments généralisées) proposées par Hansen (1982). Il existe aussi une approche plus simple qui consiste à utiliser les équations des deux premiers moments combinées avec les densités gaussienne ou gamma pour estimer les paramètres : ceci correspond à la méthode dite de quasi-maximum de vraisemblance.

Une des motivations du premier chapitre était le fait que les modèles du type VARMA n'entrent pas dans la famille des modèles affines définis en (1). Ce modèle entre dans la famille des modèles affines généralisés. Dans le deuxième chapitre nous étudions théoriquement et empiriquement la structure à terme des taux d'intérêt quand la variable d'état suit un modèle du type VARMA. Plus précisément nous discutons de la spécification de la variable d'état, taux d'intérêt, variables macroéconomiques et facteurs latents. Nous nous intéressons aussi à la spécification du facteur d'escompte stochastique, au problème d'inversion quand il y a des composantes inobservables dans la variable d'état. Nous dérivons la formule analytique des taux d'intérêt à toutes les maturités et montrons que les taux sont des fonctions affines de la variable d'état et de son espérance conditionnelle. Une généralisation du résultat à des ordres (p,q) est fournie en appendice.

Comme le démontre nos résultats, la composante moyenne mobile est particulièrement utile quand il y'a des facteurs macroéconomiques dans la variables d'état. Plusieurs articles de recherche se sont intéressés aux différents liens qui existent entre la structure à terme des taux, et les variables macroéconomiques, entre autres nous pouvons citer Ang et Piazzesi(2003), Diebold, Piazzesi et Rudebusch (2005), Ang, Piazzesi et Wei (2006), et Diebold, Rudebusch et Aruoba (2006). Toutes ces recherches ont mis en exergue l'importance de la relation entre les marchés financiers et la macroéconomie dans la prévision de la courbe des taux. Tous ces articles ont utilisé un modèle VAR pour décrire la dynamique jointe des taux et des variables macroéconomiques. Ang et

Piazzesi (2003) ont insisté sur l'importance d'introduire plusieurs retards dans le modèle affine de base afin de bien décrire la dynamique jointe de l'inflation et de la mesure de l'activité économique. Ils ont utilisé un VAR(12) à cet effet. Pourquoi ne pas utiliser un VARMA ? Étant donné que comme l'indique le premier chapitre, un calcul analytique des taux reste faisable dans ce cadre. En général, une caractéristique générale des variables macroéconomiques telle que mentionnée par la littérature est leur dépendance de plusieurs réalisations passées. Un autocorrélogramme de la mesure d'inflation est empiriquement bien approximé par un processus ARMA (voir Ang, Bekaert et Wei (2006)). La différence qualitative entre les modèles de type AR et ARMA réside sur le fait que un ARMA(1,1) essaie de séparer les composantes imprévisible et prévisible de l'inflation, ce qui n'est pas le cas des modèles AR(1).

Le principal objectif du deuxième chapitre est de construire un modèle VARMA de structure à terme avec absence d'opportunité d'arbitrage qui généralise le modèle de Ang et Piazzesi (2003). Nous voulons mettre en évidence l'importance d'une composante MA. Étant donné l'importance des variables macroéconomiques en structure à terme des taux d'intérêt, nous étudions plus en détail le cas où la variable d'état est constituée des taux d'intérêt et de variables macroéconomiques. Nous utilisons le taux à un mois (ce qui est une approximation du niveau de la courbe des taux), et la différence entre le taux à cinq ans et le taux à un mois (approximation de la pente de la courbe des taux). Nous ne considérons pas le facteur courbure, car certains auteurs ont souligné l'insignifiance de ce facteur à des fréquences d'observation mensuelle et trimestrielle. Les variables macroéconomiques sont constituées d'une mesure de l'activité économique et d'une mesure de l'inflation. D'un point de vue théorique nous montrons comment spécifier le facteur d'escompte stochastique quand la variable d'état suit un VARMA. Cette spécification généralise celle de Ang et Piazzesi (2003) et permet toujours un calcul analytique des taux à toutes les maturités.

Nous montrons que comparé au modèle $VAR(1)$, le modèle $VARMA(1,1)$ offre une meilleure représentation de la dynamique de la variable d'état. Les erreurs de prévision des composantes de la variable d'état sont moins élevées dans un modèle $VARMA(1,1)$ que dans un $VAR(1)$ et un modèle de marche aléatoire. La conclusion est valide à différents horizons de prévision et autant à l'intérieur de l'échantillon qu'à l'extérieur. Une analyse des fonctions de réponses impulsionnelles révèle des différences significatives dans les modèles VAR et VARMA, en particulier dans la réponse des variables macroéconomiques aux chocs sur le niveau et la pente de la courbe des taux. Par rapport au modèle $VAR(1)$, le modèle $VARMA(1,1)$ nécessite l'inclusion de la moyenne conditionnelle dans la spécification du prix du risque. Ainsi le prix du risque est une fonction affine de la variable d'état et de son espérance conditionnelle. Par conséquent pour une maturité donnée, le taux est une fonction affine de la variable d'état et de son espérance conditionnelle. Pour un modèle VAR, le coefficient de l'espérance conditionnelle s'annule. Nos résultats indiquent que l'information passée (résumée ici par l'espérance conditionnelle de la variable d'état) a un impact élevé pour les courtes maturités, et que cet impact diminue graduellement quand la maturité augmente. Le modèle VARMA prédit mieux toute la courbe des taux comparé à la marche aléatoire et au VAR.

Les résultats indiquent que les modèles VAR et VARMA s'accordent sur l'impact positif de l'inflation et de la mesure d'activité espérées sur la courbe des taux, mais divergent sur les composantes inattendues de ces deux agrégats. Le modèle VAR prédit un impact positif d'un choc sur l'inflation et l'activité économique sur la courbe des taux, alors que le modèle VARMA prédit le contraire. Un exercice de décomposition de la variance montre que les variables macroéconomiques expliquent à peu près 60% de la variation de la courbe de taux.

À l'aide des données utilisées dans l'article de Diebold et Li (2006), nous avons comparé les performances hors échantillon de plusieurs modèles d'absence d'opportunité d'arbitrage VAR et VARMA au modèle intitulé "Nelson-Siegel avec facteurs dynamiques AR(1)" (qui est le meilleur dans Diebold et Li (2006)). Le modèle $VARMA(1,1)$ sur les taux à 1, 24 et 96 mois est meilleur. Il fait mieux que "Nelson-Siegel avec facteurs dynamiques AR(1)", la marche aléatoire et surtout le modèle $VAR(1)$ sur les taux à 1, 24 et 96 mois et l'inflation. Nous concluons que contrairement au modèle VAR, le modèle VARMA sur le traditionnel vecteur de trois facteurs (niveau, pente et courbure) permet de prendre en compte les facteurs non financiers tels que les variables macroéconomiques.

Le troisième chapitre de cette thèse, intitulé "Évaluation des options avec hétéroscédasticité et non-Normalité conditionnelle" développe et étudie les propriétés d'un changement de probabilité en temps discret utile pour l'évaluation des produits dérivés quand le sous-jacent suis un modèle paramétrique (non spécifié) avec variation de volatilité conditionnelle et non-Normalité conditionnelle. Il diffère de l'esprit des deux premiers chapitres dans le sens où la préoccupation n'est pas l'obtention des formules analytiques des prix des options, mais on se pose la question de la spécification d'une mesure de martingale équivalente (ou de façon équivalente d'un facteur d'escompte stochastique) qui puisse s'appliquer à un plus grand nombre de modèles et qui soit compatible avec la notion d'absence d'opportunité d'arbitrage.

Un titre contingent est un actif dont la valeur future dépend de celle d'un autre actif. Une relation d'évaluation est une expression qui lie la valeur du titre contingent à celle de l'actif sous-jacent et d'autres variables. Le principe d'évaluation des titres contingents le plus populaire est la relation d'évaluation risque neutre (RNVR).

La plupart de la littérature sur les titres contingents et la plupart des applications du principe RNVR ont été faites en temps continu. Bien que l'approche en temps continu offre plusieurs avantages, l'évaluation des titres contingents en temps discret est aussi d'un intérêt certain. Par exemple, dans la couverture des positions prises sur des options, les décisions de recomposition de portefeuille se prennent en temps discret, et dans le cas des options américaines et exotiques, les décisions d'exercer avant échéance sont faites en temps discret. Cependant, la relative facilité de l'inférence économétrique demeure de loin l'atout principal du temps discret. La complexité qui résulte du problème de filtrage pour les processus qui captent de manière adéquate les faits stylisés (comme le modèle de volatilité stochastique de Heston (1993)) rend difficile l'estimation des processus en temps continu. Par contre, le filtrage est très simple pour la plupart des modèles étudiés dans ce chapitre.

Pour des raisons de convenance économétrique, la plupart des faits stylisés caractérisant les actifs sous-jacents ont été étudiés en temps discret. L'hétéroscédasticité conditionnelle est un fait important des rendements d'actif, elle a été à l'origine des modèles GARCH de Engle (1982) et de Bollerslev (1986). Vraisemblablement, grâce à cette évidence, la plupart des travaux empiriques récents en évaluation des options en temps discret se sont focalisés sur les processus GARCH. Puisque la distribution des innovations sur le rendement des actifs est continue, l'espace état du model GARCH est infini. Dans ce cas le marché est incomplet, et il est en général impossible de construire un portefeuille non risqué constitué de l'actif contingent et du sous-jacent.

Pour obtenir un RNVR, la littérature sur l'évaluation des dérivés dans un cadre GARCH a suivi l'approche de Rubinstein (1976) et Brennan (1979), qui démontrent comment obtenir un RNVR pour les distributions normale et log-normale et dans un cadre où la moyenne et la variance sont constantes. Ceci s'opère en spécifiant une économie avec agent représentatif et en caractérisant des conditions suffisantes sur les préférences. Pour une dynamique donnée de l'actif sous-jacent, des hypothèses spécifiques doivent être faites sur les préférences afin d'obtenir une risque-neutralisation. La condition de premier ordre résultant de cette économie implique une équation d'Euler qui peut-être utilisée pour évaluer tout actif. Pour des rendements d'actifs log-normaux et conditionnellement hétéroscédastique, le résultat standard est celui de Duan (1995). Le résultat de Duan s'appuie sur l'existence d'un agent représentatif avec indice relatif (ou absolu) d'aversion pour le risque constant.

A cause de la difficulté qu'on rencontre dans la caractérisation d'un modéle d'équilibre général qui sous-tend un RNVR, très peu de résultats sur l'évaluation des titres contingents sont actuellement disponibles dans le cadre des processus conditionnellement hétéroscédastiques et non normaux. Dans ce troisième chapitre, nous argumentons qu'il est possible d'investiguer l'évaluation des options pour une large classe de processus conditionnellement hétéroscédastiques et non normaux, à condition que la fonction génératrice des moments conditionnels existe. Il est aussi possible de prendre en compte une classe très large de prime de risque. Notre cadre diffère de celui de Brennan (1979) et de Duan (1995) et est plus relié à l'approche utilisée en temps continu : nous ne nous basons que sur des arguments d'absence d'opportunité d'arbitrage et sur quelques conditions techniques sur les stratégies d'investissement pour montrer l'existence d'un RNVR. Nous démontrons l'existence d'une mesure martingale équivalente (EMM) que nous caractérisons, sans faire d'hypothèse explicite sur la fonction d'utilité d'un agent représentatif. Nous montrons alors que le prix d'un actif contingent défini comme l'espérance conditionnelle des gains escomptés futurs est un prix d'absence d'opportunité d'arbitrage et nous caractérisons la dynamique risque neutre. Nous donnons des résultats pour le processus GARCH et plus généralement pour le modèle à volatilité stochastique en temps discret. Nous analysons aussi quelques limites en temps continu des modèles discrets considérés, et nous discutons des relations entre la risque-neutralisation faite en temps continu et celle faite en temps discret pour les modèles à volatilité stochastique.

Pourquoi sommes nous capables de fournir un résultat plus général sur l'évaluation des titres contingents que la littérature existante ? A notre avis, les analyses de Brennan (1979) et de Duan (1995) répondent à deux questions à la fois : premièrement, une question plus technique qui caractérise la dynamique risque neutre et l'évaluation des options ; deuxièmement, une question plus économique qui caractérise le cadre d'équilibre général qui sous-tend le principe d'évaluation. La littérature existante a pour la plupart considéré ces deux questions comme étant inextricablement liées, et s'est de ce fait largement limitée aux processus (log)normaux et quelques cas spéciaux non-normaux. Ce troisième chapitre diffère de façon subtile mais importante de la plupart des études existantes. Nous argumentons qu'il est possible et désirable de traiter ces questions séparément. Nous n'ambitionnons pas de caractériser la distribution bivariée des préférences et des rendements qui sous-tend la relation d'évaluation risque neutre. Mais, nous nous restreignons plutôt à une classe de dérivées de Radon-Nikodym et nous cherchons un EMM dans cette classe. Ceci permet de fournir des résultats généraux sur l'évaluation des options sous l'hypothèse de non-Normalité conditionnelle des rendements d'actifs sans recours à des arguments d'équilibre général. Nous montrons aussi comment le modèle normal et des modèles non-normaux existant sont des cas particuliers de notre cadre d'analyse.

Une approche similaire consistant à séparer ces deux problématiques existe aussi dans la littérature sur l'évaluation des options dans le cadre des modèles en temps continu avec volatilité stochastique, à l'instar du modèle de Heston (1993). Ces modèles impliquent différents EMMs pour différentes spécifications de la prime de risque de la volatilité. Pour une spécification donnée de la prime de risque de la volatilité, on peut trouver un EMM et caractériser la dynamique risque neutre en utilisant le théorème de Girsanov. Pour obtenir ce résultat, et pour évaluer des options, il n'est pas nécessaire de caractériser explicitement la fonction d'utilité qui sous-tend la prime de risque de la volatilité. Cette tâche est très instructive, mais est différente de la caractérisation de la dynamique risque neutre et du prix de l'option pour une dynamique donnée des rendements sur le probabilité historique. Cette dernière tâche est un exercice purement mathématique. Pourtant la première fournit le cadre économique qui sous-tend le choix particulier de la prime du risque de la volatilité, et de ce fait aide à comprendre si un choix particulier de la prime de risque, qui est souvent opéré pour des raisons de convenance mathématique, est aussi raisonnable d'un point de vue économique.

Le quatrième et dernier chapitre de cette thèse, intitulé " Modèles Affines à Asymétrie Stochastique ", développe un modèle affine à facteurs multiples en temps discret et à composantes inobservables dans lequel la variance et l'asymétrie conditionnelles des rendements sont stochastiques. Plus important encore, dans le cas du modèle à deux facteurs, le vecteur constitué par rendements, la volatilité et l'asymétrie suit un processus affine. La variation temporelle dans la volatilité des rendements trouve son origine dans les modèles autorégressifs à hétéroscédasticité conditionnelle (ARCH, Engle (1982)) ou ses extensions (GARCH, Bollerslev (1986), et EGARCH, Nelson (1991)). Alors que dans les modèles ARCH et GARCH la volatilité des rendements est complètement déterminée par l'historique des rendements observés, une approche alternative, devenue populaire dans la littérature récente, est le modèle à volatilité stochastique (SV), dans lequel la volatilité des rendements est une composante inobservable qui subit des chocs de source différente de celle générant les chocs sur les rendements. La plupart des applications des modèles GARCH et SV supposent que la distribution conditionnelle des rendements est symétrique. Même si cette hypothèse permet de générer les queues épaisses observées pour la distribution inconditionnelle des rendements, il reste encore à expliquer la variation temporelle et le signe des asymétries conditionnelles (asymétrie et effets de levier) et les queues de la distribution conditionnelle des rendements (voir Hansen (1994)). Les asymétries conditionnelles sont importantes car, pour la valorisation des options par exemple, l'hétéroscédasticité conditionnelle ne suffit pas à expliquer ce fait empirique important qui dans la littérature est qualifiée de "sourire des options". Au premier plan, nous développons un modèle affine à facteurs multiples, à volatilité stochastique dont les innovations sur les rendements sont asymétriques. Christoffersen, Heston et Jacobs (2006) étudient également un modèle semi-affine des rendements avec asymétrie variable dans le temps. Cependant, l'asymétrie conditionnelle dans leur modèle est liée de façon déterministe à la variance conditionnelle, ce qui est également le cas pour le modèle à un facteur dans notre cas. Cependant, la volatilité et l'asymétrie conditionnelles dans leur modèle subissent les mêmes chocs que les rendements puisqu'il s'agit d'une variante des modèles GARCH. Au contraire, notre modèle à un facteur est une variante des modèles à volatilité stochastique, qui nouvellement peuvent être étudiés dans un cadre affine ne supposant pas la normalité conditionnelle des rendements. Mieux encore, dans notre cas à deux facteurs ou plus, nous brisons le lien déterministe entre la volatilité et l'asymétrie conditionnelles qui se comportent dès lors comme deux facteurs linéairement indépendants caractérisant de manières différentes la dynamique temporelle des rendements et subissant des chocs de sources différentes de celle générant les chocs sur les rendements.

Harvey et Siddique (1999) considèrent également une distribution conditionnelle asymétrique des rendements dont la volatilité et l'asymétrie conditionnelles sont deux facteurs linéairement indépendants avec des dynamiques de type GARCH. Leur asymétrie conditionnelle autorégressive est une façon simple de modéliser l'asymétrie conditionnelle et fournit également une méthodologie d'estimation de l'asymétrie conditionnelle qui est facile à mettre en oeuvre précisément par l'applicabilité du maximum de vraisemblance. Cependant, un défaut d'application, et non pas le moindre, de la modélisation de Harvey et Siddique (1999) est que leur modèle est non-affine et devient coûteux en temps d'exécution pour la résolution des modèles d'évaluation d'actifs financiers, précisément à cause de la non-existence de formules analytiques entraînant une résolution numérique ou par simulations. Notre modèle est une alternative convenable au modèle de Harvey et Siddique (1999). Nous modélisons l'asymétrie par une combinaison affine de facteurs stochastiques linéairement indépendants. L'existence de la fonction génératrice des moments offre un cadre de résolution analytique des modèles d'évaluation d'actifs financiers permettant de gagner énormément en temps d'exécution.

Nous montrons aussi comment cette fonction génératrice des moments permet d'estimer le modèle par la méthode des moments généralisée en se basant sur des conditions de moments exactes. Dans notre cadre à facteurs stochastiques, nous distinguons l'information de l'agent économique de celui de l'économètre et fournissons explicitement les équivalents GARCH de la volatilité, de l'asymétrie et des effets de levier conditionnels. L'autre objectif est de développer et d'implémenter un algorithme pour le calcul analytique des moments inconditionnels exacts de la variable observable, dans un modèle semi-affine général en temps discret à facteurs multiples qui englobe notre modèle. Une étude similaire a été conduite par Jiang et Knight (2002) dans le cadre des processus affines en temps continu. Ces auteurs dérivent de manière analytique la fonction caractéristique inconditionnelle conjointe du processus de diffusion vectoriel. Cependant, cette question, bien que d'une importance à ne pas sous-estimer, n'a pas été examinée pour les processus affines en temps discret. Premièrement, les formules analytiques pour les moments inconditionnels permettent d'évaluer l'impact direct des paramètres du modèle sur des moments inconditionnels critiques tels que l'asymétrie, l'aplatissement excédentaire, l'autocorrélation des carrés des rendements et les corrélations croisées entre les rendements et les carrés des rendements. Deuxièmement, les moments inconditionnels en population peuvent être directement comparés à leurs contreparties empiriques. En plus, cette évaluation s'avère indispensable dans un exercice de calibrage où les paramètres du modèle sont fixés de sorte à reproduire les valeurs échantillonales de certains de ces moments inconditionnels. Plus important encore, cette comparaison entre moments en populations et moments empiriques permet la mise en oeuvre d'une procédure d'estimation du modèle par la méthode des moments généralisée avec l'avantage inqualifiable de se baser sur des conditions de moments exactes. Cette technique d'estimation permet également d'évaluer l'habileté du modèle à répliquer les faits empiriques connus tels que la persistance dans la volatilité des rendements à travers l'autocorrélation des carrés des rendements, l'absence d'autocorrélation des rendements, les effets de levier négatifs à travers les corrélations croisées entre les rendements et les carrés des rendements, l'aplatissement excédentaire positif et l'asymétrie négative. Chacun de ces faits stylisés est pris en compte par une ou plusieurs conditions de moments particulières faisant partie du vecteur des conditions de moments utilisé pour l'estimation du modèle.

Nous appliquons cette nouvelle procédure d'estimation des modèles semi-affines pour notre modèle à un facteur, en utilisant les séries de rendements journaliers de plusieurs portefeuilles d'actions et d'indices boursiers. Pour estimer les facteurs stochastiques, nous appliquons une variante du filtre de Kalman pour les modèles non-linéaires. Les paramètres du modèle sont tous significatifs et les implications du modèle sont frappantes. D'abord, la distribution des rendements journaliers courants conditionnellement à la volatilité courante est positivement asymétrique. De plus, cette asymétrie positive est nécessaire pour reproduire des statistiques échantillonales significatives telles que l'asymétrie inconditionnelle et les corrélations négatives entre rendements courants et carrés des rendements futurs. Ensuite, cette distribution positivement asymétrique engendre également une asymétrie positive de la distribution des rendements courants conditionnellement aux rendements passés. Ce résultat est contraire à certaines conclusions d'une large partie de la littérature existante (Forsberg et Bollerslev (2002)). Finalement, lorsque la distribution des rendements journaliers courants conditionnellement à la volatilité courante est contrainte à la normalité, alors le modèle engendre une asymétrie négative de la distribution des rendements courants conditionnellement aux rendements passés, ce qui corrobore la littérature existante. Cependant, sous cette hypothèse, le modèle ne reproduit plus l'asymétrie et les effets de levier inconditionnels. En plus, les tests de restrictions sur-identifiantes rejettent le modèle contraint aux niveaux conventionnels tandis que ces tests ne rejettent pas le modèle non contraint générant une asymétrie conditionnelle positive de la distribution des rendements courants conditionnellement aux rendements passés.

CHAPTER 1

GENERALIZED AFFINE MODELS

Abstract

Affine models are very popular in modeling financial time series as they allow for analytical calculation of prices of financial derivatives like treasury bonds and options. The main property of affine models is that the conditional cumulant function, defined as the logarithmic of the conditional characteristic function, is affine in the state variable. Consequently, an affine model is Markovian, like an autoregressive process, which is an empirical limitation. The chapter generalizes affine models by adding in the current conditional cumulant function the lagged conditional cumulant function. Hence, generalized affine models are non-Markovian, such as ARMA and GARCH processes, allowing one to disentangle the short term and long-run dynamics of the process. Importantly, the new model keeps the tractability of prices of financial derivatives. This chapter studies the statistical properties of the new model, derives its conditional and unconditional moments, as well as the conditional cumulant function of future aggregated values of the state variable, which is critical for pricing financial derivatives. It derives the analytical formulas of the term structure of interest rates and option prices. Different estimating methods are discussed including MLE, QML, GMM, and characteristic function based estimation methods. In a term structure of interest rate out-of-sample forecasting exercise, our results suggest that for a many horizons, a simple multivariate generalized affine model on observed yields predicts the whole term structure of the interest rate better than the VAR and the Nelson-Siegel's model with AR(1) factor dynamic.

1.1 Introduction

Affine models are often used when one models the short term of interest rates because they lead to closed form of the bond prices and yields whatever the maturity. In addition, these yields are linear functions of the state variables, often the short term interest rate when one considers a one-factor model, which makes the pricing and the statistical inference simple. This approach has been introduced in continuous time by Vasicek (1977) where the short term interest rate is assumed to follow a Gaussian autoregressive process of order one and extended by Duffie and Kan (1997) to more non-negative models. Discrete time versions of affine models are studied in Ang and Piazzesi (2003), Darolles, Gourieroux, Jasiak (2006) and Gourieroux, Monfort, and Polimenis (2002) among others while several papers, including Piazzesi (2005) and Ang and Piazzesi (2003), used them to characterize the term structure of interest rates and its interaction with macroeconomic variables; see Piazzesi (2003) for a survey on affine term structure models. Likewise, several authors used the affine processes for modeling the stochastic volatility of asset returns and characterized analytically the formulas of option prices; see Heston (1993) and Duffie, Pan and Singleton (2000) in continuous time and Heston and Nandi (2001) in discrete time.

A discrete time process x_t is called affine when its conditional cumulant function, denoted $\psi_t(u)$, and defined as the logarithmic of the moment generating function,[1] i.e.,

$$\psi_t(u) \equiv \log[E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t]],$$

is given by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t. \tag{1.1}$$

Any autoregressive process of order one, AR(1), with i.i.d. innovations, is affine. A consequence of (1.1) is that an affine process is Markovian, which could be a limitation for modeling some financial data. It is well known that financial data, like volatility of asset returns, exhibit serial correlation that Markov ARCH models of Engle (1982) do not describe well, which leads to the introduction of the GARCH models in Bollerslev (1986). Likewise, we do know that allowing for non-Markovian components in a model,

1. Instead of considering the moment generating function, one could use the characteristic function which exists for any random variable while the moment generating function does not exist for some random variables. The theory developed in this chapter holds for characteristic functions. However, we decided to use the moment generating function for convenience and due to its familiarity with researchers in financial economics.

like moving average (MA) components, allows one to disentangle the short-term and the long-run dynamics of the variable of interest, which could be important for some financial data like volatility of asset returns and short term of interest rates (Andersen and Lund (1997)).

Several generalizations of affine models have been introduced in order to include more memory in the basic model (1.1) and maintaining the tractability of affine models. Dai and Singleton (2003) and Dai, Singleton and Yang (2006) assumed that the coefficients driving the affine model follow a Markov switching model. The authors show the empirical usefulness of this approach, although filtering techniques are needed to price and estimate the model. Darolles, Gourieroux and Jasiak (2006) added lags of x_t in (1.1), i.e. they proposed an affine model of order $p > 1$. Monfort and Pegoraro (2007a) successfully applied this approach to the term structure of interest rates, although one could need many lags leading to the estimation of many parameters. In a different chapter, Monfort and Pegoraro (2007b) combined the two approaches describe above, i.e. they added lags and assumed that some parameters are driven by a Markov switching model. Again, such a method needs filtering techniques for both pricing and estimating the model.

In this chapter, we follow a more traditional approach by including MA component in the model. The following examples highlight our approach. Assume that the process x_t is an ARMA(1,1) given by

$$x_t = a + bx_{t-1} + \varepsilon_t - c\varepsilon_{t-1}, \quad \varepsilon_t \text{ i.i.d.}, \ |b| < 1, \ |c| < 1,$$

where the cumulant function of ε is denoted $\psi_\varepsilon(\cdot)$. One can show (see Section 2) that

$$\psi_t(u) = (ua + (1-c)\psi_\varepsilon(u)) + u(b-c)x_t + c\psi_{t-1}(u),$$

which suggests the following extension of (1.1)

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(u). \tag{1.2}$$

One could view the new model as an extension of AR models to ARMA ones. It is worth noting that Eq. (1.2) implies that any power function of x_t is an ARMA process. This is the case because the conditional cumulant function of x_t is an auto-regression.

Let us now consider a multivariate version of our model. Assume that the process x_t is an VARMA(1,1) given by

$$x_t = A + Bx_{t-1} + \varepsilon_t - C\varepsilon_{t-1}, \;\; \varepsilon_t \text{ i.i.d.}, \; |B| < 1, \; |C| < 1,$$

where the cumulant function of ε is denoted $\psi_\varepsilon(\cdot)$. One can show (see Section 2) that

$$\psi_t(u) = (u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u)) + u'(B-C)x_t + \psi_{t-1}(C'u),$$

which suggests the following extension of (1.1)

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \psi_{t-1}(\theta u). \tag{1.3}$$

Our approach has several advantages. Its involves less parameters than the approach adopted in Darolles, Gourieroux and Jasiak (2006). The pricing and estimation procedures of the model are simpler than those of a model with Markov switching factors like Dai, Singleton and Yang (2006). Another important advantage of the approach is to allow one to disentangle the short term dynamics of x_t from its long-run ones. When one considers an affine model (1.1), the function $\alpha(u)$ has to match the two dynamics, which could be restrictive. We do know from the volatility literature that GARCH models allow for more persistence than ARCH models and that this is empirically important. Our empirical examples highlight this advantage.

Several dynamic term structure models with macroeconomic variables assume latent variables in the affine state variable; see Ang and Piazzesi (2003). Such approach is often done because current values of the macroeconomic variables do not fully explain the term structure of interest. However, it is somewhat difficult to understand what exactly these latent variables are. It is well known from the time series literature that AR models with latent variables, called structural models, imply reduced form ARMA representations for the observable variables. Consequently, one could interpret the new model as a reduced from of affine models with latent factors.

We introduce a slightly more general model than (1.2) by allowing the coefficient in front of $\psi_{t-1}(u)$ to be a function of u, i.e., we study the model defined by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u). \tag{1.4}$$

We denote this model as the generalized affine model of type I. We also introduce a slightly more general model (type II) than (1.3) by allowing the argument of $\psi_{t-1}(.)$ to be a function of u, which is not necessarily linear, and by allowing a coefficient in front of $\psi_{t-1}(\theta(u))$, i.e., we study the model defined by

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(\theta(u)). \tag{1.5}$$

These extensions are theoretically important because Eq. (1.2) implies that the vector (x_t, m_t), where $m_t = E[x_{t+1} \mid x_\tau, \tau \leq t]$, is a bivariate affine model while it is not the case for models defined by Eq. (1.4) and (1.5). Likewise, we allow for several lags of x_t and $\psi_{t-1}(u)$ in Eq. (1.4) and (1.5), i.e., we consider ARMA(p,q) type models.

The chapter has several contributions. First of all, we study the statistical properties of the models and derive several conditional and unconditional moments and cumulants. We also derive the conditional cumulant function of the vector $(x_{t+1}, x_{t+2}, \ldots, x_{t+h})$. This function is critical when one wants to derive analytical formulas of yields and option prices. We then derive the Treasury yields when assuming that short term of interest rate is given by (1.4) or (1.5) under the risk neutral measure or the physical measure (the latter needs the specification of the price of the risk). Likewise, we derive the formulas of options prices when assuming a stochastic volatility where the dynamics of the stochastic variance is given by Eq. (1.4).

One can use several methods to estimate to model. Sometimes one could characterize the likelihood of the model as in our empirical analysis. Otherwise, one could follow Singleton (2001) by using the characteristic function of the process x_t and the instrumental variable approach of Hansen (1982). Actually, an efficient use of the whole characteristic function leads to an efficient estimation of the parameters comparable to the maximum likelihood estimators; see Carrasco and Florens (2001, 2006) and Carrasco, Chernov, Florens, and Ghysels (2006). It is also possible to use the conditional mean and variance of the process x_t combined with the Gaussian quasi-maximum likelihood approach to consistently estimate the parameters.

Our results suggest that using observable variables in a no-arbitrage VARMA model can do better than "Nelson-Siegel with AR(1) factor dynamic" in forecasting the entire yield curve at any horizon. Macro-economic factors add new information which are not contained into affine yields only model, but we can cope with these macro-economic factors by implementing a no arbitrage generalized affine model (here the VARMA) on the classic three factors model.

The rest of the chapter is organized as follows. Section 2 provides the simple generalized affine model and provides its statistical properties. Section 3 provides the analytical formulas of the term structure of interest rates when the short term of interest rates is a generalized affine process under the physical or the risk-neutral measure. Likewise, section 3 provides the formulas of the option prices when the volatility of the stock returns is a generalized affine process under the physical or the risk-neutral measure. We discuss several estimation methods in section 4. Section 5 provides an empirical application in the term structure of interest rate modeling where used VARMA models, while Section 6 concludes. Appendix A provides an example where the function $\beta(\cdot)$ is not constant. All the proofs are provided in Appendix B.

1.2 Generalized Affine Models

This section introduces and studies the two simple generalized affine modes.

Definition: Generalized Affine Process. A process x_t is called a generalized affine process of Type I and order (1,1) when the conditional cumulant function of x_{t+1} given its lagged values $x_t, x_{t-1}, \ldots$, is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u). \qquad (1.6)$$

A process x_t is called a generalized affine process of Type II and order (1,1) when the conditional cumulant function of x_{t+1} given its lagged values $x_t, x_{t-1}, \ldots$, is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta\,\psi_{t-1}(\theta(u)). \qquad (1.7)$$

To simplify the exposure, we combine the two types in the same framework and we provide the following general definition:

A process x_t is called a generalized affine process of order (1,1) when the conditional cumulant function of x_{t+1} given its lagged values $x_t, x_{t-1}, \ldots$, is characterized by

$$\psi_t(u) \equiv \log E[\exp(ux_{t+1}) \mid x_\tau, \tau \leq t] = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(\theta(u)). \qquad (1.8)$$

1.2.1 Examples

Several well know examples in the time series and financial literatures are generalized affine. Obviously, affine models correspond to the case $\beta(u) = 0$. Other examples are given below.

1.2.1.1 Linear and Non-Linear ARMA(1,1) Models

Assume that x_t follows a linear ARMA(1,1) whose innovation process is i.i.d., i.e.

$$x_t = a + bx_{t-1} + \varepsilon_t - c\varepsilon_{t-1}, \;\; \varepsilon_t \text{ i.i.d.}, \; |b| < 1, \; |c| < 1,$$

where the cumulant function of ε is denoted $\psi_\varepsilon(\cdot)$. Denote the conditional mean of x_{t+1} by m_t, i.e.,

$$m_t \equiv E[x_{t+1} \mid x_\tau, \tau \leq t] = a + bx_t - c\varepsilon_t.$$

Observe that

$$m_t = a + (b-c)x_t + cm_{t-1}.$$

Hence,

$$\begin{aligned}
\psi_t(u) = \log E_t[\exp(ux_{t+1})] &= um_t + \psi_\varepsilon(u) = u(a+(b-c)x_t) + \psi_\varepsilon(u) + ucm_{t-1} \\
&= u(a+(b-c)x_t) + \psi_\varepsilon(u) + c(\psi_{t-1}(u) - \psi_\varepsilon(u)) \\
&= (ua + (1-c)\psi_\varepsilon(u)) + u(b-c)x_t + c\psi_{t-1}(u),
\end{aligned}$$

i.e., any ARMA(1,1) process with i.i.d. innovations defined is a generalized affine process given in (1.6) where

$$\omega(u) = ua + (1-c)\psi_\varepsilon(u),\ \alpha(u) = u(b-c),\ \beta(u) = c.$$

Let us now assume that the conditional mean of x_t is non-linear but still has an MA(1) structure, i.e.,

$$x_t = f(x_{t-1}) + \varepsilon_t - c\varepsilon_{t-1},\ \ \varepsilon_t \text{ i.i.d.},\ |c| < 1.$$

The condition mean of x_{t+1} denoted m_t is given by

$$m_t = f(x_t) - c\varepsilon_t = f(x_t) - cx_t + cm_{t-1}.$$

Hence,

$$\begin{aligned}
\psi_t(u) = \log E_t[\exp(ux_{t+1})] &= um_t + \psi_\varepsilon(u) = u(f(x_t) - cx_t) + \psi_\varepsilon(u) + ucm_{t-1} \\
&= (1-c)\psi_\varepsilon(u) + u(f(x_t) - cx_t) + c\psi_{t-1}(u).
\end{aligned}$$

Consequently, a non-linear ARMA(1) process with i.i.d. innovations is not a generalized affine process but belongs to the family defined by

$$\psi_t(u) = \omega(u) + \alpha(u, x_t) + \beta(u)\psi_{t-1}(u). \tag{1.9}$$

This family is currently under study in a different paper and called generalized non-affine models.

1.2.1.2 Linear VARMA(1,1) Models

Assume that x_t follows a linear VARMA(1,1) where the innovation process is i.i.d., i.e.

$$x_t = A + Bx_{t-1} + \varepsilon_t - C\varepsilon_{t-1}, \ \ \varepsilon_t \text{ i.i.d.}, \ |B| < 1, \ |C| < 1,$$

where the cumulant function of ε is denoted $\psi_\varepsilon(\cdot)$. Denote the conditional mean of x_{t+1} by m_t, i.e.,

$$m_t = A + Bx_t - C\varepsilon_t.$$

Observe that

$$m_t = A + (B - C)x_t + Cm_{t-1}.$$

Hence,

$$\begin{aligned}
\psi_t(u) = \log E_t[\exp(u'x_{t+1})] &= u'm_t + \psi_\varepsilon(u) = u'(A + (B-C)x_t) + \psi_\varepsilon(u) + u'Cm_{t-1} \\
&= u'(A + (B-C)x_t) + \psi_\varepsilon(u) + \psi_{t-1}(C'u) - \psi_\varepsilon(C'u) \\
&= (u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u)) + u'(B-C)x_t + \psi_{t-1}(C'u),
\end{aligned}$$

i.e., any VARMA(1,1) process with i.i.d. innovations defined in (1.2.1.2) is a generalized affine process of Type 2 given in (1.7), where

$$\omega(u) = u'A + \psi_\varepsilon(u) - \psi_\varepsilon(C'u), \ \alpha(u) = u'(B-C), \ \beta = 1, \ \theta(u) = C'u$$

1.2.1.3 GARCH(1,1) Type Models

We start the analysis by considering the model introduced in Bollerslev (1986), i.e.,

$$x_t = \mu + \varepsilon_t = \mu + \sqrt{h_{t-1}}z_t, \ z_t \text{ i.i.d. } \mathcal{N}(0,1), \ h_t = \omega + \alpha\varepsilon_t^2 + \beta h_{t-1},$$

with $\alpha \geq 0, \beta \geq 0, \alpha + \beta < 1$. By doing the same calculations as in the ARMA example, one gets

$$\psi_t(u) = \left((1-\beta)\mu u + \omega\frac{u^2}{2}\right) + \frac{1}{2}\alpha u^2(x_t - \mu)^2 + \beta\psi_{t-1}(u). \qquad (1.10)$$

In other words, the GARCH(1,1) is not a generalized affine process as (1.6) but a generalized non-affine process given by (1.9).

It is well known that the GARCH(1,1) does not lead to closed forms of option prices. Heston and Nandi (2000) proposed a different specification for h_t that solved the problem where h_t is given by[2]

$$h_t = \omega + \alpha(z_t - \gamma\sqrt{h_{t-1}})^2 + \beta h_{t-1}.$$

Likewise, one can show that the conditional cumulant function of x_{t+1} is given by

$$\psi_t(u) = \left(u\mu(1-\beta-\alpha\gamma^2) + \frac{u^2}{2}(\omega + \alpha\frac{(x_t-\mu)^2}{h_{t-1}} - 2\gamma(x_t-\mu))\right) + (\beta + \alpha\gamma^2)\psi_{t-1}(u). \qquad (1.11)$$

Consequently, the Heston and Nandi (2000) model is a generalized non-linear model defined by (1.9) where the function $\alpha(x_t, u)$ depends x_t and h_{t-1}, i.e., the whole past of x_t.

Eq. (1.11) looks more non-linear than Eq. (1.10), which is puzzling given that the Heston and Nandi (2000) model leads to analytical formulas for option prices while the Bollerslev (1986) does not. As already mentioned, affine models lead to closed form of prices of derivatives. It turns out that the variance process h_t is affine when one considers the Heston and Nandi (2000) while it is not the case for the traditional GARCH model. More precisely, one has

$$\text{Heston \& Nandi}: \log E[\exp(uh_{t+1}) \mid h_\tau, \tau \leq t] = u\omega + \psi_{\chi^2(1)}(\alpha u) + ((\beta + \alpha\gamma^2)u - 2\gamma^2 u^2)h_t$$
$$\text{Bollerslev}: \log E[\exp(uh_{t+1}) \mid h_\tau, \tau \leq t] = u\omega + \psi_{\chi^2(1)}(\alpha u h_t) + \beta u h_t,$$

2. There is an additional coefficient γ that appears in (1.2.1.3) which captures the leverage effect. One could easily add such term in the Bollerslev's GARCH equation.

where $\psi_{\chi^2(1)}(\cdot)$ denotes the cumulant function of the $\chi^2(1)$ distribution. We will consider again the Heston and Nandi model when we will derive the option pricing formulas of generalized affine models.

1.2.1.4 ACD(1,1) type models

Engle and Russell (1997) introduced the autoregressive conditional duration (ACD) model where the duration x_i between two consecutive trades follows the process

$$x_i = \eta_{i-1}\, v_i,\; v_i \text{ i.i.d.},\; v_i > 0,\; E[v_i] = 1,\; \eta_i = \omega + \alpha x_i + \beta \eta_{i-1}.$$

If one assume that v_i follows an exponential distribution whose density function is $f_v(v) = \exp(-v)$, then one gets

$$\psi_i = E_i[\exp(ux_{i+1})] = \frac{1}{1 - u\eta_i},\quad u < \frac{1}{\eta_i},$$

which is not a generalized affine model. However, it is the case for the logarithmic duration model of Bauwens and Giot (2000) defined by

$$x_i = \exp(\eta_{i-1})\, v_i,\; v_i \text{ i.i.d.},\; v_i > 0,\; E[v_i] = 1,\; \log(\eta_i) = \omega + \alpha \log(x_i) + \beta \eta_{i-1}.$$

For this model, $\log(x_i)$ is an ARMA(1,1) and therefore a generalized affine process.

1.2.1.5 The generalized autoregressive gamma process

The autoregressive gamma process (ARG) studied in Meddahi (2001) and Gourieroux and Jasiak (2006), corresponding to the discretization of the Cox-Ingersoll-Ross diffusion process (Cox, Ingersoll, Ross (1985), is an affine model, whose cumulant function is given $\psi_t(u) = \omega(u) + \alpha(u)x_t$, where $\omega(u) = -\nu \log(1 - u\mu)$ and $\alpha(u) = \frac{\rho u}{1-u\mu}$.

One can show that it admits the following state space representation:

$$\begin{aligned} \frac{x_{t+1}}{\mu}|U_{t+1},I_t &\sim \gamma(\nu+U_{t+1}) \\ U_{t+1}|I_t &\sim P\left(\frac{\rho x_t}{\mu}\right), \end{aligned}$$

with I_t defined as the sigma algebra generated by $(x_s, s \leq t)$, $\gamma(\cdot)$ is the standard gamma distribution and $P(\cdot)$ is the Poisson distribution.

We generalize the ARG process to the GARG process which is built through the following state space representation:

$$x_{t+1} = \beta^t \theta^t \overline{Z}_{t+1} + \sum_{j=0}^{t-1} Z_{t+1}^{(j)},$$

with $\overline{Z}_{t+1}$, $Z_{t+1}^{(j)}$ for $j = 0, ..., t-1$, $t+1$ conditionally (conditional on I_t) independent random variables, $\overline{Z}_{t+1} \sim \psi_0(u)$, and

$$\begin{aligned} \frac{Z_{t+1}^{(j)}}{\mu_j}|U_{t+1}^{(j)},I_t &\sim \gamma\left(\nu_j+U_{t+1}^{(j)}\right) \\ U_{t+1}^{(j)}|I_t &\sim P\left(\frac{\rho_j x_{t-j}}{\mu_j}\right) \end{aligned}$$

where

$$\nu_j = \nu\beta^j, \ \ \mu_j = \mu\theta^j, \ \ \rho_j = \rho\beta^j\theta^j$$

We can show that

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta\psi_{t-1}(\theta u)$$

Then x_t follows a generalized affine model of Type 2 given in (1.7).

1.2.2 Existence of Generalized Affine Models

1.2.2.1 The Function $\beta(\cdot)$ is Constant.

Generalized affine models are defined recursively by their conditional cumulant function in (1.8). Therefore, one needs to show that the function $\psi_t(\cdot)$ in (1.8) is a proper cumulant function. The rest of this subsection focuses on the case where $\beta(\cdot)$ is constant, the other case being studied in section 1.2.2.2.

The first important property of cumulant function is that the sum of cumulant functions is a cumulant function. Consequently, when $\omega(u)$, $\alpha(u)x_t$, and $\beta\psi_{t-1}(\theta(u))$ are cumulant functions, the function $\psi_t(u)$ defined in (1.8) is a cumulant function. Observe that often, as in our empirical examples, $\omega(u)+\alpha(u)x_t$ is the cumulant function of an affine model. Therefore, the generalized affine model is well defined when $\beta\psi_{t-1}(\theta(u))$ is a cumulant function.

The second important property of cumulant functions is related to infinitely divisible random variables. A random variable z whose cumulant function is denoted $\psi_z(u)$, is called infinitely divisible when for any positive number c, $c\psi_z(u)$ is a cumulant function. Observe that a consequence of this definition is that $c\psi_z(u)$ is the cumulant function of an infinitely divisible random variable. Such variables appear in central limit theorems; examples of infinitely divisible random variables include normal, Poisson, and Gamma random variables. The first version of Darolles et al. (2006) provided sufficient conditions such that an affine process is infinitely divisible. In particular, popular affine models in Finance, i.e., the Gaussian and the square root processes are infinitely divisible.

The third important property of cumulant functions is still related to infinitely divisible random variables. For a given positive random variable z whose cumulant function is denoted $\psi_z(u)$, and an infinitely divisible cumulant function $\theta(u)$, $\psi_z(\theta(u))$ is a well defined cumulant function. In the characteristic function literature, this type of construction is recognized as the subordination of processes.

The second and third properties on cumulant functions are quite important for our purpose. By expanding recursively $\psi_t(u)$ given in (1.8), one gets

$$\psi_t(u) = \sum_{i=0}^{t-1} \beta^i(\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i}) + \beta^t \psi_0(\theta^{\circ t}(u))$$

where $\psi_0(u)$ is the unconditional cumulant function of x_1 and $\theta^{\circ i}(u)$ denotes function $\theta(.)$ compounded i times with itself. Consequently, when $\beta > 0$, $\theta(.)$ a cumulant of an infinitely divisible positive random variable (which implies that $\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i}$ is the cumulant function of an indivisible random variable like some affine models derived in Darolles et al. (2006)), $\beta^i(\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i})$ is a cumulant function of an infinitely divisible random variable. The definition of infinitely divisible random variables implies that the sum of infinitely divisible random variables is also an infinitely divisible random variable. Therefore, $\sum_{i=0}^{t} \beta^i(\omega(\theta^{\circ i}(u)) + \alpha(\theta^{\circ i}(u))x_{t-i})$ is the cumulant function of an infinitely divisible random variable. Consequently, $\psi_t(u)$ is the cumulant function of an infinitely divisible random variable when one assumes that this is the case for $\psi_0(u)$. In other words, sufficient conditions to guarantee that $\psi_t(u)$ defined in (1.8) is a proper cumulant function are: $\beta \geq 0$, $\theta(\cdot)$ an infinitely divisible cumulant function of a positive random variable, $\omega(u) + \alpha(u)(x)$ and $\psi_0(u)$ are cumulant functions of indivisible random variables.

The previous argument handles the case of positive random variable. To analyze random processes which have the whole real set as support, we will treat the two types of generalized affine models differently. In the Type I case, in general β need not to be positive for processes which have the whole real line as support (the Gaussian Vector Autoregressive Process for instance). The only needed conditions are $|\beta| < 1$ and function $\omega(u)$ being an infinitely divisible cumulant generating function. Indeed we can write $\psi_t(u) = \frac{\omega(u)}{1-\beta} + \sum_{i=0}^{\infty} \alpha(u)\beta^i x_{t-i}$. Since $\omega(u)$ an infinitely divisible cumulant generating function and $\frac{1}{1-\beta} > 0$ then $\frac{\omega(u)}{1-\beta}$ is a cumulant generating function (c.f.g). By denoting $y_{t-i} = \beta^i x_{t-i}$, we have that for each i $\alpha(u)y_{t-i}$ is a well defined c.f.g. Using the fact that the sum of c.f.g is a c.f.g, we get that $\psi_t(u)$ is a c.f.g.

In the case of Type II, we consider the following restrictions $\beta > 0$ and $\theta(u) = \theta u$ (which is the case of VARMA models). Then, (1.8) is well defined for processes with the whole real line as support. Indeed if $\omega(u)$, $\alpha(u)$ and $\psi_0(u)$ are infinitely divisible cumulant functions, we can show that $\omega(cu)$, $\alpha(cu)$ and $\psi_0(cu)$ are infinitely divisible cumulant function whatever the sign of c. Therefor for each i, $\omega(\theta^i u) + \alpha(\theta^i u)x_{t-i}$ is a cumulant function of an infinitely divisible function. This implies that (1.8) is well defined.

Another question not studied here is the existence of a stationary solution of (1.6). As usual, such a question is very difficult for discrete time non-linear models like GARCH models and it is left for future research. In the sequel of the chapter, we assume such existence.

1.2.2.2 The Function $\beta(\cdot)$ is Varying.

It is worth noting that in all the examples discussed in section 1.2.1, the function $\beta(u)$ given in (1.6) does not depend on u. In this section we provide two flexible classes of generalized affine models that generalize any positive affine model $\omega(u) + \alpha(u)x_t$ where $\omega(u)$ is a cumulant generating function of a positive random variable and $\alpha(u) = q(f(u) - 1)$ with $f(\cdot)$ the moment generating function of a positive random variable. We provide more details on the construction of such processes in Appendix A.

Proposition 1.2.1. *We show in Appendix A that the following model is a well defined generalized affine model of order (1,1).*

$$\psi_t(u) = \omega(u) + \alpha(u)x_t + \beta(u)\psi_{t-1}(u)$$

with

$$\beta(u) = \frac{\mu g(u)}{1 - u\mu}, \alpha(u) = u\mu + g(u) - 1, \omega(u) = u\rho - \frac{\rho}{1-\mu}\left(\frac{\beta(u)}{\mu} - 1\right) + h(u)(1 - \beta(u)),$$

where g(u) is a Laplace transform of a positive random variable, h(u) is a cumulant of a positive random variable and $\mu > 0$.

Notice that this family of generalized affine processes of order (1,1) is very flexible in the sense that it is defined up to unspecified functions $g(\cdot)$ and $h(\cdot)$. The only required conditions are that $g(\cdot)$ and $h(\cdot)$ are respectively moment and cumulant generating functions of positive random variables.

Given the properties of the functions $g(\cdot)$, $h(\cdot)$ and $(\cdot)$, it is straightforward to show that $\frac{\beta(\cdot)}{\mu}$ is a moment generating function of a positive random variable, and $\alpha(\cdot)$ is a cumulant function of an infinitely divisible positive random variable. But one can not conclude that $\omega(\cdot)$ is a cumulant generating function of a positive random variable, unless one imposes that $h(u) = \frac{q\mu\left(\frac{\beta(u)}{\mu}-1\right)}{1-\beta(u)}$ with $q > \max(0, \frac{\rho}{\mu(1-\mu)})$ and $\mu < 1$. In that case $\omega(\cdot)$ is a cumulant generating function and one can reformulate $\omega(\cdot)$ as follows:

$$\omega(u) = u\rho + \frac{q^*\alpha(u)}{1-u\mu},$$

with $q^* = q - \frac{\rho}{\mu(1-\mu)}$. The following proposition gives a generalization of Proposition (1.2.1) to order (2,1).

Proposition 1.2.2. *We show in Appendix A that the following model is a well defined generalized affine model of order (2,1).*

$$\psi_t(u) = \omega(u) + \alpha_1(u)x_t + \alpha_2(u)x_{t-1} + \beta(u)\psi_{t-1}(u)$$

with

$$\beta(u) = \mu f(u), \qquad \alpha_1(u) = u\mu + g(u) - 1,$$
$$\alpha_2(u) = \mu((1-\mu u)f(u) - g(u)),\ \omega(u) = (1-\beta(u))\left(u\rho + \frac{\rho}{1-\mu} + h(u)\right) - \rho g(u),$$

where $f(u)$ is a moment generating function of a positive infinitely divisible random, $g(u)$ is moment generating function of a positive random variable, $h(u)$ is a cumulant of a positive random variable and $\mu > 0$.

The order (1,1) example built in Proposition (1.2.1) is obtained by imposing $\alpha_2(u) = 0$. The following affine process of order (1,1) is obtained by imposing $\mu = \rho = 0$

$$\psi_t(u) = h(u) + (g(u) - 1)x_t.$$

We are then able to prove the following general theorem.

Theorem 1. *The following affine model $\psi_t(u) = h(u) + q(g(u) - 1)x_t$ defined on positive process x_t, with $q \geq 0$, $h(u)$ a positive c.f.g, $g(u)$ a positive m.g.f, can be extended to a generalized affine model of order (1,1) or (2,1) with non-constant $\beta(.)$ as shown in Propositions 1.2.1 and 1.2.2. The corresponding generalized model is*

$$\psi_t(u) = \omega(u) + \alpha_1(u)x_t + \alpha_2(u)x_{t-1} + \beta(u)\psi_{t-1}(u)$$

with

$$\beta(u) = \mu f(u), \qquad \alpha_1(u) = uq\mu + q(g(u) - 1),$$

$$\alpha_2(u) = q\mu((1 - \mu u)f(u) - g(u)),\ \omega(u) = (1 - \beta(u))\left(u\rho + \frac{\rho}{1-\mu} + h(u)\right) - \rho g(u).$$

where $\mu, \rho \geq 0$.

This family of affine processes which have been generalized is well known in the literature as the family of compound processes with positive value, and it contains the autoregressive gamma process of Gourieroux and Jasiak (2006) and the autoregressive inverse gaussian process.

1.2.3 Conditional Cumulants and Moments Structures

We now derive some conditional moments and cumulants implied by the generalized affine model. Given that the process x_t is defined by its conditional cumulant function, it is more convenient to derive the conditional cumulants of x_{t+1} and then the conditional moments. The conditional cumulant of x_{t+1} of order n denoted $\kappa_{n,t}$, is given by

$$\kappa_{n,t} = \psi_t^{(n)}(0),$$

where $f^{(n)}(\cdot)$ denotes the n-th derivative function of $f(\cdot)$. We will also use the notation

$$\overline{\kappa}_{n,t} \equiv \left(\kappa_{1,t}, \kappa_{2,t}, ..., \kappa_{n,t}\right)^\top. \tag{1.12}$$

From the following paragraph to the end of this chapter, we will use the following definition of function $\beta_{n,i}(u)$.

$$\beta_{n,i}(u)=\sum_{j=0}^{n-i}\binom{n}{j}\beta^{(j)}(u)\,B_{n-j,i}\left(\theta^{(1)}(u),\theta^{(2)}(u),...,\theta^{(n-j-i+1)}(u)\right),$$

where $B_{n,k}(a_1,a_2,...,a_{n-k+1})$ denote the Bell polynomials

Proposition 1.2.3. *Let x_t be a generalized affine process defined by (1.8). Then,*

$$\kappa_{n,t}=\omega^{(n)}(0)+\alpha^{(n)}(0)x_t+\sum_{i=1}^{n}\beta_{n,i}(0)\,\kappa_{i,t-1}, \tag{1.13}$$

and

$$\bar{\kappa}_{n,t}=\bar{\omega}_n+\bar{\alpha}_n x_t+\bar{\beta}_n\bar{\kappa}_{n,t-1}, \tag{1.14}$$

where

$$\bar{\omega}_n=\begin{pmatrix}\omega^{(1)}(0)\\ \omega^{(2)}(0)\\ \vdots\\ \vdots\\ \omega^{(n)}(0)\end{pmatrix},\ \bar{\alpha}_n=\begin{pmatrix}\alpha^{(1)}(0)\\ \alpha^{(2)}(0)\\ \vdots\\ \vdots\\ \alpha^{(n)}(0)\end{pmatrix},$$

and

$$\bar{\beta}_n=\begin{pmatrix}\beta_{1,1}(0) & 0 & 0 & \cdots & 0\\ \beta_{2,1}(0) & \beta_{2,2}(0) & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ \vdots & \vdots & \vdots & \cdots & \vdots\\ \beta_{n,1}(0) & \beta_{n,2}(0) & \cdots & \beta_{n,n-1}(0) & \beta_{n,n}(0)\end{pmatrix}.$$

A direct consequence of Proposition (1.2.3) is that $\bar{\kappa}_{n,t}$ is a $VAR(1)$. Indeed using relation (1.14), it can be established that

$$E_{t-1}\left[\bar{\kappa}_{n,t}\right]=\bar{\omega}_n+\bar{\rho}_n\bar{\kappa}_{n,t-1}, \tag{1.15}$$

where

$$\bar{\rho}_n = \begin{pmatrix} \alpha^{(1)}(0)+\beta_{1,1}(0) & 0 & 0 & \cdot\cdot & 0 \\ \alpha^{(2)}(0)+\beta_{2,1}(0) & \beta_{2,2}(0) & 0 & \cdot\cdot & 0 \\ \vdots & \vdots & \vdots & \cdot\cdot & \vdots \\ \vdots & \vdots & \vdots & \cdot\cdot & \vdots \\ \alpha^{(n)}(0)+\beta_{n,1}(0) & \beta_{n,2}(0) & \cdot\cdot & \beta_{n,n-1}(0) & \beta_{n,n}(0) \end{pmatrix}.$$

An important implication of Proposition (1.2.3) is that any conditional cumulant $\kappa_{n,t}$ is a linear combination of x_t and its lagged values. This property is a characteristic of generalized affine type models.

One has different forms when one considers generalized non-affine models defined in (1.9). Another consequence of the VAR representation is that when $\beta(\cdot)$ is not constant or $\theta(\cdot)$ is not linear, a conditional cumulant variable admits an ARMA representation of higher order. On the one hand when $\beta(\cdot)$ is constant and $\theta(u)=u$, one has a GARCH(1,1) type equation for $\kappa_{n,t}$

$$\kappa_{n,t} = \omega^{(n)}(0)+\alpha^{(n)}(0)x_t+\beta\kappa_{n,t-1}.$$

Likewise, when $\beta(\cdot)=1$ and $\theta(u)=\theta u$, one has the following GARCH(1,1) type equation for $\kappa_{n,t}$

$$\kappa_{n,t} = \omega^{(n)}(0)+\alpha^{(n)}(0)x_t+\theta^n\kappa_{n,t-1}.$$

We will show below that x_t admits an ARMA(1,1) representation, implying that $k_{n,t}$ admits an ARMA(2,1) representation.

There is a mapping between cumulants and moments of a random variable, which allows one to derive the conditional moments of x_{t+1} from its conditional cumulants. Denote the conditional moments by $m_{n,t}$, i.e.,

$$m_{n,t} = E[x_{t+1}^n \mid x_\tau, \tau \le t],$$

Then, the conditional moment $m_{n,t}$ is related to conditional cumulant $\kappa_{1,t},...\kappa_{n,t}$ through the complete Bell polynomial by

$$m_{n,t} = B_n\left(\kappa_{1,t}, ..., \kappa_{n,t}\right), \tag{1.16}$$

where $Bn(a_1;...;a_n)$ is the the nth complete Bell polynomial. Using (1.16) we can deduce the following relations for the first 6 moments.

$$\begin{aligned}
m_{1,t} &= \kappa_{1,t} \\
m_{2,t} &= \kappa_{2,t} + \kappa_{1,t}^2 \\
m_{3,t} &= \kappa_{3,t} + 3\kappa_{2,t}\kappa_{1,t} + \kappa_{1,t}^3 \\
m_{4,t} &= \kappa_{4,t} + 4\kappa_{3,t}\kappa_{1,t} + 3\kappa_{2,t}^2 + 6\kappa_{2,t}\kappa_{1,t}^2 + \kappa_{1,t}^4 \\
m_{5,t} &= \kappa_{5,t} + 5\kappa_{4,t}\kappa_{1,t} + 10\kappa_{3,t}\kappa_{2,t} + 10\kappa_{3,t}\kappa_{1,t}^2 + 15\kappa_{2,t}^2\kappa_{1,t} + 10\kappa_{2,t}\kappa_{1,t}^3 + \kappa_{1,t}^5 \\
m_{6,t} &= \kappa_{6,t} + 6\kappa_{5,t}\kappa_{1,t} + 15\kappa_{4,t}\kappa_{2,t} + 15\kappa_{4,t}\kappa_{1,t}^2 + 10\kappa_{3,t}^2 + 60\kappa_{3,t}\kappa_{2,t}\kappa_{1,t} \\
&\quad + 20\kappa_{3,t}\kappa_{1,t}^3 + 15\kappa_{2,t}^3 + 45\kappa_{2,t}^2\kappa_{1,t}^2 + 15\kappa_{2,t}\kappa_{1,t}^4 + \kappa_{1,t}^6.
\end{aligned}$$

Therefore, by using the results of Proposition 1.2.3, one gets the conditional moments of x_{t+1}.

1.2.4 Unconditional Cumulants and Moments Structures: Conditions for stationarity

1.2.4.1 Unconditional first and second moments of process x_t and cumulant $\overline{\kappa}_{n,t}$

As in affine models, we can compute unconditional moments which are useful to understand the dynamics of the model and to estimate unknown parameters. We start by focusing on the covariance structure of the process x_t which will allow us to show that x_t is an ARMA(1,1) with possibly heteroskedastic innovations. At every step we will provide required conditions on stationarity which are needed to compute time-independent unconditional moments.

Let x_t be a generalized affine process of order (1,1) defined in (1.8). By using Eq. (1.14) for $n = 1$ and by taking the unconditional expectation of both sides of the equation we have,

$$E(x_t) = E(\kappa_{1,t}) = \frac{\bar{\omega}_1}{1-\bar{\alpha}_1-\bar{\beta}_1} \equiv \mu.$$

It is important to mention that this result is valid if only if we made a first-order stationarity assumption. The unconditional mean process is stationary if only if

$$E(\kappa_{1,0}) = \mu \quad \text{and} \quad |\bar{\alpha}_1 + \bar{\beta}_1| < 1.$$

In general the unconditional expectation of the vector of cumulants $\bar{\kappa}_{n,t}$ is straightforwardly computed by using Eq. (1.15) and by taking the unconditional expectation of both sides of the equation. Indeed, we have

$$E(\bar{\kappa}_{n,t}) = (I_n - \bar{\rho}_n)^{-1} \bar{\omega}_n \equiv \bar{\mu}_n. \tag{1.17}$$

This time-invariant unconditional expectation of $\bar{\kappa}_{n,t}$ is obtained under the following necessaries and sufficient conditions:

$$E(\bar{\kappa}_{n,0}) = \bar{\mu}_n$$

The largest eigenvalue of $\bar{\rho}_n$ has a modulus smaller than one. (1.18)

For the case $n = 1$, condition 1.18 coincides with condition 1.17. For the case $n = 2$, condition 1.18 is satisfied if and only if condition 1.17 is satisfied and $|\beta_{2,2}(0)| < 1$.

For the second order unconditional moments, we first look at the unconditional variance. Consider Eq. (1.14) for $n = 1$, and square the two sides of the equation and take the unconditional expectation. Let Denote $\phi = \bar{\alpha}_1 + \bar{\beta}_1$. Then, if $|\phi| < 1$ we have:

$$E(\kappa_{1,t}^2) = \frac{\bar{\omega}_1^2 + \bar{\alpha}_1^2 E(\kappa_{2,t}) + 2\bar{\omega}_1 \phi E(\kappa_{1,t})}{1-\phi^2}.$$

In addition if $|\beta_{2,2}(0)| < 1$ then $E(x_t^2)$ and $Var(x_t)$ are time invariants, and are computed using the following relations $E(x_t^2) = E(\kappa_{1,t}^2) + E(\kappa_{2,t})$ and $Var(x_t) = E(x_t^2) - E(x_t)^2$.

The unconditional variance of the vector of cumulants $(\bar{\kappa}_{n,t})$ is obtained by taking the unconditional variance of the two sides of Eq. (1.14). By doing so, one gets

$$Var(\bar{\kappa}_{n,t}) = Var(x_t)\bar{\alpha}_n\bar{\alpha}_n^\top + \bar{\beta}_n Var(\bar{\kappa}_{n,t})\bar{\beta}_n^\top + \bar{\alpha}_n cov(x_t,\bar{\kappa}_{n,t-1})\bar{\beta}_n^\top + \bar{\alpha}_n cov(\bar{\kappa}_{n,t-1},x_t)\bar{\alpha}_n^\top.$$

Using the fact that $cov(\bar{\kappa}_{n,t-1},x_t) = cov(\bar{\kappa}_{n,t-1},\kappa_{1,t-1})$ and by replacing $\bar{\kappa}_{n,t-1}$ and $\bar{\kappa}_{n,t-1}$ by their recursions given in Eq. 1.14, one gets

$$\begin{aligned} cov(\bar{\kappa}_{n,t-1},x_t) &= Var(x_t)\bar{\alpha}_1\bar{\alpha}_n + \left(\bar{\alpha}_1\bar{\beta}_n + \bar{\beta}_1\bar{\rho}_n\right) cov(\bar{\kappa}_{n,t-2},x_{t-1}) \\ &\equiv Var(x_t)\bar{\alpha}_1\bar{\alpha}_n + \rho_n cov(\bar{\kappa}_{n,t-2},x_{t-1}). \end{aligned}$$

Consequently if we assume the following conditions

$$\begin{gathered} cov(\bar{\kappa}_{n,0},x_1) = cov(\bar{\kappa}_{n,0},\kappa_{1,0}) = Var(x_t)\bar{\alpha}_1\left(I_n-\rho_n\right)^{-1}\bar{\alpha}_n \equiv \lambda_n \\ \text{The largest eigenvalue of } \rho_n \text{ has a modulus smaller than one,} \end{gathered} \tag{1.19}$$

then one has

$$cov(\bar{\kappa}_{n,t-1},x_t) = \lambda_n,$$

which implies that

$$\begin{aligned} Var(\bar{\kappa}_{n,t}) &= Var(x_t)\bar{\alpha}_n\bar{\alpha}_n^\top + \bar{\beta}_n Var(\bar{\kappa}_{n,t-1})\bar{\beta}_n^\top + \bar{\alpha}_n\lambda_n^\top\bar{\beta}_n^\top + \bar{\beta}_n\lambda_n\bar{\alpha}_n^\top \\ &\equiv \bar{\beta}_n Var(\bar{\kappa}_{n,t-1})\bar{\beta}_n^\top + \theta_n. \end{aligned}$$

By using the formula

$$vec(AXB) = \left(B^\top \otimes A\right) vec(X),$$

one gets

$$vec\left[Var\left(\bar{\kappa}_{n,t}\right)\right] = vec(\theta_n) + \bar{\beta}_n \otimes \bar{\beta}_n vec\left[Var\left(\bar{\kappa}_{n,t-1}\right)\right].$$

Consequently, under the following assumption

$$vec\left[Var\left(\bar{\kappa}_{n,0}\right)\right] = \left(I_n - \bar{\beta}_n \otimes \bar{\beta}_n\right)^{-1} vec(\theta_n) \equiv \Omega_n$$
$$\text{the largest eigenvalue of } \bar{\beta}_n \otimes \bar{\beta}_n \text{ has a modulus smaller than one,} \qquad (1.20)$$

one gets the time invariant unconditional variance-covariance matrix of $\bar{\kappa}_{n,t}$

$$vec\left[Var\left(\bar{\kappa}_{n,t}\right)\right] = \Omega_n.$$

The unconditional autocovariance $cov(x_{t+h}, x_t)$ is obtained using the following recursion

$$cov(x_{t+h}, x_t) = \phi cov(x_{t+h-1}, x_t),$$

which is a consequence of Eq. (1.14) applied at $n = 1$. This implies that

$$cov(x_{t+h}, x_t) = \phi^h Var(x_t).$$

Hence, x_t is an ARMA(1,1) whose autoregressive root equals ϕ. Similarly to the unconditional mean and variance results, this new result on unconditional covariance of process x_t can be generalized to cumulant $\bar{\kappa}_{n,t}$.

Since $\bar{\kappa}_{n,t}$ is a $VAR(1)$, by using induction arguments and the double expectation formula and by making the following two assumptions:

1. the largest eigenvalue of $\bar{\rho}_n$ is smaller than one,
2. $Var(\bar{\kappa}_{n,t}) < \infty$ and constant in time,

one gets that

$$\bar{\rho}_{n,h} = Corr\left[\bar{\kappa}_{n,t}, \bar{\kappa}_{n,t-1}\right] = \bar{\rho}_n^h.$$

We can also comute the unconditional covariance between x_t and $\bar{\kappa}_{n,t+h}$. Indeed, by using again Eq.(1.14), $E(x_t \bar{\kappa}_{n,t+h})$ is computed recursively as follows

$$E(x_t \bar{\kappa}_{n,t+h}) = E(x_t)\bar{\omega}_n + E(x_t^2)\bar{\alpha}_n + \bar{\beta}_n E(x_t \bar{\kappa}_{n,t+h-1}), \qquad (1.21)$$

with the first term of the recursion given by

$$E(x_t\overline{\kappa}_{n,t}) = E(x_t)\overline{\omega}_n + E(x_t^2)\overline{\alpha}_n + \overline{\beta}_n E(\kappa_{1,t}\overline{\kappa}_{n,t}),$$

where $E(\kappa_{1,t}\overline{\kappa}_{n,t})$ is deduced from the variance-covariance matrix of $\overline{\kappa}_{n,t}$.

1.2.4.2 Higher order covariance

In this subsection, we use results on first and second unconditional moments of x_t and $\overline{\kappa}_{n,t}$ derived in the previous subsection to compute the third and fourth unconditional moments of x_t. We begin by deriving the third moments.

Using the relation between moments and cumulants, we have

$$E(x_t^3) = E(m_{3,t}) = E(\kappa_{3,t}) + 3E(\kappa_{2,t}\kappa_{1,t}) + E(\kappa_{1,t}^3). \tag{1.22}$$

From the previous subsection, all the terms in the right hand side of Eq. (1.22) are known, except $E(\kappa_{1,t}^3)$. $E(\kappa_{1,t}^3)$ is obtained by cubing the two sides of Eq. (1.14) for $n = 1$, which implies that

$$\begin{aligned}(1-\phi^3)E(\kappa_{1,t}^3) &= \overline{\omega}_1^3 + \overline{\alpha}_1^3 E(\kappa_{3,t}) + 3\overline{\alpha}_1^2\phi E(\kappa_{1,t}\kappa_{2,t}) + 3\overline{\omega}_1^2\phi E(\kappa_{1,t}) \\ &+ 3\overline{\omega}_1(\overline{\alpha}_1^2 + \overline{\beta}_1^2)E(\kappa_{1,t}^2) + 3\overline{\omega}_1\overline{\alpha}_1^2 E(\kappa_{2,t}).\end{aligned}$$

Consequently we can deduce the closed form expression of $E(x_t^3)$.

We now consider $E(x_t x_{t+h}^2)$. Using the relation between moments and cumulants, we have

$$E(x_t x_{t+h}^2) = E(x_t\kappa_{2,t+h-1}) + E(x_t\kappa_{1,t+h-1}^2). \tag{1.23}$$

$E(x_t\kappa_{2,t+h-1})$ is deduced from Eq. (1.21). We need to evaluate $E(x_t\kappa_{1,t+h-1}^2)$. By squaring the two sides of Eq. (1.14) for $n = 1$ and multiply by x_t, we show that

$$\begin{aligned}E(x_t\kappa_{1,t}^2) &= \overline{\omega}_1^2 E(x_t) + 2\overline{\omega}_1\overline{\alpha}_1 E(x_t^2) + \overline{\alpha}_1^2 E(x_t^3) + (\overline{\beta}_1^2 + 2\overline{\omega}_1\overline{\beta}_1)E(\kappa_{1,t}^2) \\ &+ 2\overline{\alpha}_1\overline{\beta}_1 E(\kappa_{1,t}\kappa_{2,t}) + 2\overline{\alpha}_1\overline{\beta}_1 E(\kappa_{1,t}^3).\end{aligned}$$

In general, for any integer h, when one knows $E(x_t\kappa_{1,t+h-1}^2)$, one can compute $E(x_t\kappa_{1,t+h}^2)$ as follows:

$$E(x_t\kappa_{1,t+h}^2)=\bar{\omega}_1^2E(x_t)+\bar{\alpha}_1^2E(x_t\kappa_{2,t+h-1})+2\bar{\omega}_1\phi E(x_tx_{t+h})+\phi^2E(x_t\kappa_{1,t+h-1}^2). \quad (1.24)$$

All the terms of the right hand side of Eq. (1.24) are known, consequently we can deduce the closed form expression of $E(x_tx_{t+h}^2)$.

We will conclude on this section on third order moments by evaluating $E(x_t^2\bar{\kappa}_{n,t+h})$. To achieve this purpose, we need to know $E(\kappa_{1,t}^2\bar{\kappa}_{n,t})$. By using the relation (1.14), we have:

$$\begin{aligned}(I_n-\phi^2\bar{\beta}_n)E(\kappa_{1,t}^2\bar{\kappa}_{n,t}) &= \bar{\omega}_1^2E(\bar{\kappa}_{n,t})+\phi^2E(\kappa_{1,t}^2)\bar{\omega}_n+\bar{\alpha}_1^2E(\kappa_{2,t})\bar{\omega}_n+\bar{\alpha}_1^2E(\kappa_{3,t})\bar{\alpha}_n\\ &+ \bar{\alpha}_1^2\bar{\beta}_nE(\kappa_{2,t}\bar{\kappa}_{n,t})+\phi^2E(\kappa_{1,t}^3)\bar{\alpha}_n+(3\bar{\alpha}_1^2+\bar{\alpha}_1\bar{\beta}_1)E(\kappa_{2,t}\kappa_{1,t})\bar{\alpha}_n\\ &+ 2\bar{\omega}_1\bar{\alpha}_1E(x_t\bar{\kappa}_{n,t})+2\bar{\omega}_1\bar{\beta}_1E(\kappa_{1,t-1}\bar{\kappa}_{n,t}).\end{aligned}$$

Hence,

$$E(x_t^2\bar{\kappa}_{n,t})=E(x_t^2)\bar{\omega}_n+E(x_t^3)\bar{\alpha}_n+\bar{\beta}_nE(\kappa_{2,t}\bar{\kappa}_{n,t})+\bar{\beta}_nE(\kappa_{1,t}^2\bar{\kappa}_{n,t}).$$

We are now able to compute $E(x_t^2\bar{\kappa}_{n,t+h})$ using the following recursion:

$$E(x_t^2\bar{\kappa}_{n,t+h})=E(x_t^2)\bar{\omega}_n+E(x_t^3)\bar{\alpha}_n+\bar{\beta}_nE(x_t^2\bar{\kappa}_{n,t+h-1}).$$

We will now compute the fourth moments of x_t. By using the relation between moments and cumulants, we have

$$E(x_t^4)=E(m_{4,t})=E(\kappa_{4,t})+4E(\kappa_{3,t}\kappa_{1,t})+3E(\kappa_{2,t}^2)+6E(\kappa_{2,t}\kappa_{1,t}^2)+E(\kappa_{1,t}^4).$$

Except $E(\kappa_{1,t}^4)$, all the terms in the right hand side are known. $E(\kappa_{1,t}^4)$ is obtained by raising the two sides of Eq. (1.14) for $n = 1$ to the fourth power. Indeed, one has

$$(1-\phi^4)E(\kappa_{1,t}^4) = \bar{\omega}_1^4 + 4\bar{\omega}_1^3\phi E(\kappa_{1,t}) + 6\bar{\omega}_1^2\bar{\alpha}_1^2 E(\kappa_{2,t}) + 4\bar{\alpha}_1^3\bar{\omega}_1 E(\kappa_{3,t}) + 6\bar{\omega}_1^2\phi^2 E(\kappa_{1,t}^2)$$
$$+12\bar{\alpha}_1^2\bar{\omega}_1\phi E(\kappa_{1,t}\kappa_{2,t}) + 4\bar{\alpha}_1^3\phi E(\kappa_{3,t}\kappa_{1,t}) + 3\bar{\alpha}_1^4 E(\kappa_{2,t}^2) + 6\bar{\alpha}_1^2\phi^2 E(\kappa_{2,t}\kappa_{1,t}^2) + 4\bar{\omega}_1\phi^3 E(\kappa_{1,t}^3).$$

Consequently, we can deduce the closed form expression of $E(x_t^4)$.

Let us consider the autocorrelation of x_t^2. One has

$$E(x_t^2 x_{t+h}^2) = E(x_t^2\kappa_{2,t+h-1}) + E(x_t^2\kappa_{1,t+h-1}^2).$$

From the results derived in the calculations of the third order moments, one knows $E(x_t^2\kappa_{2,t+h-1})$. Hence, one can derive $E(x_t^2\kappa_{1,t+h-1}^2)$ recursively as follows:

$$E(x_t^2\kappa_{1,t+h-1}^2) = \bar{\omega}_1 E(x_t^2) + \bar{\alpha}_1^2 E(x_t^2\kappa_{2,t+h-2}) + \phi^2 E(x_t^2\kappa_{1,t+h-2}) + 2\bar{\omega}_1\phi E(x_t^2 x_{t+h-2}),$$

with the first term of the recursion being

$$E(x_t^2\kappa_{1,t-1}^2) = E(\kappa_{1,t}^4) + E(\kappa_{1,t}^2\kappa_{2,t}).$$

Then we obtain the closed form expression of the autocorrelation of x_t^2

1.2.5 Forecasting and Conditions for stationarity

An important formula used in the analytical calculation of the term structure of interest rates and option prices is the conditional distribution function of $\sum_{i=1}^h u_i x_{t+i}$ for given real numbers u_i. Affine models allow one to derive the conditional cumulant function of $(x_{t+1}, x_{t+2}, \ldots, x_{t+h})$ and consequently the one of $\sum_{i=1}^n u_i x_{t+i}$. It turns out that this is the case for generalized affine models.

Let us denote $\bar{u}_h = (u_1, u_2, \ldots, u_h)^\top$ and $\psi_{t,h}(\bar{u}_h) = \log E_t\left[\exp\left(\sum_{i=1}^h u_i x_{t+i}\right)\right]$.

Proposition 1.2.4. *Assume that the process x_t is generated by (1.8), then the conditional* *cumulant function of* $(x_{t+1}, x_{t+2}, \ldots, x_{t+h})$ *is given by*

$$\psi_{t,h}(\bar{u}_h) = b_h(\bar{u}_h) + a_h(\bar{u}_h)x_t + \sum_{k=1}^{h} \bar{\beta}_k \left(v_{k,h}\right) \psi_{t-1} \left(\theta_k \left(v_{k,h}\right)\right), \tag{1.25}$$

with

$$\begin{aligned} a_h(\bar{u}_h) &= \sum_{k=1}^{h} \bar{\beta}_{k-1} \left(v_{k,h}\right) \alpha \left(\theta_{k-1} \left(v_{k,h}\right)\right), \\ b_h(\bar{u}_h) &= \sum_{k=1}^{h} \left[\sum_{j=1}^{k} \bar{\beta}_{j-1} \left(v_{k,h}\right) \omega \left(\theta_{j-1} \left(v_{k,h}\right)\right) \right]. \end{aligned}$$

The sequence $(v_{k,h})$ is defined recursively as follows: $v_{h,h} = u_h$, and for $k \leq h-1$:

$$v_{k,h} = u_k + \sum_{j=1}^{h-k} \bar{\beta}_{j-1} \left(v_{k+j,h}\right) \alpha \left(\theta_{j-1} \left(v_{k+j,h}\right)\right),$$

where functions $\bar{\beta}_j(\cdot)$ and $\theta_j(\cdot)$ are related to $\beta(\cdot)$ and $\theta(\cdot)$ as follows:

$$\begin{aligned} \theta_j(u) &= \theta^{\circ j}(u) \\ \bar{\beta}_0(u) &= 1 \\ \bar{\beta}_j(u) &= \prod_{k=1}^{j} \beta\left(\theta_{k-1}(u)\right), \quad for \;\; j \geq 1. \end{aligned}$$

The proof is provided in Appendix B. We will often use Eq. (1.25) in the next section when we derive yields and option prices. When $\beta(u) = 0$ or $\theta(u) = 0$, we obtain the well known affine model's result which stipulates that the conditional log-moment generating function of $(x_{t+1}, x_{t+2}, \ldots, x_{t+h})$ is affine in x_t. We can rewrite $\psi_{t,h}(\bar{u}_h)$ in term of the present value of process x, x_t, and all the past realizations x_s, $1 \leq s \leq t-1$. Indeed, one has

$$\psi_{t,h}(\bar{u}_h) = b_{t,h}(\bar{u}_h) + \sum_{j=1}^{t-1} a_{j,h}(\bar{u}_h)x_{t-j}, \tag{1.26}$$

with

$$
\begin{aligned}
a_{j,h}(\bar{u}_h) &= \sum_{k=1}^{h} \bar{\beta}_{k+j-1}(v_{k,h})\, \alpha\left(\theta_{k+j-1}(v_{k,h})\right) \\
b_{t,h}(\bar{u}_h) &= \sum_{k=1}^{h} \left\{ \bar{\beta}_{t+k-1}(v_{k,h})\, \psi_0\left(\theta_{t+k-1}(v_{k,h})\right) + \sum_{j=1}^{t+k-1} \left[\bar{\beta}_{j-1}(v_{k,h})\, \omega\left(\theta_{j-1}(v_{k,h})\right)\right] \right\}.
\end{aligned}
$$

Eq. (1.26) can be useful to provide some conditions on the existence of a stationary unconditional distribution. A stationary unconditional cumulant generating function (denoted here by $\psi(\cdot)$), when it exists, has to be the limit of $\psi_{t,h}(u) = \log E_t\left[\exp\left(ux_{t+h}\right)\right]$ when $h \longrightarrow \infty$. Using Eq. (1.26), we conclude that when $\psi(\cdot)$ exists, one has

$$\psi(u) = \lim_{h \longrightarrow \infty} b_{t,h}(u).$$

This limit have been obtained by making the following assumptions, which are in fact conditions on the existence of the stationary distribution

$$\lim_{h \longrightarrow \infty} a_{j,h}(u) = 0 \, , \forall j.$$

1.3 Analytical Formulas of Prices of Financial Derivatives

This section characterizes the yields and option prices when one assumes a generalized affine model for the interest rate and the stochastic volatility respectively. For each of them, we follow two approaches. We assume the generalized affine model under the physical measure (P-measure) and specify the price of risk and then derive the price of the financial derivatives (bonds or options). The second approach consists on specifying the generalized affine model under the risk neutral measure (Q-measure) and then derive the prices of the financial derivatives. We start the analysis by studying the term structure model

1.3.1 The Term Structure of Interest Rates

1.3.1.1 Generalized Affine Model Under the P-Measure

We assume that under the P-measure, the short term of interest rate denoted r_t follows a generalized affine process given in (1.6), i.e.,

$$\begin{aligned}\psi_t^P(u) &\equiv \ln E_t^P\left[\exp\left(u r_{t+1}\right)\right] \\ &= \omega^P(u) + \alpha^P(u) r_t + \beta^P(u)\,\psi_{t-1}^P\left(\theta^P(u)\right).\end{aligned}$$

When $\beta^P(u) = 0$ or $\theta^P(u) = 0$, one gets affine models like those of Vasicek (1977), Cox et al. (1985) and Duffie and Kan (1996) who derived the term structure of interest rates.

In order to derive the dynamics of r_t under the Q-measure and the yield curve, one needs to specify the stochastic discount factor denoted here $M_{t,t+1}$ or the price of risk. We follow the general approach of Gourieroux and Monfort (2007) who proposed the following formulation:

$$M_{t,t+1} = \exp(\gamma r_{t+1} + \theta_t). \tag{1.27}$$

Given the restriction

$$\exp(-r_t) = E_t^P[M_{t,t+1}], \tag{1.28}$$

one gets

$$\theta_t = -r_t - \psi_t(\gamma) \text{ and } M_{t,t+1} = \exp(\gamma r_{t+1} - r_t - \psi_t(\gamma)). \tag{1.29}$$

In the sequel, we define $B(t,h)$ and $r_{t,h}$ as

$$B(t,h) = E_t^P\left[\prod_{i=1}^{h} M_{t+i-1,t+i}\right], \quad r_{t,h} = -\frac{\log(B(t,h))}{h}. \tag{1.30}$$

We are now able to derive the term structure of interest rates, i.e., the formula of $r_{t,h}$ when h varies.

Proposition 1.3.1.

$$r_{t,h} = d_h + c_{h,0} r_t + z_{t-1}^{(h)},$$

with

$$
\begin{aligned}
d_h &= \frac{1}{h}\sum_{k=1}^{h-1}\left(\sum_{i=1}^{k}\bar{\beta}_{i-1}^{P}(\gamma)\,\omega^{P}(\theta_{i-1}^{P}(\gamma))-\bar{\beta}_{i-1}(d_k)\,\omega(\theta_{i-1}(d_k))\right),\\
c_{h,0} &= \frac{1}{h}\left[1+\sum_{k=1}^{h-1}\left(\bar{\beta}_{k-1}(\gamma)\,\alpha^{P}\left(\theta_{k-1}^{P}(\gamma)\right)-\bar{\beta}_{k-1}(d_k)\,\alpha\left(\theta_{k-1}(d_k)\right)\right)\right],\\
z_{t-1}^{(h)} &= \sum_{k=1}^{h-1}\left[\frac{\bar{\beta}_k(\gamma)}{h}\psi_{t-1}^{P}\left(\theta_k^{P}(\gamma)\right)-\frac{\bar{\beta}_k(d_k)}{h}\psi_{t-1}^{P}\left(\theta_k^{P}(d_k)\right)\right].
\end{aligned}
$$

The sequence d_k for $k \leq h-1$ satisfies the following backward recursion: $d_{h-1} = \gamma - 1$ and for $k \leq h-2$

$$
d_k = \gamma - 1 + \sum_{j=1}^{h-1-k}\left[\bar{\beta}_{j-1}\left(d_{k+j}\right)\alpha\left(\theta_{j-1}\left(d_{k+j}\right)\right)-\bar{\beta}_{j-1}(\gamma)\,\alpha\left(\theta_{j-1}(\gamma)\right)\right].
$$

We will study now the special case case of constant function $\beta(\cdot)$ and linear function $\theta(\cdot)$. One has the following model for the short term rate:

$$
\psi_t^{P}(u) = \omega(u) + \alpha(u) r_t + \beta\,\psi_{t-1}^{P}(\theta u)
$$

where $\beta,\ \theta \geq 0,$. Since r_t is a positive random variable, $\omega(u)$ and $\alpha(u)$ are cumulant functions of positive random variables. The yield to maturity n formula is given by:

$$
y_t^{(n)} = d_{n,t} + \sum_{j=0}^{t-1} c_{n,j} r_{t-j},
$$

where

$$
\begin{aligned}
d_{n,t} &= \frac{1}{n}\sum_{k=1}^{n-1}\left(\beta^{t+k-1}\left[\psi_0^{P}\left(\theta^{t+k-1}\gamma\right)-\psi_0^{P}\left(\theta^{t+k-1}c_k\right)\right]+\sum_{i=1}^{t+k-1}\beta^{i-1}\left(\omega(\theta^{i-1}\gamma)-\omega\left(\theta^{i-1}c_k\right)\right)\right),\\
c_{n,0} &= \frac{1}{n}\left[1+\sum_{k=1}^{n-1}\beta^{k-1}\left(\alpha\left(\theta^{k-1}\gamma\right)-\alpha\left(\theta^{k-1}c_k\right)\right)\right],
\end{aligned}
$$

and for $j \geq 1$

$$c_{n,j} = \sum_{k=1}^{n-1} \frac{\beta^{j+k-1}}{n} \left(\alpha\left(\theta^{j+k-1}\gamma\right) - \alpha\left(\theta^{j+k-1}c_k\right)\right),$$

with $c_{n-1} = \gamma - 1$ and for $0 \leq k \leq n-2$

$$c_k = \gamma - 1 + \sum_{j=1}^{n-1-k} \beta^{j-1}\left(\alpha\left(\theta^{j-1}c_{k+j}\right) - \alpha\left(\theta^{j-1}\gamma\right)\right).$$

One issue of interest is the positivity of yield. In this simple particular case, we are able to specify some conditions under which yields generated by our model are always positive. The functions $\omega(u)$ and $\alpha(u)$ are defined on $u \leq 0$ and are absolutely monotone, in particular $\omega'(u)$, $\alpha'(u) \geq 0$. This implies that $\omega(u)$ and $\alpha(u)$ are increasing functions. The model is well defined if we choose the price of risk $\gamma \leq 0$, it implies that $c_{n-1} < \gamma \leq 0$. Since $\alpha(u)$ is an increasing function, it implies that $c_{n-2} < \gamma \leq 0$. By using the same type of arguments, we deduce that $c_k < \gamma \leq 0$ for $k \leq n-1$.

Since $c_k < \gamma \leq 0$ for $k \leq n-1$, it follows that $c_{n,0} \geq 0$, $d_n \geq 0$ and

$$\sum_{k=1}^{n-1} \frac{\beta^k}{n}\left[\psi_{t-1}^P\left(\theta^k\gamma\right) - \psi_{t-1}^P\left(\theta^k c_k\right)\right] \geq 0.$$

Consequently our model is coherent because it generates positive yield $y_t^{(n)}$ at any given maturity.

We will now characterize the dynamics of r_t under the Q-measure. We denote by $\psi_t^Q(u)$ the conditional cumulant function of r_{t+1} under the Q-measure, i.e.,

$$\psi_t^Q(u) \equiv \log E_t^Q[\exp(ur_{t+1})]. \tag{1.31}$$

We will restrict our analysis to the generalized affine of type I (i.e., we restrict function so that $\theta^P(u) = u$).

Proposition 1.3.2. Dynamics of r_t under the Q-measure. *One has*

$$\psi_t^Q(u) = \psi_t^P(u+\gamma) - \psi_t^P(\gamma). \tag{1.32}$$

Hence if $\theta^P(u) = u$,

$$\psi_t^Q(u) = \omega^Q(u) + \alpha^Q(u) r_t + \beta^Q(u)\psi_{t-1}^Q(u) + [\beta(u+\gamma) - \beta(\gamma)]\psi_t(\gamma), \tag{1.33}$$

where

$$\omega^Q(u) = \omega^P(u+\gamma) - \omega^P(\gamma),\ \alpha^Q(u) = \alpha^P(u+\gamma) - \alpha^P(\gamma),\ \beta^Q(u) = \beta^P(u+\gamma). \tag{1.34}$$

Eq. (1.32) is model free, i.e., it does not depend on our generalized affine specification. In particular, the same equation appears in affine models; see Gourieroux and Monfort (2007) and Monfort and Pegoraro (2006a). An additional term appears in (1.33) which vanishes when $\beta(\cdot)$ is constant, as in our empirical examples. When this term does not vanish, the short term of interest rate is not a generalized affine under the Q-measure. However, the following proposition characterizes the conditional cumulant of $(r_{t+1}, \psi_{t+1}^P(\gamma))$ which will allow us to understand the dynamics of r_{t+1} under the Q-measure. In the sequel, $\psi_{r,\psi(\gamma),t}^Q(u,v)$ denotes the conditional cumulant function of $(r_{t+1}, \psi_{t+1}^P(\gamma))$ under the Q-measure.

Proposition 1.3.3. $\theta^P(u) = u$ *implies*

$$\begin{aligned}\psi_{r,\psi(\gamma),t}^Q(u,v) = {} & \omega_1^Q(u,v) + \left(\alpha_1^Q(u,v) r_t + \alpha_2^Q(u,v)\psi_t^P(\gamma)\right) + \beta_1^Q(u,v)\psi_{r,\psi(\gamma),t-1}^Q(u,v) \\ & - \alpha_2^Q(u,v)\beta_1^Q(u,v)\psi_{t-1}^Q(\gamma)\end{aligned}, \tag{1.35}$$

where

$$\begin{aligned}
\omega_1^Q(u,v) &= v\omega^P(\gamma)\left[1-\beta^P\left(u+v\alpha^P(\gamma)+\gamma\right)\right]+\omega^P\left(u+v\alpha^P(\gamma)+\gamma\right),\\
\alpha_1^Q(u,v) &= \alpha^P\left(u+v\alpha^P(\gamma)+\gamma\right),\ \alpha_2^Q(u,v)=v\beta^P(\gamma)-1,\\
\beta_1^Q(u,v) &= \beta^P\left(u+v\alpha^P(\gamma)+\gamma\right).
\end{aligned}$$

While the definition of generalized affine models (1.6) is given for univariate processes and of order (1,1), the extensions to multivariate and higher order is not very difficult. Eq. (1.35) means that the bivariate vector $(r_{t+1}, \psi_{t+1}^P(\gamma))$ is a generalized affine process of order (2,1). Consequently, one can characterize formulas of financial derivatives, including yields, by using the generalized affine dynamics of $(r_{t+1}, \psi_{t+1}^P(\gamma))$ under the Q-measure.

1.3.1.2 Generalized Affine Model Under the Q-Measure

We now assume that the short term of interest rate r_t follows a generalized affine process given in (1.8) under the Q-measure, i.e.,

$$\log E_t^Q[\exp(ur_{t+1})] \equiv \psi_t^Q(u) = \omega^Q(u) + \alpha^Q(u)r_t + \beta^Q(u)\psi_{t-1}^Q(\theta^Q(u)).$$

The following proposition provides the formula of the yield curve.

Proposition 1.3.4. *The yield at horizon n is given by*

$$r_{t,n} = b_n + a_n r_t - \frac{\sum_{k=1}^{n-1}\bar{\beta}_k(d_k)\,\psi_{t-1}^Q(\theta_k(d_k))}{n}, \tag{1.36}$$

where

$$\begin{aligned}
a_n &= \frac{1-\sum_{k=1}^{n-1}\bar{\beta}_{k-1}(d_k)\,\alpha^Q(\theta_{k-1}(d_k))}{n},\\
b_{n-1} &= \frac{-\sum_{k=1}^{n-1}\left[\sum_{j=1}^{k}\bar{\beta}_{j-1}(d_k)\,\omega^Q\left(\theta_{j-1}(d_k)\right)\right]}{n},
\end{aligned}$$

$$d_k = -1 + \sum_{j=1}^{n-1-k} \bar{\beta}_{j-1}\left(d_{k+j}\right)\alpha^{Q}\left(\theta_{j-1}\left(d_{k+j}\right)\right)\ ,\ for\ k<n-1,\ d_{n-1}=-1.$$

One could also characterize the dynamics of r_t under the P-measure if one assumes a stochastic discount factor. We assume again that the stochastic discount factor is given by[3] (1.27). Hence, one gets

$$\psi_t^P(u) = \psi_t^Q(u-\gamma) - \psi_t^Q(-\gamma). \tag{1.37}$$

Again, this equation is model free and appears in affine models (Gourieroux and Monfort (2007), Monfort and Pegoraro (2006a)). Likewise, r_{t+1} is not a generalized affine process under the P-measure. However in the case of generalized affine model of type I, the vector $(r_{t+1}, \psi_{t+1}(-\gamma))$ is a generalized affine process of order (2,1) as shown in the following proposition. In the sequel, $\psi^P_{r,\psi(-\gamma),t}$ denotes the conditional cumulant function of $(r_{t+1}, \psi_{t+1}(-\gamma))$ under the P-measure:

Proposition 1.3.5. *$\theta^P(u) = u$ implies that*

$$\begin{aligned}\psi^P_{r,\psi(-\gamma),t}(u,v) = {} & \omega_1^P(u,v) + \left(\alpha_1^P(u,v)\, r_t + \alpha_2^P(u,v)\,\psi_t^P(-\gamma)\right) + \beta_1^P(u,v)\,\psi^Q_{r,\psi(-\gamma),t-1}(u,v) \\ & - \alpha_2^Q(u,v)\beta_1^P(u,v)\,\psi_{t-1}^P(-\gamma),\end{aligned} \tag{1.38}$$

where

$$\begin{aligned}\omega_1^P(u,v) &= v\omega^Q(-\gamma)\left[1-\beta^Q\left(u+v\alpha^Q(-\gamma)-\gamma\right)\right] + \omega^Q\left(u+v\alpha^Q(-\gamma)-\gamma\right), \\ \alpha_1^P(u,v) &= \alpha^Q\left(u+v\alpha^Q(-\gamma)-\gamma\right),\ \ \alpha_2^P(u,v) = v\beta^Q(-\gamma)-1, \\ \beta_1^P(u,v) &= \beta^Q\left(u+v\alpha^Q(-\gamma)-\gamma\right).\end{aligned}$$

3. Observe that when one specifies the dynamics of r_t under the Q-measure as a generalized affine process, one could allow γ in (1.27) to be time-varying and adapted to the information available at time t. A consequence is that the short term of interest rate will not be a generalized affine process under the P-measure; see Gourieroux and Monfort (2007) for the same discussion about affine models.

1.3.2 Option Pricing

We now consider models of stock returns where we assume that the conditional variance of the returns is time-varying and is generalized affine. In what follows r_t denotes the log-returns of the stock price, i.e.

$$r_{t+1} = \ln\left(\frac{S_{t+1}}{S_t}\right)$$

The key approach behind the analytical calculations of Heston (1993), Duffie, Pan, and Singleton (2000), and Heston and Nandi (2000) is the possibility to write the joint process (r_t, h_t) as an affine process, where h_{t+1} is the conditional variance of r_{t+1} given an information set that contains r_t and its lagged values and possibly another variables, possibly latent ones as in stochastic volatility models. In what follows, we will allow for both cases. We will write the joint model of (r_{t+1}, h_{t+1}). The variable h_{t+1} could be the conditional variance of r_{t+2} given $\{r_\tau, h_\tau, \tau \leq t\}$ (including GARCH type models). The variable h_{t+1} could be an observable variable like realized volatility as in a different paper we are writing.

In the rest of this section the information set I_t is the sigma algebra generated by $\{r_\tau, h_\tau, \tau \leq t\}$. The conditional expectation operator $E[\cdot \mid I_t]$ will be denoted $E_t[\cdot]$. Likewise, we restrict ourself to the generalized affine model of type I.

1.3.2.1 Generalized Affine process under the P-Measure

We denote the conditional cumulant function of (r_{t+1}, h_{t+1}) under the P measure by $\psi_t^P(u,v)$:

$$\psi_t^P(u,v) = \log E_t^P[\exp(ur_{t+1} + vh_{t+1})] = \omega^P(u,v) + \alpha^P(u,v)h_t + \beta^P(u,v)\psi_{t-1}^P(u,v).$$

When one assumes that h_t is exactly the conditional variance of r_{t+1}, one needs to impose the following restrictions on the cumulant function in order to guarantee this assumption:

$$\frac{\partial^2 \omega^P}{\partial u^2}(0,0) = 0,\ \frac{\partial^2 \alpha^P}{\partial u^2}(0,0) = 1,\ \frac{\partial \beta^P}{\partial u}(0,0) = 0,\ \beta^P(0,0) = 0, \tag{1.39}$$

which implies

$$\frac{\partial^2 \psi_t}{\partial u^2}(0,0) = Var_t^P[r_{t+1}] = h_t.$$

We denote by r the short term interest rate supposed to be constant for simplicity. We consider the following stochastic discount factor

$$M_{t,t+1} = \exp(\gamma r_{t+1} + \lambda h_{t+1} + \theta_t). \tag{1.40}$$

Observe that both Heston and Nandi (2000) and Christoffersen et al. (2006) assumed that $\lambda = 0$. There is no theoretical foundation for such assumption other than simplicity. In other words, we allow the volatility to be priced.

In addition, one needs to impose restrictions in order to guarantee that $M_{t,t+1}$ is a stochastic discount factor, which implies that prices under the Q-measure are martingales. This is the purpose of the following proposition.

Proposition 1.3.6. *The parameters γ and λ are restricted by the following system of equations*

$$\begin{aligned} \omega(1+\gamma,\lambda) - \omega(\gamma,\lambda) &= r(1-\beta(\gamma,\lambda)), \\ \alpha(1+\gamma,\lambda) &= \alpha(\gamma,\lambda), \\ \beta(1+\gamma,\lambda) &= \beta(\gamma,\lambda). \end{aligned}$$

Observe that when $\beta(\cdot)$ is a constant function, the third equation in the previous system holds, which leads to a fully identified system.

We will now characterize the dynamics of (r_{t+1}, h_{t+1}) under the Q-measure by deriving its conditional cumulant function denoted $\psi_t^Q(u,v)$.

Proposition 1.3.7. *We have*

$$\Psi_t^Q(u,v) = \Psi_t^P(u+\gamma, v+\lambda) - \Psi_t^P(\gamma,\lambda), \tag{1.41}$$

and

$$\begin{aligned}\Psi_t^Q(u,v) &= (\omega^P(u+\gamma,v+\lambda)-\omega^P(\gamma,\lambda)) \\ &\quad + (\alpha^P(u+\gamma,v+\lambda)-\alpha^P(\gamma,\lambda))h_t + \beta(u+\gamma,v+\lambda)\Psi_{t-1}^Q(u,v) \\ &\quad + (\beta(u+\gamma,v+\lambda)-\beta(\gamma,\lambda))\psi_{t-1}(\gamma,\lambda).\end{aligned} \tag{1.42}$$

Several remarks are in order. Similarly to the term structure of interest rates, an additional term appears in (1.42), implying that the process (r_{t+1},h_{t+1}) is not generalized affine process under the Q-measure. Likewise, this additional term vanishes when the function $\beta(\cdot)$ is constant. Again, one can still prove that a particular vector is a generalized affine model of higher order, which will allow us to derive option prices. Indeed, one can show that the vector $(r_{t+1},h_{t+1},\psi_{t+1}(\gamma,\lambda))$ is a generalized affine process of order. We now provide the formula of the option prices.

Proposition 1.3.8. *The price at time t of a European call option with payoff $(S_{t+h}-X)^+$ at time $t+h$ is given by*

$$C_t = \exp(-rh)S_tC_{1,t} - \exp(-rh)XC_{2,t}, \tag{1.43}$$

where

$$\begin{aligned} C_{1,t} &= \frac{\exp(rh)}{2} + \int_0^{+\infty}\frac{1}{\pi u} Im\left[\exp\left(\Psi_{t,t+h}^Q(1+iu) - iu\ln\left(\frac{X}{S_t}\right)\right)\right]du, \\ C_{2,t} &= \frac{1}{2} + \int_0^{+\infty}\frac{1}{\pi u} Im\left[\exp\left(-iu\ln\left(\frac{X}{S_t}\right) + \Psi_{t,t+h}^Q(iu)\right)\right]du, \end{aligned}$$

and

$$\Psi_{t,t+h}^Q(u) = a_h(u)h_t + b_h(u) + \sum_{k=1}^{h}\left(\beta^P(d_k)^k\psi_{t-1}^P(d_k) - \beta^P(\gamma,\lambda)^k\psi_{t-1}^P(\gamma,\lambda)\right),$$

with

$$a_h(u) = \sum_{k=1}^{h}\left(\beta^P(d_k)^k\alpha^P(d_k) - \beta^P(\gamma,\lambda)^k\alpha^P(\gamma,\lambda)\right),$$

$$b_h(u) = \sum_{k=1}^{h} \left(\frac{1-\beta^P(d_k)^k}{1-\beta^P(d_k)} \omega^P(d_k) - \frac{1-\beta^P(\gamma,\lambda)^k}{1-\beta^P(\gamma,\lambda)} \omega^P(\gamma,\lambda) \right),$$

and

$$\begin{aligned} d_k &= (u+u_k, v_k) + \sum_{j=k}^{h-1} \beta\left(d_{j+1}\right)^{j-k} \left(0, \alpha\left(d_{j+1}\right)\right) \textit{ for } k \leq h-1, \\ d_h &= (u+u_h, v_h), \end{aligned} \tag{1.44}$$

where

$$\begin{aligned} u_h &= \gamma,\ v_h = \lambda, \\ u_j &= \gamma - \alpha(\gamma,\lambda) \frac{1-\beta(\gamma,\lambda)^{h-j}}{1-\beta(\gamma,\lambda)} \textit{ for } 1 \leq j < h, \\ v_j &= \lambda - \alpha(\gamma,\lambda) \frac{1-\beta(\gamma,\lambda)^{h-j}}{1-\beta(\gamma,\lambda)} \textit{ for } 1 \leq j < h, \end{aligned}$$

$$\begin{aligned} \psi_{t-1}^P(d_k) &= \omega^P(d_k) + \alpha^P(d_k) h_{t-1} + \beta^P(d_k) \psi_{t-2}^P(d_k) \\ &= \frac{\omega^P(d_k)}{1-\beta^P(d_k)} + \alpha^P(d_k) \left(\sum_{i=0}^{\infty} \beta^P(d_k)^i h_{t-1-i} \right). \end{aligned}$$

This proposition uses Fourier transforms, which is a traditional approach in affine models. It is important to note that, for this purpose, we had to use the logarithmic of the characteristic function instead of the logarithmic of the moment generating function. A simple modification of the notation is sufficient to do this change.

1.3.3 Generalized Affine process under the Q-Measure

This subsection specifies the dynamics of (r_{t+1}, h_{t+1}) under the Q-measure, $\Psi_t^Q(u,v)$, and derives the option prices. We assume that

$$\Psi_{t+1}^Q(u,v) = \omega(u,v) + \alpha(u,v) h_{t+1} + \beta(u,v) \Psi_t^Q(u,v). \tag{1.45}$$

A well defined risk-neutral distribution for log-returns must satisfy

$$\exp(r) = E^{Q}\left[\exp\left(r_{t+1}\right)|I_t\right],$$

where r is the risk-free rate. Thus $\Psi_t^Q(1,0)$ must satisfy

$$\Psi_t^Q(1,0) = r.$$

Proposition 1.3.9. *Eq. (1.45) is a valid risk-neutral model if and only if*

$$\begin{aligned} \frac{\omega(1,0)}{1-\beta(1,0)} &= r, \\ \alpha(1,0) &= 0. \end{aligned} \tag{1.46}$$

The result is an implication of the following representation:

$$\Psi_{t+1}^Q(u,v) = \frac{\omega(u,v)}{1-\beta(u,v)} + \alpha(u,v)\sum_{i=0}^{\infty}\beta(u,v)^i h_{t-i+1}.$$

We are now able to characterize the option prices.

Proposition 1.3.10. *When* (1.46) *holds, the price at time t of European call option with payoff $(S_{t+h}-X)^+$ at time $t+h$ is given by*

$$C_t = \exp(-rh)S_tC_{1,t} - \exp(-rh)XC_{2,t},$$

where

$$\begin{aligned} C_{1,t} &= \frac{\exp(rh)}{2} + \int_0^{+\infty}\frac{1}{\pi u}Im\left[\exp\left(\Psi_{t,t+h}^Q(1+iu) - iu\ln\left(\frac{X}{S_t}\right)\right)\right]du, \\ C_{2,t} &= \frac{1}{2} + \int_0^{+\infty}\frac{1}{\pi u}Im\left[\exp\left(-iu\ln\left(\frac{X}{S_t}\right) + \Psi_{t,t+h}^Q(iu)\right)\right]du, \end{aligned}$$

$$\Psi_{t,t+h}^Q(u) = a_h(u)h_t + b_h(u) + \sum_{k=1}^{h}\left(\beta(d_k)^k\psi_{t-1}(d_k)\right),$$

with

$$a_h(u)=\sum_{k=1}^{h}\left(\beta(d_k)^{k-1}\alpha(d_k)\right),\ b_h(u)=\sum_{k=1}^{h}\left(\frac{1-\beta(d_k)^k}{1-\beta(d_k)}\omega(d_k)\right),$$

$$\psi_{t-1}(d_k)=\frac{\omega(d_k)}{1-\beta(d_k)}+\alpha(d_k)\left(\sum_{i=0}^{\infty}\beta(d_k)^i h_{t-1-i}\right),$$

and the sequence $(d_k)_{1\leq k\leq h}$ *is defined as follows:*

$$\begin{aligned} d_k &= (u,0)+\sum_{j=k}^{h-1}\beta\left(d_{j+1}\right)^{j-k}\left(0,\alpha\left(d_{j+1}\right)\right) \text{ for } k\leq h-1, \\ d_h &= (u,0). \end{aligned} \tag{1.47}$$

We will use these formulas in the empirical section.

1.4 Estimation of generalized affine models

In general we can distinguish between parametric and non-parametric estimation methods. In the 2002 version of Darolles et al.(2006), a detailed discussion have been made on non-parametric estimation of discrete time affine models. In this section we focus on parametric methods. This implies that we consider the general affine model defined by (1.8) where the functions $\omega(\cdot)$, $\alpha(\cdot)$, $\beta(\cdot)$, and $\theta(\cdot)$ are specified up to unknown vector of parameters, which we aim to estimate. Among parametric methods we will consider empirical characteristic function method (ECF) which has been used by Singleton (2001), the generalized method of moment (GMM), the quasi-maximum likelihood and the maximum likelihood.

1.4.1 Empirical characteristic function

Our modeling strategy differs fundamentally from the "classic" approach which consists in writing down an equation for each component (it could be the mean, the variance, the intensity of the jump component...) of the conditional distribution of the process of interest. We model directly the conditional characteristic function of the process. For this reason the ECF seems to be the most natural approach to estimate efficiently our generalized affine models. The basic idea behind ECF is to match the theoretical characteristic function given by the model and the empirical characteristic function obtained from data. This approach has several advantages. It avoids difficulties inherent in calculating and maximizing the likelihood function. Although the likelihood function can be unbounded, its Fourier transform (which is the characteristic function) is always bounded. The important theorem which establishes the one-to-one correspondence between characteristic function (CF) and the cumulative distribution function (CDF) suggests that estimation and inference via the ECF should be as efficient as the likelihood-based approaches (Carrasco and Florens (2002)).

Let first recall the ECF principle in the case of affine models. It means that we are considering $\beta(u) = 0$ or $\theta(u) = 0$. Let denote λ_0 the unknown parameter to estimate. We can rewrite the affine model (1.1) as follows:

$$E\left[\exp\left(ux_{t+1}\right) - \exp\left(\omega(u;\lambda_0) + \alpha(u;\lambda_0)x_t\right)|x_s, s \leq t\right] = 0 \;\forall u. \tag{1.48}$$

This implies that for any weighting function, w (often termed instruments in the GMM literature), we have:

$$E\left[\left(\exp\left(ux_{t+1}\right) - \exp\left(\omega(u;\lambda_0) + \alpha(u;\lambda_0)x_t\right)\right)w(x_t, v)\right] = 0 \;\forall u, v. \tag{1.49}$$

This leads to continuum of moments restrictions; hence we can estimate by applying the GMM to a continuum of moments restrictions (see Carrasco, Florens (2000)). The ML efficiency is achieved by choosing Carrasco et al. (2002) weighting function, i.e $w(x_t, v) = e^{vx_t}$.

Let us consider now the generalized affine model (1.8), the model can be rewrite as follow:

$$\psi_t(u) = \bar{\beta}_j(u)\,\psi_{t-j}\left(\theta_j(u)\right) + \sum_{k=1}^{j} \bar{\beta}_{k-1}(u)\left[\omega(\theta_{k-1}(u)) + \alpha(\theta_{k-1}(u))x_{t-k+1}\right]. \tag{1.50}$$

This implies that for any weighting function, $w(\cdot,\cdot)$

$$E\left[\left(e^{ux_{t+1}} - e^{\bar{\beta}_t(u;\lambda_0)\psi_0(\theta_t(u;\lambda_0);\lambda_0)+\Sigma_{k=1}^{t}\bar{\beta}_{k-1}(u;\lambda_0)[\omega(\theta_{k-1}(u;\lambda_0);\lambda_0)+\alpha(\theta_{k-1}(u;\lambda_0);\lambda_0)x_{t-k+1}]}\right) w(\underline{x}_t, v)\right] = 0 \tag{1.51}$$

$\forall u, v$, where $\underline{x}_t = \left(x_{t-q+1}, ..., x_t\right)'$ and $v = \left(v_1, ..., v_q\right)'$. Similarly to the affine model results, we specify the following weighting function $w(\underline{x}_t, v) = e^{\underline{x}_t' v}$.

1.4.2 Generalized method of moments

One of the advantages of the discrete time affine models is that we can compute unconditional moments of any component of the process of interest. This point have been studied in detail in Feunou and Tedongap (2009). This is an important result for estimation purpose because even when there are some unobserved components in the process of interest, we can still compute the moments of observed components and use them to implement a GMM estimation routine. It turns out that we keep this advantage in the generalized affine model as shown in section 2. In the case of observable variable of interest, there is no need to compute the unconditional moments. We can use the conditional moments equations which have been derived in section 2. The derivation of conditional moments equations has nothing to do with the affine structure of the model, but it is the consequence of the fact that we have written a model directly on conditional characteristic function. This means that in the case of generalized non-affine models (1.9), we can still compute conditional moments equations.

The following moment conditions have been used in the literature by Bollerslev and zhou (2002) to estimate one factor and two factors stochastic volatility models by means of conditional moments of realized variance:

$$\begin{aligned}
E\left[x_{t+1}-\mu_{1,0}\right]=0\,,\ E\left[x_{t+1}^2-\mu_{2,0}\right]&=0,\\
E\left[(x_{t+1}-\mu_{1,t})x_t\right]=0\,,\ E\left[\left(x_{t+1}^2-\mu_{2,t}\right)x_t\right]&=0,\\
E\left[(x_{t+1}-\mu_{1,t})x_t^2\right]=0\,,\ E\left[\left(x_{t+1}^2-\mu_{2,t}\right)x_t^2\right]&=0,\\
E\left[(x_{t+1}-\mu_{1,t})x_{t-1}\right]=0\,,\ E\left[\left(x_{t+1}^2-\mu_{2,t}\right)x_{t-1}\right]&=0,\\
E\left[(x_{t+1}-\mu_{1,t})x_{t-1}^2\right]=0\,,\ E\left[\left(x_{t+1}^2-\mu_{2,t}\right)x_{t-1}^2\right]&=0,
\end{aligned}$$

where $\mu_{1,0}=E(x_{t+1})$, $\mu_{2,0}=E(x_{t+1}^2)$, $\mu_{1,t}=E_t\left[x_{t+1}\right]$ and $\mu_{2,t}=E_t\left[x_{t+1}^2\right]$. We simulate the generalized autoregressive gamma (GARG) model built section (1.2.1.5) with $\mu=2.784E-05$, $\nu=0.1394$, $\rho=0.1125$, $\beta=0.9227$ and $\theta=0.9066$. These parameters have been obtained by estimating the GARG on realized variance data. For different sample sizes (T) and number of replications (N) we estimate the GARG and report in table 1.1 different statistics (mean, median and root mean square errors (RMSE)) across different replication sizes. The GMM does well if we consider the longest sample size (1000) and the biggest number of replications (4000).

1.4.3 Maximum and quasi-maximum likelihood method

In general, the conditional likelihood can be obtained from the conditional characteristic function using the following inversion formula:

$$\begin{aligned}
f\left(x_{t+1}|I_t\right) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\exp\left(-iux_{t+1}+\psi_t(iu)\right)du \qquad (1.52)\\
&= \frac{1}{\pi}\int_{0}^{+\infty}R_e\left[\exp\left(-iux_{t+1}+\psi_t(iu)\right)\right]du.
\end{aligned}$$

Most of the time, we do not have the likelihood function in closed form, except for some specific generalized affine processes like VARMA, GARG with $\theta(u)=u$. When $\theta(u)\neq u$, we estimate the GARG on interest rate and realized variance data by maximizing (1.53) (using numerical integration tools), we obtain significant estimators of parameters β and θ. We did not report the results in this chapter. To circumvent the numerical integration difficulties, we can derive the first two conditional moments equations (see section 2 for details on conditional moments structure) and then implement a quasi-maximum likelihood. In the case of processes with positive support, we can use the standard gamma density (with two parameters), a Gaussian density being used when the support is the whole real line. Finally we run the same Monte-Carlo exercise as in section 1.4.2, but we use the QMLE method (with gamma density) instead of the GMM. Results are summarized in table 1.2. Compared to GMM, the QMLE have some difficulties in estimating the parameters ν and μ.

1.5 Empirical Application: No-Arbitrage VARMA Term Structure Models

We give more details on VARMA term structure models in a companion paper (Feunou (2009)). Ang and Piazzesi (2003) argued that macro-economic variables add something new in traditional three factors term structure models. One motivation of generalized affine models studied in this chapter is the possibility that they allow one to reduce the dimension of the state vector usually used in affine models. In particular we argue that an affine model on a vector of dimension $n+1$ is less parsimonious and would yield poor out-of sample forecasts compared to a generalized affine model of order (1,1) on n components of the same vector. In the context of this example, we will compare for instance a VARMA model on three observed yields to a VAR on the same three observed yields and one macroeconomic variable. We will run an out-of-sample forecasting exercise to compare the considered models. In the literature the "Nelson-Sielgel with AR(1) factor dynamic" is known as one of the best performer in forecasting the entire yield curve at different horizons (Diebold and Li (2006)). Our competitors are then the "Nelson-Sielgel with AR(1) factor dynamic" and the random walk model (Duffee(2002)) where the random walk model is shown to provide better out-of-sample forecasting results compared to a many of affine models.

The state vector is denoted by Z_{t+1}, we consider the following VARMA(1,1) dynamic under the historical probability measure P:

$$Z_{t+1} = \mu + \phi Z_t + \Sigma(\varepsilon_{t+1} - \Theta\varepsilon_t), \;\; \varepsilon \sim i.i.d. \mathcal{N}(0,I).$$

As shown in section 1 and 2, this model is part of the generalized affine class. Indeed the conditional cumulant function of Z_{t+1} denoted $\psi_t(u)$ satisfies the following recursive relation:

$$\psi_t(u) = \omega(u) + \alpha(u)'Z_t + \psi_{t-1}(\theta(u)),$$

with

$$\omega(u) = u'\mu + \frac{1}{2}u'\Sigma\left(I_4 - \Theta\Theta'\right)\Sigma'u, \;\; \alpha(u) = u'\left(\phi - \Sigma\Theta\Sigma^{-1}\right), \; \theta(u) = \left(\Sigma\Theta\Sigma^{-1}\right)'u.$$

The specification of the pricing kernel is similar to one used in Ang and Piazzesi (2003),

$$M_{t+1} = \exp\left(-y_t^{(1)} - \frac{1}{2}\lambda_t^\top\lambda_t - \lambda_t^\top\varepsilon_{t+1}\right).$$

To maintain the tractability of the model, affine price of risk (λ_t) is often used $\lambda_t = \lambda_0 + \lambda_1 Z_t$, (see Ang and Piazzesi (2003)). In the context of our VARMA model, Feunou(2009) discusses this specification and adds the conditional expectation of the state vector $E_t(Z_{t+1})$, i.e., he assumes

$$\lambda_t = \lambda_0 + \lambda_1 Z_t + \lambda_2 E_t(Z_{t+1}). \tag{1.53}$$

Several interpretations can be given to Eq. (2.3). First, we can reformulate it as follows:

$$\begin{aligned} \lambda_t &= \lambda_0 + \lambda_1 Z_t + \lambda_2\left(\mu + \left(\phi - \Sigma\Theta\Sigma^{-1}\right)Z_t + \Sigma\Theta\Sigma^{-1}E_{t-1}(Z_t)\right) \\ &= \lambda_0^* + \lambda_1^* Z_t + \lambda_2^* E_{t-1}(Z_t), \end{aligned}$$

where

$$\lambda_0^* = \lambda_0 + \lambda_2\mu, \;\; \lambda_1^* = \lambda_1 + \lambda_2\left(\phi - \Sigma\Theta\Sigma^{-1}\right), \;\text{ and }\; \lambda_2^* = \lambda_2\Sigma\Theta\Sigma^{-1}.$$

Thus, the parameter λ_2 captures the past information impact on the current market price of risk. Another way of rewriting the price of risk is to express it in terms of the expected variable $E_{t-1}(Z_t)$ and the unexpected news $\Sigma\varepsilon_t$:

$$\lambda_t = \lambda_0^* + \lambda_1^* \Sigma\varepsilon_t + (\lambda_1^* + \lambda_2^*) E_{t-1}(Z_t).$$

Feunou (2009) shows that bond yields (with maturity n) are no longer affine of the state variable Z_t, but are rather affine function of the state variable Z_t and it lagged conditional expectation $E_{t-1}(Z_t)$, i.e.,

$$y_t^{(n)} = a_n + b_{1,n}^\top Z_t + b_{2,n}^\top E_{t-1}(Z_t). \tag{1.54}$$

Another representation derived in Feunou (2009) is

$$y_t^{(n)} = a_n + (b_{1,n} + b_{2,n})^\top E_{t-1}(Z_t) + b_{1,n}^\top \Sigma\varepsilon_t, \tag{1.55}$$

where the coefficients a_n, $b_{1,n}$, and $b_{2,n}$ are given in Feunou (2009) (next chapter of the thesis).

The estimation of the unknown parameters, i.e., the parameters of the historical distribution and those of the price of risk, is done in two steps. The first step estimates the parameters of the historical distribution of the state vector by using the maximum likelihood method. By taking the parameters of the historical dynamic to their estimated values (results of the first step), we estimate in the second step the pricing kernel's parameters by minimizing the squared difference between the model implied yields and the observed yields (in practice, the maturities are 3, 12, 36, 60 and 120 months). Since there are observed yields in the state vector, Feunou (2009) (see also Ang et al (2006)) argued that the second step is a constrained optimization problem.

The yields data used have been obtained from unsmoothed Fama-Bliss forward rates (see Diebold and Li (2006) for full details on the construction and description of these yields data) and they are the same used in Diebold and Li (2006). The Macroeconomic Data are the two factors termed "inflation" and "real activity" used in Ang and Piazzesi

(2006). We estimate and forecast recursively, using data from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. In Tables (1.3) to (1.5), we compare h-months-ahead out-of sample forecasting results from "Nelson–Siegel with AR(1) factor dynamic" to no arbitrage VAR and VARMA models, for maturities of 3, 12, 36, 60 and 120 months, and forecast horizons of $h = 1, 6$ and 12 months.

To summarize the RMSE results, the best model in DL(2006) ("Nelson–Siegel with AR(1) factor dynamic model") performs better only at horizon 1 for 3 and 120 months yield. In general the best performer is the VARMA(1,1) model on 1, 24, and 96 months yield, except for the 10 years yield at horizon 6 months where the VAR(1) model on 1, 24, and 96 months yield and inflation performs better.

In conclusion, by using observable variables in either a no-arbitrage VAR or VARMA model we can do better than "Nelson–Siegel with AR(1) factor dynamic" in forecasting the entire yield curve at any horizon. Macro-economic factors add new information which are not contained in affine yields only model, but we can cope with these macro-economic factors by implementing a no arbitrage generalized affine model (here the VARMA) on the classic three factors model. In other words, we find out empirically that the generalized affine model is a reduced form of an affine model with more variables.

1.6 Conclusion

This chapter extends affine models by introducing moving average type components in the conditional cumulant function. The extension is important theoretically because important models like ARMA are not affine, beside that we show how we can build parsimonious infinite order affine models. The extension is also empirically important as shown in the empirical example. In particular, the term structure exercise shows that a generalized affine model on traditional three factors term structure model forecasts better the entire yield curve than an affine model on three factors and macroeconomic variable.

We are currently implementing the same model in two applications using realized volatility (Feunou and Meddahi (2009), and Feunou, Christoffersen, Jacobs, and Meddahi (2009)). The preliminary results are quite promising.

There is an alternative approach that leads to non-Markov affine processes. It uses the conditional Laplace transform of the process x_t defined as $\mathscr{L}_t(u) = \exp(\psi_t(u))$ instead of the cumulant function. The traditional affine models are characterized by

$$\mathscr{L}_t(u) = \exp(\omega(u) + \alpha(u)x_t).$$

In a companion paper, we are currently studying the process defined by

$$\mathscr{L}_t(u) = \gamma(u) + \exp(\omega(u) + \alpha(u)x_t) + \beta(u)\mathscr{L}_{t-1}(u).$$

BIBLIOGRAPHY

[1] Ang, A., and M. Piazzesi (2003), A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, Journal of Monetary Economics, 50, 745-787.

[2] Ang, A., M. Piazzesi, and M. Wei (2006), What does the yield curve tell us about GDP growth, Journal of Econometrics, 131, 359-403.

[3] Barndorff-Nielsen, OE. (1997), Normal inverse Gaussian distributions and stochastic volatility modelling, Scandinavian Journal of Statistics 24, 1-13.

[4] Bollerslev, T. (1986), Generalized autoregressive conditional heteroscedasticity, Journal of Econometrics 31, 307-327.

[5] Bollerslev, T., and L. Forsberg (2002), Bridging the gap between the distribution of realized (ECU) and ARCH modelling (of the euro): The GARCH-NIG model, Journal of Applied Econometrics 17, 535-548.

[6] Bollerslev, T. and H. Zhou (2002), Estimating stochastic volatility diffusion using conditional moments of integrated volatility, Journal of Econometrics, 109, 33-65.

[7] Carrasco, M., and J-P. Florens (2002), Efficient GMM estimation using the empirical characteristic function, Manuscript, University of Rochester.

[8] Carrasco, M., and J-P. Florens (2000), Generalization of GMM to a continuum of moment conditions, Econometric Theory, 16 797-834.

[9] Carrasco, M., M. Chernov, J-P. Florens, and E. Ghysels (2007). Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions, Journal of Econometrics, 140 529-573.

[10] Christoffersen, P. (1998), Valuating interval forecasts, International Economic Review 39, 841-862.

[11] Christoffersen, P., R. Elkamhi, B. Feunou, and K. Jacobs (2009), Option valuation with conditional heteroskedasticity and non-normality, Review of Financial Studies, Forthcoming.

[12] Christoffersen, P., S. Heston, and K. Jacobs (2006), Option valuation with conditional skewness, Journal of Econometrics 131 253-284.

[13] Christoffersen, P., and K. Jacobs (2004), Which GARCH model for option valuation?, Management Science, 50, 1204-1221.

[14] Christoffersen, P., C. Dorion, K. Jacobs, and Y. Wang (2005), Volatility components: affine restrictions and non-normal innovations, Journal of Business and Economic Statistics, Forthcoming.

[15] Cox, J.C., J.E. Ingersoll, and S.A. Ross (1985), A theory of the term structure of interest rates, Econometrica 53, 385-407.

[16] Dai, Q., and K. Singleton (2000), Specification analysis of affine term structure models, Journal of Finance 55, 1943-1978.

[17] Darolles, S., C. Gourieroux, and J. Jasiak (2006), Structural laplace transform and compound autoregressive models, Journal of time serie analysis, 27, 477-503.

[18] Diebold, F., M. Piazzesi, and D. Rudebusch (2005), Modeling bond yields in finance and macroeconomics, American Economic Review, 95, 415-420.

[19] Diebold, F., D. Rudebusch, and B. Aruoba (2006), The macroeconomy and the yield curve: a dynamic latent factor approach, Journal of Econometrics 131 309-338.

[20] Diebold, F., T.A. Gunther, and T.-S. Tay (1998), Evaluating density forecasts with applications to financial risk management. International Economic Review, 39, 863-883.

[21] Diebold, F., and C. Li (2006), Forecasting the term structure of government bond yields, Journal of Econometrics, 130 337-364.

[22] Duan, J.-C. (1995), The GARCH option pricing model, Mathematical Finance, 5, 13-32.

[23] Duffee, G.-R. (2002), Term premia and interest rate forecasts in affine models, The Journal of Finance, 57, 405-443.

[24] Duffie, D., J. Pan, and K. Singleton (2000), Transform analysis and asset pricing for affine jump diffusions, Econometrica, 68, 1343-1376.

[25] Duffie, D., and R. Kan(1996), A yield-factor model of interest rates, Mathematical Finance, 6, 379-406.

[26] Engle, R. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, Econometrica, 50, 987-1008.

[27] Feller, W. (1968), An introduction to probability theory and its applications, John Wiley & Sons.

[28] Feuerverger, A. and R.-A. Mureika (1977), The empirical characteristic function and its applications, The annals of Statistics, 5, 88-97.

[29] Feuerverger, A., and P. McDunnough (1981), On the efficiency of empirical characteristic function procedure, Journal of the Royal Statistical Society, Series B, 43, 20-27.

[30] Feunou, B. (2007), No-arbitrage VARMA term structure models with macroeconomic variables, Working Paper, Université de Montréal.

[31] Feunou, B., and N. Meddahi (2007), Realized term structure of risk, Work in Progress.

[32] Feunou, B., P. Christoffersen, K. Jacobs, and N. Meddahi (2007), Realized option pricing models, Work in progress.

[33] Feunou, B., and R. Tedongap (2007). Affine stochastic Sskewness models, Working Paper, Université de Montréal.

[34] Gourieroux, C. (2006), Continuous time Wishart process for stochastic risk, Econometric Reviews, 25, 177-217.

[35] Gourieroux, C., and J. Jasiak (2002), Value at risk, Handbook of financial Econometric.

[36] Gourieroux, C., and J. Jasiak (2006), Autoregressive gamma processes, Journal of Forecasting, 25, 129-152.

[37] Gourieroux, C., A. Monfort, and V. Polimenis (2002), Affine term structure models, Working Paper, CREST.

[38] Gourieroux, C., and A. Monfort (2006), Econometric specifications of stochastic discount factor models, Journal of Econometrics, 136, 509-530.

[39] Hansen, L.-P. (1982), Large sample properties of generalized method of moment estimators, Econometrica, 50, 1029-1053.

[40] Heston, S. (1993), A Closed-form solution for options with stochastic volatility with applications to bond and currency options, Review of Financial Studies, 6,327-343.

[41] Heston, S., and S. Nandi (2000), A closed-form GARCH option pricing model, Review of Financial Studies, 13, 585-626.

[42] Johnson, N.-L., S. Kotz, and N. Balakrishnan, Continuous univariate distributions, Volume1, Second Edition, Wiley series in probability and mathematical statistics.

[43] Monfort, A., and F. Pegoraro (2006), Switching VARMA term structure models, Journal of Financial Econometrics, 5, 105-153.

[44] Monfort, A. and F., Pegoraro (2006), Multi-lag term structure models with stochastic risk premia, Working paper, CREST.

[45] Lukacs, E. (1970), Characteristic functions, Griffin 1820 London.

[46] Singleton, K. (2001), Estimation of affine asset pricing models using the empirical characteristic function, Journal of Econometrics, 102, 111-141.

[47] Vasicek, O. (1977), An equilibrium characterization of term structure, Journal of Financial Economics, 5, 177-188.

[48] Yu, J. (2004), Empirical characteristic function estimation and its applications, Econometric Reviews, 23, 93-123.

Appendix A

In this appendix, we build a generalized affine model of type I where the function $\beta(\cdot)$ varies. Let us consider a positive process X_t with conditional cumulant function Ψ_t.

$$E_t\left[\exp\left(uX_{t+1}\right)\right] = \exp\left(\Psi_t\left(u\right)\right).$$

We define $\Psi_t\left(u\right)$ recursively as follows:

$$\begin{aligned}
\Psi_0\left(u\right) &= \omega\left(u\right) + a_0\left(u, X_0\right) \\
\Psi_1\left(u\right) &= \omega\left(u\right) + a_0\left(u, X_1\right) + a_1\left(u, X_0\right) \\
\Psi_2\left(u\right) &= \omega\left(u\right) + a_0\left(u, X_2\right) + a_1\left(u, X_1\right) + a_2\left(u, X_0\right),
\end{aligned}$$

and generally, we have

$$\Psi_t\left(u\right) = \omega\left(u\right) + \sum_{i=0}^{t} a_i\left(u, X_{t-i}\right). \tag{1.56}$$

The first issue is to give some conditions on sequence functions $a_i(u,x)$ and $\omega\left(u\right)$ such that $\Psi_t(u)$ is a well defined cumulant function.

If $\omega\left(u\right)$ and $a_i\left(u,x\right)$ are cumulant functions $\forall\, i$, then $\Psi_t\left(u\right)$ is a well defined cumulant function. Indeed, the sum of cumulant function is a cumulant function.

Consequently we will choose $\omega\left(u\right)$ and $a_i\left(u,x\right)$ such that they will be always cumulant functions. Another consequence is the fact that we can write X_{t+1} as follows

$$X_{t+1} = \eta_{t+1} + \sum_{i=0}^{t} Z_{i,t+1},$$

where η_{t+1} and $Z_{i,t+1}$ are mutually conditionally independent with cumulant function $\omega\left(u\right)$ and $a_i\left(u, X_{t-i}\right)$. This give us a simple approach to simulate X_{t+1}.

The final goal is to rewrite definition of $\Psi_t\left(u\right)$ given by (1.56) recursively. To achieve this goal the following expression is given to cumulant function $a_i\left(u,x\right)$

$$\begin{aligned}
a_i\left(u,x\right) &= P_i\left(x\right)\left[\exp\left(a\left(u\right)i + b\left(u\right)\right) - 1\right] \\
P_i\left(x\right) &= \exp\left(\lambda_0 + \lambda_1 i\right)x.
\end{aligned} \tag{1.57}$$

As it was the case with $\Psi_t\left(u\right)$, we need to make sure that (1.57) is a valid cumulant function.

This is done using Lemma 5.4.1 of Lukacs (1970) (page 111) where it is shown that $p(g(u)-1)$ is an infinitely divisible cumulant function whenever $g(u)$ is a characteristic function and $p>0$.

Thus if $a(u)$ and $b(u)$ are cumulant functions and X a positive process, then $a_i(u,x)$ is a cumulant function. Since process X_t is built using cumulant generating function, it is hard to simulate. We give an answer in the following lines. Proposition 1.6.1 shows how a random variable with cumulant function $a_i(u,x)$ can be simulated.

Proposition 1.6.1. *$p(g(u)-1)$ is the cumulant function of Z iff*

$$Z=\sum_{n=0}^{N} Y_n,$$

where random variables N and Y_n are mutually independent, N follows Poisson distribution of parameter p and the moment generating function of Y_n is $g(u)$.

Since $\Psi_t(u)$ is the conditional cumulant function of X_{t+1} (which is a positive random variable), we must then choose $\omega(u)$, $a(u)$ and $b(u)$ such that $\Psi_t(u)$ is a cumulant function of a positive random variable. The following proposition addresses this issue.

Proposition 1.6.2. *If $a(u)$, $b(u)$ and $\omega(u)$ are cumulant functions of positive random variable, then $\Psi_t(u)$ is a well defined conditional cumulant function of positive random variable X_{t+1}*

We are now ready to write $\Psi_t(u)$ recursively.

$$\begin{aligned}
\Psi_t(u) &= \omega(u)+\sum_{i=0}^{t} P_i(X_{t-i})\left[\exp(a(u)i+b(u))-1\right] \\
&= \omega(u)+\sum_{i=0}^{t} P_i(X_{t-i})\exp(a(u)i+b(u))-\sum_{i=0}^{t} P_i(X_{t-i}) \\
&= \omega(u)+\sum_{i=0}^{t}\exp((a(u)+\lambda_1)i+\lambda_0+b(u))X_{t-i}-\sum_{i=0}^{t}\exp(\lambda_0+\lambda_1 i)X_{t-i}.
\end{aligned}$$

Proposition 1.6.3. *$\omega(u)$ can always be reformulated as following*

$$\omega(u)=\frac{c(u)}{1-\exp(a(u)+\lambda_1)}-\frac{c(0)}{1-\exp(\lambda_1)}.$$

As shown below, the proof of Proposition 1.6.3 is a direct consequence of the fact that $\omega(u)$ is a cumulant function.

We can then rewrite $\Psi_t(u)$ as following.

$$\Psi_t(u) = f_t(u) - f_t(0),$$

with

$$\begin{aligned} f_t(u) &= \frac{c(u)}{1-\exp(a(u)+\lambda_1)} + \sum_{i=0}^{t} \exp((a(u)+\lambda_1)i + \lambda_0 + b(u))X_{t-i} \\ &= \frac{c(u)}{1-\exp(a(u)+\lambda_1)} + \exp(\lambda_0 + b(u)) \sum_{i=0}^{t} \exp(a(u)+\lambda_1)^i X_{t-i}. \end{aligned}$$

Proposition 1.6.4. *$f_{t+1}(u)$ evolves recursively as follows:*

$$f_{t+1}(u) = c(u) + \exp(\lambda_0 + b(u))X_{t+1} + \exp(a(u)+\lambda_1) f_t(u)$$

An immediate consequence of Proposition 1.6.4 is the recursive formulation of $\Psi_t(u)$ given by proposition 1.6.5.

Proposition 1.6.5.

$$\Psi_t(u) = \omega_0(u) + \alpha_1(u) f_t(0) + \alpha_2(u) f_{t-1}(0) + \beta(u)\Psi_{t-1}(u), \tag{1.58}$$

where

$$\begin{aligned} \omega_0(u) &= c(u) - c(0)e^{b(u)} \\ \alpha_1(u) &= e^{b(u)} - 1 \\ \alpha_2(u) &= e^{\lambda_1}\left[e^{a(u)} - e^{b(u)}\right] \\ \beta(u) &= e^{\lambda_1 + a(u)}. \end{aligned}$$

Note that in the right hand side of equation (1.58), we have $f_t(0)$ instead of X_t. For this reason the conditional cumulant generating function of $f_t(0)$ is evaluated. The joint conditional moment generating function of X_{t+1} and $f_{t+1}(0)$ is:

$$\begin{aligned} E_t\left[\exp\left(uX_{t+1}+vf_{t+1}(0)\right)\right] &= E_t\left[\exp\left(uX_{t+1}+v\left(c(0)+e^{\lambda_0}X_{t+1}+e^{\lambda_1}f_t(0)\right)\right)\right] \\ &= \exp\left(vc(0)+ve^{\lambda_1}f_t(0)+\Psi_t(u+ve^{\lambda_0})\right). \end{aligned}$$

Thus if we denote $\Psi_t^c(u,v)=\ln\left(E_t\left[\exp\left(uX_{t+1}+vf_{t+1}(0)\right)\right]\right)$, we have

$$\begin{aligned} \Psi_t^c(u,v) &= vc(0)+ve^{\lambda_1}f_t(0)+\Psi_t(u+ve^{\lambda_0}) \\ &= vc(0)+ve^{\lambda_1}f_t(0)+\omega_0\left(u+ve^{\lambda_0}\right)+\alpha_1\left(u+ve^{\lambda_0}\right)f_t(0) \\ &\quad +\alpha_2\left(u+ve^{\lambda_0}\right)f_{t-1}(0)+\beta\left(u+ve^{\lambda_0}\right)\Psi_{t-1}(u+ve^{\lambda_0}) \\ &= vc(0)+ve^{\lambda_1}f_t(0)+\omega_0\left(u+ve^{\lambda_0}\right)+\alpha_1\left(u+ve^{\lambda_0}\right)f_t(0) \\ &\quad +\alpha_2\left(u+ve^{\lambda_0}\right)f_{t-1}(0)+\beta\left(u+ve^{\lambda_0}\right)\left[\Psi_{t-1}^c(u,v)-vc(0)-ve^{\lambda_1}f_{t-1}(0)\right]. \end{aligned}$$

The whole expression of $\Psi_t^c(u,v)$ is summarized in the following proposition.

Proposition 1.6.6.

$$\Psi_t^c(u,v)=W(u,v)+A_1(u,v)f_t(0)+A_2(u,v)f_{t-1}(0)+B(u,v)\Psi_{t-1}^c(u,v),$$

where

$$\begin{aligned} W(u,v) &= vc(0)\left(1-\beta\left(u+ve^{\lambda_0}\right)\right)+\omega_0\left(u+ve^{\lambda_0}\right) \\ A_1(u,v) &= ve^{\lambda_1}+\alpha_1\left(u+ve^{\lambda_0}\right) \\ A_2(u,v) &= \alpha_2\left(u+ve^{\lambda_0}\right)-ve^{\lambda_1}\beta\left(u+ve^{\lambda_0}\right) \\ B(u,v) &= \beta\left(u+ve^{\lambda_0}\right). \end{aligned}$$

In conclusion the vector $(X_{t+1},f_{t+1}(0))$ is a generalized affine of order (2,1), implying a univariate generalized affine for $f_{t+1}(0)$ as stated in the following corollary.

Corollary 1. *Notice that by imposing* $u=0$ *we have a generalized affine model of order* $(2,1)$ *for* $f_t(0)$.

Indeed

$$E_t\left[\exp\left(vf_t(0)\right)\right] \equiv \exp\left(\Psi_t^f(v)\right) = \exp\left(\Psi_t^c(0,v)\right),$$

with

$$\Psi_t^c(0,v) = W(0,v) + A_1(0,v)f_t(0) + A_2(0,v)f_{t-1}(0) + B(0,v)\Psi_{t-1}^c(0,v).$$

$$\psi_t^f(u) = \omega^f(u) + \alpha_1^f(u)f_t + \alpha_2^f(u)f_{t-1} + \beta^f(u)\psi_{t-1}^f(u),$$

where

$$\begin{aligned}
\psi^f(u) &= \ln\left[E_t\left(\exp(uf_{t+1})\right)\right] \\
\beta^f(u) &= \mu f(u) \\
\alpha_1^f(u) &= u\mu + g(u) - 1 \\
\alpha_2^f(u) &= \mu((1-\mu u)f(u) - g(u)) \\
\omega^f(u) &= (1-\beta^f(u))\left(u\rho + \frac{\rho}{1-\mu} + h(u)\right) - \rho g(u),
\end{aligned}$$

with $\mu = e^{\lambda_1}$, $f(u) = e^{a(ue^{\lambda_0})}$ a moment generating function of a positive infinitely divisible random, $g(u) = e^{b(ue^{\lambda_0})}$ a moment generating function of a positive random variable and $h(u) = \omega(ue^{\lambda_0})$ a cumulant generating function of a positive random variable.

Hence we get the result stated in proposition (1.2.2).

Proposition 1.6.7. *Generally, for any given s, $(f_t(0), f_t(s))$ is a generalized affine of order* $(2,1)$

We can restrict $f_t(0)$ to be positive by just imposing $c(0)$ to be positive and considering positive initial value $f_0(0)$. On the other hand $f_t(0)$ can take any sign if any restriction is made on $c(0)$ and $f_0(0)$. All these assertions are consequences of the recursive definition of $f_{t+1}(0)$

$$f_{t+1}(0) = c(0) + \exp(\lambda_0)X_{t+1} + \exp(\lambda_1)f_t(0).$$

Since X_{t+1} is a positive random variable, if $f_t(0) \geq 0$ and $c(0) \geq 0$, then $f_{t+1}(0) \geq 0$. $c(0)$ is an undetermined parameter with undetermined sign. This implies that if the sign of $c(0)$ and $f_t(0)$ are undetermined then $f_{t+1}(0)$'s sign is also undetermined.

Generalized affine of order $(1,1)$ (for $f_t(0)$) can be obtained by restricting functions a and b to satisfy $A_2(0,v)=0$. Solving $A_2(0,v)=0$ implies

$$b(v)=a(v)+\ln\left(1-ve^{\lambda_1-\lambda_0}\right),$$

which in fact is equivalent to imposing the following restriction to the moment generating function f.

$$f(u)=\frac{g(u)}{1-\mu u}$$

We then get the result of proposition (1.2.1).

Proof of Proposition 1.6.1: The proof is quite easy, in fact it is done by realizing that if G is the distribution function corresponding to characteristic function g (or moment generating function), then $F=e^{-p}\sum_0^\infty \frac{p^n}{n!}G^{n\star}$ is the distribution function corresponding to characteristic function (or moment generating function $\exp(p(g(u)-1))$). In this expression $G^{n\star}$ means the convolution of n identical distribution function G. The simulation of random variable corresponding to distribution function F is also easy to deal with. Let consider a sequence of *iid* random variable $(Y_i)_{i=1,2...}$, and a discrete random variable N which is independent to $(Y_i)_{i=1,2...}$ and which follows a Poisson distribution with parameter p. The following random variable X has F as distribution function:

$$Z=\sum_{n=0}^{N}Y_n$$

where Y_0 is a constant.

Proof of Proposition 1.6.2: The result is the consequence of the fact that $p(g(u)-1)$ is the cumulant function of positive random variable when $g(u)$ is the moment generating function of a positive random variable. This result is deduced from the previous Proposition, indeed since $p(g(u)-1)$ is the cumulant function of $Z=\sum_{n=0}^{N}Y_n$, and $g(u)$ the moment generating function of Y_n. $Y_n\geq 0\Rightarrow Z\geq 0$

Proof of Proposition 1.6.3: In fact, for any given choice of a cumulant function of positive random variable $\omega(u)$, choose $c(u)$ as follows

$$c(u) = (1-\exp(a(u)+\lambda_1))\left[\omega(u)+\frac{\delta}{1-\exp(\lambda_1)}\right],$$

for any real δ. Since $\omega(u)$ and $a(u)$ are a cumulant functions, thus $\omega(0)=a(0)=0$, which implies that

$$c(0)=\delta.$$

Proof of Proposition 1.6.4: Indeed

$$\begin{aligned}
& c(u)+\exp(\lambda_0+b(u))X_{t+1}+\exp(a(u)+\lambda_1)f_t(u) \\
= \; & c(u)+\exp(\lambda_0+b(u))X_{t+1}+\frac{\exp(a(u)+\lambda_1)c(u)}{1-\exp(a(u)+\lambda_1)}+\exp(\lambda_0+b(u))\sum_{i=0}^{t}\exp(a(u)+\lambda_1)^{i+1}X_{t-i} \\
= \; & \frac{c(u)}{1-\exp(a(u)+\lambda_1)}+\exp(\lambda_0+b(u))\sum_{i=0}^{t+1}\exp(a(u)+\lambda_1)^{i}X_{t+1-i} \\
= \; & f_{t+1}(u).
\end{aligned}$$

Proof of Proposition 1.6.5:

$$f_{t+1}(u)=c(u)+\exp(\lambda_0+b(u))X_{t+1}+\exp(a(u)+\lambda_1)f_t(u),$$

and

$$\Psi_t(u)=f_t(u)-f_t(0),$$

imply that

$$\begin{aligned}
\Psi_t(u)+f_t(0) \;=\; & c(u)+e^{b(u)}\left[f_t(0)-c(0)-e^{\lambda_1}f_{t-1}(0)\right] \\
& +e^{\lambda_1+a(u)}\left[\Psi_{t-1}(u)+f_{t-1}(0)\right].
\end{aligned}$$

Proof of Proposition 1.6.7: Indeed

$$
\begin{aligned}
E_t\left[\exp\left(uf_{t+1}(0)+vf_{t+1}(s)\right)\right] &= E_t\left[\exp\left(\begin{array}{c} u\left(c(0)+e^{\lambda_0}X_{t+1}+e^{\lambda_1}f_t(0)\right) \\ +v\left(c(s)+e^{\lambda_0+b(s)}X_{t+1}+e^{\lambda_1+a(s)}f_t(s)\right)\end{array}\right)\right] \\
&= \exp\left(\begin{array}{c} uc(0)+vc(s)+ue^{\lambda_1}f_t(0) \\ +ve^{\lambda_1+a(s)}f_t(s)+\Psi_t(ue^{\lambda_0}+ve^{\lambda_0+b(s)})\end{array}\right) \\
&\equiv \exp\left(\Psi_t^{c,s}(u,v)\right).
\end{aligned}
$$

$$
\begin{aligned}
\Psi_t^{c,s}(u,v) &= uc(0)+vc(s)+ue^{\lambda_1}f_t(0)+ve^{\lambda_1+a(s)}f_t(s)+\Psi_t(ue^{\lambda_0}+ve^{\lambda_0+b(s)}) \\
&= uc(0)+vc(s)+ue^{\lambda_1}f_t(0)+ve^{\lambda_1+a(s)}f_t(s)+\omega_0\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right) \\
&\quad +\alpha_1\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)f_t(0)+\alpha_2\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)f_{t-1}(0) \\
&\quad +\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)\Psi_{t-1}(ue^{\lambda_0}+ve^{\lambda_0+b(s)}) \\
&= uc(0)+vc(s)+ue^{\lambda_1}f_t(0)+ve^{\lambda_1+a(s)}f_t(s)+\omega_0\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right) \\
&\quad +\alpha_1\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)f_t(0)+\alpha_2\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)f_{t-1}(0) \\
&\quad +\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)\left[\Psi_{t-1}^{c,s}(u,v)-uc(0)-vc(s)-ue^{\lambda_1}f_{t-1}(0)-ve^{\lambda_1+a(s)}f_{t-1}(s)\right].
\end{aligned}
$$

Hence

$$
\Psi_t^{c,s}(u,v)=W^s(u,v)+A_1^s(u,v)'f_t(0,s)+A_2^s(u,v)'f_{t-1}(0,s)+B^s(u,v)\Psi_{t-1}^{c,s}(u,v),
$$

where

$$
f_t(0,s)=\begin{pmatrix} f_t(0) \\ f_t(s)\end{pmatrix}.
$$

$$
W^s(u,v)=(uc(0)+vc(s))\left(1-\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)\right)+\omega_0\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right).
$$

$$
\begin{aligned}
A_1^s(u,v) &= \begin{pmatrix} ue^{\lambda_1}+\alpha_1\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right) \\ ve^{\lambda_1+a(s)}\end{pmatrix} \\
A_2^s(u,v) &= \begin{pmatrix} \alpha_2\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)-ue^{\lambda_1}\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right) \\ -ve^{\lambda_1+a(s)}\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right)\end{pmatrix}.
\end{aligned}
$$

$$
B^s(u,v)=\beta\left(ue^{\lambda_0}+ve^{\lambda_0+b(s)}\right).
$$

Appendix B

This appendix provides the proofs of Section 2 and Section 3. **Proof of Proposition 1.2.4:**

From (1.8) we have

$$\psi_t(u) = \bar{\beta}_j(u)\,\psi_{t-j}(\theta_j(u)) + \sum_{k=1}^{j} \bar{\beta}_{k-1}(u)\left[\omega(\theta_{k-1}(u)) + \alpha(\theta_{k-1}(u))x_{t-k+1}\right].$$

$$\begin{aligned}
\exp\left(V_{t,h}(\bar{u}_h)\right) &= E_t\left[\exp\left(\sum_{i=1}^{h} u_i x_{t+i}\right)\right] \\
&= E_t\left[\exp\left(\sum_{i=1}^{h-1} u_i x_{t+i} + \psi_{t+h-1}(u_h)\right)\right] \\
&= E_t\left[\begin{array}{c} \exp\left(\sum_{i=1}^{h-1} u_i x_{t+i}\right)\times \\ \times\exp\left(\begin{array}{c} \bar{\beta}_h(u_h)\,\psi_{t-1}(\theta_h(u_h)) \\ +\sum_{k=1}^{h} \bar{\beta}_{k-1}(u_h)\left[\omega(\theta_{k-1}(u_h)) + \alpha(\theta_{k-1}(u_h))x_{t+h-k}\right] \end{array}\right)\end{array}\right] \\
&= \exp\left(\begin{array}{c} \bar{\beta}_h(u_h)\,\psi_{t-1}(\theta_h(u_h)) + \bar{\beta}_{h-1}(u_h)\,\alpha(\theta_{h-1}(u_h))x_t \\ +\sum_{k=1}^{h} \bar{\beta}_{k-1}(u_h)\,\omega(\theta_{k-1}(u_h)) \end{array}\right)\times \\
&\quad \times E_t\left[\exp\left(\sum_{i=1}^{h-1} u_i^{(h-1)} x_{t+i}\right)\right],
\end{aligned}$$

thus

$$\begin{aligned}
V_{t,h}(\bar{u}_h) &= \bar{\beta}_h\left(u_h^{(h)}\right)\psi_{t-1}\left(\omega(\theta_h\left(u_h^{(h)}\right)\right) + \bar{\beta}_{h-1}\left(u_h^{(h)}\right)\alpha\left(\theta_{h-1}\left(u_h^{(h)}\right)\right)x_t \\
&\quad + \sum_{k=1}^{h} \bar{\beta}_{k-1}\left(u_h^{(h)}\right)\omega\left(\theta_{k-1}\left(u_h^{(h)}\right)\right) + V_{t,h-1}(\bar{u}_{h-1}),
\end{aligned}$$

where

$$\begin{aligned}
\bar{u}_{h-1} &= \left(u_1^{(h-1)}, u_2^{(h-1)}, \ldots, u_{h-1}^{(h-1)}\right)^{\top} \\
u_j^{(h-1)} &= u_j^{(h)} + \bar{\beta}_{h-1-j}\left(u_h^{(h)}\right)\alpha\left(\theta_{h-1-j}\left(u_h^{(h)}\right)\right) \\
u_j^{(h)} &= u_j.
\end{aligned}$$

Let denote

$$d_k = u_k^{(k)},$$

we have $d_h = u_h$ and for $k \leq h-1$

$$\begin{aligned} d_k &= u_k^{(k)} = u_k^{(k+1)} + \bar{\beta}_0\,(d_{k+1})\,\alpha\,(\theta_0\,(d_{k+1})) \\ &= u_k^{(k+2)} + \bar{\beta}_1\,(d_{k+2})\,\alpha\,(\theta_1\,(d_{k+2})) + \bar{\beta}_0\,(d_{k+1})\,\alpha\,(\theta_0\,(d_{k+1})) \\ &= u_k^{(h)} + \sum_{j=1}^{h-k} \bar{\beta}_{j-1}\left(d_{k+j}\right)\alpha\left(\theta_{j-1}\left(d_{k+j}\right)\right) \\ &= u_k + \sum_{j=1}^{h-k} \bar{\beta}_{j-1}\left(d_{k+j}\right)\alpha\left(\theta_{j-1}\left(d_{k+j}\right)\right). \end{aligned}$$

Thus

$$\begin{aligned} V_{t,h}(\bar{u}_h) &= \bar{\beta}_h\,(d_h)\,\psi_{t-1}\,(\theta_h\,(d_h)) + \bar{\beta}_{h-1}\,(d_h)\,\alpha\,(\theta_{h-1}\,(d_h))x_t + \\ &\quad \sum_{k=1}^{h} \bar{\beta}_{k-1}\,(d_h)\,\omega\,(\theta_{k-1}\,(d_h)) + V_{t,h-1}(\bar{u}_{h-1}) \\ &= \bar{\beta}_h\,(d_h)\,\psi_{t-1}\,(\theta_h\,(d_h)) + \bar{\beta}_{h-1}\,(d_{h-1})\,\psi_{t-1}\,(\omega(\theta_{h-1}\,(d_{h-1})) \\ &\quad + \left[\bar{\beta}_{h-1}\,(d_h)\,\alpha\,(\theta_{h-1}\,(d_h)) + \bar{\beta}_{h-2}\,(d_{h-1})\,\alpha\,(\theta_{h-2}\,(d_{h-1}))\right]x_t \\ &\quad + \sum_{k=1}^{h} \bar{\beta}_{k-1}\,(d_h)\,\omega\,(\theta_{k-1}\,(d_h)) \\ &\quad + \sum_{k=1}^{h-1} \bar{\beta}_{k-1}\,(d_{h-1})\,\omega\,(\theta_{k-1}\,(d_{h-1})) + V_{t,h-2}(\bar{u}_{h-2}) \\ &= a_h\,(\bar{u}_h)x_t + b_h\,(\bar{u}_h) + \sum_{k=1}^{h} \bar{\beta}_k\,(d_k)\,\psi_{t-1}\,(\theta_k\,(d_k)), \end{aligned}$$

where

$$\begin{aligned} a_h\,(\bar{u}_h) &= \sum_{k=1}^{h} \bar{\beta}_{k-1}\,(d_k)\,\alpha\,(\theta_{k-1}\,(d_k)) \\ b_h\,(\bar{u}_h) &= \sum_{k=1}^{h}\left[\sum_{j=1}^{k} \bar{\beta}_{j-1}\,(d_k)\,\omega\,(\theta_{j-1}\,(d_k))\right]. \end{aligned}$$

Proof of Proposition 1.3.4. The yield to maturity $n\left(y_t^{(n)}\right)$ is related to the short term rate as following *for* $n=1$ we have $y_t^{(1)}=r_t$, for n≥ 2

$$
\begin{aligned}
y_t^{(n)} &= -\frac{1}{n}\ln\left\{E_t^Q\left[\exp\left(-\sum_{k=0}^{n-1}y_{t+k}^{(1)}\right)\right]\right\} \\
&= -\frac{1}{n}\ln\left\{E_t^P\left[\left[\prod_{k=0}^{n-2}\frac{M_{t+k,t+k+1}}{E_{t+k}^P\left[M_{t+k,t+k+1}\right]}\right]\exp\left(-\sum_{k=0}^{n-1}y_{t+k}^{(1)}\right)\right]\right\} \\
&= -\frac{1}{n}\ln\left\{E_t^P\left[\exp\left(-\sum_{k=0}^{n-1}y_{t+k}^{(1)}+\sum_{k=0}^{n-2}\left(\gamma y_{t+k+1}^{(1)}-\psi_{t+k}^P(\gamma)\right)\right)\right]\right\} \\
&= -\frac{1}{n}\ln\left\{E_t^P\left[\exp\left(-\sum_{k=0}^{n-1}y_{t+k}^{(1)}+\sum_{k=1}^{n-1}\left(\gamma y_{t+k}^{(1)}-\psi_{t+k-1}^P(\gamma)\right)\right)\right]\right\} \\
&= \frac{y_t^{(1)}+\psi_t^P(\gamma)}{n}-\frac{1}{n}\ln\left\{E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\psi_{t+k}^P(\gamma)\right)\right]\right\}.
\end{aligned}
$$

$$
\begin{aligned}
& E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\psi_{t+k}^P(\gamma)\right)\right] \\
=\; & E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\left(\begin{array}{c}\bar{\beta}_k^P(\gamma)\,\psi_t^P\left(\theta_k^P(\gamma)\right)+ \\ \sum_{i=1}^{k}\bar{\beta}_{i-1}^P(\gamma)\left[\omega^P(\theta_{i-1}^P(\gamma))+\alpha^P(\theta_{i-1}^P(\gamma))y_{t+k-i+1}^{(1)}\right]\end{array}\right)\right)\right] \\
=\; & \exp\left(-\sum_{k=1}^{n-2}\left[\bar{\beta}_k^P(\gamma)\,\psi_t^P\left(\theta_k^P(\gamma)\right)+\sum_{i=1}^{k}\bar{\beta}_{i-1}^P(\gamma)\,\omega^P(\theta_{i-1}^P(\gamma))\right]\right)\times \\
& E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\left(\sum_{i=1}^{k}\bar{\beta}_{i-1}^P(\gamma)\,\alpha^P(\theta_{i-1}^P(\gamma))y_{t+k-i+1}^{(1)}\right)\right)\right].
\end{aligned}
$$

$$
\begin{aligned}
& E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\left(\sum_{i=1}^{k}\bar{\beta}_{i-1}^P(\gamma)\,\alpha^P(\theta_{i-1}^P(\gamma))y_{t+k-i+1}^{(1)}\right)\right)\right] \\
=\; & E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{k=1}^{n-2}\left(\sum_{j=1}^{k}\bar{\beta}_{k-j}^P(\gamma)\,\alpha^P(\theta_{k-j}^P(\gamma))y_{t+j}^{(1)}\right)\right)\right] \\
=\; & E_t^P\left[\exp\left(\sum_{k=1}^{n-1}(\gamma-1)y_{t+k}^{(1)}-\sum_{j=1}^{n-2}\left(\sum_{k=j}^{n-2}\bar{\beta}_{k-j}^P(\gamma)\,\alpha^P(\theta_{k-j}^P(\gamma))\right)y_{t+j}^{(1)}\right)\right] \\
=\; & E_t^P\left[\exp\left(\sum_{k=1}^{n-1}u_k y_{t+k}^{(1)}\right)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
u_{n-1} &= \gamma - 1 \\
u_j &= \gamma - 1 - \sum_{k=j}^{n-2} \bar{\beta}^P_{k-j}(\gamma)\,\alpha^P(\theta^P_{k-j}(\gamma)) \quad for \quad j < n-1.
\end{aligned}
$$

$$
\begin{aligned}
y_t^{(n)} &= \frac{y_t^{(1)} + \psi_t^P(\gamma) + \sum_{k=1}^{n-2}\left[\bar{\beta}^P_k(\gamma)\,\psi_t^P\left(\theta^P_k(\gamma)\right) + \sum_{i=1}^{k}\bar{\beta}^P_{i-1}(\gamma)\,\omega^P(\theta^P_{i-1}(\gamma))\right]}{n} \\
&\quad - \frac{a_{n-1}(\bar{u}_{n-1})\,y_t^{(1)} + b_{n-1}(\bar{u}_{n-1}) + \sum_{k=1}^{n-1}\bar{\beta}_k(d_k)\,\psi^P_{t-1}\left(\theta^P_k(d_k)\right)}{n} \\
&= \frac{y_t^{(1)} + \sum_{k=1}^{n-1}\left(\sum_{i=1}^{k}\bar{\beta}^P_{i-1}(\gamma)\,\omega^P(\theta^P_{i-1}(\gamma))\right) + \sum_{k=1}^{n-1}\bar{\beta}_k(\gamma)\,\psi^P_{t-1}\left(\theta^P_k(\gamma)\right)}{n} \\
&\quad + \frac{y_t^{(1)}\sum_{k=1}^{n-1}\bar{\beta}_{k-1}(\gamma)\,\alpha^P\left(\theta^P_{k-1}(\gamma)\right)}{n} \\
&\quad - \frac{a_{n-1}(\bar{u}_{n-1})\,y_t^{(1)} + b_{n-1}(\bar{u}_{n-1}) + \sum_{k=1}^{n-1}\bar{\beta}_k(d_k)\,\psi^P_{t-1}\left(\theta^P_k(d_k)\right)}{n} \\
&= d_n + c_{n,0}y_t^{(1)} + \sum_{k=1}^{n-1}\left[\frac{\bar{\beta}_k(\gamma)}{n}\psi^P_{t-1}\left(\theta^P_k(\gamma)\right) - \frac{\bar{\beta}_k(d_k)}{n}\psi^P_{t-1}\left(\theta^P_k(d_k)\right)\right] \\
&= d_n + c_{n,0}y_t^{(1)} + z_{t-1}^{(n)},
\end{aligned}
$$

where

$$
\begin{aligned}
d_n &= \frac{1}{n}\sum_{k=1}^{n-1}\left(\sum_{i=1}^{k}\bar{\beta}^P_{i-1}(\gamma)\,\omega^P(\theta^P_{i-1}(\gamma)) - \bar{\beta}_{i-1}(d_k)\,\omega(\theta_{i-1}(d_k))\right) \\
c_{n,0} &= \frac{1}{n}\left[1 + \sum_{k=1}^{n-1}\left(\bar{\beta}_{k-1}(\gamma)\,\alpha^P\left(\theta^P_{k-1}(\gamma)\right) - \bar{\beta}_{k-1}(d_k)\,\alpha\left(\theta_{k-1}(d_k)\right)\right)\right].
\end{aligned}
$$

$d_{n-1} = \gamma - 1$ and for $k \leq n-2$

$$
d_k = \gamma - 1 + \sum_{j=1}^{n-1-k}\left[\bar{\beta}_{j-1}(d_{k+j})\,\alpha\left(\theta_{j-1}(d_{k+j})\right) - \bar{\beta}_{j-1}(\gamma)\,\alpha\left(\theta_{j-1}(\gamma)\right)\right].
$$

Proof of Proposition 1.3.6. One has

$$\begin{aligned} E_t[M_{t+1}] &= \exp(-r) \\ E_t[M_{t+1}\exp(r_{t+1})] &= 1, \end{aligned}$$

which leads to

$$\begin{aligned} \theta_t + \Psi_t(\gamma,\lambda) &= -r \\ \theta_t + \Psi_t(1+\gamma,\lambda) &= 0. \end{aligned}$$

Hence,

$$\begin{aligned} \theta_t &= -r - \Psi_t(\gamma,\lambda) \\ \Psi_t(1+\gamma,\lambda) - \Psi_t(\gamma,\lambda) &= r. \end{aligned}$$

By using the following expression of the model:

$$\Psi_{t+1}(u,v) = \frac{\omega(u,v)}{1-\beta(u,v)} + \alpha(u,v)\sum_{i=0}^{\infty}\beta(u,v)^i h_{t-i+1},$$

one gets,

$$\frac{\omega(1+\gamma,\lambda)}{1-\beta(1+\gamma,\lambda)} - \frac{\omega(\gamma,\lambda)}{1-\beta(\gamma,\lambda)} + \sum_{i=0}^{\infty}\left[\beta(1+\gamma,\lambda)^i\alpha(1+\gamma,\lambda) - \beta(\gamma,\lambda)^i\alpha(\gamma,\lambda)\right]h_{t-i} = r,$$

which implies

$$\begin{aligned} \frac{\omega(1+\gamma,\lambda)}{1-\beta(1+\gamma,\lambda)} - \frac{\omega(\gamma,\lambda)}{1-\beta(\gamma,\lambda)} &= r \\ \beta(1+\gamma,\lambda)^i\alpha(1+\gamma,\lambda) - \beta(\gamma,\lambda)^i\alpha(\gamma,\lambda) &= 0,\ \forall i \geq 0. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\omega(1+\gamma,\lambda)}{1-\beta(1+\gamma,\lambda)} - \frac{\omega(\gamma,\lambda)}{1-\beta(\gamma,\lambda)} &= r \\ \beta(1+\gamma,\lambda) &= \beta(\gamma,\lambda) \\ \alpha(1+\gamma,\lambda) &= \alpha(\gamma,\lambda). \end{aligned}$$

Proof of Proposition 1.3.7.

$$E_t^Q\left[\exp\left(ur_{t+1}+v\Psi_{t+1}(\gamma)\right)\right]=E_t^Q\left[\exp\left(ur_{t+1}+v\left(\omega(\gamma)+\alpha(\gamma)r_{t+1}+\beta(\gamma)\Psi_t(\gamma)\right)\right)\right],$$

thus

$$\begin{aligned}\Psi_t^*(u,v) &= v\omega(\gamma)+v\beta(\gamma)\Psi_t(\gamma)+\Psi_t^Q(u^*)\\ &= v\omega(\gamma)+v\beta(\gamma)\Psi_t(\gamma)+\Psi_t(u^*+\gamma)-\Psi_t(\gamma),\end{aligned}$$

where

$$u^*=u+v\alpha(\gamma),$$

this implied that

$$\begin{aligned}\Psi_{t+1}^*(u,v) &= v\omega(\gamma)+(v\beta(\gamma)-1)\Psi_{t+1}(\gamma)+\Psi_{t+1}(u^*+\gamma)\\ &= v\omega(\gamma)+(v\beta(\gamma)-1)\Psi_{t+1}(\gamma)+\omega(u^*+\gamma)\\ &\quad +\alpha(u^*+\gamma)r_{t+1}+\beta(u^*+\gamma)\Psi_t(u^*+\gamma)\\ &= v\omega(\gamma)+(v\beta(\gamma)-1)\Psi_{t+1}(\gamma)+\omega(u^*+\gamma)+\\ &\quad +\alpha(u^*+\gamma)r_{t+1}+\beta(u^*+\gamma)\left[\Psi_t^*(u,v)-v\omega(\gamma)-(v\beta(\gamma)-1)\Psi_t(\gamma)\right].\end{aligned}$$

Table 1.1: **Monte carlo exercise for the GMM.**

Mean, Median and RMSE of parameters estimates across N samples (of T observations). Parameters used in the simulation of GARG model are μ =2.78E-05, ν =0.139, ρ =0.112, β =0.922 and θ =0.906

Par	Mean	Median	RMSE	Mean	Median	RMSE	Mean	Median	RMSE
	N=1000, T=250			N=1000, T=500			N=1000, T=1000		
μ	2.58E-05	3.66E-05	1.68E-05	2.61E-05	3.66E-05	1.66E-05	2.64E-05	3.66E-05	1.64E-05
ν	0.077	0.109	0.079	0.077	0.109	0.078	0.078	0.109	0.077
ρ	0.104	0.115	0.018	0.104	0.115	0.018	0.105	0.115	0.018
β	0.695	0.946	0.437	0.702	0.946	0.430	0.710	0.946	0.423
θ	1.318	0.879	0.780	1.307	0.879	0.770	1.293	0.879	0.755
	N=2000, T=250			N=2000, T=500			N=2000, T=1000		
μ	2.80E-05	3.67E-05	1.55E-05	2.90E-05	3.66E-05	1.49E-05	2.97E-05	3.66E-05	1.44E-05
ν	0.083	0.109	0.072	0.086	0.109	0.069	0.088	0.109	0.066
ρ	0.106	0.115	0.017	0.107	0.115	0.016	0.108	0.115	0.015
β	0.744	0.946	0.395	0.766	0.946	0.372	0.784	0.946	0.353
θ	1.248	0.879	0.733	1.208	0.878	0.693	1.175	0.878	0.657
	N=4000, T=250			N=4000, T=500			N=4000, T=1000		
μ	3.20E-05	3.66E-05	1.28E-05	3.24E-05	3.66E-05	1.25E-05	3.27E-05	3.66E-05	1.24E-05
ν	0.095	0.109	0.056	0.096	0.109	0.054	0.097	0.109	0.053
ρ	0.110	0.115	0.012	0.111	0.115	0.012	0.111	0.115	0.011
β	0.837	0.946	0.291	0.847	0.946	0.278	0.852	0.946	0.270
θ	1.084	0.879	0.557	1.066	0.878	0.533	1.056	0.879	0.517

Table 1.2: **Monte carlo exercise for the QMLE.**

Mean, Median and RMSE of parameters estimates across N samples (of T observations). Parameters used in the simulation of GARG model are μ =2.78E-05, ν =0.139, ρ =0.112, β =0.922 and θ =0.906

Par	Mean	Median	RMSE	Mean	Median	RMSE	Mean	Median	RMSE
	N=1000, T=250			N=1000, T=500			N=1000, T=1000		
μ	6.00E-05	5.13E-05	5.37E-05	5.97E-05	5.20E-05	5.19E-05	6.13E-05	5.31E-05	5.43E-05
ν	0.142	0.051	0.453	0.154	0.051	0.415	0.148	0.050	0.427
ρ	0.110	0.109	0.029	0.108	0.109	0.027	0.108	0.106	0.028
β	1.028	0.959	0.356	1.048	0.960	0.415	1.066	0.964	0.488
θ	0.879	0.909	0.189	0.875	0.908	0.204	0.870	0.909	0.206
	N=2000, T=250			N=2000, T=500			N=2000, T=1000		
μ	5.45E-05	5.02E-05	4.20E-05	5.46E-05	4.99E-05	4.12E-05	5.56E-05	5.15E-05	4.23E-05
ν	0.216	0.057	0.963	0.220	0.055	0.915	0.179	0.052	0.711
ρ	0.109	0.109	0.023	0.107	0.107	0.023	0.107	0.107	0.024
β	0.998	0.937	0.270	0.993	0.941	0.243	1.001	0.947	0.255
θ	0.891	0.918	0.164	0.893	0.918	0.158	0.888	0.915	0.161
	N=4000, T=250			N=4000, T=500			N=4000, T=1000		
μ	5.30E-05	5.07E-05	3.82E-05	5.26E-05	5.17E-05	3.57E-05	5.30E-05	5.31E-05	3.57E-05
ν	0.244	0.048	0.942	0.373	0.049	3.038	0.275	0.050	2.347
ρ	0.108	0.109	0.022	0.108	0.107	0.021	0.107	0.106	0.021
β	0.969	0.940	0.168	0.963	0.941	0.146	0.966	0.964	0.147
θ	0.901	0.914	0.126	0.905	0.913	0.121	0.903	0.909	0.122

Table 1.3: **Out-of-sample 1-month-ahead forecasting results.**

We present the results of out-of-sample 1-month-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+1 as $y_{t+1}(\tau) - \hat{y}_{t+1}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their first and 12th sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$	Mean	Std. Dev.	RMSE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.033	0.176	0.179	0.220	0.053	-0.045	0.170	0.176	0.247	0.017
1 year	0.021	0.240	0.241	0.340	-0.153	0.023	0.235	0.236	0.425	-0.213
3 years	0.007	0.279	0.279	0.341	-0.133	-0.056	0.273	0.279	0.332	-0.117
5 years	-0.003	0.276	0.276	0.275	-0.131	-0.091	0.277	0.292	0.333	-0.116
10 years	-0.011	0.254	0.254	0.215	-0.145	-0.062	0.252	0.260	0.259	-0.115
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.043	0.196	0.200	0.126	0.320	-0.021	0.231	0.230	0.014	0.299
1 year	-0.011	0.235	0.234	0.380	-0.139	0.004	0.230	0.229	0.055	-0.006
3 years	0.018	0.269	0.268	0.358	-0.153	0.023	0.250	0.249	0.143	-0.095
5 years	-0.014	0.281	0.280	0.375	-0.154	-0.007	0.267	0.265	0.266	-0.100
10 years	-0.163	0.274	0.318	0.386	-0.094	-0.152	0.264	0.304	0.348	-0.066
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.078	0.196	0.210	0.120	0.295	-0.082	0.202	0.217	0.143	0.316
1 year	0.088	0.227	0.243	0.328	-0.144	-0.057	0.239	0.245	0.411	-0.149
3 years	0.134	0.261	0.292	0.294	-0.099	0.024	0.265	0.264	0.332	-0.139
5 years	0.052	0.273	0.276	0.315	-0.107	-0.031	0.278	0.278	0.351	-0.136
10 years	-0.050	0.269	0.272	0.354	-0.100	-0.107	0.277	0.296	0.403	-0.092

Table 1.4: **Out-of-sample 6-months-ahead forecasting results.**

We present the results of out-of-sample 6-months-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+6 as $y_{t+6}(\tau) - \hat{y}_{t+6/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their sixth and eighteenth sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$	Mean	Std. Dev.	RMSE	$\hat{\rho}(6)$	$\hat{\rho}(18)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.220	0.564	0.605	0.381	-0.214	0.083	0.510	0.517	0.301	-0.190
1 year	0.181	0.758	0.779	0.139	-0.150	0.131	0.656	0.669	0.168	-0.174
3 years	0.099	0.873	0.879	0.018	-0.211	-0.052	0.748	0.750	0.049	-0.189
5 years	0.048	0.860	0.861	0.008	-0.249	-0.173	0.758	0.777	0.069	-0.273
10 years	-0.020	0.758	0.758	0.019	-0.271	-0.251	0.676	0.721	0.058	-0.288
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.074	0.494	0.496	0.193	-0.109	0.001	0.531	0.528	0.312	-0.163
1 year	-0.040	0.696	0.693	0.085	-0.142	0.015	0.665	0.661	0.208	-0.168
3 years	-0.089	0.777	0.777	-0.014	-0.197	-0.028	0.725	0.721	0.047	-0.200
5 years	-0.180	0.789	0.805	-0.006	-0.220	-0.103	0.735	0.738	0.049	-0.223
10 years	-0.388	0.735	0.827	-0.019	-0.186	-0.297	0.681	0.739	0.018	-0.198
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.590	0.464	0.750	0.143	0.116	-0.184	0.508	0.537	0.156	-0.042
1 year	0.562	0.671	0.872	0.028	-0.025	-0.151	0.712	0.723	0.067	-0.113
3 years	0.396	0.769	0.861	-0.045	-0.135	-0.116	0.787	0.791	-0.020	-0.191
5 years	0.186	0.789	0.806	-0.023	-0.178	-0.218	0.798	0.822	-0.007	-0.219
10 years	-0.040	0.727	0.723	-0.060	-0.187	-0.341	0.739	0.810	-0.019	-0.187

Table 1.5: **Out-of-sample 12-months-ahead forecasting results.**

We present the results of out-of-sample 12-months-ahead forecasting using eight models, as described in detail in the text. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+12 as $y_{t+12}(\tau) - \hat{y}_{t+12/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their their 12th and 24th sample autocorrelation coefficients.

Maturity	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$	Mean	Std. Dev.	RMSE	$\hat{\rho}(12)$	$\hat{\rho}(24)$
	Random walk					Nelson–Siegel with AR(1) factor dynamics				
3 months	0.416	0.930	1.019	-0.118	-0.109	0.150	0.724	0.739	-0.288	0.001
1 year	0.388	1.132	1.197	-0.268	-0.019	0.173	0.823	0.841	-0.332	-0.004
3 years	0.236	1.214	1.237	-0.419	0.060	-0.123	0.910	0.918	-0.408	0.015
5 years	0.130	1.184	1.191	-0.481	0.072	-0.337	0.918	0.978	-0.412	0.003
10 years	-0.033	1.051	1.052	-0.508	0.069	-0.531	0.825	0.981	-0.433	-0.003
	VAR(1) on 1, 24, 96 months yields					VARMA(1,1) on 1, 24, 96 months yields				
3 months	-0.152	0.792	0.801	-0.214	-0.076	-0.137	0.708	0.716	-0.059	-0.123
1 year	-0.188	0.913	0.926	-0.307	-0.027	-0.170	0.794	0.807	-0.143	-0.076
3 years	-0.325	0.953	1.001	-0.393	0.001	-0.267	0.832	0.868	-0.269	-0.043
5 years	-0.459	0.956	1.055	-0.413	-0.006	-0.371	0.849	0.921	-0.305	-0.049
10 years	-0.710	0.875	1.123	-0.440	-0.006	-0.596	0.775	0.974	-0.357	-0.053
	VAR(1) on 1,24, 96 months yields and inflation					VAR(1) on 1,24, 96 months yields and real activity				
3 months	0.854	0.851	1.202	0.020	-0.045	-0.329	0.852	0.908	-0.245	-0.081
1 year	0.743	1.002	1.242	-0.137	-0.005	-0.359	0.994	1.051	-0.344	-0.030
3 years	0.436	1.044	1.125	-0.286	0.016	-0.390	1.018	1.084	-0.420	-0.008
5 years	0.153	1.050	1.054	-0.323	0.001	-0.528	1.007	1.131	-0.430	-0.014
10 years	-0.143	0.933	0.937	-0.414	-0.001	-0.682	0.901	1.125	-0.455	-0.011

CHAPTER 2

NO-ARBITRAGE VARMA TERM STRUCTURE MODELS WITH MACROECONOMIC VARIABLES

Abstract

This chapter combines VARMA processes with the no-arbitrage restrictions and studies the forecastability of yields and macroeconomic variables. The chapter shows that adding a Moving Average [MA] component to a standard VAR process offers substantial improvements in forecasting future yields, inflation, real activity and future interest rate risk premia where our benchmarks are either a standard VAR model or a dynamic version of the Nelson-Siegel model. An important hindsight from our results is that using VARMA processes breaks the tight link between current value of the state variable and the current conditional expectation of the future realization of the state variable, implicit in VAR models. Moreover, we show that the state variable follows a VARMA process under the risk-neutral probability measure only if the price of risk is linear in the current value of the state variable and the current conditional expectation of the future value of the state variable.

2.1 Introduction

How can we combine tractability and flexibility when pricing financial instruments like bonds and derivatives? Affine models are considered as the ideal set-up to answer this question. To cope with non-Markovian state variable, the dimension of the state variable is often increased by adding several past observations. Non-linearities are handled by introducing unobserved component in the state variable. All these approaches introduce difficulties in the estimation, among which high number of unknown parameters (for multi-lags affine models) and an unobserved state component's filtering. Feunou and Meddahi (2007) introduce a new class of models, the generalized affine model, which is a parsimonious infinite order affine model (just like the GARCH process is an infinite order ARCH process), and which still leads to closed form solutions to term structure problems. Within the class of generalized affine models, ARMA models and more generally VARMA models are the most popular, but surprisingly to our knowledge VARMA models have not yet been studied theoretically and empirically in the term structure of interest rate literature. They are the main focus of this chapter.

Interest rates combine expectations of future short rates, inflation and real activity as well as an adjustment for risk. Therefore, our results are important empirically and economically for two reasons. First, estimating a VARMA specification for the historical dynamics of the state vector delivers better in-sample and out-of-sample forecasts of real activity, short rates and term spread compared to standard VAR models. This implies that a VARMA specification for the short rate better captures the expectation component of interest rates. We use impulse responses and conclude that the results depend crucially on the ability of the MA component to filter the time-varying conditional mean of the state vector from its past and current noisy realizations. Intuitively, innovations to a given state variable are allowed for different correlation structures with the next period realization of the state vector and with its conditional mean.

Second, the more flexible specification of prices of risk allowed by the VARMA structure delivers substantial improvement of in-sample fit and out-of-sample forecasts of interest rates across the term structure. Again, the results suggest that allowing for a different impact of the state, z_t, and its conditional expectation, $E_t[z_{t+1}]$, on the evolution of the risk premia is crucial. Intuitively, the impact of shocks to a state variable, say inflation, on its own or other variables' prices of risk, depends on whether the current conditional expectations of other state variables, e.g. the short rate, is high or low. Together, the improvements in out-of-sample forecasting are a significant contribution to the literature given the strong presumption that more flexible models suffer from over-parametrization and offer inferior forecasts.

Term structure models often treat the determinants of interest rates as latent and use a filtering approach to estimate model parameters from observed yields. This approach offers a parsimonious fit of the data and three factors combined within a simple VAR dynamics are generally thought to capture many stylized facts about interest rates. Intuitively, latent factors can approximate non-linearities and Markov processes of higher order in the dynamics of the true economic variable. However, latent factors introduce many estimation challenges and, more importantly, offer only few hindsights about the underlying economic structure. In contrast, the MA component in a VARMA process can capture some stationary process of possibly infinite order but can still be estimated via standard MLE. We check the relative contribution of using latent variables in a simple setup, increasing the dimension of the state vector in VAR(1)-based model, and using a VARMA specification on yields only, and find that the latter offer the best out-of-sample fit.

The chapter is organized as follows. Section 2 summarizes the data used, a discussion on state vector specification and the dynamic under the historical probability measure is performed, the VARMA(1,1) model is compared to a VAR(1) in term of likelihood. In section 3 we discuss state vector's forecasting. Section 4 presents the pricing kernel and gives closed form expression of yield in function of state vector and its conditional expectation. Estimation of price of risk as well as results are discussed in Section 5. Different out of sample forecasting exercises are implemented in section 6 while section 7 concludes.

2.2 VARMA Model of Yields, Inflation and Real Activity

This section introduces the k-dimensional vector of state variable z_t and discusses the benefits of a VARMA specification under the historical probability measure. We focus on the case where z_t combines yields and macroeconomic variables and estimate its dynamics directly from observed data. The results indicate that the Moving Average component is significant and we show through the impulse function analysis and out-of-sample forecasting how this component improves upon the simpler VAR model.

2.2.1 Model Specification

In principle, a VAR dynamic combined with a sufficiently high number of latent factors, f_t within z_t, can capture a rich set of dynamics, including Markov processes of high order. Nonetheless, the curse of dimensionality raises many estimation challenges and the literature has settled around a small number of factors combined with a flexible specification of the price of risk. This approach potentially captures stylized properties of the term structure of interest rates but does not shed light on the linkages between the latent factors and the underlying macroeconomic structure. For this purpose, the state vector must be observable and a VAR(1) may be too restrictive.

Consider the general reduced-form VARMA(1,1) model

$$z_{t+1} = \mu + \phi z_t + \Sigma\left(\varepsilon_{t+1} - \Theta\varepsilon_t\right), \tag{2.1}$$

where $\varepsilon \sim i.i.d.N(0, I_k)$, I_k is the identity matrix. This VARMA(1,1) is equivalent to the following more familiar form

$$z_{t+1} = \mu + \phi z_t + u_{t+1} - \Theta_1 u_t,$$

indeed, we have $u_{t+1} = \Sigma\varepsilon_{t+1}$ and $\Theta_1 = \Sigma\Theta\Sigma^{-1}$. (2.1) is more convenient for pricing purpose. The conditional expectation of z_{t+1} is

$$m_t = \mu + \phi z_t - \Sigma\Theta\varepsilon_t,$$

we have

$$z_{t+1} = m_t + \Sigma\varepsilon_{t+1},$$

and, if $|\Sigma| \neq 0$ (where $|\cdot|$ denotes the determinant operator), we have that

$$\varepsilon_{t+1} = \Sigma^{-1}\left(z_{t+1} - m_t\right).$$

This implies the following recursive representation for m_{t+1}

$$m_{t+1} \quad = \quad \mu + \left(\phi - \Sigma\Theta\Sigma^{-1}\right) z_{t+1} + \Sigma\Theta\Sigma^{-1} m_t,$$

and, the following VAR(1) representation of the conditional mean

$$m_{t+1} = \mu + \phi m_t + \left(\phi - \Sigma\Theta\Sigma^{-1}\right)\varepsilon_{t+1}.$$

This implies that, contrary to VAR(1) model, the components of a VARMA(1,1) model are univariate representations of a "two-component model". To see this, note that if $\Theta = 0$, and that we have a VAR(1) process, then the evolution of the conditional expectations, m_t, coincides with the evolution of z_t. Otherwise, if $\Theta \neq 0$, the state vector is not a sufficient statistic of the process.

2.2.2 Data

In practice, we will use the following observable state variables. Define $x_t = (x_{1,t}, x_{2,t})^\top$ and $y_t = (y_{1,t}, s_t)^\top$ where $x_{1,t}$ is a measure of inflation, $x_{2,t}$ is a measure of real activity, $y_{1,t}$ is the yield on a zero-coupon bond with one month to maturity, and s_t is the term spread $y_t^{(60)} - y_t^{(1)}$. The state vector is then $z_t = (x_t^\top, y_t^\top)^\top$. Rather than restricting ourselves to specific measures of inflation or output, we use the inflation factor and the real activity factor constructed by Ang and Piazzesi (2003). [1] Each of these factors is the first principal component of a group of variables: an inflation group and a real activity group. The data on zero-coupon yields spans the period from June 1952 to December 2000. We use yields at maturities of 12, 24, 36, 48 and 60 months from the Fama-Bliss CRSP files and the one-month rate is from the CRSP Treasury Bill files. Table 2.1 displays some summary statistics. As expected, the average yield curve is upward sloping, and yield standard deviations decrease with maturity. Also, yields are highly autocorrelated and the autocorrelations increase longer maturity. The yields and macroeconomic factors exhibit mild excess kurtosis and right-skew.

1. The first principal components of each group of macro variables were graciously provided by Monika Piazzesi.

2.2.3 Estimation Results

This observed VARMA model can be estimated using standard MLE techniques but requires further identification restrictions. We use the following notation

$$\phi = (\phi_{ij})$$
$$\Sigma = (\sigma_{ij})$$
$$\Theta = (\theta_{ij}),$$

and assume

$$\begin{aligned}\sigma_{ij} &= 0, \text{ for } j \geq i+1 \\ \theta_{ij} &= 0, \text{ for } j < i+1.\end{aligned}$$

The restrictions on Σ are common in the VAR literature, and are due to the fact that the likelihood depends only on the variance matrix $\Sigma\Sigma^\top$. In consequence, only $\Sigma\Sigma^\top$ is identifiable; besides $\Sigma\Sigma^\top$ is symmetric positive definite if and only if it can be written as the product of a lower triangular matrix and its transpose. For this reason we choose to restrict Σ to a lower triangular matrix. This implies that innovations to inflation are not correlated with innovations in real activity and yields and that innovations in real activity are not correlated with innovations in yields. On the other hand, innovations in yields are correlated with innovations in macro variables. These restrictions reflect the fact that while yields react to current economic conditions, the reverse is not true.

Contrary to Σ, restrictions on Θ are not common in the literature, in fact in principle Θ is fully identifiable. We impose restrictions on Θ in order to reduce the number of parameters. These restrictions on Θ imply that the conditional expectation of macro variables are affected by the evolutions of yields through the moving average component. This captures the fact that the endogenous response of yields, and in particular of monetary policy, influences the future path of macroeconomic quantities. Finally, we set μ_1 and μ_2 equal to zero since the inflation and real activity factors are centered around zero and we impose that $\mu_1 = -\phi_{13}\bar{y}_{1,T} - \phi_{14}\bar{s}_T$ and $\mu_2 = -\phi_{23}\bar{y}_{1,T} - \phi_{24}\bar{s}_T$, where $\bar{y}_{1,T}$

and $\bar{s}_T$ are the sample means of $y_{1,t}$ and s_t, respectively.

Parameter estimates from VAR and VARMA specifications are provided in Table 2.2, which displays estimates of μ and of the autoregressive matrix ϕ. Table 2.3 displays estimates of the standard deviation matrix, Σ, and of the moving average matrix, Θ. Not surprisingly, results from the VAR model indicate that inflation, real activity and the short rate interest rate of are persistent while the spread is slightly less persistent. Estimates of the autoregressive coefficients are 0.9988, 0.9725, 0.9407 and 0.8244, respectively. Looking at interactions between state variables, we see that increases in current real activity lead to higher future inflation while higher current inflation is associated with lower future real activity. Furthermore, an increase in the short rate leads to lower inflation but, surprisingly, this coefficient is not significant. Finally, the short rate is expected to rise and the term structure to flatten whenever current inflation or real activity rises.

An increase in the short rate has a weak effect on expected inflation and may be considered as evidence against this VAR specification. Introducing a Moving Average component increases the likelihood significantly (i.e. the LR-statistic is close to 70, with 5 more significant parameters). Also, the autoregressive matrix is substantially the same but the slope of the term structure is more persistent (0.8817) and yield variables are less responsive to economic conditions through the auto-regressive component. However, Table 2.3 indicates that the short rate has a significant impact on future inflation through the Θ matrix. In other words, an increase in the short rate impacts future inflation only through its conditional expectation, $m_{x_1,t}$. The VARMA model breaks the link between inflation expectations and current inflation. Moreover, it implies a different impact of interest rates on each component. While the short rate has close to no impact on current, noisy, measures of inflation, an increase in the short rate has a lasting impact on inflation expectations. Similarly, while the short rate has close to no impact on real activity through the autoregressive component, an increase in the short rate decreases the conditional expectation of output. This is the evidence that the VARMA process correctly captures the true relationship between these variables.

Note that the results imply that it is the evolution of the short rate that affects real activity and inflation. This contrasts with the common predictive results that document a significant effect from the spread on future output and inflation. This is consistent with Ang and Piazzesi (2003). Interestingly, their results stem from imposing no-arbitrage restrictions but we show that a more flexible specification captures these interactions without the need for these restrictions.

2.2.4 Impulse Responses

Another way to interpret these results is through impulse responses analysis. Figures 2.2 to 2.5 present the response of the state vector to a shock to each of the state variables in turn. Figure 2.2 shows that estimates from the VAR and VARMA models imply similar responses of real activity, the short rate and the term spread to an inflation shock. Similarly, Figure 2.3 shows similar responses of each state variable to a real activity shock.

However, in line with the observations made above, inflation and real activity respond differently to yield factors in the VARMA model than in the VAR model. In the VAR model, current inflation rises following a short rate shock. This reflects the endogenous correlation induced by the monetary policy response. However, the impact on future inflation simply decays toward zero suggesting that policy has no lasting impact on inflation. In contrast, in the VARMA model, the impact is eventually large and negative. Inflation decreases by close to 1.5% below the initial level around 4 years following a tightening in policy. Real activity also exhibit these contrasting pattern to interest rate shocks in each model. Turning to term spread shocks, the VAR model implies that inflation and real activity rise following an increase in the slope of the term structure. However, the impact is more pronounced in the VARMA model in part because of the higher persistence of the spread, but also because a higher spread implies lower conditional expectations of the short rate through the MA component (i.e. $\theta_{y_1,s} = 0.2053$).

2.3 Forecasting

As a last check of the VARMA model to capture the true, conditional, correlation structure of the state vector, we compare its out-of-sample forecasting performance with a Random Walk [RW] and VAR(1) process. Clearly, the observations above suggest that improvements are expected. This is because the h-horizon forecast can be written as

$$m_{t,h} \equiv \left(I_4 - \phi^{h-1}\right)(I_4 - \phi)^{-1}\mu + \phi^{h-1}m_t,$$

which shows that the forecasting performance will differ across models whenever the implied processes for m_t differ.

In practice, we estimate each model using the first 384 observations and forecast each variable h periods ahead up to $h = 12$. We then extend the estimation sample by one, using 385 observations, and repeat the forecasting exercise until we reach the end of the sample. We measure the forecasting performance using Root Mean Squared Errors

$$RMSE\,(l) = \sqrt{\frac{1}{T}\left(E_t\left(l_{t+h}\right) - l_{t+h}\right)^2}.$$

Out-of-sample results are displayed in Table 2.4. We also display in-sample results for comparison. The VARMA model provides large improvements in predicting the term structure at all horizons. In the case of the short rate forecasts RMSEs decrease from 0.54 to 0.51, from 0.76 to 0.68, from 0.98 to 0.88 and from 1.38 to 1.12 at horizons of 1, 3, 6 and 12 months, respectively. Similar results are obtained for the term spread, but the VAR model provides results that are only marginally inferior. Finally, the VAR and VARMA models provide similar improvements in real activity forecasts, and the improvements increase with the forecasting horizon. There, the VARMA model appears to be doing marginally better. However, neither the VAR or the VARMA model improves significantly over the random walk to forecast inflation. This reflects the extreme persistence of the inflation process and the impact of a large noisy component in measures of inflation.

2.3.1 Discussion

The inability of the VAR model to disentangle conditional expectations of inflation and output from their current realization leads to the conclusion that monetary policy has only a weak effect on future economic conditions. On the other hand, impulse responses from the VARMA model indicate that the short rate has a lasting negative impact on inflation and real activity. Conversely, the VARMA model provides much improved forecast of the short rate in response to variation in inflation or real activity. The improved out-of-sample performance shows that a VARMA model does not suffer from over-parametrization.

This has important implications for term structure modeling. The specification of the short interest rate process is a key building block of any term structure model. In the context of macro-finance term structure models, the results above suggest that a simple VAR process is not able to correctly capture the observed response of monetary policy to inflation and real activity and, conversely, the response of future inflation and output to current changes in the short interest rate. This raises an important empirical question which the following sections turns to. In the following we specify, and evaluate, a no-arbitrage model of interest rates. At this point, the missing building block is the specification of the pricing kernel. In particular, in the VARMA context, the kernel will not only depend on the current state vector, z_t, but also on its conditional expectation, m_t.

2.4 Term Structure Model

2.4.1 The Pricing Kernel

This section introduces the pricing kernel and discusses the specification of the price of risk when the state vector follows a VARMA processes. We consider the following conditionally log-normal pricing kernel

$$M_{t+1} = \exp\left(-y_t^{(1)} - \frac{1}{2}\lambda_t^\top \lambda_t - \lambda_t^\top \varepsilon_{t+1}\right), \tag{2.2}$$

where λ_t contains the market prices of risk for each shock. Then, the moment generating function of the state vector z_t under the risk-neutral probability measure is

$$E_t^Q\left[\exp\left(u^\top z_{t+1}\right)\right] = \exp\left(u^\top(m_t - \Sigma\lambda_t) + \frac{1}{2}u^\top\Sigma\Sigma^\top u\right),$$

which shows that λ_t must be linear to obtain an affine process under the risk-neutral measure. Moreover, the vector of prices of risk must be linear in the state vector, z_t, *and* the conditional expectations, m_t, for the dynamics under the risk-neutral probability measure to remain within the family of VARMA process. This is because the state vector and the conditional mean, *together*, are sufficient statistics of VARMA processes. We discuss this in the next section. Therefore, we assume a price of risk vector, λ_t, of the form

$$\lambda_t = \tilde{\lambda}_0 + \tilde{\lambda}_1 z_t + \tilde{\lambda}_2 m_t, \tag{2.3}$$

where $\tilde{\lambda}_0$ is a 4×1 vector while $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are 4×4 matrices.

This framework generalizes the risk premia specification of Ang and Piazzesi (2003) which is obtained by imposing $\tilde{\lambda}_2 = 0$. The price of risk associated with one component of the state variable, say $\lambda_{i,t}$, depends not only on the current state, z_t, but also on its current conditional expectation, m_t. This will have a large impact on the results. An alternative way to interpret $\tilde{\lambda}_2$ follows if we re-write the price of risk, λ_t, as

$$\lambda_t = \tilde{\lambda}_0^* + \tilde{\lambda}_1^* z_t + \tilde{\lambda}_2^* m_{t-1},$$

where

$$\begin{aligned}
\tilde{\lambda}_0^* &= \tilde{\lambda}_0 + \tilde{\lambda}_2\mu \\
\tilde{\lambda}_1^* &= \tilde{\lambda}_1 + \tilde{\lambda}_2\left(\phi - \Sigma\Theta\Sigma^{-1}\right) \\
\tilde{\lambda}_2^* &= \tilde{\lambda}_2\Sigma\Theta\Sigma^{-1},
\end{aligned}$$

which shows that $\tilde{\lambda}_2$ controls the impact of past realizations of state variable on the current prices of risk. Finally, note that the risk-neutral conditional expectation of z_t is shifted, that is $m_t^Q \equiv E_t^Q[z_{t+1}] = m_t - \Sigma\lambda_t$.

2.4.2 Risk-Neutral Dynamics

The dynamic of the state vector z_t under the risk-neutral probability is given in the following proposition.

Proposition 2.4.1. *If the state vector z_t follows a Gaussian VARMA(1,1) under the historical measure, and given the pricing kernel specified in Equations 2.2 and 2.3, then the state vector z_t follows a VARMA(2,1) process under the risk-neutral probability measure*

$$z_t = \mu^Q + \phi_1^Q z_{t-1} + \phi_2^Q z_{t-2} + \Sigma\left(\varepsilon_t^Q - \Theta^Q \varepsilon_{t-1}^Q\right)$$

where

$$\begin{aligned}
\Theta^Q =& \Sigma^{-1}(I - \Sigma\tilde{\lambda}_2)\Sigma\Theta\Sigma^{-1}(I - \Sigma\tilde{\lambda}_2)^{-1}\Sigma \\
\mu^Q =& (I - \Sigma\tilde{\lambda}_2)\mu - \Sigma(I - \Theta^Q)\tilde{\lambda}_0 \\
\phi_1^Q =& \phi - \Sigma\left(\tilde{\lambda}_1 + \tilde{\lambda}_2(\phi - \Theta^Q)\right) \\
\phi_2^Q =& \Sigma\Theta^Q\tilde{\lambda}_1,
\end{aligned}$$

with $\varepsilon_t^Q = \varepsilon_t + \lambda_{t-1}$ and $\varepsilon_t^Q \sim i.i.d.N(0, I_k)$.

Moreover, we show in Appendix B that for the general case of a VARMA(p,q) process we have a VARMA(max(p,q+1),q) process under the risk-neutral measure and we provide the mapping between parameters from the historical to the risk-neutral measure. Proposition 4.1 has some important corollaries. First, $\tilde{\lambda}_2$ breaks the link between the moving average coefficients under each probability measure (i.e. Θ^Q and Θ). Second, $\tilde{\lambda}_1 = 0$ implies that the state vector follows a VARMA(1,1) dynamic under Q. Finally, if z_t follows a VAR(1) under the historical probability measure then z_t also follow a VAR(1) under the risk neutral probability measure.

This last remark has implications for the identification of the price of risk parameters.

Note that the conditional moment generating function of the VAR(1) process under Q is

$$E_t^Q\left[\exp\left(u^\top z_{t+1}\right)\right] = \exp\left(u^\top\left(\mu-\Sigma\left(\tilde\lambda_0+\tilde\lambda_2\mu\right)\right)+\frac{1}{2}u^\top\Sigma\Sigma^\top u+u^\top\left(\phi-\Sigma(\tilde\lambda_1-\tilde\lambda_2\phi)\right)z_t\right).$$

Hence we can only identify $\tilde\lambda_0+\tilde\lambda_2\mu$ and $\tilde\lambda_1+\tilde\lambda_2\phi$. Therefore, in the case of a VAR(1) process we set $\tilde\lambda_2=0$ but, more generally, λ_2 can be identified in the case of VAR(p) processes with $p>1$ (see Appendix B). Finally, we need the following result to compute bond prices. The conditional moment generating function under the risk-neutral probability measure is

$$E_t^Q\left[\exp\left(u^\top z_{t+1}+v^\top m_{t+1}\right)\right]=\exp\left(a(u,v)+b_1(u,v)^\top z_t+b_2(u,v)^\top m_t\right),$$

where

$$\begin{aligned} a(u,v) &= -w^\top\Sigma\tilde\lambda_0+\frac{1}{2}w^\top\Sigma\Sigma^\top w+v^\top\mu \\ b_1(u,v) &= -\tilde\lambda_1^\top\Sigma^\top w \\ b_2(u,v) &= w+\left(\Sigma\Theta\Sigma^{-1}\right)^\top v-\tilde\lambda_2^\top\Sigma^\top w \\ w &= u+\left(\phi-\Sigma\Theta\Sigma^{-1}\right)^\top v. \end{aligned}$$

2.4.3 Bond Prices

The price at time t of a zero coupon bond with maturity n is

$$\begin{aligned} P_t^{(n)} &= E_t^Q\left[\exp\left(-\sum_{i=0}^{n-1}y_{t+i}^{(1)}\right)\right] \\ &= \exp\left(-y_t^{(1)}+A_n+B_{1,n}^\top z_t+B_{2,n}^\top m_t\right) \end{aligned}$$

where the scalar A_n and the 4×1 coefficient vectors, $B_{1,n}$ and $B_{1,n}$, are functions of the maturity n. These can be computed from the following recursions

$$\begin{aligned}
A_{n+1} &= A_n + a\left(-e_3 + B_{1,n}\,;\, B_{2,n}\right) \\
B_{1,n+1} &= b_1\left(-e_3 + B_{1,n}\,;\, B_{2,n}\right) \\
B_{2,n+1} &= b_2\left(-e_3 + B_{1,n}\,;\, B_{2,n}\right),
\end{aligned}$$

with initial conditions $A_1 = 0$, $B_{1,1} = (0,0,0,0)^\top$ and $B_{2,1} = (0,0,0,0)^\top$ and where $e_3 = (0,0,1,0)^\top$. It follows that bond yields are affine in m_t and z_t

$$y_t^{(n)} = a_n + b_{1,n}^\top z_t + b_{2,n}^\top m_t, \tag{2.4}$$

and we can see the importance of the MA component and the split between the conditional expectation and current innovation, if we rewrite Equation (2.4) in terms of $m_{t-1} = E_{t-1}(z_t)$ and an unexpected component $\Sigma\varepsilon_t$

$$y_t^{(n)} = a_n + (b_{1,n} + b_{2,n})^\top E_{t-1}(z_t) + b_{1,n}^\top \Sigma\varepsilon_t. \tag{2.5}$$

Finally, the loadings are computed recursively

$$\begin{aligned}
b_{2,n+1} &= \frac{n}{n+1}(\Sigma\Theta\Sigma^{-1})'\left[(I-\Sigma\tilde{\lambda}_2)'b_{1,n} + b_{2,n}\right] \\
b_{1,n+1} &= \frac{e_3}{n+1} + \frac{n}{n+1}\left[\left((I-\Sigma\tilde{\lambda}_2)(\phi - \Sigma\Theta\Sigma^{-1}) - \Sigma\tilde{\lambda}_1\right)' b_{1,n} + (\phi - \Sigma\Theta\Sigma^{-1})'b_{2,n}\right] \\
a_{n+1} &= \frac{n}{n+1}\left[a_n + ((I-\Sigma\tilde{\lambda}_2)\mu - \Sigma\tilde{\lambda}_0)'b_{1,n} + \mu' b_{2,n} - \frac{n}{2}b_{1,n}'\Sigma\Sigma' b_{1,n}\right],
\end{aligned}$$

with initial conditions

$$\begin{aligned}
a_1 &= 0 \\
b_{2,1} &= (0,0,0,0)^\top \\
b_{1,1} &= (0,0,1,0)^\top.
\end{aligned}$$

Appendix B presents similar results for the case where the state vector z_t follows a VARMA(p,q) process and considers latent state vectors.

2.5 Estimation of Risk Premia Parameters

This sections presents the estimation of risk premia parameters. We follows Ang et al. (2006) and use a two-step procedure. We take as given VARMA estimates obtained above, and estimate the risk-neutral parameters from observed yields. Note that we impose the appropriate restrictions so that the dynamics of the spread are consistent with the pricing equations. The results indicate that the break between the current state vector, z_t, and its conditional expectation, m_t, induce significant changes in the behavior of the price of risk between the VAR and VARMA specifications.

2.5.1 Estimation Method

As in Ang et al. (2006), we use a two-step procedure to estimate the model. Parameters of the VARMA process under the historical probability measure (i.e. μ, ϕ, Σ and Θ) are estimated in the first step using Maximum Likelihood. In the second step, we minimize the sum of squared fitting errors for yields by a choice of parameters $\tilde{\lambda}_0$, $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ given estimates obtained in the first stage. We use the model-implied yields, $\hat{y}_t^{(n)} = a_n + b_{1,n}^{\top} z_t + b_{2,n}^{\top} m_{t-1}$, and minimize

$$\min_{\{\tilde{\lambda}_0, \tilde{\lambda}_1, \tilde{\lambda}_2\}} \sum_{t=1}^{T} \sum_{n=1}^{N} \left(\hat{y}_t^{(n)} - y_t^{(n)} \right)^2, \tag{2.6}$$

where N is the number of yields used at estimation and T is the number of observation periods. Next, we must ensure the consistency between the implications of the historical dynamics for yields and the implications from the pricing equation. That is,

$$y_t^{(60)} = a_{60} + b_{1,60}^{\top} z_t + b_{2,60}^{\top} m_{t-1},$$

which implies that

$$y_t^{(60)} - y_t^{(1)} = a_{60} + \left(b_{1,60} - e_3\right)^\top z_t + b_{2,60}^\top m_{t-1},$$

which is true if we impose that

$$a_{60} = 0 \, , \, b_{1,60} = e_3 + e_4 \, , \, b_{2,60} = 0. \tag{2.7}$$

Note that, Equation (2.7) and the fact that $a_1 = 0$, $b_{1,1} = e_3$ and $b_{2,1} = 0$, imply that $y_t^{(1)}$ and $y_t^{(60)}$ are measured without errors.

2.5.2 Latent Factors: Rotation

A two-step procedure is feasible whenever the state vector is observable. We now consider the case where some components of the state vector are latent. Assume that the first $m-l$ components of z_t are observed and that the l last components are latent. We can invert l yields $y_t^{(n_1)},...,y_t^{(n_l)}$ for the values of the latent factors under the standard assumption that these yields are measured without error.

Equation (2.4) implies that

$$\overline{y_t} = \overline{a_l} + \overline{b}_{1,l} z_t + \overline{b}_{2,l} m_{t-1}$$

where $\overline{y_t} = \left(y_t^{(n_1)},...,y_t^{(n_l)}\right)'$, $\overline{a_l} = (a_{n_1},...,a_{n_l})'$, $\overline{b}_{1,l} = (b_{1,n_1},...,b_{1,n_l})'$ and $\overline{b}_{2,l} = (b_{2,n_1},...,b_{2,n_l})'$. Denote $\overline{z_t} = (z_1,...,z_{m-l},\overline{y_t})'$, we then have

$$\overline{z_t} = \overline{a} + \overline{b}_1 z_t + \overline{b}_2 m_{t-1} \tag{2.8}$$

where $\overline{a} = (0,...,0,\overline{a_l})'$, $\overline{b}_1 = (e_1,...,e_{n-l},\overline{b}_{1,l}')'$ and $\overline{b}_2 = (0_m,...,0_m,\overline{b}_{2,l}')'$. We note e_i the $(m \times 1)$ vector whose component i equals to 1, and 0 elsewhere, 0_m is a $(m \times 1)$ vector of zeros.

We can then use (2.8) to write the latent variables as a function of observable yields

$$z_t = \bar{b}_1^{-1}\left(\bar{z}_t - \bar{a} - \bar{b}_2 m_{t-1}\right), \tag{2.9}$$

and substitute the results in Equation (2.4) to obtain yields in term of observable variables only

$$y_t^{(n)} = \hat{a}_n + \hat{b}_{1,n}^{\top}\bar{z}_t + \hat{b}_{2,n}^{\top} m_{t-1}, \tag{2.10}$$

where $\hat{a}_n = a_n - b'_{1,n}\bar{b}_1^{-1}\bar{a}$, $\hat{b}'_{1,n} = b'_{1,n}\bar{b}_1^{-1}$ and $\hat{b}_{2,n} = b_{2,n} - b'_{1,n}\bar{b}_1^{-1}\bar{b}_2$. Finally, the conditional expectation m_t can be filtered recursively from the observed state vector $\bar{z}_t$

$$m_{t+1} = \hat{\mu} + \hat{\phi}\bar{z}_{t+1} + \hat{\Theta} m_t,$$

where

$$\begin{aligned}
\hat{\mu} &= \mu - \left(\phi - \Sigma\Theta\Sigma^{-1}\right)\bar{b}_1^{-1}\bar{a} \\
\hat{\phi} &= \left(\phi - \Sigma\Theta\Sigma^{-1}\right)\bar{b}_1^{-1} \\
\hat{\Theta} &= \Sigma\Theta\Sigma^{-1} - \left(\phi - \Sigma\Theta\Sigma^{-1}\right)\bar{b}_1^{-1}\bar{b}_2.
\end{aligned}$$

2.5.3 Risk Premium Estimates

This section presents parameter estimates for the price of risk, λ_t, when z_t follows a VAR(1) and a VARMA(1,1) process. In the case of a VAR process, we impose $\tilde{\lambda}_2 = 0$ which leaves 20 parameters to be estimated with 5 constraints. In the case of a VARMA process, we have 36 parameters with 9 constraints. Tables 2.5 and 2.6 present the results and we discuss the evolution of the price of risk for each variable in turn. In particular, we contrast the difference between estimates from the VAR model and estimates from the VARMA model. One important conclusion is that imposing a tight link between the current state, z_t, and its conditional expectation, m_t, as in the VAR(1) is not innocuous for the purpose of term structure modeling. Figure 2.6 illustrates the path of the price of risk for each variable in the case of VAR and the case of VARMA.

2.5.3.1 Inflation

The price of inflation risk is negative on average implying higher valuations for assets that tend to have higher payoffs in states of the world with higher inflation. Moreover, the price of inflation risk becomes more negative when current inflation is higher. Inflation is also riskier when current real activity, the current short interest rate or the current term spread is higher. When we allow for a Moving Average component, the impacts of the short rate and the term spread increase. Future inflation is much riskier when the current short rate or term spread is higher in a VARMA framework. On the other hand, the impact of the expected short rate and term spread is of the opposite sign, so that the net effect of yield shock is ambiguous. In other word the price of inflation risk may be lower following an increase in the short interest rate if future monetary policy is expected to be tighter (and credible). This result depends on the connection between the current realization of each of these variables and their conditional expectations.

2.5.3.2 Real Activity

The price of real activity risk is positive on average, implying higher valuations for assets that have higher payoffs in a state of the world with lower real activity. As in the case of inflation, estimate from the VAR model implies that higher current inflation and higher real activity increase the price of real activity risk. Surprisingly, the impact of the short rate is very small while the impact of the term spread is modest, but this does not hold in the VARMA model. An increase in the value of the current short rate decreases the risk of real activity, while an increase in expected short rates leads to a large increase in the price of real activity risk. On the other hand, the impact of the term spread becomes insignificant. Again, a VARMA process disentangles the contemporaneous changes in interest, possibly due to higher output, wealth and, hence, higher intertemporal substitution from the endogenous response of future monetary policy.

2.5.3.3 Short Term Interest Rate

The price of short rate risk is negative on average, implying higher valuations for assets that have higher payoffs in states of the world with a higher short rate. Surprisingly, estimates from the VAR model imply little or no variations in the price of interest rate risk. Again, estimates from the VARMA model paint a very different picture. First, an increase in real activity increases the price of short rate risk while an increase in expected real activity leads to a large reduction in the price of short rate risk. Second, an increase in current inflation does not have an impact on the price of short rate risk but an increase in expected inflation decreases the price of short rate risk. Looking at the impact of yield variables, we see that short rate variations do not have a significant impact on the own price of risk. However we find that a higher term spread is associated with higher price of risk as expected, but an increase in the expected slope leads to a large decline in the price of short rate risk.

2.5.3.4 Term Spread

Estimates from the VAR model imply that the price of risk associated with the uncertain future term spread is negative but small. In contrast, estimates from the VARMA model imply that it is negative and large. Again, the difference is due to the difference between the current value and the expected value of the state vector. A higher current real activity, a lower short rate and a higher term spread raise the price of term spread risk. In contrast, higher expected real activity, a lower expected short rate and slope lead to substantial decrease of the price of term spread's risk.

2.5.4 Term Structure Loadings

Figures 2.7 and 2.8 plot the constant a_n, the factor loadings b_{1n} and the conditional mean loadings b_{2n} across maturities. Estimates from the VAR and VARMA models imply a similar constant a_n and, hence, a similar average term premium across maturities. However, the factor loadings are very different across the two models. In the VAR model, higher inflation generally implies higher yields with the maximum loadings around a maturity of one year. In contrast, higher current inflation leads to lower yields for short maturities, with a maximum impact at the shortest maturity. What matters for longer maturities is an increase in expected inflation which is associated with higher yields across all maturities. Next, the loadings on real activity and expected real activity follow a similar pattern in each model. Finally, the loadings on the short rate and the term spread are generally higher in the VARMA model and the loadings on the expected term spread induce a steeper term structure.

2.5.5 Impulses Responses

The state vector z_t has the following infinite order MA representation

$$z_t = (I_4 - \Phi)^{-1} \mu + \Sigma\varepsilon_t + \sum_{k=1}^{\infty} \Phi^{k-1} [\Phi\Sigma - \Sigma\Theta] \varepsilon_{t-k},$$

which implies that yields have the following infinite order MA representation

$$\begin{aligned} y_t^{(n)} &= a_n + (b_{1,n} + b_{2,n})^\top (I_4 - \Phi)^{-1} \mu + b_{1,n}^\top \Sigma\varepsilon_t \\ &\quad + (b_{1,n} + b_{2,n})^\top \sum_{k=1}^{\infty} \Phi^{k-1} \left[\Phi - \Sigma\Theta\Sigma^{-1}\right] \Sigma\varepsilon_{t-k}, \end{aligned}$$

which allows us to analyze the response of yields to variations in the state vector. Figures 2.9 to 2.14 display the results.

In both models, the responses of the yield curve to an inflation and real activity shocks are positive and hump-shaped. The responses to a short rate shock are initially high, but decay with maturity and with time. The responses to term spread shock are, not surprisingly, a steeper term structure, but the effect decays very quickly with time. Contrasting both models, we find that responses from shock to inflation are higher in the VARMA model while responses to real activity shocks are lower in the VARMA model. Responses to short rate shocks are similar across models. However, the responses to a term spread shock are different across models.

2.5.6 Variance decompositions

The h-step ahead forecast of the state variable is

$$E_t[z_{t+h}] = \left(I_4 - \Phi^{h-1}\right)(I_4 - \Phi)^{-1}\mu + \Phi^{h-1}E_t[z_{t+1}],$$

the infinite order MA representation of the h-step forecasting error is

$$z_{t+h} - E_t[z_{t+h}] = \Sigma\varepsilon_{t+h} + \sum_{k=1}^{h-1}\Phi^{h-k-1}[\Phi\Sigma - \Sigma\Theta]\,\varepsilon_{t+k},$$

and, therefore, the relation between the h-step ahead yield forecast error and the h-step ahead state forecast error is

$$\varepsilon_{t+h}^{(n)} = y_{t+h}^{(n)} - E_t\left[y_{t+h}^{(n)}\right] = (b_{1,n} + b_{2,n})^{\top}(z_{t+h} - E_t[z_{t+h}]) - b_{2,n}^{\top}\Sigma\varepsilon_{t+h}.$$

Then, using that $\varepsilon_t \sim i.i.d.\mathcal{N}(0, I_4)$, the variance of forecast errors is

$$var\left(\varepsilon_{t+h}^{(n)}\right) = \sum_{k=1}^{h}\Psi_{k,n}^{\top}\Psi_{k,n} = \sum_{j=1}^{4}\sum_{k=1}^{h}\left[\Psi_{k,n}(j)\right]^2$$

where $\Psi_{h,n} = \Sigma^{\top}b_{1,n}$ and for $k \leq h-1$

$$\Psi_{k,n} = [\Phi\Sigma - \Sigma\Theta]^{\top}\left(\Phi^{\top}\right)^{h-k-1}(b_{1,n} + b_{2,n})$$

and, finally, the proportion of variance explained by state vector component j is

$$\frac{\sum_{k=1}^{h}\left[\Psi_{k,n}(j)\right]^2}{var\left(\varepsilon_{t+h}^{(n)}\right)}.$$

Tables 2.7 and 2.8 display the proportions of variance explained by macroeconomic variables and yield factors, respectively. The fraction of the variance explained by macroeconomic variables increases with the forecasting horizon. Overall for infinite horizon forecasts, real activity explains the highest proportion (40%) of the variance across maturities. Inflation and real activity together explain roughly 60% of the variance. These proportions are stable across maturities. This contrasts with results from Ang and Piazzesi (2003) where the proportion of variance explained by macroeconomic variables decreases with the maturity. The fraction of the variance explained by the short rate decays quickly with time and with the yield maturity. Finally, the fraction of yield variances explained by the term spread is highest at long maturity and decays only slowly with the forecasting horizon.

The proportion of yield variances explained by the short rate and by inflation is higher (more than 2%) in the VARMA compared to VAR, while the proportion explained by real activity and the term spread are lower in the VARMA model.

2.5.7 Forecasts

In this section, we follow Duffee (2002) and compare the relative performance of different models at forecasting future yields. We use the random walk process as a benchmark, because Duffee (2002) shows that it performs better than existing affine models. Table 2.9 presents the forecasting errors by maturity and model. We also perform an out-of-sample forecasting exercise. We apply the two-step estimation procedure to a progressively longer sample, estimating from the $384+i$ observations and forecasting the $n = 12, 24, 36, 48$ horizons. Table 2.10 summarizes the out-of-sample forecasting errors. We use the following relation to forecast yields

$$\begin{aligned} E_t\left[y_{t+h}^{(n)}\right] &= E_t\left[\hat{y}_{t+h}^{(n)}\right] = E_t\left[a_n + b_{1,n}^{\top} z_{t+h} + b_{2,n}^{\top} m_{t+h-1}\right] \\ &= a_n + \left(b_{1,n} + b_{2,n}\right)^{\top} m_{t,h}. \end{aligned}$$

In-sample, the VAR and VARMA models improves upon the random walk for horizons of 3, 6 and 12 months and the improvements are higher for longer maturities. However, the VARMA model generally offers some improvement over the VAR model. Out-of-sample, the VAR model improves upon the random walk only for the longest maturities. On the other hand, the VARMA model still improves upon the random walk at horizons of 3 months or more. Moreover, the improvement are relatively larger compared to the in-sample results. The VARMA improves upon the VAR and the random walk while using a much more flexible specification of the dynamics and the price of risk is a striking result. More flexible models generally fail this test. This is a strong evidence supporting the specification proposed in this chapter.

2.5.8 Discussion

Overall, while risk premium parameters from the VAR model appear to be precisely estimated, they lead to weak or counterintuitive results. The results from the VARMA model suggest that imposing a tight link between m_t and z_t, as in the VAR model, confuses the effect of the current shock to the state vector with the effect of expected change in the state vector. The VARMA model appears to disentangle the endogenous nature of the short interest rate and of the term structure. The impact of current shocks on the price of risk depends on their relative influence on the state vector and the conditional expectation vector. In particular, the price of real activity risk is higher when the current short rate is low but lower when the short interest rate is expected to be lower and the term structure steeper in the future. The former is likely a reflection of the monetary policy responding to bad economic outcomes with a lower short interest rate while the latter is likely due to the effectiveness of a lasting, credible loosening in monetary policy. In a VAR model, the price of the risk associated with short rate fluctuations does not seem to vary significantly with variations in the state vector within the VAR model. This is a standard result (see Ang and Piazzasi (2006)) even with a VAR with longer lags. This reinforces the flexibility of the MA component to capture the relevant stylized facts from the data and support the importance of the MA model to fit the cross-section of yields.

2.6 Robustness Checks: comparison with the Nelson-Siegel model

In this section we compare results from the VARMA term structure model against different specifications of the VAR model. The main question we seek to answer is whether increasing the dimension of the state vector in a VAR model suffices to capture the forecasting improvement observed in the case of the VARMA model. In this section, we focus on forecasting performance. Therefore, we allow for maximum flexibility and use a VARMA(1,1) model on yields. We also evaluate the performance of the following model. First, the Nelson-Siegel [NS] model with AR(1) factor dynamics of Diebold and Li (2006) is our benchmark. This approach is based on latent factors and displays in-sample and out-of-sample performance improvements compared to a standard VAR term structure model. Diebold and Li (2006) find that this model provides the best forecasting performance at horizons of 6 and 12 months. Then, we consider VAR models with three yield factors, and then successively add inflation, real activity, and both to the state vector.

We use unsmoothed Fama-Bliss forward rates which differ from the yields used above. This provides a further check of the robustness of our results. Also, all the VAR and VARMA term structure models are estimated following the two-step procedure described above using yields with 3 months and 1, 3, 5 and 10 years to maturity. We estimate and forecast recursively, using data from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. Tables 2.11 to 2.13 compare h-month ahead out-of-sample forecasts for $h = 1, 6$ and 12 months. Overall, the VARMA models with yields offer the best performance at horizons of 6 and 12 months, surpassing models based on latent factor or richer state vector.

2.7 Conclusion

We study a no-arbitrage VARMA model of the term structure. From a theoretical view we show how to extend a no-arbitrage VAR model to a no-arbitrage VARMA model. In particular, the price of risk is now linear in the state vector and its conditional expectation and, thus, depend on the entire history of the state vector. The model is easily estimated through a two-step procedure and we show that disentangling the impact of innovations on the current state vector and its expectations improves out-of-sample forecasting of yields and of the risk premium compared to standard VAR-based models.

BIBLIOGRAPHY

[1] Ang, A., G. Bekaert, and W. Min (2007), The term structure of real rates and expected inflation, Journal of Finance, 63, 2, 797-849.

[2] Ang, A. and, M. Piazzesi (2003), A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, Journal of Monetary Economics, 50, 745-787.

[3] Ang, A., M. Piazzesi, and M. Wei (2006), What does the yield curve tell us about GDP growth, Journal of Econometrics, 131, 359-403.

[4] Cox, J.C., J.E. Ingersoll, and S.A. Ross (1985), A theory of the term structure of interest rates, Econometrica 53, 385-407.

[5] Dai, Q., and K. Singleton (2000), Specification analysis of affine term structure models, Journal of Finance 55, 1943-1978.

[6] Darolles, S., C. Gourieroux, and J. Jasiak (2006), Structural laplace transform and compound autoregressive models, Journal of Time Serie Analysis, 27, 477-503.

[7] Diebold, F., M. Piazzesi, and D. Rudebusch (2005), Modeling bond yields in finance and macroeconomics, American Economic Review, 95, 415-420.

[8] Diebold, F., D. Rudebusch, and B. Aruoba (2006), The macroeconomy and the yield curve: a dynamic latent factor approach, Journal of Econometrics 131 309-338.

[9] Diebold, F., T.A. Gunther, and T.-S. Tay (1998), Evaluating density forecasts with applications to financial risk management, International Economic Review, 39, 863-883.

[10] Diebold, F., and C. Li (2006), Forecasting the term structure of government bond yields, Journal of Econometrics, 130 337-364.

[11] Diebold, F., and R. Mariano (1995), Comparing predictive accuracy, Journal of Business and Economic Statistics, 13, 253-263.

[12] Duffee, G.-R. (2002), Term premia and interest rate forecasts in affine models, The Journal of Finance, 57, 405-443.

[13] Duffie, D., J. Pan, and K. Singleton (2000), Transform analysis and asset pricing for affine jump diffusions, Econometrica, 68, 1343-1376.

[14] Duffie, D., and R. Kan(1996), A yield-Factor model of interest rates, Mathematical Finance, 6, 379-406.

[15] Feunou, B., and N. Meddahi (2007), Generalized affine models, Working Paper, Université de Montréal.

[16] Gourieroux, C., A. Monfort, and V. Polimenis (2002), Affine term structure models, Working Paper, CREST.

[17] Gourieroux, C., and A. Monfort (2006), Econometric specifications of stochastic discount factor models, Journal of Econometrics, 136, 509-530.

[18] Monfort, A., and F. Pegoraro (2006), Switching VARMA term structure models, Journal of Financial Econometrics, 5, 105-153.

[19] Monfort, A. and F., Pegoraro (2006), Multi-lag term structure models with stochastic risk premia, Working paper, CREST.

[20] Vasicek, O. (1977), An equilibrium characterization of term structure, Journal of Financial Economics, 5, 177-188.

Appendix A

In this Appendix, we will show that the state vector z_t is a VARMA(2,1) under the risk neutral measure Q.

It will be useful to notice that a state vector z_t has the following Gaussian VARMA(2,1) dynamic $z_{t+1} = \mu + \phi_1 z_t + \phi_2 z_{t-1} + \Sigma(\varepsilon_{t+1} - \Theta\varepsilon_t)$, if and only if the conditional mean m_t evolves recursively as $m_{t+1} = \mu + (\phi_1 - \Sigma\Theta\Sigma^{-1})z_{t+1} + \phi_2 z_t + \Sigma\Theta\Sigma^{-1}m_t$.

The risk neutral conditional mean of the state vector z_{t+1} (say m_t^Q) is related to the historical conditional mean (m_t) as follows $m_t^Q = m_t - \Sigma\lambda_t$. Using the expression of the time varying price of risk $\lambda_t = \tilde{\lambda}_0 + \tilde{\lambda}_1 z_t + \tilde{\lambda}_2 m_t$, we can establish that

$$m_t = (I - \Sigma\tilde{\lambda}_2)^{-1}m_t^Q + (I - \Sigma\tilde{\lambda}_2)^{-1}\Sigma(\tilde{\lambda}_0 + \tilde{\lambda}_1 z_t).$$

Using the fact that

$$m_{t+1} = \mu + (\phi - \Sigma\Theta\Sigma^{-1})z_{t+1} + \Sigma\Theta\Sigma^{-1}m_t,$$

one gets easily,

$$m_{t+1}^Q = \mu^Q + (\phi_1^Q - \Sigma\Theta^Q\Sigma^{-1})z_{t+1} + \phi_2^Q z_t + \Sigma\Theta^Q\Sigma^{-1}m_t^Q$$

This implies that z_t is a VARMA(2,1) under Q.

Appendix B

In this Appendix, we will derive the yield curve formula when the state vector z_t is a VARMA(p,q) under the historical probability measure P. The dynamic of z_t under P is:

$$z_{t+1} = \mu + \sum_{j=1}^{p}\phi_j z_{t+1-j} + \Sigma(\varepsilon_{t+1} - \sum_{j=1}^{q}\Theta_j\varepsilon_{t+1-j})$$

where $\varepsilon_t \sim iid\mathcal{N}(0,I)$. The conditional mean of z_{t+1} ($E_t(z_{t+1}) \equiv m_t$) is:

$$m_t = \mu + \sum_{j=1}^{p}\phi_j z_{t+1-j} - \sum_{j=1}^{q}\Sigma\Theta_j\varepsilon_{t+1-j}.$$

We can show that this conditional mean evolves recursively as follows:

$$m_{t+1} = \mu + \sum_{j=1}^{p} \phi_j z_{t+2-j} - \sum_{j=1}^{q} \Sigma\Theta_j\Sigma^{-1} z_{t+2-j} + \sum_{j=1}^{q} \Sigma\Theta_j\Sigma^{-1} m_{t+1-j}.$$

Using the following relationship between historical and risk neutral conditional mean:

$$m_t = (I - \Sigma\tilde{\lambda}_2)^{-1} m_t^Q + (I - \Sigma\tilde{\lambda}_2)^{-1}\Sigma(\tilde{\lambda}_0 + \tilde{\lambda}_1 z_t),$$

one can show that:

$$m_{t+1}^Q = \mu^Q + \sum_{j=1}^{max(p,q+1)} \phi_j^Q z_{t+2-j} - \sum_{j=1}^{q} \Sigma\Theta_j^Q\Sigma^{-1} z_{t+2-j} + \sum_{j=1}^{q} \Sigma\Theta_j^Q\Sigma^{-1} m_{t+1-j}^Q,$$

where

$$\begin{aligned}
\Theta_j^Q &= \Sigma^{-1}(I - \Sigma\tilde{\lambda}_2)\Sigma\Theta_j\Sigma^{-1}(I - \Sigma\tilde{\lambda}_2)^{-1}\Sigma, \\
\mu^Q &= (I - \Sigma\tilde{\lambda}_2)\mu - \Sigma(I - \sum_{j=1}^{q}\Theta_j^Q)\tilde{\lambda}_0, \\
\phi_1^Q &= \phi_1 - \Sigma\left(\tilde{\lambda}_1 + \tilde{\lambda}_2(\phi_1 - \Theta_1^Q)\right),
\end{aligned}$$

and

$$\begin{aligned}
\phi_j^Q &= \Sigma(\Theta_{j-1}^Q\tilde{\lambda}_1 + \Theta_j^Q\tilde{\lambda}_2) + (I - \Sigma\tilde{\lambda}_2)\phi_j \text{ for } 2 \le j \le min(p,q), \\
\phi_j^Q &= \Sigma(\Theta_{j-1}^Q\tilde{\lambda}_1 + \Theta_j^Q\tilde{\lambda}_2) \text{ for } min(p,q) < j \le q, \\
\phi_{q+1}^Q &= \Sigma\Theta_q^Q\tilde{\lambda}_1 + (I - \Sigma\tilde{\lambda}_2)\phi_{q+1}\mathbf{1}_{q+1} \le p, \\
\phi_j^Q &= (I - \Sigma\tilde{\lambda}_2)\phi_j \text{ for } q+1 < j \le max(p,q+1),
\end{aligned}$$

In conclusion z_t is a VARMA(max(p,q+1),q) under the risk neutral probability measure Q. The dynamic of z_t under Q is:

$$z_{t+1} = \mu + \sum_{j=1}^{max(p,q+1)} \phi_j^Q z_{t+1-j} + \Sigma(\varepsilon_{t+1}^Q - \sum_{j=1}^{q}\Theta_j^Q\varepsilon_{t+1-j}^Q),$$

where $\varepsilon_t^Q \sim^Q i.i.d.\mathcal{N}(0,I)$. In order to derive the pricing relationship, we need to compute the multi-horizon risk-neutral conditional mean ($E_t^Q(z_{t+h}) \equiv m_{t,h}^Q$) and the multi-horizon risk-neutral conditional variance ($Var_t^Q(\sum_{k=1}^h z_{t+k}) \equiv \Omega_h^Q$) . We can show that this risk-neutral conditional mean can be computed using the following recursion:

$$
\begin{aligned}
m_{t,1}^Q &= m_t^Q, \\
m_{t,h}^Q &= \mu^Q + \sum_{j=h}^{max(p,q+1)} \phi_j^Q z_{t+2-j} - \sum_{j=h}^{q} \Sigma\Theta_j^Q\Sigma^{-1} z_{t+2-j} + \sum_{j=h}^{q} \Sigma\Theta_j^Q\Sigma^{-1} m_{t+1-j}^Q + \sum_{j=1}^{h-1} \phi_j^Q m_{t,h-j}^Q \text{ for } 2 \leq h \leq q, \\
m_{t,h}^Q &= \mu^Q + \sum_{j=h}^{max(p,q+1)} \phi_j^Q z_{t+2-j} + \sum_{j=1}^{h-1} \phi_j^Q m_{t,h-j}^Q \text{ for } q+1 \leq h \leq max(p,q+1), \\
m_{t,h}^Q &= \mu^Q + \sum_{j=1}^{max(p,q+1)} \phi_j^Q m_{t,h-j}^Q \text{ for } h > max(p,q+1).
\end{aligned}
$$

To compute the variance Ω_h^Q we will decompose the state vector z_{t+h} in function of risk neutral *iid* shocks $\varepsilon_{t+1}^Q,...,\varepsilon_{t+h}^Q$. Assume that,

$$
z_{t+h} = A_{t,h} + \sum_{j=1}^{h} \Sigma_j^{(h)} \varepsilon_{t+j}
$$

then using the following equation,

$$
z_{t+h+1} = A_{t,h+1} + \sum_{j=1}^{max(p,q+1)} \phi_j^Q \left(\sum_{i=1}^{h+1-j} \Sigma_i^{(h+1-j)} \varepsilon_{t+i}^Q \right) + \Sigma\left(\varepsilon_{t+h+1}^Q - \sum_{j=1}^{q} \Theta_{j=1}^Q \varepsilon_{t+h+1-j}^Q\right),
$$

we can easily establish the following recurrence relationship between the $\Sigma_j^{(h)}$ for $1 \leq j \leq h$.

$$
\begin{aligned}
\Sigma_h^{(h)} &= \Sigma, \\
\text{If } 2 \leq h \leq q \text{ then } \Sigma_i^{(h+1)} &= \sum_{j=1}^{h+1-i} \phi_j^Q \Sigma_i^{(h+1-j)} + \Sigma\Theta_{h+1-i}^Q \text{ for } 1 \leq i \leq h, \\
\text{If } q+1 \leq h \leq max(p,q+1) \text{ then } \Sigma_i^{(h+1)} &= \sum_{j=1}^{h+1-i} \phi_j^Q \Sigma_i^{(h+1-j)} \text{ for } i \leq h-q, \\
\text{If } q+1 \leq h \leq max(p,q+1) \text{ then } \Sigma_i^{(h+1)} &= \sum_{j=1}^{h+1-i} \phi_j^Q \Sigma_i^{(h+1-j)} + \Sigma\Theta_{h+1-i}^Q \text{ for } h+1-q \leq i \leq h, \\
\text{If } h > max(p,q+1) \text{ then } \Sigma_i^{(h+1)} &= \sum_{j=1}^{h+1-i} \phi_j^Q \Sigma_i^{(h+1-j)} \text{ for } i \leq h-q, \\
\text{If } h > max(p,q+1) \text{ then } \Sigma_i^{(h+1)} &= \sum_{j=1}^{max(p,q+1)} \phi_j^Q \Sigma_i^{(h+1-j)} + \Sigma\Theta_{h+1-i}^Q \text{ for } h+1-q \leq i \leq h.
\end{aligned}
$$

We can deduce that

$$\Omega_h^Q = \sum_{j=1}^{h} \left(\sum_{k=j}^{h} \Sigma_j^{(k)} \right) \left(\sum_{k=j}^{h} \Sigma_j^{(k)} \right)^\top$$

The yield to maturity n ($y_t^{(n)}$) is an affine function of the state vector z_t and the risk neutral conditional mean $m_{t,h}^Q$

$$y_t^{(n)} = \frac{e_3^\top}{n} \left(z_t + \sum_{k=1}^{n-1} m_{t,k}^Q \right) - \frac{e_3^\top \Omega_{n-1}^Q e_3}{2n}$$

Table 2.1: **Summary statistics of data**

The 1, 12, 24, 36, 48 and 60 month yields are annual zero coupon bond yields from the Fama–Bliss CRSP bond files. The inflation measure is the first component of CPI, PCOM and PPI which refer to CPI inflation, PCOM spot market commodity price inflation, and PPI (Finished Goods) inflation respectively. The real activity measure is the first component of HELP, EMPLOY, IP and UE, which refer to the Index of Help Wanted Advertising in Newspapers, the growth rate of employment, the growth rate in industrial production and the unemployment rate respectively. For the macro variables, the sample period is 1952:01 to 2000:12. For the bond yields, the sample period is 1952:06 to 2000:12.

	Central moments				Autocorrelations		
	Mean	Stdev	Skew	Kurt	Lag1	Lag2	Lag3
1 month	5.1482	2.7893	1.0690	4.6051	0.9720	0.9512	0.9273
12 months	5.8810	2.8436	0.8566	3.9065	0.9841	0.9643	0.9474
24 months	6.0689	2.8112	0.7852	3.6376	0.9878	0.9715	0.9574
36 months	6.2199	2.7624	0.7442	3.5209	0.9893	0.9760	0.9643
48 months	6.3317	2.7482	0.7043	3.4103	0.9899	0.9783	0.9686
60 months	6.3970	2.7245	0.6857	3.2839	0.9912	0.9807	0.9715
Inflation	0.0109	0.9979	1.3342	4.5088	0.9906	0.9760	0.9584
Real Activity	0.0036	1.0041	-1.0523	3.6937	0.9692	0.9235	0.8606

Table 2.2: **Autoregressive Matrix of VAR and VARMA**

		VAR Panel A				VARMA Panel A			
parameter		$x_{1,t-1}$	$x_{2,t-1}$	$y_{1,t-1}$	s_{t-1}	$x_{1,t-1}$	$x_{2,t-1}$	$y_{1,t-1}$	s_{t-1}
μ				0.1207	0.0264			0.1747	-0.0159
				(0.0796)	(0.0709)			(0.0737)	(0.0579)
ϕ	$x_{1,t}$	0.9988	0.0364	-0.0052	0.0004	0.9929	0.0356	-0.0041	-0.0039
		(0.0080)	(0.0060)	(0.0028)	(0.0058)	(0.0096)	(0.0070)	(0.0033)	(0.0077)
	$x_{2,t}$	-0.0343	0.9725	0.0007	0.0143	-0.0229	0.9663	-0.0039	0.0213
		(0.0138)	(0.0103)	(0.0048)	(0.0099)	(0.0165)	(0.0121)	(0.0057)	(0.0135)
	$y_{1,t}$	0.1790	0.0948	0.9407	0.1461	0.1435	0.0748	0.9472	0.0762
		(0.0364)	(0.0273)	(0.0128)	(0.0262)	(0.0354)	(0.0255)	(0.0123)	(0.0285)
	s_t	-0.1215	-0.0781	0.0372	0.8244	-0.0913	-0.0633	0.0313	0.8817
		(0.0324)	(0.0243)	(0.0113)	(0.0233)	(0.0273)	(0.0199)	(0.0095)	(0.0217)

Table 2.3: **Unconditional variance and MA coefficients**
sample period is 1952:06 to 2000:12.

		Panel A: VAR				Panel B: VARMA			
		$x_{1,t-1}$	$x_{2,t-1}$	$y_{1,t-1}$	s_{t-1}	$x_{1,t-1}$	$x_{2,t-1}$	$y_{1,t-1}$	s_{t-1}
Σ	$x_{1,t}$	0.1426				0.1412			
		(0.0041)				(0.0041)			
	$x_{2,t}$	0.0174	0.0924			0.0127	0.0873		
		(0.0100)	(0.0263)			(0.0098)	(0.0243)		
	$y_{1,t}$	0.2436	0.6293	-0.0159		0.2371	0.6164	-0.0126	
		(0.0071)	(0.0184)	(0.0238)		(0.0069)	(0.0180)	(0.0210)	
	s_t	0.0872	0.0124	-0.4594	0.3393	0.0754	0.0110	-0.4451	0.3388
		(0.0264)	(0.0236)	(0.0194)	(0.0099)	(0.0257)	(0.0227)	(0.0192)	(0.0099)
Θ	$x_{1,t}$					-0.1247	0.0203	-0.1642	0.0188
						(0.0386)	(0.0426)	(0.0356)	(0.0496)
	$x_{2,t}$						-0.1245	0.1139	-0.0467
							(0.0506)	(0.0482)	(0.0447)
	$y_{1,t}$							-0.0501	-0.2053
								(0.0449)	(0.0449)
	s_t								0.0483
									(0.0468)
Lik		-450.3976				-415.8590			
Bic		1.0783				1.1283			

Table 2.4: **State variable forecasting errors: RMSE**
We measure the difference between the model forecast of state variable z_t, for a given horizon h ($E_t[z_{t+h}]$) and the observed state variable z_{t+h}. $RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(E_t[z_{t+h}]-z_{t+h}\right)^2}$ In sample period is 1952:06 to 2000:12. Out of sample exercise is conducted by successively estimating on 200+i th first observations and forecasting the 200+i+1

	$x_{1,t}$	$x_{2,t}$	$y_t^{(1)}$	$y_t^{(60)}-y_t^{(1)}$	$x_{1,t}$	$x_{2,t}$	$y_t^{(1)}$	$y_t^{(60)}-y_t^{(1)}$
	IS RMSE 1 month horizon				IS RMSE 3 months horizon			
RW	0.1368	0.2491	0.6601	0.6049	0.2880	0.5309	1.0632	0.8731
VAR	0.1312	0.2439	0.6271	0.5705	0.2651	0.4993	0.9722	0.7678
VARMA	0.1296	0.2372	0.6116	0.5586	0.2631	0.4858	0.9692	0.7615
	OS RMSE 1 month horizon				OS RMSE 3 months horizon			
RW	0.1248	0.1367	0.5424	0.5874	0.2782	0.2557	0.7559	0.7818
VAR	0.1244	0.1367	0.5161	0.5684	0.2799	0.2545	0.7264	0.7515
VARMA	0.1212	0.1375	0.5076	0.5681	0.2779	0.2493	0.6829	0.7254
	IS RMSE 6 months horizon				IS RMSE 12 months horizon			
RW	0.4461	0.9014	1.3589	1.0510	0.7319	1.4017	1.7494	1.2651
VAR	0.3915	0.8034	1.2137	0.8711	0.6142	1.1155	1.5490	0.9487
VARMA	0.3885	0.7796	1.2012	0.8586	0.6096	1.0646	1.5406	0.9437
	OS RMSE 6 months horizon				OS RMSE 12 months horizon			
RW	0.4154	0.3983	0.9770	0.9765	0.6149	0.6192	1.3765	1.2150
VAR	0.4197	0.3819	0.9556	0.9078	0.6182	0.5488	1.1998	0.9481
VARMA	0.4186	0.3722	0.8761	0.8702	0.6187	0.5272	1.1204	0.9336

Table 2.5: **VAR Risk premia parameters**
The sample period is 1952:06 to 2000:12. Standard error in parenthesis. They have been calculated using GMM procedure on first order conditions

	Inflation	Real Activity	Short Rate	Spread
λ_0	1.3301	1.7022	-0.3790	-0.1686
	(0.0008)	(0.0170)	(0.0035)	(0.0017)
λ_1	-0.9654	-0.7687	-0.1974	-0.8370
	(0.0140)	(0.0079)	(0.0057)	(0.0234)
	1.6296	1.0599	0.0068	0.8881
	(0.0193)	(0.0093)	(0.0004)	(0.0184)
	0.0969	0.0566	-0.0682	0.0348
	(0.0013)	(0.0002)	(0.0011)	(0.0006)
	0.0379	0.0098	0.0278	-0.1186
	(0.0004)	(0.0003)	(0.0011)	(0.0023)
RMSE : 0.2466				

Table 2.6: **VARMA Risk premia parameters**

The sample period is 1952:06 to 2000:12. Standard error in parenthesis. They have been calculated using GMM procedure on first order conditions

Inflation	Real Activity	Short Rate	Spread	Inflation	Real Activity	Short Rate	Spread
λ_0							
0.8619	2.3873	-0.0750	-0.7245				
(0.6254)	(0.3146)	(0.0255)	(0.1026)				
λ_1				λ_2			
-0.4302	-0.3456	-1.6447	-1.7944	-0.4157	-0.4043	1.4552	1.2905
(0.1155)	(0.4860)	(0.5558)	(0.2073)	(0.2206)	(0.1505)	(0.4470)	(0.5139)
0.5910	0.3732	-2.2473	-1.9016	0.6980	0.5745	2.2894	2.4499
(0.2666)	(0.1251)	(0.6730)	(1.9740)	(0.3121)	(0.3791)	(0.8684)	(2.4225)
-0.0579	-3.3396	-0.4000	-2.1077	0.5557	3.7152	0.2608	2.3644
(0.0378)	(0.0221)	(0.4561)	(0.1949)	(0.0379)	(0.0370)	(0.4676)	(0.1434)
0.0091	1.7152	-1.8345	-1.0487	-0.3323	-1.9232	1.9472	0.8613
(0.0100)	(0.0851)	(0.0577)	(0.5883)	(0.1073)	(0.0268)	(0.0674)	(0.7392)
RMSE : 0.2348							

Table 2.7: **Proportion of variance explained by macroeconomic variables**

Proportion of variance explained by Inflation					Proportion of variance explained by Real Activity			
	Forecast horizon h				Forecast horizon h			
	1 month	12 months	60 months	∞	1 month	12 months	60 months	∞
VAR								
1 month	1.846	14.562	20.034	18.128	2.071	12.842	41.816	44.888
12 months	5.609	17.380	21.467	19.086	5.082	14.458	41.386	45.224
24 months	6.438	15.487	21.555	18.994	5.462	11.717	37.814	42.804
60 months	6.220	12.563	21.978	19.239	3.661	6.789	32.401	39.168
VARMA								
1 month	1.446	12.432	21.045	20.006	1.938	12.467	39.455	41.794
12 months	3.947	14.440	22.340	21.034	2.691	13.767	38.858	41.843
24 months	4.293	13.095	22.476	21.058	3.499	11.354	35.504	39.390
60 months	4.759	11.268	23.022	21.480	3.547	6.888	30.510	35.812

This table displays the contribution of the macroeconomic variables to the h-step ahead forecast variance of the 1, 12, 24 and 60 month yield for the VAR and VARMA models.

Table 2.8: **Proportion of variance explained by yield factors**

	Proportion of variance explained by Short Rate				Proportion of variance explained by Term Spread			
	Forecast horizon				Forecast horizon			
	1 month	12 months	60 months	∞	1 months	12 months	60 months	∞
VAR								
1 month	96.081	60.566	21.103	19.117	0	12.028	17.045	17.864
12 months	78.048	39.052	12.675	11.341	11.258	29.107	24.469	24.346
24 months	54.985	29.361	10.300	9.081	33.113	43.433	30.329	29.119
60 months	18.065	17.572	7.858	6.810	72.051	63.074	37.761	34.782
VARMA								
1 month	96.614	60.781	23.609	22.135	0	14.318	15.888	16.063
12 months	64.762	41.019	15.323	14.093	28.598	30.772	23.477	23.028
24 months	45.146	31.243	12.577	11.348	47.061	44.306	29.441	28.202
60 months	18.669	18.521	9.575	8.426	73.022	63.320	36.891	34.280

This table displays the contribution of level and slope to the h-step ahead forecast variance of the 1, 12, 24 and 60 month yield for the VAR and VARMA models.

Table 2.9: **Cross Section Root Mean Squared Errors**

We measure the difference between model-yields $\hat{y}_t^{(n)}$ and observed yield $y_t^{(n)}$. $RMSE^{(n)} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t^{(n)} - y_t^{(n)}\right)^2}$ In sample period is 1952:06 to 2000:12. Out of sample exercise is conducted by successively estimating on 200+i th first observations and forecasting the 200+i+1

	$y_t^{(12)}$	$y_t^{(24)}$	$y_t^{(36)}$	$y_t^{(48)}$	Total
In Sample					
VAR	0.3669	0.2519	0.1774	0.1774	0.2466
VARMA	0.3529	0.2433	0.1740	0.1740	0.2385
Out of Sample					
VAR	0.3867	0.2832	0.1866	0.0846	0.2607
VARMA	0.3636	0.2679	0.1781	0.0830	0.2462

Table 2.10: **Yield curve forecasting errors by horizon**
We measure the difference between model forecast of yield to maturity n, for a given horizon m ($E_t[\hat{y}^{(n)}_{t+h}]$) and observed yield $y^{(n)}_{t+h}$. $RMSE^{(n)}(h) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(E_t[\hat{y}^{(n)}_{t+h}] - y^{(n)}_{t+h}\right)^2}$ In sample period is 1952:06 to 2000:12. Out of sample exercise is conducted by successively estimating on 200+i th first observations and forecasting the 200+i+1

	$y_t^{(12)}$	$y_t^{(24)}$	$y_t^{(36)}$	$y_t^{(48)}$	$y_t^{(12)}$	$y_t^{(24)}$	$y_t^{(36)}$	$y_t^{(48)}$
	IS RMSE 1 month horizon				IS RMSE 3 months horizon			
RW	0.5065	0.4395	0.4034	0.3898	0.9221	0.8201	0.7377	0.6884
VAR	0.5650	0.4624	0.4125	0.3829	0.8638	0.7619	0.6876	0.6479
VARMA	0.5493	0.4533	0.4083	0.3814	0.8524	0.7570	0.6858	0.6481
	OS RMSE 1 month horizon				OS RMSE 3 months horizon			
RW	0.3297	0.3386	0.3434	0.3487	0.6299	0.6671	0.6551	0.6472
VAR	0.5024	0.4356	0.3808	0.3464	0.7368	0.6971	0.6438	0.6117
VARMA	0.4656	0.4115	0.3675	0.3414	0.6783	0.6636	0.6239	0.6039
	IS RMSE 6 months horizon				IS RMSE 12 months horizon			
In Sample								
RW	1.2328	1.1003	0.9938	0.9375	1.6257	1.4786	1.3614	1.3059
VAR	1.1274	1.0021	0.9086	0.8562	1.4826	1.3475	1.2284	1.1645
VARMA	1.1106	0.9929	0.9031	0.8541	1.4694	1.3378	1.2202	1.1601
	OS RMSE 6 months horizon				OS RMSE 12 months horizon			
RW	0.9455	0.9785	0.9531	0.9318	1.4762	1.4753	1.4156	1.3777
VAR	1.0187	0.9651	0.8976	0.8491	1.3991	1.3413	1.2597	1.1914
VARMA	0.9440	0.9180	0.8676	0.8339	1.3237	1.2831	1.2166	1.1622

Table 2.11: **Out-of-sample 1-month-ahead forecasting results**

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
Nelson–Siegel with AR(1) factor dynamics					
3 months	-0.045	0.170	0.176	0.247	0.017
1 year	0.023	0.235	0.236	0.425	-0.213
3 years	-0.056	0.273	0.279	0.332	-0.117
5 years	-0.091	0.277	0.292	0.333	-0.116
10 years	-0.062	0.252	0.260	0.259	-0.115
VAR(1) on 1, 24, 96 months yields					
3 months	-0.043	0.196	0.200	0.126	0.320
1 year	-0.011	0.235	0.234	0.380	-0.139
3 years	0.018	0.269	0.268	0.358	-0.153
5 years	-0.014	0.281	0.280	0.375	-0.154
10 years	-0.163	0.274	0.318	0.386	-0.094
VAR(1) on 1, 24, 96 months yields and inflation					
3 months	0.078	0.196	0.210	0.120	0.295
1 year	0.088	0.227	0.243	0.328	-0.144
3 years	0.134	0.261	0.292	0.294	-0.099
5 years	0.052	0.273	0.276	0.315	-0.107
10 years	-0.050	0.269	0.272	0.354	-0.100
VAR(1) on 1, 24, 96 months yields and real activity					
3 months	-0.082	0.202	0.217	0.143	0.316
1 year	-0.057	0.239	0.245	0.411	-0.149
3 years	0.024	0.265	0.264	0.332	-0.139
5 years	-0.031	0.278	0.278	0.351	-0.136
10 years	-0.107	0.277	0.296	0.403	-0.092
VAR(1) on 1, 96 months yields and inflation					
3 months	0.263	0.252	0.363	0.186	0.287
1 year	0.366	0.361	0.513	0.318	-0.035
3 years	0.243	0.402	0.468	0.104	-0.074
5 years	0.087	0.355	0.364	0.108	-0.131
10 years	-0.158	0.318	0.354	0.383	0.012

Table 2.11: **Out-of-sample 1-month-ahead forecasting results (continued)**

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(1)$	$\widehat{\rho}(12)$
VARMA(1,1) on 1, 96 months yields and inflation					
3 months	0.224	0.271	0.351	0.118	0.152
1 year	0.333	0.293	0.443	0.362	-0.086
3 years	0.223	0.269	0.348	0.263	-0.086
5 years	0.073	0.267	0.276	0.236	-0.098
10 years	-0.158	0.269	0.311	0.314	-0.073
VAR(1) on 1, 96 months yields inflation and real activity					
3 months	0.196	0.236	0.306	0.208	0.311
1 year	0.365	0.292	0.467	0.493	-0.032
3 years	0.208	0.280	0.348	0.394	-0.122
5 years	0.077	0.277	0.286	0.335	-0.092
10 years	-0.085	0.275	0.286	0.376	-0.099
VARMA(1,1) on 1, 24, 96 months yields					
3 months	-0.021	0.231	0.230	0.014	0.299
1 year	0.004	0.230	0.229	0.055	-0.006
3 years	0.023	0.250	0.249	0.143	-0.095
5 years	-0.007	0.267	0.265	0.266	-0.100
10 years	-0.152	0.264	0.304	0.348	-0.066

Note: We present the results of out-of-sample 1-month-ahead forecasting using eight models, as described in detail in section 2.6. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+1 as $y_{t+1}(\tau) - \hat{y}_{t+1}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their first and 12th sample autocorrelation coefficients.

Table 2.12: **Out-of-sample 6-months-ahead forecasting results**

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(6)$	$\widehat{\rho}(18)$
Nelson-Siegel with AR(1) factor dynamics					
3 months	0.083	0.510	0.517	0.301	-0.190
1 year	0.131	0.656	0.669	0.168	-0.174
3 years	-0.052	0.748	0.750	0.049	-0.189
5 years	-0.173	0.758	0.777	0.069	-0.273
10 years	-0.251	0.676	0.721	0.058	-0.288
VAR(1) on 1, 24, 96 months yields					
3 months	-0.074	0.494	0.496	0.193	-0.109
1 year	-0.040	0.696	0.693	0.085	-0.142
3 years	-0.089	0.777	0.777	-0.014	-0.197
5 years	-0.180	0.789	0.805	-0.006	-0.220
10 years	-0.388	0.735	0.827	-0.019	-0.186
VAR(1) on 1, 24, 96 months yields and inflation					
3 months	0.590	0.464	0.750	0.143	0.116
1 year	0.562	0.671	0.872	0.028	-0.025
3 years	0.396	0.769	0.861	-0.045	-0.135
5 years	0.186	0.789	0.806	-0.023	-0.178
10 years	-0.040	0.727	0.723	-0.060	-0.187
VAR(1) on 1, 24, 96 months yields and real activity					
3 months	-0.184	0.508	0.537	0.156	-0.042
1 year	-0.151	0.712	0.723	0.067	-0.113
3 years	-0.116	0.787	0.791	-0.020	-0.191
5 years	-0.218	0.798	0.822	-0.007	-0.219
10 years	-0.341	0.739	0.810	-0.019	-0.187
VAR(1) on 1, 96 months yields and inflation					
3 months	0.768	0.506	0.918	0.284	0.161
1 year	0.716	0.705	1.002	0.125	0.025
3 years	0.373	0.832	0.907	-0.016	-0.087
5 years	0.117	0.839	0.841	-0.019	-0.146
10 years	-0.228	0.781	0.809	-0.065	-0.207

Table 2.12: **Out-of-sample 6-months-ahead forecasting results (continued)**

VARMA(1,1) on 1, 96 months yields and inflation					
3 months	0.519	0.469	0.698	0.231	0.134
1 year	0.512	0.639	0.816	0.098	0.007
3 years	0.250	0.754	0.790	-0.048	-0.134
5 years	0.031	0.792	0.787	-0.042	-0.185
10 years	-0.269	0.754	0.796	-0.063	-0.188
VAR(1) on 1, 96 months yields inflation and real activity					
3 months	0.563	0.538	0.776	0.238	0.157
1 year	0.567	0.730	0.921	0.098	0.012
3 years	0.281	0.800	0.843	-0.036	-0.144
5 years	0.051	0.834	0.830	-0.024	-0.190
10 years	-0.193	0.779	0.798	-0.054	-0.192
VARMA(1,1) on 1, 24, 96 months yields					
3 months	0.001	0.531	0.528	0.312	-0.163
1 year	0.015	0.665	0.661	0.208	-0.168
3 years	-0.028	0.725	0.721	0.047	-0.200
5 years	-0.103	0.735	0.738	0.049	-0.223
10 years	-0.297	0.681	0.739	0.018	-0.198

Note: We present the results of out-of-sample 6-months-ahead forecasting using eight models, as described in detail in section 2.6. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+6 as $y_{t+6}(\tau) - \hat{y}_{t+6/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their sixth and eighteenth sample autocorrelation coefficients.

Table 2.13: **Out-of-sample 12-months-ahead forecasting results**

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(12)$	$\widehat{\rho}(24)$
Nelson–Siegel with AR(1) factor dynamics					
3 months	0.150	0.724	0.739	-0.288	0.001
1 year	0.173	0.823	0.841	-0.332	-0.004
3 years	-0.123	0.910	0.918	-0.408	0.015
5 years	-0.337	0.918	0.978	-0.412	0.003
10 years	-0.531	0.825	0.981	-0.433	-0.003
VAR(1) on 1, 24, 96 months yields					
3 months	-0.152	0.792	0.801	-0.214	-0.076
1 year	-0.188	0.913	0.926	-0.307	-0.027
3 years	-0.325	0.953	1.001	-0.393	0.001
5 years	-0.459	0.956	1.055	-0.413	-0.006
10 years	-0.710	0.875	1.123	-0.440	-0.006
VAR(1) on 1, 24, 96 months yields and inflation					
3 months	0.854	0.851	1.202	0.020	-0.045
1 year	0.743	1.002	1.242	-0.137	-0.005
3 years	0.436	1.044	1.125	-0.286	0.016
5 years	0.153	1.050	1.054	-0.323	0.001
10 years	-0.143	0.933	0.937	-0.414	-0.001
VAR(1) on 1, 24, 96 months yields and real activity					
3 months	-0.329	0.852	0.908	-0.245	-0.081
1 year	-0.359	0.994	1.051	-0.344	-0.030
3 years	-0.390	1.018	1.084	-0.420	-0.008
5 years	-0.528	1.007	1.131	-0.430	-0.014
10 years	-0.682	0.901	1.125	-0.455	-0.011
VAR(1) on 1, 96 months yields and inflation					
3 months	0.952	0.889	1.299	0.077	-0.055
1 year	0.800	1.050	1.314	-0.097	-0.005
3 years	0.359	1.102	1.152	-0.280	0.035
5 years	0.058	1.083	1.077	-0.342	0.026
10 years	-0.337	0.958	1.010	-0.428	-0.009

Table 2.13: **Out-of-sample 12-months-ahead forecasting results (continued)**

Maturity	Mean	Std. Dev.	RMSE	$\widehat{\rho}(12)$	$\widehat{\rho}(24)$
VARMA(1,1) on 1, 96 months yields and inflation					
3 months	0.534	0.797	0.955	0.079	-0.074
1 year	0.447	0.942	1.037	-0.099	-0.021
3 years	0.123	1.001	1.002	-0.303	0.011
5 years	-0.122	1.016	1.016	-0.359	0.002
10 years	-0.454	0.920	1.021	-0.446	0.001
VAR(1) on 1, 96 months yields inflation and real activity					
3 months	0.483	1.004	1.108	-0.083	-0.012
1 year	0.400	1.168	1.228	-0.222	0.009
3 years	0.127	1.150	1.150	-0.363	0.018
5 years	-0.122	1.140	1.139	-0.391	0.006
10 years	-0.388	1.006	1.071	-0.465	0.007
VARMA(1,1) on 1, 24, 96 months yields					
3 months	-0.137	0.708	0.716	-0.059	-0.123
1 year	-0.170	0.794	0.807	-0.143	-0.076
3 years	-0.267	0.832	0.868	-0.269	-0.043
5 years	-0.371	0.849	0.921	-0.305	-0.049
10 years	-0.596	0.775	0.974	-0.357	-0.053

Note: We present the results of out-of-sample 12-months-ahead forecasting using eight models, as described in detail in section 2.6. We estimate all models recursively from 1985:1 to the time that the forecast is made, beginning in 1994:1 and extending through 2000:12. We define forecast errors at t+12 as $y_{t+12}(\tau) - \hat{y}_{t+12/t}(\tau)$, and we report the mean, standard deviation and root mean squared errors of the forecast errors, as well as their their 12th and 24th sample autocorrelation coefficients.

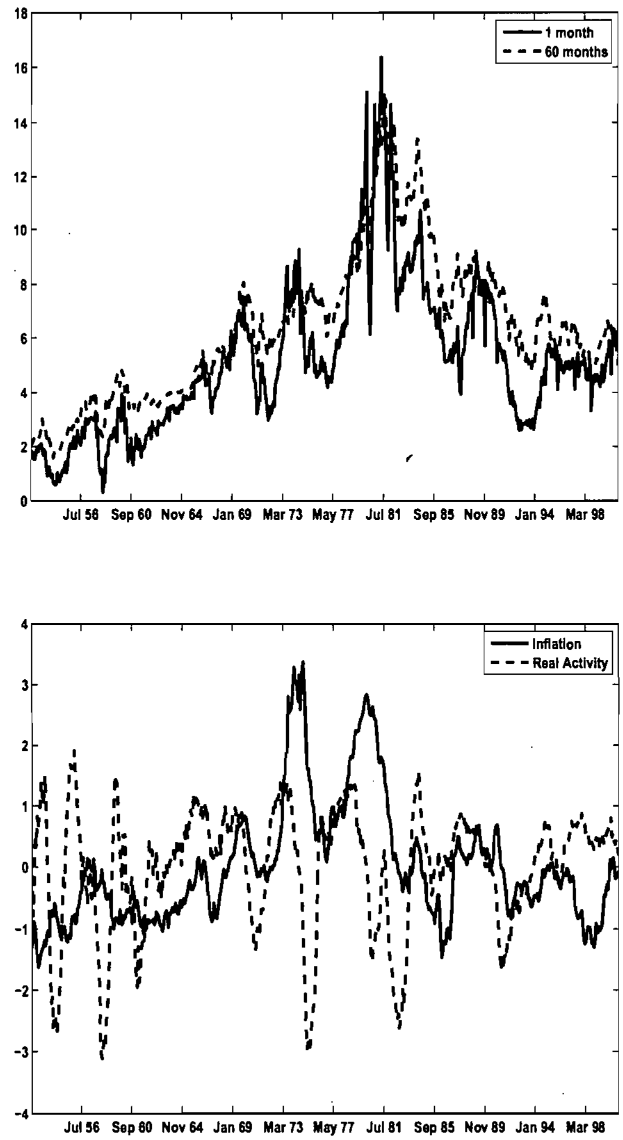


Figure 2.1: **Bond yields and macro principal components**

The top panel shows a plot of (annualized) monthly yields of maturity at 1 month and 60 months. The bottom panel plots the two macro factors representing inflation and real activity. The sample period is 1952:06 to 2000:12

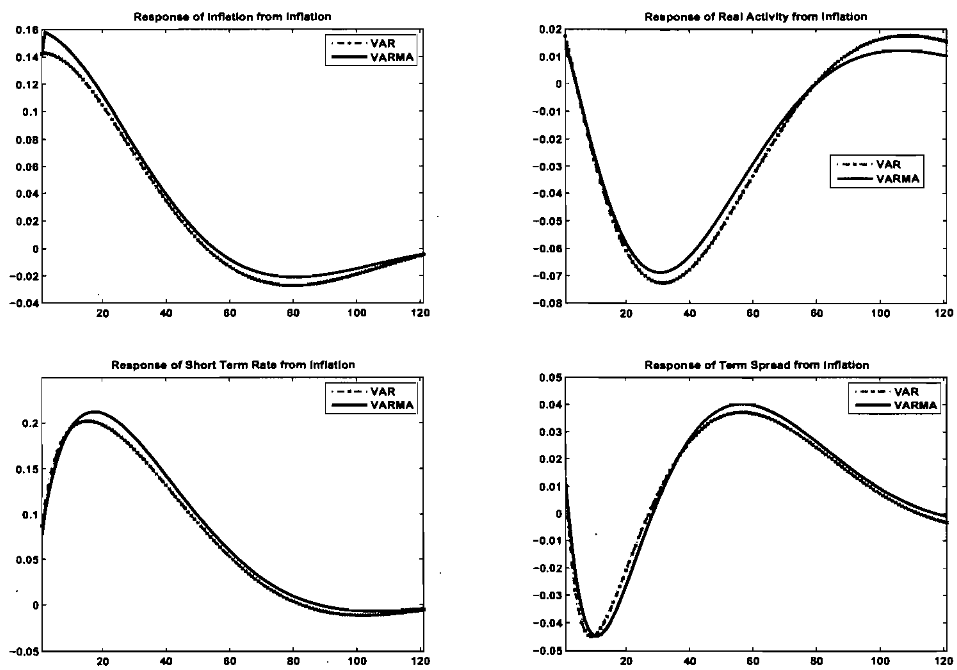


Figure 2.2: **Impulse responses from the VAR and VARMA on yields and macro factors** We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to inflation. Time is measured in months on the x-axis.

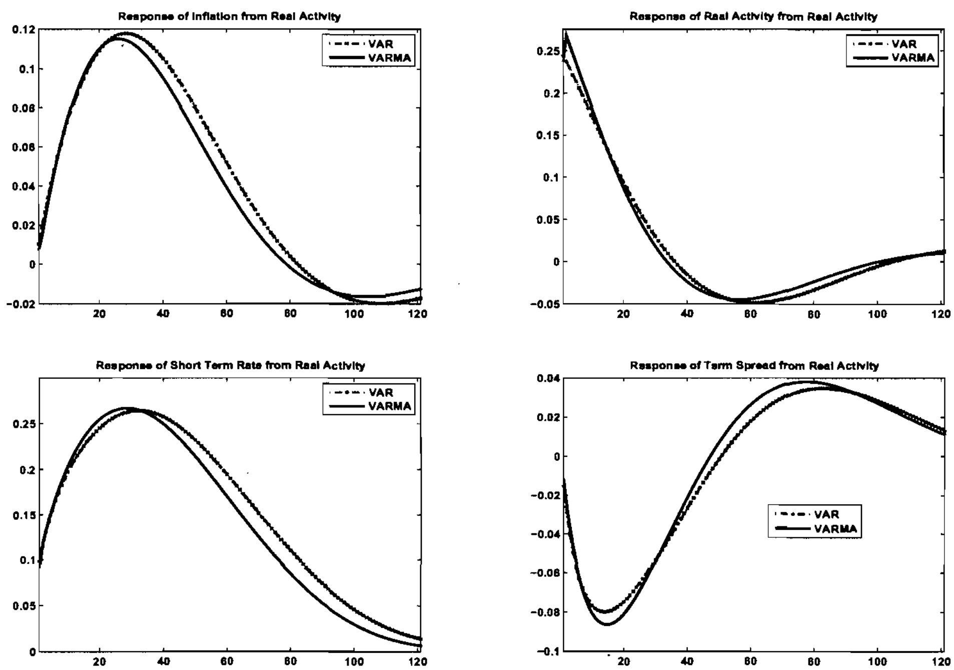


Figure 2.3: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to real activity. Time is measured in months on the x-axis.

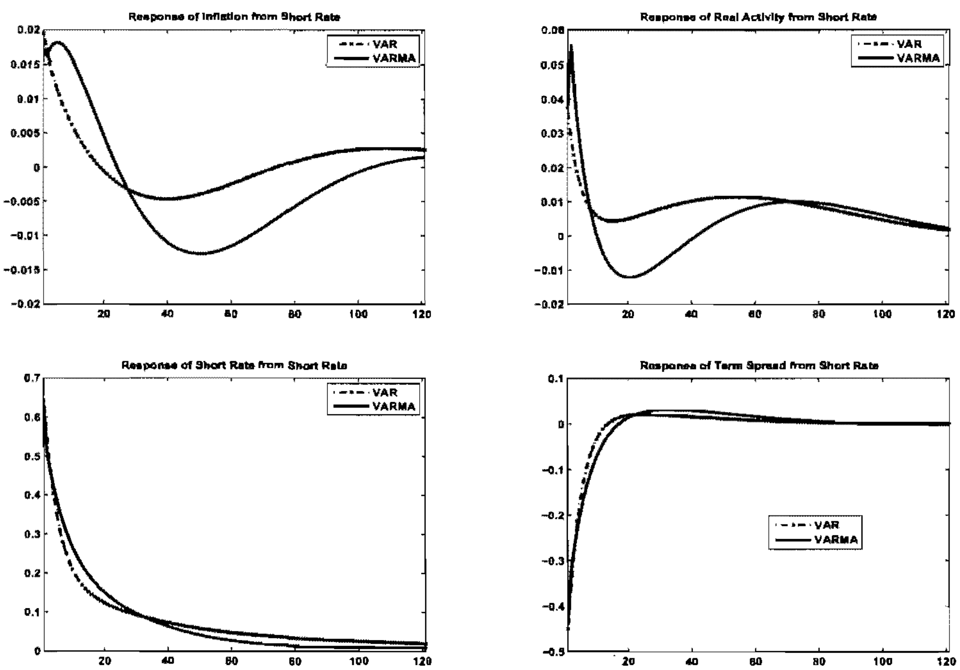


Figure 2.4: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to short rate. Time is measured in months on the x-axis

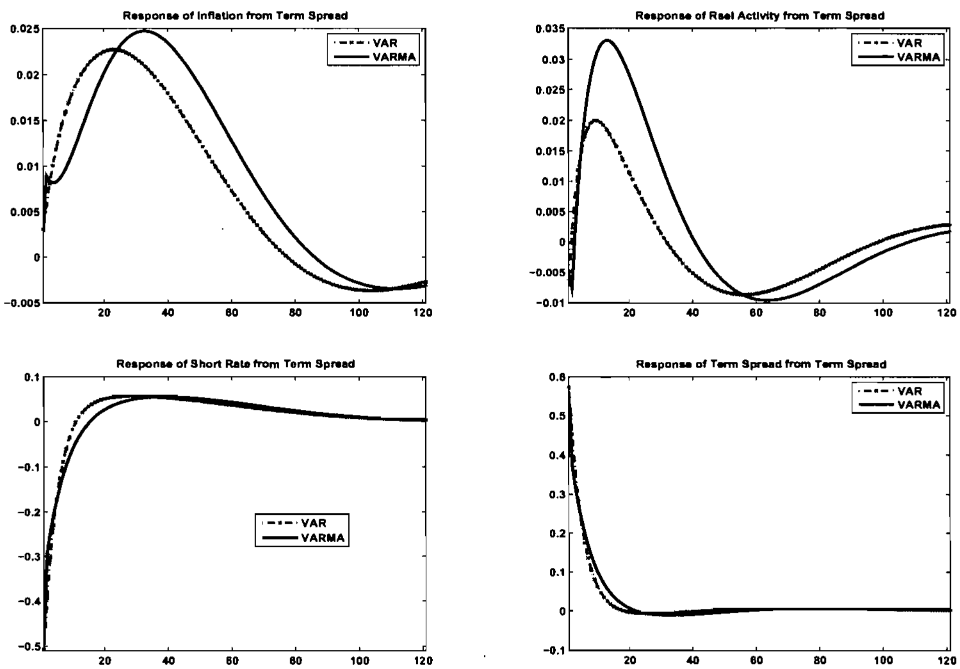


Figure 2.5: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to short rate. Time is measured in months on the x-axis

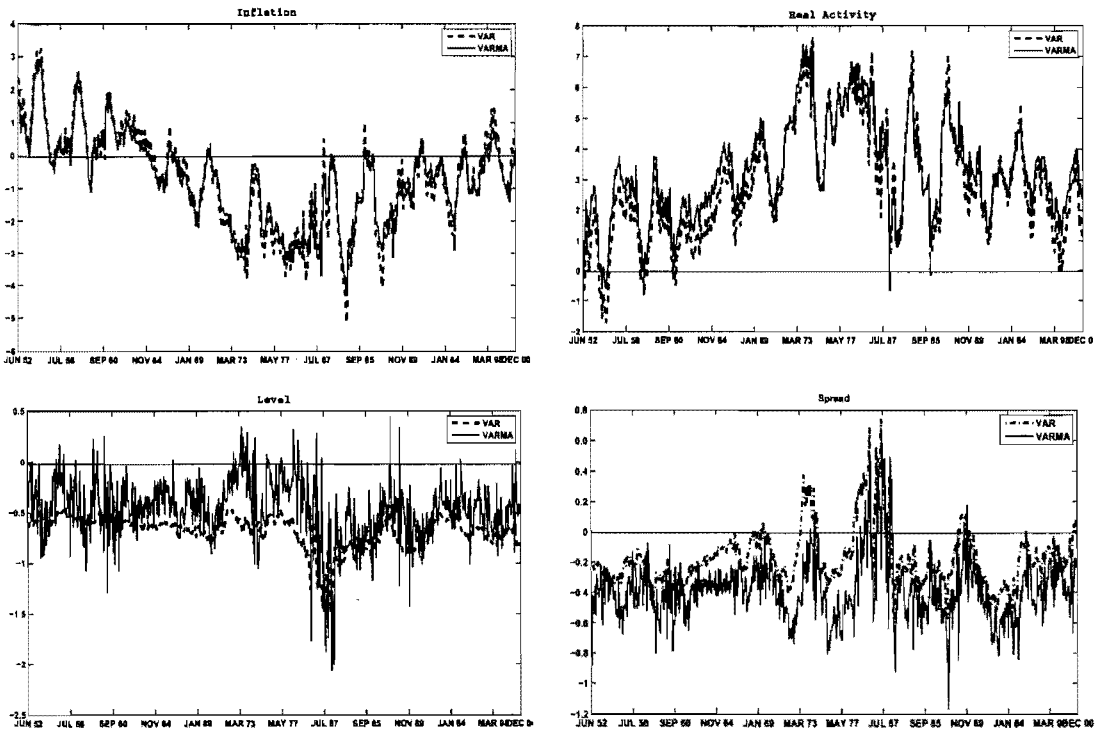


Figure 2.6: **Price of risk**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the evolution of price of risk component through time. Time is measured in months on the x-axis

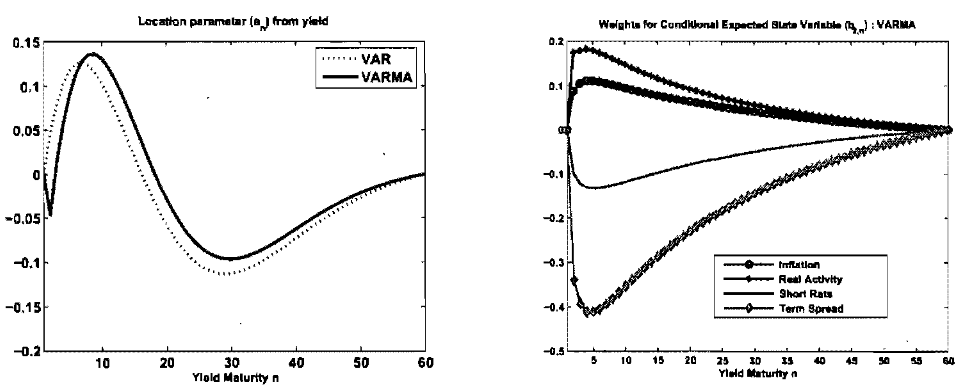


Figure 2.7: a_n **and** $b_{2,n}$ **for the VAR and VARMA models**

The figure displays a_n and $b_{2,n}$ as a function of maturity n

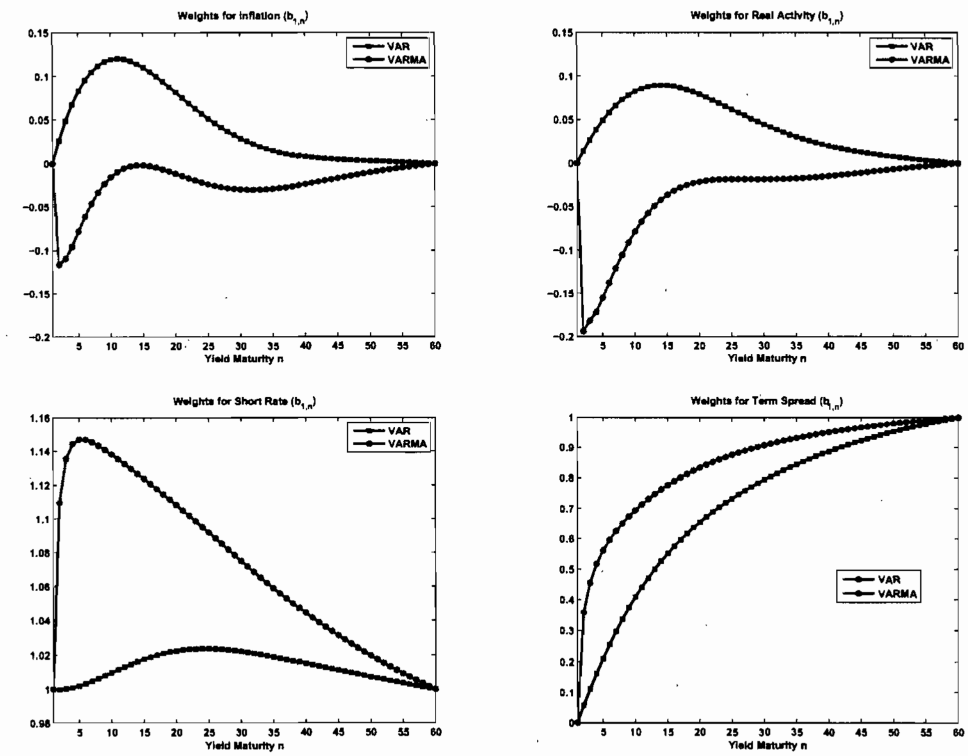


Figure 2.8: $b_{1,n}$ **for the VAR and VARMA models**

The figure displays $b_{1,n}$ as a function of maturity n

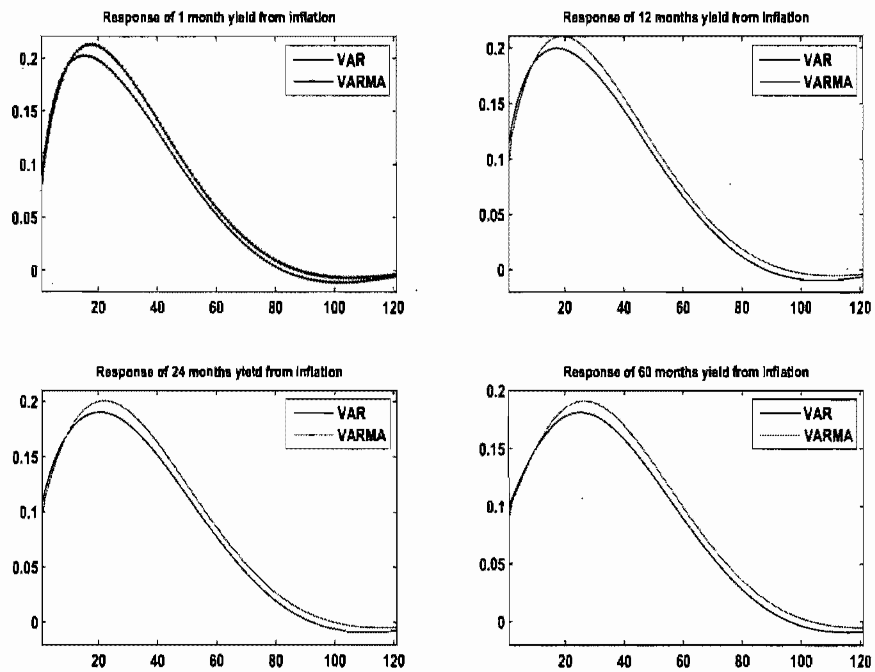


Figure 2.9: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to inflation. Time is measured in months on the x-axis

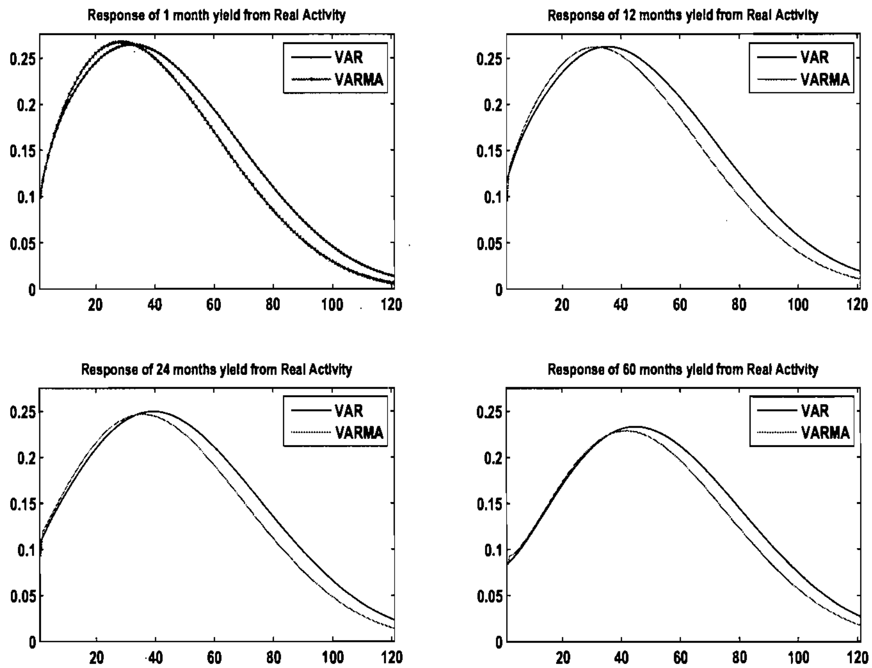


Figure 2.10: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to real activity. Time is measured in months on the x-axis

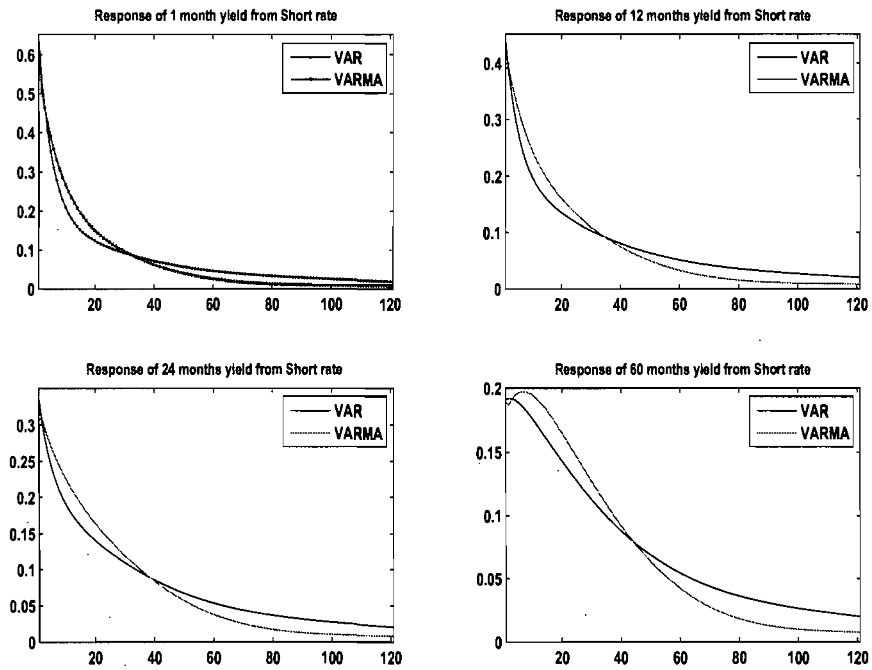


Figure 2.11: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to short rate. Time is measured in months on the x-axis

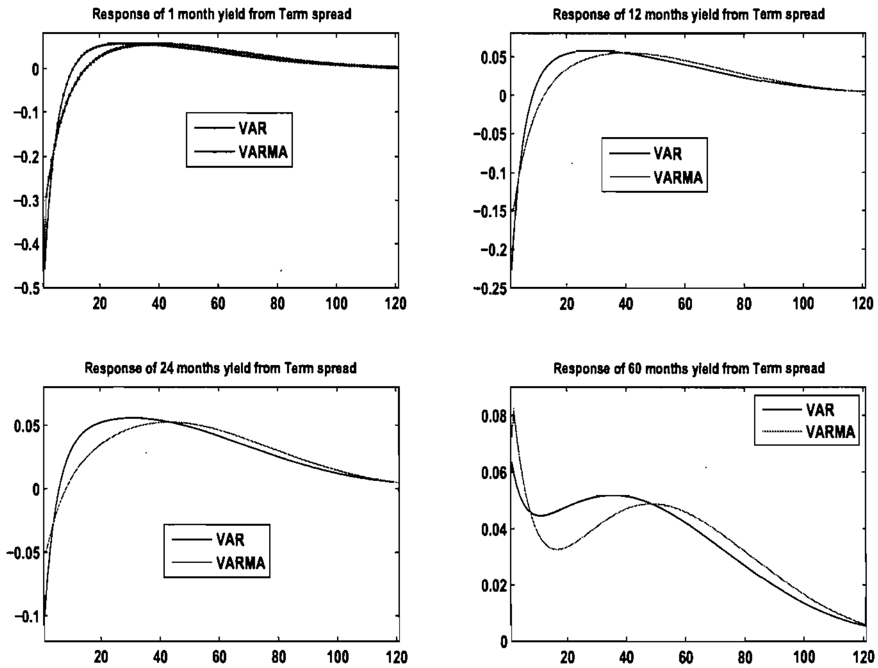


Figure 2.12: **Impulse responses from the VAR and VARMA on yields and macro factors**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to term spread. Time is measured in months on the x-axis

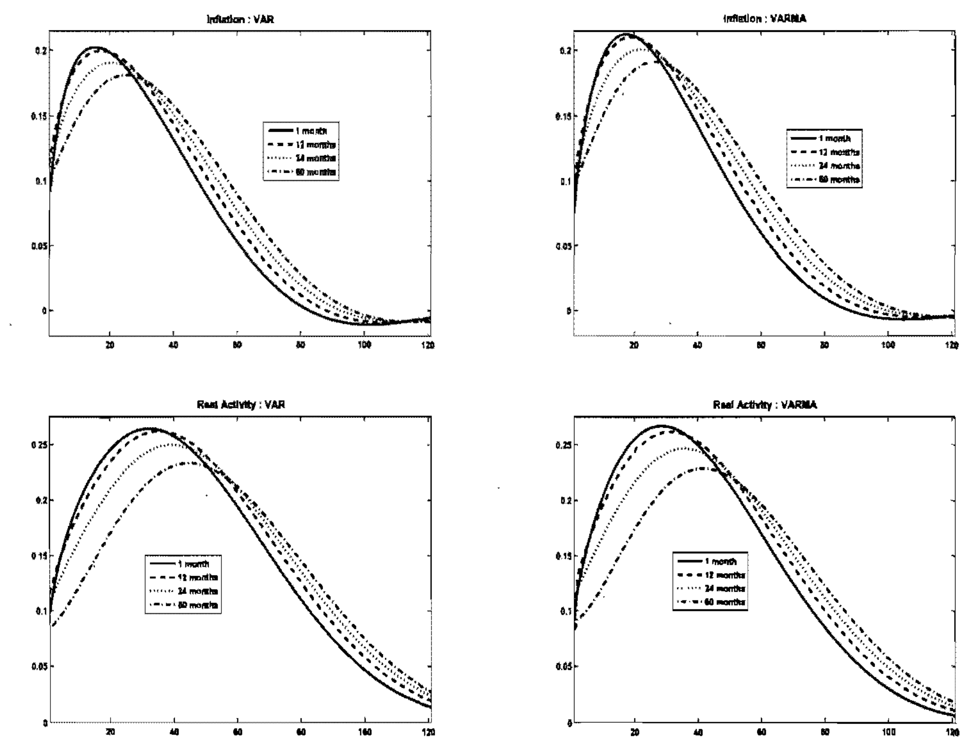


Figure 2.13: **Responses of yields from Inflation and real activity**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to inflation. Time is measured in months on the x-axis

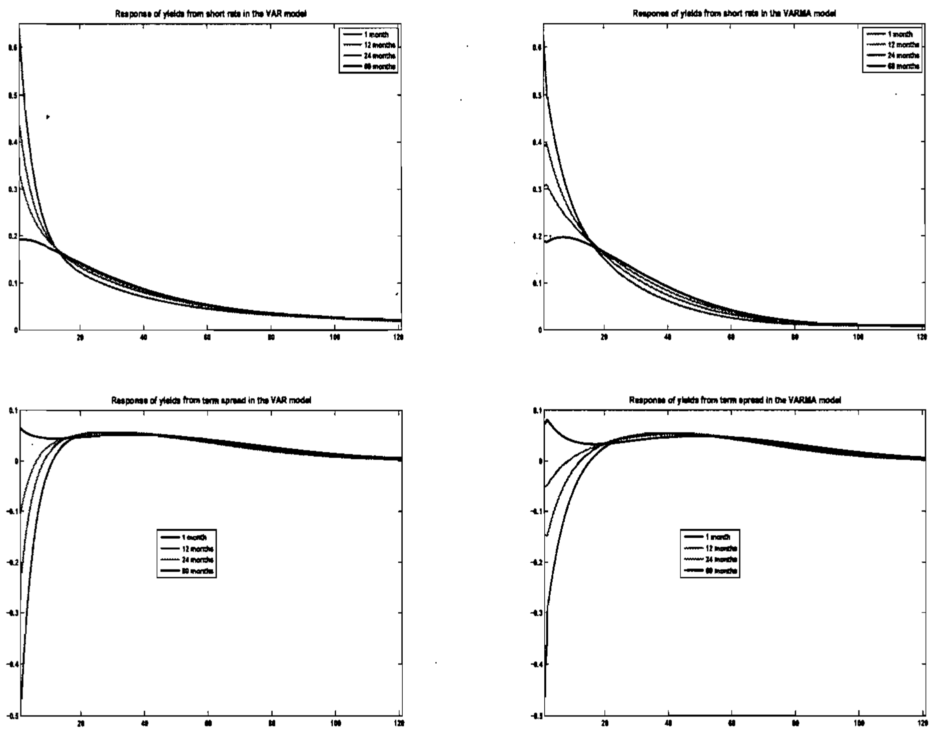


Figure 2.14: **Responses of yields from level and slope**

We fit these models to inflation, real activity, short rate and term spread. The plot shows the impulse responses to a Cholesky one standard deviation innovation to short rate. Time is measured in months on the x-axis

CHAPTER 3

OPTION VALUATION WITH CONDITIONAL HETEROSKEDASTICITY AND NON-NORMALITY

Abstract

We provide results for the valuation of European style contingent claims for a large class of specifications of the underlying asset returns. Our valuation results obtain in a discrete time, infinite state-space setup using the no-arbitrage principle and an equivalent martingale measure. Our approach allows for general forms of heteroskedasticity in returns, and valuation results for homoskedastic processes can be obtained as a special case. It also allows for conditional non-normal return innovations, which is critically important because heteroskedasticity alone does not suffice to capture the option smirk. We analyze a class of equivalent martingale measures for which the resulting risk-neutral return dynamics are from the same family of distributions as the physical return dynamics. In this case, our framework nests the valuation results obtained by Duan (1995) and Heston and Nandi (2000) by allowing for a time-varying price of risk and non-normal innovations. We provide extensions of these results to more general equivalent martingale measures and to discrete time stochastic volatility models, and we analyze the relation between our results and those obtained for continuous time models.

3.1 Introduction

A contingent claim is a security whose payoff depends upon the value of another underlying security. A valuation relationship is an expression that relates the value of the contingent claim to the value of the underlying security and other variables. The most popular approach for valuing contingent claims is the use of a Risk Neutral Valuation Relationship (RNVR).

Most of the literature on contingent claims and most of the applications of the RNVR have been cast in continuous time. While the continuous-time approach offers many advantages, the valuation of contingent claims in discrete time is also of substantial interest. For example, when hedging option positions, rebalancing decisions must be made in discrete time, and in the case of American and exotic options, early exercise decisions must be made in discrete time. However, by far the most important advantage of working in discrete time is econometric convenience. It is difficult to estimate continuous-time processes, because of the complexity of the resulting filtering problem for processes that adequately capture stylized facts, such as Heston's (1993a) stochastic volatility model. In contrast, for many of the models we study in this chapter, the resulting filtering problem is extremely simple.

Because of the econometric convenience, most of the stylized facts characterizing underlying securities have been studied in discrete-time models. One very important feature of returns is conditional heteroskedasticity, which can be addressed in the GARCH framework of Engle (1982) and Bollerslev (1986).[1] Presumably, because of this evidence, most of the recent empirical work on discrete-time option valuation has also focused on GARCH processes.[2] The GARCH model amounts to an infinite state space setup, with the innovations for underlying asset returns described by continuous distributions. In this case the market is incomplete, and it is in general not possible to construct a portfolio containing combinations of the contingent claim and the underlying asset that make the resulting portfolio riskless.[3]

To obtain a RNVR, the GARCH option valuation literature builds on the approach of Rubinstein (1976) and Brennan (1979), who demonstrate how to obtain RNVRs for lognormal and normal returns in the case of constant mean return and volatility, by specifying a representative agent economy and characterizing sufficient conditions on pref-

1. See for example French, Schwert and Stambaugh (1987) and Schwert (1989) for early studies on stock returns. The literature is far too voluminous to cite all relevant papers here. See Bollerslev, Chou and Kroner (1992) and Diebold and Lopez (1995) for reviews on GARCH modeling.

2. See Bollerslev and Mikkelsen (1996), Satchell and Timmermann (1996), Garcia and Renault (1998), Heston and Nandi (2000), Christoffersen and Jacobs (2004), Christoffersen, Heston and Jacobs (2006), and Barone-Adesi, Engle and Mancini (2008) for applications to option valuation.

3. In a discrete time finite state space setting, Harrison and Pliska (1981) provide the mathematical framework to obtain the existence of the risk neutral probability measure, to demonstrate uniqueness in the case of complete markets, and to get a RNVR for any contingent claim. See also Harrison and Kreps (1979), Cox, Ross and Rubinstein (1979) and Cox and Ross (1976) for discrete-time finite state-space approaches.

erences. For a given dynamic of the underlying security, specific assumptions have to be made on preferences in order to obtain a risk neutralization result.[4] The first order condition resulting from this economy yields an Euler equation that can be used to price any asset. For lognormal stock returns and a conditionally heteroskedastic (GARCH) volatility dynamic, the standard result is the one in Duan (1995). Duan's result relies on the existence of a representative agent with constant relative risk aversion or constant absolute risk aversion.[5]

Because it is difficult to characterize the general equilibrium setup underlying a RNVR, very few valuation results are currently available for heteroskedastic processes with non-normal innovations.[6] In this chapter, we argue that it is possible to investigate option valuation for a large class of conditionally non-normal heteroskedastic processes, provided that the conditional moment generating function (MGF) exists. It is also possible to accommodate a large class of time-varying risk premia. Our framework differs from the approach in Brennan (1979) and Duan (1995), and is more intimately related to the approach adopted in continuous-time option valuation: we only use the no-arbitrage assumption and some technical conditions on the investment strategies to show the existence of an RNVR. We demonstrate the existence of an EMM and characterize it, without first making an explicit assumption on the utility function of a representative agent. We then show that the price of the contingent claim defined as the expected value of the discounted payoff at maturity is a no-arbitrage price and characterize the risk-neutral dynamic. We provide results for GARCH processes and for more general discrete-time stochastic volatility models. We also analyze several important limit results

4. Brennan (1979) characterizes the bivariate distribution of returns on aggregate wealth and the underlying asset under which a risk-neutral valuation relationship obtains in the homoskedastic case. Camara (2003) uses this approach to obtain valuation results for transformed normal dynamics of returns and state variables. See also Schroder (2004).

5. See also Amin and Ng (1993) who study the heteroskedastic case by making an assumption on the bivariate distribution of the stochastic discount factor and the underlying return process.

6. Duan, Ritchken and Sun (2005) analyze a heteroskedastic model with Poisson-normal innovations and Duan (1999) analyzes a conditionally fat-tailed heteroskedastic model. Christoffersen, Heston and Jacobs (2006) use a heteroskedastic return dynamic with inverse Gaussian innovations. Other studies analyze non-normal innovations. Madan and Seneta (1990) use the symmetric and i.i.d. variance gamma distribution. Heston (1993b) presents results for the gamma distribution and Heston (2004) analyzes a number of infinitely divisible distributions.

for the discrete-time processes we consider, and we discuss the relationships between risk-neutralization in these models and continuous-time stochastic volatility models.

Why are we able to provide more general valuation results than the existing literature? In our opinion, the analysis in Brennan (1979) and Duan (1995) addresses two important questions simultaneously: First, a mostly technical question that characterizes the risk-neutral dynamic and the valuation of options; second, a more economic one that characterizes the equilibrium underlying the valuation procedure. The existing discrete-time literature for the most part has viewed these two questions as inextricably linked, and has therefore largely limited itself to (log)normal return processes as well as a few special non-normal cases. Our chapter differs in a subtle but important way from most existing studies. We argue that it is possible and desirable to treat these questions one at a time. We do not attempt to characterize the preferences underlying the risk-neutral valuation relationship. Instead, we assume a class of Radon-Nikodym derivatives and search for an EMM within this class. This allows us to provide some general results on the valuation of options under conditionally non-normal asset returns without fully characterizing the economic environment. We also show how the normal model and available conditional non-normal models are special cases of our setup.

The same approach of separating these two questions occurs in the literature on option valuation using continuous-time stochastic volatility models, such as for instance in Heston's (1993a) model. These models yield different equivalent martingale measures for different specifications of the volatility risk premium. For a given specification of the volatility risk premium, one can find an EMM and characterize the risk-neutral dynamic using Girsanov's theorem. To derive this result, and to value options, there is no need to explicitly characterize the utility function underlying the volatility risk premium. The latter task is very interesting in its own right, but differs from characterizing the risk-neutral dynamic and the option value for a given physical return dynamic.[7] The latter is a purely mathematical exercise. The former provides the economic background behind a particular choice of volatility premium, and therefore helps us understand whether a particular choice of functional form for the risk premium, which is often made for con-

7. See for instance Heston (1993a) and Bates (1996, 2000) for a discussion.

venience, is also reasonable from an economic perspective.

The chapter proceeds as follows. In Section 2 we define a class of heteroskedastic stock return processes, and we characterize the condition for an EMM for this class of processes. We then show sufficient conditions for an EMM to exist and we derive the risk neutral distribution of returns. In Section 3 we further discuss the choice of EMM in Section 2, and introduce a more general class of EMMs. Section 4 derives the no-arbitrage option price corresponding to the EMM. Section 5 discusses several special cases of return dynamics that can be analyzed using our approach. Section 6 provides continuous-time limits of a number of important models. Section 7 introduces an extension to discrete-time stochastic volatility models and compares it with the benchmark continuous time model. Section 8 concludes.

3.2 Conditionally heteroskedastic models

In Section 3.2.1 we define the stock price process that we use in Sections 3.2 through 3.6. This process is able to accommodate the class of ARCH and GARCH processes. In Sections 3.2.2-3.2.6, we then analyze the risk-neutralization of this stock price process using a particularly convenient candidate Radon-Nikodym derivative.

We use P to describe the physical distribution of the states of nature. The financial market consists of a zero-coupon risk-free bond index and a stock. The dynamics of the bond are described by the process $\{B_t\}_{t=0}^{T}$ normalized to $B_0 = 1$ and the dynamics of the stock price by $\{S_t\}_{t=0}^{T}$. The information structure is given by the filtration $\mathbb{F} = \{F_t | t = 0, ..., T\}$ generated by the stock and the bond process.

3.2.1 The stock price process

The underlying stock price process is assumed to follow the conditional distribution D under the physical measure P. We write

$$R_t \equiv \ln\left(\frac{S_t}{S_{t-1}}\right) = \mu_t - \gamma_t + \varepsilon_t \qquad \varepsilon_t | F_{t-1} \sim D(0, \sigma_t^2), \tag{3.1}$$

where S_t is the stock price at time t, and σ_t^2 is the conditional variance of the log return in period t. The mean correction factor, γ_t, is defined from

$$\exp(\gamma_t) \equiv E_{t-1}\left[\exp(\varepsilon_t)\right],$$

and it serves to ensure that the conditional expected gross rate of return, $E_{t-1}[S_t/S_{t-1}]$, is equal to $\exp(\mu_t)$. More explicitly,

$$\begin{aligned} E_{t-1}[S_t/S_{t-1}] &= E_{t-1}\left[\exp(\mu_t - \gamma_t + \varepsilon_t)\right] = \exp(\mu_t) \\ &\Longleftrightarrow \exp(\gamma_t) = E_{t-1}\left[\exp(\varepsilon_t)\right]. \end{aligned}$$

Note that our specification (3.1) does not restrict the risk premium in any way, nor does it assume conditional normality.

For now, we follow most of the existing discrete-time empirical finance literature by focusing on conditional means μ_t and conditional variances σ_t^2 that are F_{t-1} measurable. We will relax this assumption in Section 3.7. We do not constrain the interest rate r_t to be constant. It is instead assumed to be an element of F_{t-1} as well. This setup is able to accommodate the class of ARCH and GARCH processes proposed by Engle (1982) and Bollerslev (1986) and used for option valuation by Amin and Ng (1993), Duan (1995, 1999), and Heston and Nandi (2000). Our results also hold for different types of GARCH specifications, such as the EGARCH model of Nelson (1991) or the specification of Glosten, Jagannathan and Runkle (1993).

In the following, we show that we can find an EMM by defining a probability measure that makes the discounted security process a martingale. We derive more general results on option valuation for heteroskedastic processes compared to the available literature, because we focus on the narrow question of option valuation while ignoring the economic question regarding the preferences of the representative agent that support this valuation argument in equilibrium.

We use a no-arbitrage argument that is similar to the one used in the continuous-time literature. We first prove the existence of an EMM. Subsequently we demonstrate the existence of a RNVR by demonstrating that the price of the contingent claim, defined as the expected value of the discounted payoff at maturity, is a no-arbitrage price under this EMM. [8] The proof uses an argument similar to the one used in the continuous-time literature, but is arguably more straightforward as it avoids the technical issues involved in the analysis of local and super martingales.

3.2.2 Specifying an equivalent martingale measure

The objective in this section is to find a measure equivalent to the physical measure P that makes the price of the stock discounted by the riskless asset a martingale. An EMM is defined as long as the Radon-Nikodym derivative is defined. We start by specifying a candidate Radon-Nikodym derivative of a probability measure. We then show that this Radon-Nikodym derivative defines an EMM that makes the discounted stock price process a martingale. This result in turn allows us to obtain the distribution of the stock return under this EMM.

For a given predetermined sequence, $\{v_t\}$, we define the following candidate Radon-Nikodym derivative

$$\left.\frac{dQ}{dP}\right| F_t = \exp\left(-\sum_{i=1}^{t}\left(v_i\varepsilon_i + \Psi_i\left(v_i\right)\right)\right), \tag{3.2}$$

where $\Psi_t\left(u\right)$ is defined as the natural logarithm of the moment generating function

$$E_{t-1}\left[\exp(-u\varepsilon_t)\right] \equiv \exp\left(\Psi_t\left(u\right)\right).$$

Note that we can think of the mean correction factor in (3.1) as $\gamma_t = \Psi_t\left(-1\right)$. Note also that in the normal case we have $\Psi_t\left(u\right) = \frac{1}{2}\sigma_t^2 u^2$ and $\gamma_t = \Psi_t\left(-1\right) = \frac{1}{2}\sigma_t^2$.

We can now show the following lemma

8. Duan (1995) refers to RNVR as Local RNVR in the case of GARCH. The reason for the distinction is that (under normality) the conditional volatility is identical under the two measures only one period ahead. In the remainder of the chapter we will drop this distinction for ease of exposition. We emphasize that the result that the conditional volatility differs between the two measures for more than one period ahead is to be expected as volatility is random in this case. This feature is very similar to the continuous time case, which has random volatility for any horizon.

Lemma 1. *$\frac{dQ}{dP}\Big| F_t$ is a Radon-Nikodym derivative*

Proof. We need to show that $\frac{dQ}{dP}\Big| F_t > 0$ which is immediate. We also need to show that $E_0^P\left[\frac{dQ}{dP}\Big| F_t\right] = 1$. We have

$$E_0^P\left[\frac{dQ}{dP}\Big| F_t\right] = E_0^P\left[\exp\left(-\sum_{i=1}^{t}(\nu_i\varepsilon_i + \Psi_i(\nu_i))\right)\right].$$

Using the law of iterative expectations we can write

$$\begin{aligned}
E_0^P\left[\frac{dQ}{dP}\Big| F_t\right] &= E_0^P\left[E_1^P \ldots E_{t-1}^P \exp\left(-\sum_{i=1}^{t}(\nu_i\varepsilon_i + \Psi_i(\nu_i))\right)\right] \\
&= E_0^P\left[E_1^P \ldots E_{t-2}^P \exp\left(-\sum_{i=1}^{t-1}\nu_i\varepsilon_i - \sum_{i=1}^{t}\Psi_i(\nu_i)\right) E_{t-1}^P \exp(-\nu_t\varepsilon_t)\right] \\
&= E_0^P\left[E_1^P \ldots E_{t-2}^P \exp\left(-\sum_{i=1}^{t-1}\nu_i\varepsilon_i - \sum_{i=1}^{t}\Psi_i(\nu_i)\right) \exp(\Psi_t(\nu_t))\right] \\
&= E_0^P\left[E_1^P \ldots E_{t-2}^P \exp\left(-\sum_{i=1}^{t-1}\nu_i\varepsilon_i - \sum_{i=1}^{t-1}\Psi_i(\nu_i)\right)\right].
\end{aligned}$$

Iteratively, using this result we get

$$\begin{aligned}
E_0^P\left[\frac{dQ}{dP}\Big| F_t\right] &= E_0^P\left[\exp(-\nu_1\varepsilon_1 - \Psi_1(\nu_1))\right] \\
&= \exp(-\Psi_1(\nu_1))\exp(\Psi_1(\nu_1)) = 1,
\end{aligned}$$

and the lemma obtains. □

We are now ready to show that we can specify an EMM using this Radon-Nikodym derivative.

Proposition 3.2.1. *The probability measure Q defined by the Radon-Nikodym derivative in (3.2) is an EMM if and only if*

$$\Psi_t(\nu_t - 1) - \Psi_t(\nu_t) - \gamma_t + \phi_t\sigma_t^2 = 0, \tag{3.3}$$

where $\phi_t = \dfrac{\mu_t - r_t}{\sigma_t^2}$.

Proof. We need $E^Q\left[\dfrac{S_t}{B_t}\middle|F_{t-1}\right] = \dfrac{S_{t-1}}{B_{t-1}}$ or equivalently $E^Q\left[\dfrac{S_t}{S_{t-1}}/\dfrac{B_t}{B_{t-1}}\middle|F_{t-1}\right] = 1$. We have

$$\begin{aligned}
E^Q\left[\frac{S_t}{S_{t-1}}/\frac{B_t}{B_{t-1}}\middle|F_{t-1}\right] &= E^P\left[\left(\frac{\frac{dQ}{dP}\big|F_t}{\frac{dQ}{dP}\big|F_{t-1}}\right)\frac{S_t}{S_{t-1}}/\frac{B_t}{B_{t-1}}\middle|F_{t-1}\right] \\
&= E^P\left[\left(\frac{\frac{dQ}{dP}\big|F_t}{\frac{dQ}{dP}\big|F_{t-1}}\right)\frac{S_t}{S_{t-1}}\exp(-r_t)\middle|F_{t-1}\right] \\
&= E^P\left[\exp\left(-\nu_t\varepsilon_t - \Psi_t(\nu_t)\right)\exp(\mu_t - \gamma_t + \varepsilon_t)\exp(-r_t)|F_{t-1}\right] \\
&= \exp\left(-\Psi_t(\nu_t) + \mu_t - r_t - \gamma_t\right)E^P\left[\exp\left((1-\nu_t)\varepsilon_t\right)|F_{t-1}\right] \\
&= \exp\left(-\Psi_t(\nu_t) + \mu_t - r_t - \gamma_t + \Psi_t(\nu_t - 1)\right).
\end{aligned}$$

Thus Q is a probability measure that makes the stock discounted by a riskless asset a martingale if and only if

$$\Psi_t(\nu_t - 1) - \Psi_t(\nu_t) - \gamma_t + \phi_t\sigma_t^2 = 0. \tag{3.4}$$

This result implies that we can construct an EMM by choosing the sequence $\{\nu_t\}$ to make (3.4) hold.[9] □

3.2.3 Solving for the EMM

In this section we develop various results on the existence of a solution to (3.4), conditional on our assumption regarding the family of Radon-Nikodym derivatives.

Note first that in the conditional normal special case we get the solution to be the well-known price of risk $\nu_t = \phi_t = (\mu_t - r_t)/\sigma_t^2$. Note also that if we additionally specify the conditional mean of the excess return to be affine in σ_t^2, so that $\mu_t = r_t + \lambda\sigma_t^2$, then ν_t is simply a constant λ.

9. See Shiryaev (1999) for an introduction to the conditional use of the Radon-Nikodym derivative.

When allowing for conditional non-normal returns, we need to put some structure on $\Psi_t(.)$ in order to analyze the existence of a solution to (3.4). In Section 3.5 below we consider some important non-normal special cases where an explicit solution for v_t can be found. More generally, we provide the following result.

Proposition 3.2.2. *If Ψ is strictly convex, twice differentiable, and tends to infinity at the boundaries of its domain (u_1, u_2) where $u_1 + 1 < u_2$, then there exists a solution to equation (3.4). This solution is unique. Note that u_1 and u_2 are not restricted to be finite.*

Proof. See the Appendix. □

Proposition 3.2.2 provides a set of sufficient, not necessary, conditions for a unique solution to exist within the class of Radon-Nikodym derivatives defined by (3.2). A similar result can be obtained assuming that Ψ is strictly concave. However, the parametric examples we consider below are part of the class of infinitely divisible distributions, thus ensuring that strict convexity holds (Feller, 1968), and therefore the strict convexity assumption in Proposition 2 is more realistic for our purposes. Gourieroux and Monfort (2007) provide similar conditions in a setup with a stochastic discount factor. They do not relate their result to the class of inifinitely divisible distributions. In Section 3.5 below, we discuss the other conditions in Proposition 2 on a case-by-case basis, and thus verify that overall these conditions are very reasonable.

In the absence of sufficient conditions, an approximate solution to the EMM equation in (3.4) can be obtained from the second-order approximations

$$\begin{aligned}\Psi_t(v_t - 1) &\approx \Psi_t(0) + \Psi_t'(0)(v_t - 1) + \tfrac{1}{2}\Psi_t''(0)(v_t - 1)^2 \\ \Psi_t(v_t) &\approx \Psi_t(0) + \Psi_t'(0)\,v_t + \tfrac{1}{2}\Psi_t''(0)\,v_t^2.\end{aligned}$$

From the definition of the mean-zero shock ε_t we have that $\Psi_t'(0) = E_{t-1}[\varepsilon_t] = 0$, and $\Psi_t''(0) = Var_{t-1}[\varepsilon_t] = \sigma_t^2$, so that the approximation along with the EMM condition (3.4) gives us

$$v_t \approx \frac{\mu_t - r_t}{\sigma_t^2} + \frac{1}{2} - \frac{\gamma_t}{\sigma_t^2}. \tag{3.5}$$

Notice that this approximation is exact in the normal case, where $\gamma_t = \frac{1}{2}\sigma_t^2$ and $v_t = (\mu_t - r_t)/\sigma_t^2$. This approximate solution can be used in place of the exact solution, or it can be used as a starting value in a numerical search for the exact v_t.

Note finally that (3.4) suggests that the problem of finding a solution for v_t can be circumvented altogether if one is willing to put more structure on the return process in (3.1). If the conditional return mean is specified as follows

$$\mu_t = r_t + \Psi_t(v_t) - \Psi_t(v_t - 1) + \gamma_t, \tag{3.6}$$

then the EMM condition in (3.4) is trivially satisfied for any value of v_t. Thus v_t can be set to a constant v, to be estimated as part of the return dynamic. This approach is viable but suffers from the drawback that the return mean dynamic is now motivated by convenience rather than empirical relevance. Note that in the normal case this approach yields

$$\begin{aligned} R_t &= r_t + \Psi_t(v) - \Psi_t(v-1) + \varepsilon_t \qquad\qquad \varepsilon_t|F_{t-1} \sim N(0,\sigma_t^2) \\ &= r_t + v\sigma_t^2 - \tfrac{1}{2}\sigma_t^2 + \varepsilon_t, \end{aligned}$$

which corresponds to an affine risk premium.

We emphasize that the uniqueness result in Proposition 3.2.2 and the solution strategies in (3.5) and (3.6) are conditional on the assumption on the Radon-Nikodym derivative in (3.2) and are therefore not fully general. In Section 3.3 we present a more general result, but we are not able to completely characterize the class of all possible RN derivatives.

3.2.4 Characterizing the risk-neutral distribution

When pricing options using Monte Carlo simulation, knowing the risk neutral distribution is valuable. In this section, we derive an important result that shows that for the class of models we investigate and using the class of Radon-Nikodym derivatives in (3.2), the risk neutral distribution is from the same family as the original physical distribution.

We first need the following lemma where we recall that $\Psi_t(u)$ denotes the one-day log conditional moment generating function

Lemma 2.

$$E_{t-1}^{Q}\left[\exp\left(-u\varepsilon_t\right)\right] = \exp\left(\Psi_t\left(v_t+u\right) - \Psi_t\left(v_t\right)\right).$$

Proof.

$$\begin{aligned} E_{t-1}^{Q}\left[\exp\left(-u\varepsilon_t\right)\right] &= E^{P}\left[\left(\frac{\frac{dQ}{dP}|F_t}{\frac{dQ}{dP}|F_{t-1}}\right)\exp(-u\varepsilon_t)|F_{t-1}\right] \\ &= E^{P}\left[\exp\left(-v_t\varepsilon_t - \Psi_t\left(v_t\right)\right)\exp(-u\varepsilon_t)|F_{t-1}\right] \\ &= \exp\left(\Psi_t\left(v_t+u\right) - \Psi_t\left(v_t\right)\right). \end{aligned}$$

□

From this lemma, if we define $\Psi_t^Q(u)$ to be the log conditional moment generating function under the risk neutral probability measure, then we have

$$\Psi_t^Q(u) = \Psi_t\left(v_t+u\right) - \Psi_t\left(v_t\right). \tag{3.7}$$

While other candidate risk-neutral log MGFs are available, corresponding to other choices of Radon-Nikodym derivatives, this particular specification is extremely convenient because for many physical innovation distributions, it provides a tractable risk-neutral distribution, building on the work of Esscher (1932).[10] From this we can derive

$$E_{t-1}^{Q}\left[\varepsilon_t\right] = \left.\frac{\partial\Psi_t^Q(-u)}{\partial u}\right|_{u=0} = -\Psi_t'\left(v_t\right).$$

Define the risk neutral innovation

$$\varepsilon_t^* \equiv \varepsilon_t - E_{t-1}^{Q}\left[\varepsilon_t\right] = \varepsilon_t + \Psi_t'\left(v_t\right). \tag{3.8}$$

10. For applications of the Esscher transform in option valuation, see Buhlmann, Delbaen, Embrechts and Shiryaev (1996, 1998), Gerber and Shiu (1994), and Siu, Tong and Yang (2004). See Dai and Singleton (2006) for an application to term structure models.

The risk-neutral log conditional moment generating function of ε_t^*, labeled $\Psi_t^{Q*}(u)$, is then

$$\Psi_t^{Q*}(u) = -u\Psi_t'(\nu_t) + \Psi_t^{Q}(u). \tag{3.9}$$

We are now ready to show the following

Proposition 3.2.3. *If the physical conditional distribution of ε_t is an infinitely divisible distribution with finite second moment, then the risk-neutral conditional distribution of ε_t^* is also an infinitely divisible distribution with finite second moment.*

Proof. See the Appendix. □

In the special case of the normal distribution we get simply

$$\varepsilon_t^* = \varepsilon_t + \Psi_t'(\nu_t) = \varepsilon_t + \mu_t - r_t,$$

and $\Psi_t^{Q*}(u) = \frac{1}{2}\sigma_t^2 u^2$ so that the risk-neutral innovations are normal and correspond to the physical innovations shifted by the equity risk premium. In the more general case, the relationship between physical and risk-neutral innovations is not necessarily this simple.

Because of the one-to-one mapping between moment generating functions and distribution functions, the proposition can be used to derive specific parametric risk-neutral distributions consistent with the parametric physical distributions assumed by the researcher.

3.2.5 Characterizing the risk-neutral conditional variance

The conditional risk-neutral variance, σ_t^{*2}, is of particular interest in the dynamic heteroskedastic models we consider. It can be obtained by taking the second derivative of the risk neutral log conditional moment generating function $\Psi_t^{Q*}(u)$ and evaluating it at $u = 0$. Using equations (3.9) and (3.7) we get

$$\sigma_t^{*2} = \left.\frac{\partial^2 \Psi_t^{Q*}(-u)}{\partial u^2}\right|_{u=0} = \Psi_t''(\nu_t).$$

Recall that by definition, the conditional variance under the physical measure is $\sigma_t^2 = \Psi_t''(0)$. Thus in general we have the following relationship between the (one day ahead) conditional variances under the two measures

$$\sigma_t^{*2} = \sigma_t^2 \frac{\Psi_t''(\nu_t)}{\Psi_t''(0)}.$$

For conditionally normal returns, we have $\Psi_t(u) = \frac{1}{2}\sigma_t^2 u^2$ and $\nu_t = (\mu_t - r_t)/\sigma_t^2$, so that $\Psi_t''(\nu_t) = \Psi_t''(0)$ and thus $\sigma_t^{*2} = \sigma_t^2$, but this will not generally be the case for non-normal distributions. Non-normality drives an additional wedge between the physical and risk-neutral conditional variances. Interestingly, this phenomenon is often observed empirically, as physical volatility measures from historical returns are systematically lower than risk-neutral volatilities implied from options. See for example Carr and Wu (2007).

We can use our results to provide some more insight into this wedge between one day ahead physical and risk-neutral conditional variances. Consider the following approximation to the risk-neutral variance

$$\sigma_t^{*2} = \Psi_t''(\nu_t) \approx \Psi_t''(0) + \Psi_t'''(0)\nu_t + \frac{\Psi_t''''(0)}{2}\nu_t^2.$$

Denoting conditional skewness by $skew_t$ and conditional excess kurtosis by $kurt_t$, we have $\Psi_t'''(0) = -skew_t\sigma_t^3$ and $\Psi_t''''(0) = kurt_t\sigma_t^4$. Therefore

$$\sigma_t^{*2} \approx \sigma_t^2 - skew_t\sigma_t^3\nu_t + \frac{kurt_t}{2}\sigma_t^4\nu_t^2. \tag{3.10}$$

From (3.5), ν_t can be thought of as a modified Sharpe ratio, and will generally be positive. Therefore, from (3.10), the risk neutral variance will always be larger than the historical variance if conditional skewness is negative and/or excess kurtosis is positive.

Furthermore, we can characterize the risk-neutral conditional variance dynamic. As an example, start from the simple GARCH(1,1) dynamic of Bollerslev (1986) for the physical conditional variance

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-1}^2 \varepsilon_{t-1}^2, \tag{3.11}$$

which can be shown to lead to the risk-neutral variance dynamic

$$\sigma_t^{*2} = \beta_{0,t} + \beta_{1,t} \sigma_{t-1}^{*2} + \beta_{2,t} \left(\varepsilon_{t-1}^* - \Psi'(\nu_{t-1}) \right)^2,$$

where

$$\beta_{0,t} = \beta_0 \frac{\Psi_t''(\nu_t)}{\Psi_t''(0)}, \beta_{1,t} = \beta_1 \frac{\Psi_t''(\nu_t)}{\Psi_t''(0)} \frac{\Psi_{t-1}''(0)}{\Psi_{t-1}''(\nu_{t-1})}, \beta_{2,t} = \beta_2 \frac{\Psi_t''(\nu_t)}{\Psi_t''(0)}.$$

Under normality $\beta_{0,t} = \beta_0$, $\beta_{1,t} = \beta_1$, and $\beta_{2,t} = \beta_2$, and therefore

$$\sigma_t^{*2} = \beta_0 + \beta_1 \sigma_{t-1}^{*2} + \beta_2 \left(\varepsilon_{t-1}^* - \Psi'(\nu_{t-1}) \right)^2. \tag{3.12}$$

Taking into account that under normality we also have $\sigma_t^{*2} = \sigma_t^2$, this can be re-written as

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \beta_2 \left(\varepsilon_{t-1}^* - \Psi'(\nu_{t-1}) \right)^2. \tag{3.13}$$

Note that (3.13) can also be derived by using the expression for the risk-neutral innovation (3.8) in (3.11). This derivation does not depend on normality. Therefore, (3.13) holds in general but it is only under normality that the risk-neutral variance (3.12) follows the same dynamic with the same coefficients, which is consistent with the finding that $\sigma_t^{*2} = \sigma_t^2$ for conditionally normal returns. We will discuss the implications of conditionally non-normal returns further below, and give explicit examples of non-normal distributions that generate the interesting and important empirical feature that physical and risk-neutral one day ahead conditional variances differ.

3.2.6 Characterizing Risk-Neutral Conditional Skewness

We can also derive a useful result on risk-neutral skewness. Using

$$\Psi_t'''(0) = -skew_t \sigma_t^3 \text{ and } \Psi_t'''(\nu_t) = -skew_t^* \sigma_t^{*3},$$

as well as

$$\Psi_t'''(\nu_t) \approx \Psi_t'''(0) + \Psi_t''''(0)\,\nu_t \text{ and } \Psi_t''''(0) = kurt_t\sigma_t^4,$$

we get that

$$\begin{aligned} -skew_t^*\sigma_t^{*3} &\approx -skew_t\sigma_t^3 + kurt_t\sigma_t^4\nu_t \\ skew_t^* &\approx skew_t\left(\frac{\sigma_t}{\sigma_t^*}\right)^3 - kurt_t\frac{\sigma_t^4\nu_t}{\sigma_t^{*3}}. \end{aligned}$$

Note that for the empirically relevant case where $\sigma_t \leq \sigma_t^*$, we have $skew_t\left(\frac{\sigma_t}{\sigma_t^*}\right)^3 \leq skew_t$. Therefore $skew_t^* \leq skew_t$ for the empirically relevant case where the price of risk $\nu_t \geq 0$ and $kurt_t \geq 0$.

3.3 Generalized EMMs and Option Price Bounds

While the one-shock stock price processes in Section 3.2.1, and the GARCH processes nested in it, imply an incomplete-markets setup, we do obtain a unique price conditional on the choice of Radon-Nikodym derivative. Clearly therefore there have to be other valid prices corresponding to other choices of Radon-Nikodym derivative. We now characterize EMMs corresponding to other classes of Radon-Nikodym derivatives.

3.3.1 Generalized EMMs in GARCH models

We use the dynamic of the stock price process under the physical measure in (3.1), with $\Psi_t(u)$ the natural logarithm of the moment generating function. In order to allow for as much generality as possible while still staying in our framework, we define a class of Radon-Nikodym derivatives defined by a general log-MGF under Q, call it $\Omega_t(u)$. We then show which restrictions need to be placed on $\Omega_t(u)$ in order for it to result in a proper EMM.

First, define the following candidate Radon-Nikodym derivative for a given predetermined sequence of log moment generating functions $\{\Omega_t(u)\}$, which is F_{t-1} adapted,

$$\left.\frac{dQ}{dP}\right| F_t = \prod_{j=1}^{t} \frac{\int_{-\infty}^{+\infty} \exp\left(-iu\varepsilon_j + \Omega_j(-iu)\right) du}{\int_{-\infty}^{+\infty} \exp\left(-iu\varepsilon_j + \Psi_j(-iu)\right) du}. \tag{3.14}$$

Lemma 3. *$\left.\frac{dQ}{dP}\right| F_t$ is a Radon-Nikodym derivative*

Proof. We need to show that $\left.\frac{dQ}{dP}\right| F_t > 0$. For each j, $\exp\left(\Omega_j(-iu)\right)$ is a characteristic function which is absolutely integrable over (-∞, +∞). Using the inversion formula (Lukacs (1970, p. 33)), $q_j(\varepsilon_j) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-iu\varepsilon_j + \Omega_j(-iu)\right) du$ is the corresponding density function. Similarly $p_j(\varepsilon_j) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-iu\varepsilon_j + \Psi_j(-iu)\right) du$ is a density function. Therefore

$$\left.\frac{dQ}{dP}\right| F_t = \prod_{j=1}^{t} \frac{q_j(\varepsilon_j)}{p_j(\varepsilon_j)}.$$

We have $\left.\frac{dQ}{dP}\right| F_t > 0$ because density functions are always positive. We also need to show $E_0^P\left[\left.\frac{dQ}{dP}\right| F_t\right] = 1$. We have

$$E_0^P\left[\left.\frac{dQ}{dP}\right| F_t\right] = E_0^P\left[\prod_{j=1}^{t} \frac{q_j(\varepsilon_j)}{p_j(\varepsilon_j)}\right].$$

Using the law of iterated expectations we have

$$\begin{aligned} E_0^P\left[\left.\frac{dQ}{dP}\right| F_t\right] &= E_0^P\left[E_1^P \ldots E_{t-1}^P \prod_{j=1}^{t} \frac{q_j(\varepsilon_j)}{p_j(\varepsilon_j)}\right] \\ &= E_0^P\left[E_1^P \ldots E_{t-2}^P \prod_{j=1}^{t-1} \frac{q_j(\varepsilon_j)}{p_j(\varepsilon_j)} E_{t-1}^P \frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)}\right]. \end{aligned}$$

Note

$$E_{t-1}^P \frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)} = \int \frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)} p_t(\varepsilon_t)\, d\varepsilon_t.$$

Therefore $E^P_{t-1}\frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)} = \int q_t(\varepsilon_t)\,d\varepsilon_t = 1$ and

$$E_0^P\left[\frac{dQ}{dP}\middle|F_t\right] = E_0^P\left[E_1^P \ldots E_{t-2}^P \prod_{j=1}^{t-1}\frac{q_j(\varepsilon_j)}{p_j(\varepsilon_j)}\right].$$

Iteratively using this result we get

$$E_0^P\left[\frac{dQ}{dP}\middle|F_t\right] = E_0^P\left[\frac{q_1(\varepsilon_1)}{p_1(\varepsilon_1)}\right] = 1,$$

and the lemma obtains. □

We are now ready to show the restriction required on $\Omega_t(u)$ so that we can specify an EMM using this Radon-Nikodym derivative.

Proposition 3.3.1. *The probability measure Q defined by the Radon-Nikodym derivative in (3.14) is an EMM if and only if*

$$\Omega_t(-1) - \gamma_t + \phi_t\sigma_t^2 = 0, \tag{3.15}$$

where $\phi_t = \dfrac{\mu_t - r_t}{\sigma_t^2}$.

Proof. We need $E^Q\left[\frac{S_t}{B_t}\middle|F_{t-1}\right] = \frac{S_{t-1}}{B_{t-1}}$ or equivalently $E^Q\left[\frac{S_t}{S_{t-1}}/\frac{B_t}{B_{t-1}}\middle|F_{t-1}\right] = 1$. We have

$$E^Q\left[\frac{S_t}{S_{t-1}}/\frac{B_t}{B_{t-1}}\middle|F_{t-1}\right] = E^P\left[\left(\frac{\frac{dQ}{dP}\big|F_t}{\frac{dQ}{dP}\big|F_{t-1}}\right)\frac{S_t}{S_{t-1}}/\frac{B_t}{B_{t-1}}\middle|F_{t-1}\right].$$

$$
\begin{aligned}
&= E^P\left[\left(\frac{\frac{dQ}{dP}\Big|F_t}{\frac{dQ}{dP}\Big|F_{t-1}}\right)\frac{S_t}{S_{t-1}}\exp(-r_t)\Bigg|F_{t-1}\right] \\
&= E^P\left[\frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)}\exp(\mu_t-\gamma_t+\varepsilon_t)\exp(-r_t)\Bigg|F_{t-1}\right] \\
&= \exp(\mu_t-r_t-\gamma_t)E^P\left[\exp(\varepsilon_t)\frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)}|F_{t-1}\right] \\
&= \exp(\mu_t-r_t-\gamma_t)\int\exp(\varepsilon_t)\frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)}p_t(\varepsilon_t)\,d\varepsilon_t \\
&= \exp(\mu_t-r_t-\gamma_t)\int\exp(\varepsilon_t)\,q_t(\varepsilon_t)\,d\varepsilon_t \\
&= \exp(\mu_t-r_t-\gamma_t+\Omega_t(-1)),
\end{aligned}
$$

since by definition $\Omega_t(u)$ is the log-MGF which corresponds to the density $q_t(\varepsilon_t)$. By taking logs the lemma obtains. □

This result shows that a Radon-Nikodym derivative can be defined such that any log-MGF $\Omega_t(u)$ satisfying equation (3.15) will provide a suitable EMM. The result implies that a wide class of EMMs are possible.

Note that while (3.15) characterizes a more general class of EMMs compared with the result in (3.3), it is still conditional on the choice of Radon-Nikodym derivative in (3.14). We are not able to completely characterize the class of potential Radon-Nikodym derivatives for the general class of distributions considered in this chapter.

3.3.2 Nesting the Linear EMM

We now demonstrate how the class of Radon-Nikodym derivatives in Section 3.2.2, which is linear in the stock return innovation, is nested in the class of Radon-Nikodym derivatives discussed above. For a given sequence, $\{\nu_t\}$, we restrict the function $\Omega_t(u)$ in (3.14) as follows

$$\Omega_t(u) = \Psi_t(u+\nu_t) - \Psi_t(\nu_t). \tag{3.16}$$

Note that this particular risk-neutral log MGF $\Omega_t(u)$ corresponds to the $\Psi_t^Q(u)$ defined in (3.7). The condition (3.15) becomes

$$\Psi_t(\nu_t - 1) - \Psi_t(\nu_t) - \gamma_t + \phi_t \sigma_t^2 = 0, \tag{3.17}$$

which is equal to (3.3). Substituting (3.16) in (3.14) gives

$$\begin{aligned}
\frac{\frac{dQ}{dP}|F_t}{\frac{dQ}{dP}|F_{t-1}} &= \frac{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu + \nu_t) - \Psi_t(\nu_t))\,du}{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu))\,du} \\
&= \exp(-\Psi_t(\nu_t)) \frac{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu + \nu_t))\,du}{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu))\,du} \\
&= \exp(-\nu_t\varepsilon_t - \Psi_t(\nu_t)) \frac{\int_{-\infty}^{+\infty} \exp(-i(u + i\nu_t)\varepsilon_t + \Psi_t(-i(u + i\nu_t)))\,du}{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu))\,du} \\
&= \exp(-\nu_t\varepsilon_t - \Psi_t(\nu_t)) \frac{\int_{-\infty}^{+\infty} \exp(-iu^*\varepsilon_t + \Psi_t(-iu^*))\,du^*}{\int_{-\infty}^{+\infty} \exp(-iu\varepsilon_t + \Psi_t(-iu))\,du} \\
&= \exp(-\nu_t\varepsilon_t - \Psi_t(\nu_t)),
\end{aligned}$$

where we have used the fact that $i^2 = -1$, as well as a change of measure, $u^* = u + i\nu_t$. Note that this result corresponds exactly to the assumption on the Radon-Nikodym derivative in (3.2).

We have thus demonstrated how the class of Radon-Nikodym derivatives in (3.2) obtains as a special case of the general characterization of the class of Radon-Nikodym derivatives in (3.14). In Section 3.2.4 above, and below in Section 3.5, we demonstrate that this special case is of great interest because it allows us to characterize the risk-neutral dynamics in closed form for a large class of return innovations. Such characterizations are as a rule not possible with the more general class of Radon-Nikodym derivatives. However, given that Radon-Nikodym derivatives typically used in empirical work are of the form in (3.2), and that the resulting risk-neutralizations have some empirical shortcomings, it may be of interest to analyze richer specifications of the Radon-Nikodym derivative.

3.3.3 A Quadratic EMM Under Conditional Normality

We now analyze a somewhat more general case that still allows for some analytical results. Specifically, we analyze the case of a quadratic rather than linear EMM, but we restrict ourselves to normally distributed innovations.

For a given sequence $\{\nu_{1,t}, \nu_{2,t}\}$, consider the following candidate Radon-Nikodym derivative

$$\left.\frac{dQ}{dP}\right| F_t = \exp\left(-\sum_{i=1}^{t} \left(\nu_{1,i}\varepsilon_i + \nu_{2,i}\varepsilon_i^2 + g\left(\nu_{1,i}, \nu_{2,i}, \sigma_i^2\right)\right)\right). \tag{3.18}$$

By solving the EMM equation, $E^Q\left[\frac{S_t}{S_{t-1}} / \frac{B_t}{B_{t-1}} \middle| F_{t-1}\right] = 1$, we can show that the probability measure Q defined by the Radon-Nikodym derivative in (II.3) is an EMM if and only if

$$g\left(\nu_{1,i}, \nu_{2,i}, \sigma_i^2\right) = \frac{1}{2}\left(\nu_{1,i}^2\sigma_i^{*2} - \ln\left(\sigma_i^2/\sigma_i^{*2}\right)\right), \text{ where} \tag{3.19}$$

$$\sigma_i^{*2} = Var_{i-1}^Q\left(\varepsilon_i\right) = \frac{\sigma_i^2}{1+2\nu_{2,i}\sigma_i^2}, \text{ and} \tag{3.20}$$

$$\nu_{1,i} = \left[\frac{\mu_i}{\sigma_i^2} - \frac{r_i}{\sigma_i^{*2}}\right] + 2\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\nu_{2,i}. \tag{3.21}$$

An interesting feature of this EMM is that we get a wedge between the physical and risk-neutral variance–an empirically observed fact–even when assuming conditional normality of returns. In this case the wedge is driven by the quadratic term, $\nu_{2,t}$, in the pricing kernel. Recall that in Section 3.2.5 above a wedge was created by non-normality in the conditional return distribution.

Note that we have two EMM parameters, $\nu_{1,i}$ and $\nu_{2,i}$, but only one equation defining $\nu_{1,i}$ as a function of $\nu_{2,i}$. In order to complete the model we could impose that the proportional wedge between σ_t^{*2} and σ_t^2 is constant. If we for example set $\sigma_t^2/\sigma_t^{*2} = \pi_\sigma$, we get $\nu_{2,t} = \frac{1}{2}\left(\pi_\sigma - 1\right)/\sigma_t^2$.

Next we consider how this quadratic case fits into our general setup discussed in Section 3.3.1. Since we are working with normal innovations, we can use the inversion formula to write

$$\begin{aligned}\frac{\frac{dQ}{dP}|F_t}{\frac{dQ}{dP}|F_{t-1}} &= \frac{q_t(\varepsilon_t)}{p_t(\varepsilon_t)} = \frac{\frac{1}{\sigma_t^*\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(\varepsilon_t+\delta_t^*)^2}{\sigma_t^{*2}}\right)}{\frac{1}{\sigma_t\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}\right)} = \exp\left(-\frac{1}{2}\frac{(\varepsilon_t+\delta_t^*)^2}{\sigma_t^{*2}} + \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2} + \ln\left(\frac{\sigma_t}{\sigma_t^*}\right)\right) \\ &= \exp\left(-v_{1,t}\varepsilon_t - v_{2,t}\varepsilon_t^2 + \ln\left(\frac{\sigma_t}{\sigma_t^*}\right) - \frac{\delta_t^{*2}}{2\sigma_t^{*2}}\right),\end{aligned}$$

where δ_t^* is the risk-neutral mean of ε_t and where

$$v_{1,t} = \frac{\delta_t^*}{\sigma_t^{*2}}, \text{ and } v_{2,t} = \frac{1}{2}\left(\frac{1}{\sigma_t^{*2}} - \frac{1}{\sigma_t^2}\right).$$

From normality we have that $\Omega_t(-1) = \frac{1}{2}\sigma_t^{*2} - \delta_t^{*2}$ and from the EMM condition in (3.15) we have that $\Omega_t(-1) = \mu_t - r_t - \frac{1}{2}\sigma_t^2$. These equations provide an expression for the risk-neutral mean of ε_t in the quadratic model

$$\delta_t^* = \mu_t - r_t + \frac{1}{2}\left(\sigma_t^{*2} - \sigma_t^2\right). \tag{3.22}$$

Using this equation for δ_t^* and the equation for $v_{2,t}$ in the equation for $v_{1,t}$ yields (*II*.6).

We have thus shown how in the normal case the quadratic EMM in (II.3) is a special case of the general class of EMMs defined by (3.14). Note also that by setting $v_{2,t} = 0$, we obtain the affine EMM as a special case.

3.3.4 Market Incompleteness and Bounds on Option Prices

Market incompleteness results in a wide range of available Radon-Nikodym derivatives and thus multiple EMMs and option prices. In order to illustrate this incompleteness consider Figure 1. We use the linear and quadratic EMMs to compute the price of a one-month-to-maturity, at-the-money call option with an underlying asset price of 100. We assume a risk-free rate of 5%, an underlying mean asset return of 10% and a physical asset volatility of 20% per year. In the quadratic EMM we let the ratio of the physical to risk-neutral variance, $\sigma^2/\sigma^{*2} = \pi_\sigma$ vary from 0.5 to 1. Figure 1 shows how the option price from the quadratic EMM depends critically on π_σ and thus v_2 in (II.3). The horizontal line shows the option price from the linear EMM where $\pi_\sigma = 1$ and $v_2 = 0$. Figure 1 shows that the range of option prices can be wide even when staying within the quadratic class of EMMs. This illustrates the potential of non-linear EMMs to explain outstanding empirical puzzles such as the high prices of deep out-of-the-money index put options.

The literature on option pricing bounds provides ways to quantify the degree of market incompleteness. Key early papers in this literature include Perrakis and Ryan (1984), Levy (1985), and Ritchken (1985) who all applied single-period models. Perrakis (1986) and Ritchken and Kuo (1988) extended this work to a multi-period setting, and Constantinides, Jackwerth and Perrakis (2009) contain a recent application to S&P500 index options. These papers proceed by considering a portfolio of an option, an underlying asset and a risk-free bond and derive bounds on the option price without assuming a particular EMM but instead relying only on the principle of stochastic dominance. The bounds are defined so that observing an option price outside the bounds would induce a stochastically dominating trading strategy.

While the work in this literature has evolved to allow for trading costs and other frictions (see Constantinides and Perrakis, 2002, 2007) until recently the results were developed in an i.i.d. setting, thus ruling out the GARCH effects considered in this chapter. However, current work by Oancea and Perrakis (2007) extends the stochastic dominance approach to derive intervals of admissible option prices using bounds allowing for GARCH effects. In contrast with the i.i.d. case, in the GARCH case it is necessary to assume that the representative investor has constant relative risk aversion.

The recent so-called good-deal bounds approach of Cochrane and Saa-Requejo (2000) presents another interesting venue for generating option pricing bounds.[11] Good-deal bounds are derived using a distance measure between a given stochastic discount factor (SDF) and a benchmark SDF. This approach has been adapted to option pricing under continuous-time stochastic volatility by Bondarenko and Longarela (2004). We can show that it is possible in the discrete GARCH framework to derive good-deal bounds

11. See Bjork and Slinko (2006) for a generalization, and Bernardo and Ledoit (2000) for a related approach.

on option prices when using a quadratic EMM. [12]

3.4 The valuation of European style contingent claims

In a general return model with time-varying conditional mean and volatility and non-normal shocks, we have characterized conditions under which there exists an EMM Q that makes the stock discounted by the riskless asset a martingale.

We now turn our attention to the pricing of European style contingent claims. Existing papers on the pricing of contingent claims in a discrete-time infinite state space setup, such as the literature on GARCH option pricing in Duan (1995), Amin and Ng (1993) and Heston and Nandi (2000) value such contingent claims by making an assumption on the bivariate distribution of the stock return and the endowment, or an equivalent assumption. While this approach, which most often amounts to the characterization of the equilibrium that supports the pricing, is an elegant way to deal with the incompleteness that characterizes these markets, we argue that it is not strictly necessary to characterize the equilibrium. Instead, we adopt an approach which is more prevalent in the continuous-time literature, and proceed to pricing derivatives using a no-arbitrage argument alone.

To understand our approach, the analogy with option valuation for the stochastic volatility model of Heston (1993a) is particularly helpful. In this incomplete markets setting, an infinity of no-arbitrage contingent claims prices exist, one for every different specification of the price of risk. When one fixes the price of volatility risk, however, there is a unique no-arbitrage price. For the purpose of option valuation, one can simply pick a price of volatility risk, and the resulting valuation exercise is purely mechanical.

The question whether a particular price of risk is reasonable is of substantial interest in its own right, and an analysis of the representative agent utility function that support a particular price of risk is very valuable. However, this question can be analyzed separately from the option valuation problem. For the heteroskedastic discrete-time models we consider, a similar remark applies. The link between our approach and the utility-

12. This result is available in appendix II.

based approach in Brennan (1979), Rubinstein (1976) and Duan (1995) is that assumptions on the utility function are implicit in the specification of the risk premium in the return dynamic in our case. [13] The representative agent preferences underlying this assumption are of interest, but it is not necessary to analyze them in order to value options.

We have already found an EMM Q. We therefore want to demonstrate that the price at time t is defined as

$$C_t = E^Q\left[\frac{C_T(S_T)}{B_T}B_t \middle| F_t\right].$$

The proof proceeds in a number of steps and requires defining a number of concepts that are well-known in the literature. Fortunately, even though our methodology closely follows the continuous-time case, we economize on the number of technical conditions in the continuous-time setup, such as admissibility, and avoid the concepts of local martingale and super martingale. The reason is that the integration over an infinite number of trading times in the continuous-time case is replaced by a finite sum over the trading days in discrete time.

3.4.0.0.1 Definitions

1. We denote by η_t, δ_t and ψ_t the units of the stock, the contingent claim and the bond held at date t. We refer to the F_t predictable processes η_t, δ_t and ψ_t as investment strategies.
2. The value process
$$V_t = \eta_t S_t + \delta_t C_t + \psi_t B_t,$$
describes the total dollar amount available for investments at date t.
3. The gain process
$$G_t = \sum_{i=0}^{t-1} \eta_i(S_{i+1} - S_i) + \sum_{i=0}^{t-1} \delta_i(C_{i+1} - C_i) + \sum_{i=0}^{t-1} \psi_i(B_{i+1} - B_i),$$
captures the total financial gains between dates 0 and t.

13. See Bick (1990) and He and Leland (1993) for a discussion of assumptions on the utility function implicit in the specification of the return dynamic for the market portfolio. We proceed along the lines of Jacod and Shiryaev (1998), and Shiryaev (1999).

4. We call the process $\{\eta_t, \delta_t, \psi_t\}_{t=0}^{T-1}$ a self financing strategy if and only if $V_t = G_t$ $\forall t = 1, ..., T$.
5. The definition of an arbitrage opportunity is standard: we have an arbitrage opportunity if a self financing strategy exists with either $V_0 < 0$, $V_T \geq 0$ a.s. or $V_0 \leq 0$, $V_T > 0$ a.s.
6. We denote the discounted stock price at time t as $S_t^B = \frac{S_t}{B_t}$ and the discounted contingent claim as $C_t^B = \frac{C_t}{B_t}$. Similarly, the discounted value process is denoted $V_t^B = \frac{V_t}{B_t}$ and the discounted gain process $G_t^B = \frac{G_t}{B_t}$.

Note that for a self financing strategy, we have $V_t^B = G_t^B$ because $V_t = G_t$ and $B_t > 0$. Furthermore, we can show the following.

Lemma 4. *For a self financing strategy we have*

$$G_t^B = \sum_{i=0}^{t-1} \eta_i (S_{i+1}^B - S_i^B) + \sum_{i=0}^{t-1} \delta_i (C_{i+1}^B - C_i^B) \quad \forall t = 1, ..., T.$$

Proof. The proof involves straightforward but somewhat cumbersome algebraic manipulations of the above definitions. See the Appendix for the details. □

We know that under the EMM we defined, the stock discounted by the risk free asset is a martingale. We now need to show that the contingent claims prices obtained by computing the expected value of the final payoff discounted by the risk free asset also constitute a martingale under this EMM.

Lemma 5. *The stochastic process defined by the discounted values of the candidate contingent claims prices is an F_t martingale under the EMM.*

Proof. We defined our candidate process for the contingent claims price under the EMM as $C_t = E^Q \left[\frac{C_T(S_T)}{B_T} B_t \middle| F_t \right]$. The process for the discounted values of the contingent claims prices is then defined as

$$C_t^B \equiv \frac{C_t}{B_t} = E^Q \left[\frac{C_T(S_T)}{B_T} \middle| F_t \right].$$

We use the fact that the conditional expectation itself is a Q martingale. This in turn follows from the law of iterated expectations and the European style payoff function. Taking conditional expectations with respect to F_s on both sides of the above equation yields

$$E^Q\left[\frac{C_t}{B_t}\bigg|F_s\right] = E^Q\left[E^Q\left[\frac{C_T(S_T)}{B_T}\bigg|F_t\right]\bigg|F_s\right] \qquad \forall t > s.$$

Now using the law of iterated expectations we get

$$E^Q\left[\frac{C_t}{B_t}\bigg|F_s\right] = E^Q\left[\frac{C_T(S_T)}{B_T}\bigg|F_s\right] = \frac{C_s}{B_s} = C_s^B \qquad \forall t > s,$$

which gives the desired result. □

Lemma 6. *Under the EMM defined by (3.2), the discounted gain process is a martingale.*

Proof. Under the EMM Q, the process $\{S_t^B\}_{t=1}^T$ is a Q martingale. Using a standard property of martingales the process defined as $SS_t^B = \sum_{i=0}^{t-1} \eta_i(S_{i+1}^B - S_i^B)$ then is a Q martingale, since the investment strategy η_t is included in the information set.[14] Furthermore, from Lemma 5 we get that $\{C_t^B\}_{t=1}^T$ is also a Q martingale. Then using the fact that δ_t is an F_t predetermined process and using the same martingale property as above we get that the process $CC_t^B = \sum_{i=0}^{t-1} \delta_i(C_{i+1}^B - C_i^B)$ is a Q martingale. Then since from Lemma 4 the discounted gain process $\{G_t^B\}_{t=1}^T$ is the sum of two Q martingales, SS_t^B and CC_t^B, it is itself a Q martingale. □

At this stage, we have all the ingredients to show the following result.

Proposition 3.4.1. *If we have an EMM that makes the discounted price of the stock a martingale, then defining the price of any contingent claim as the expected value of its payoff, taken under this EMM and discounted at the riskless interest rate, constitutes a no-arbitrage price.*

14. Note that because we are working in discrete time there is no need to investigate the integrability of SS_t^B.

Proof. From Lemma 6 G_t^B is a Q martingale. Because we are considering self financing strategies we get that V_t^B is a martingale. We prove the absence of arbitrage by contradiction. If we assume the existence of an arbitrage opportunity, then there exists a self financing strategy with type 1 arbitrage ($V_0 < 0, V_T \geq 0$ a.s.) or type 2 arbitrage ($V_0 \leq 0, V_T > 0$ a.s.). Both cases lead to a clear contradiction. Consider type 1 arbitrage: we start from the existence of a self financing strategy with $V_0 < 0$ that ends up with a positive final value. $V_0 < 0$ implies that $V_0^B < 0$ since the numeraire is always positive by definition. Also since $V_T \geq 0$ we have $V_T^B \geq 0$. Taking expectations and using the fact that V_t^B is a Q martingale yields $V_0^B = E_0^Q[V_T^B] \geq 0$. This is a contradiction because we assumed that we start with a negative value $V_0 < 0$. A similar argument works for type 2 arbitrage. Thus, the C_t from the EMM Q must be a no-arbitrage price. □

In summary, we have demonstrated that in a discrete-time infinite state space setting, if we have an EMM that makes the underlying asset price a martingale, then the expected value of the payoff of the contingent claim taken under this EMM, discounted at the riskless asset, is a no-arbitrage price. In Section 3.2.2, we derived such an EMM. Altogether, we have therefore demonstrated that for any contingent claim paying a final payoff $C_T(S_T)$ the current price C_t can be computed as

$$C_t = E^Q \left[\frac{C_T(S_T)}{B_T} B_t \middle| F_t \right].$$

3.5 Important special cases

In this section we demonstrate how a number of important existing models are nested in our setup, using the class of linear Radon-Nikodym derivatives in (3.2). We first consider various specifications of the equity risk premium in the conditional normal setting. We then consider two conditional non-normal specifications relying on inverse Gaussian shocks and Poisson jumps respectively.

3.5.1 Flexible risk premium specifications

One of the advantages of our approach is that we can allow for general specifications of the time-varying equity risk premium. Here we discuss some potentially interesting ways to specify the risk premium in the return process for the underlying asset. In order to demonstrate the link with the available literature and for computational simplicity, we assume conditional normal returns, although this assumption is by no means necessary.

The conditional normal models in the Duan (1995) and Heston and Nandi (2000) models are special cases of our set-up. In our notation, Duan (1995) assumes

$$r_t = r, \text{ and } \mu_t = r + \lambda \sigma_t,$$

which in our framework corresponds to a Radon-Nikodym derivative of

$$\frac{dQ}{dP}\bigg| F_t = \exp\left(-\sum_{i=1}^{t} \left(\frac{\varepsilon_i}{\sigma_i}\lambda + \frac{1}{2}\lambda^2 \right) \right),$$

and risk neutral innovations of the form

$$\varepsilon_t^* = \varepsilon_t + \lambda \sigma_t.$$

Heston and Nandi (2000) instead assume

$$r_t = r, \text{ and } \mu_t = r + \lambda \sigma_t^2 + \frac{1}{2}\sigma_t^2,$$

which in our framework corresponds to a Radon-Nikodym derivative of

$$\frac{dQ}{dP}\bigg| F_t = \exp\left(-\sum_{i=1}^{t} \left(\left(\lambda + \frac{1}{2} \right) \varepsilon_i + \frac{1}{2} \left(\lambda + \frac{1}{2} \right)^2 \sigma_i^2 \right) \right),$$

and risk neutral innovations of the form

$$\varepsilon_t^* = \varepsilon_t + \lambda \sigma_t^2 + \frac{1}{2}\sigma_t^2.$$

However, many empirically relevant cases are not covered by existing theoretical results. For example, in the original ARCH-M paper, Engle, Lilien and Robins (1987) find the strongest empirical support for a risk premium specification of the form

$$\mu_t = r_t + \lambda \ln(\sigma_t) + \tfrac{1}{2}\sigma_t^2,$$

which cannot be used for option valuation using the available theory. In our framework it simply corresponds to a Radon-Nikodym derivative of

$$\frac{dQ}{dP}\bigg| F_t = \exp\left(-\sum_{i=1}^{t} \left(\frac{\lambda \ln(\sigma_i) + \frac{1}{2}\sigma_i^2}{\sigma_i^2} \varepsilon_i + \frac{1}{2} \left(\frac{\lambda \ln(\sigma_i) + \frac{1}{2}\sigma_i^2}{\sigma_i^2} \right)^2 \sigma_i^2 \right) \right),$$

and risk neutral innovations

$$\varepsilon_t^* = \varepsilon_t + \lambda \ln(\sigma_t) + \tfrac{1}{2}\sigma_t^2.$$

Our approach allows for option valuation under such specifications whereas the existing literature does not.

3.5.2 Conditionally inverse Gaussian returns

Christoffersen, Heston and Jacobs (2006) analyze a GARCH model with an inverse Gaussian innovation, $y_t \sim IG(\sigma_t^2/\eta^2)$. We can write their return dynamic as

$$R_t = r + (\zeta + \eta^{-1})\,\sigma_t^2 + \varepsilon_t, \text{ where} \tag{3.23}$$

$$\varepsilon_t = \eta y_t - \eta^{-1}\sigma_t^2, \tag{3.24}$$

and where the conditional return variance, σ_t^2, is of the GARCH form. The inverse Gaussian belongs to the class of infinitely divisible distributions, which yields the strict convexity in Proposition 2, and the other conditions of Proposition 2 are also satisfied.

From the MGF of an inverse Gaussian variable, we can derive the conditional log

MGF

$$\Psi_t(u) = \left(u + \frac{1-\sqrt{1+2u\eta}}{\eta}\right)\frac{\sigma_t^2}{\eta}.$$

The EMM condition

$$\Psi_t(\nu_t - 1) - \Psi_t(\nu_t) - \Psi_t(-1) + \phi_t\sigma_t^2 = 0,$$

is now solved by the constant

$$\nu_t = \nu = \frac{1}{2\eta}\left[\frac{(2+\zeta\eta^3)^2}{4\zeta^2\eta^2} - 1\right], \forall t,$$

which in turn implies that the EMM is given by

$$\begin{aligned} \frac{dQ}{dP}\bigg|F_t &= \exp\left(-\sum_{i=1}^{t}\left(\nu\varepsilon_i + \left(\nu + \frac{1-\sqrt{1+2\nu\eta}}{\eta}\right)\frac{\sigma_i^2}{\eta}\right)\right) \\ &= \exp\left(-\nu t\overline{\varepsilon}_t - \delta t\overline{\sigma_t^2}\right), \end{aligned}$$

where $\overline{\varepsilon}_t = \frac{1}{t}\sum_{i=1}^{t}\varepsilon_i$, $\overline{\sigma_t^2} = \frac{1}{t}\sum_{i=1}^{t}\sigma_i^2$, and $\delta = \frac{\nu}{\eta} + \frac{1-\sqrt{1+2\nu\eta}}{\eta^2}$.

These expressions can be used to obtain the risk-neutral distribution from Christoffersen, Heston and Jacobs (2006) using the results in Section 3.2. Recall that in general the risk neutral log MGF is

$$\Psi_t^{Q*}(u) = -u\Psi_t'(\nu) + \Psi_t(\nu+u) - \Psi_t(\nu).$$

In the GARCH-IG case we can write

$$\Psi_t^{Q*}(u) = \left(u + \frac{1-\sqrt{1+2u\eta^*}}{\eta^*}\right)\frac{\sigma_t^{*2}}{\eta^*},$$

where

$$\eta^* = \frac{\eta}{1+2\nu\eta} \text{ and } \sigma_t^{*2} = \frac{\sigma_t^2}{(1+2\nu\eta)^{3/2}},$$

which indicates that generally the risk-neutral variance will be different from the physical variance. The risk neutral return model can be written as

$$R_t \equiv \ln\left(\frac{S_t}{S_{t-1}}\right) = r - \Psi_t^{Q*}(-1) + \varepsilon_t^* = r + \left(\zeta^* + \eta^{*-1}\right)\sigma_t^{*2} + \varepsilon_t^*,$$

where

$$\zeta^* = \frac{1-2\eta^* - \sqrt{1-2\eta^*}}{\eta^{*2}} \text{ and } \varepsilon_t^* = \eta^* y_t^* - \eta^{*-1}\sigma_t^{*2}.$$

The risk neutral process thus takes the same form as the physical process, confirming Proposition 3.2.3 in Section 3.2.4.

3.5.3 Conditionally Poisson-normal jumps

Another interesting model that can be nested in our framework is the heteroskedastic model with Poisson-normal innovations in Duan, Ritchken and Sun (2005).[15] For expositional simplicity, we consider the simplest version of the model. More complex models, for instance with time-varying Poisson intensities, can also be accommodated. The conditions of Proposition 2 can again readily be verified, in part because the Poisson-normal is part of the class of infinitely divisible distributions.

We can write the underlying asset return as

$$\begin{aligned} R_t &= \kappa_t + \varepsilon_t, \text{ where} \\ \varepsilon_t &= \sigma_t\left(J_t - \vartheta\bar{\mu}\right), \end{aligned}$$

where J_t is a Poisson jump process with N_t jumps each with distribution $N\left(\bar{\mu}, \bar{\varphi}^2\right)$ and jump intensity ϑ. The conditional return variance equals $\left(1+\vartheta\left(\bar{\mu}^2+\bar{\varphi}^2\right)\right)\sigma_t^2$, where σ_t^2 is of the GARCH form. The log return mean κ_t is a function of σ_t^2 as well as the jump and risk premium parameters.

15. Maheu and McCurdy (2004) consider a different discrete-time jump model but do not use it for option valuation.

We can derive the conditional log MGF as

$$\begin{aligned}\Psi_t(u) &= \ln(E_{t-1}[\exp(-u\sigma_t(J_t-\vartheta\overline{\mu}))]) \\ &= u\vartheta\overline{\mu}\sigma_t+\frac{1}{2}u^2\sigma_t^2+\vartheta\left[\exp\left(-\overline{\mu}u\sigma_t+\frac{1}{2}\overline{\varphi}^2u^2\sigma_t^2\right)-1\right].\end{aligned}$$

The approach taken in Duan et al (2005) corresponds to fixing $v_t = v$ and setting

$$\kappa_t = r+\Psi_t(v)-\Psi_t(v-1),$$

which in turn implies that the EMM is given by

$$\frac{dQ}{dP}\bigg|F_t = \exp\left(-vt\overline{\varepsilon_t}-v\vartheta\overline{\mu}t\overline{\sigma_t}-\frac{1}{2}tv^2\overline{\sigma_t^2}+\vartheta t-\vartheta\sum_{i=1}^{t}\exp\left(-\overline{\mu}v_i\sigma_i+\frac{1}{2}\overline{\varphi}^2v_i^2\sigma_i^2\right)\right),$$

where $\overline{\varepsilon_t}$ and $\overline{\sigma_t^2}$ are the historical averages as above.

We can again show that the risk-neutral distribution is from the same family as the physical distribution

$$\begin{aligned}\Psi_t^{Q*}(u) &= \ln E_{t-1}^{Q}[\exp(-u\varepsilon_t^*)] \\ &= u\vartheta_t^*\overline{\mu}_t^*\sigma_t+\frac{1}{2}u^2\sigma_t^2+\vartheta_t^*\left[\exp\left(-\overline{\mu}_t^*u\sigma_t+\frac{1}{2}\overline{\varphi}^2u^2\sigma_t^2\right)-1\right],\end{aligned}$$

where

$$\vartheta_t^* = \vartheta\exp\left(-\overline{\mu}v\sigma_t+\frac{1}{2}\overline{\varphi}^2v^2\sigma_t^2\right) \text{ and } \overline{\mu}_t^* = \overline{\mu}-\overline{\varphi}^2\sigma_t v.$$

Note that in this model the mapping between the risk-neutral and physical returns is

$$\varepsilon_t^* = \varepsilon_t+\Psi_t'(v) = \varepsilon_t+\sigma_t(\vartheta\overline{\mu}-\vartheta_t^*\overline{\mu}_t^*),$$

and the mapping between the physical and risk-neutral conditional variance is

$$\sigma_t^{*2} = \sigma_t^2+\vartheta_t^*\sigma_t^2(\overline{\varphi}^2+\overline{\mu}_t^{*2}).$$

3.6 Some continuous-time limits

In order to anchor our work in the continuous-time literature we now explore the links between some of the discrete-time models we have analyzed above and standard continuous-time models. We study three important cases: a homoskedastic model with normal innovations, a homoskedastic model with non-normal (inverse Gaussian) innovations, and a heteroskedastic model with normal innovations.

3.6.1 Homoskedastic normal returns

Consider the homoskedastic i.i.d. normal model for a given discrete-time interval Δ,

$$R_t = \ln(S_t) - \ln(S_{t-\Delta}) = \mu\Delta - \tfrac{1}{2}\sigma^2\Delta + \sigma\sqrt{\Delta}z_t \qquad z_t|F_{t-1} \sim N(0,1), \tag{3.25}$$

and for simplicity also consider a constant risk-free rate. The EMM condition (3.4) is solved by choosing a constant $\nu = (\mu - r)/\sigma^2$, and the discrete-time risk-neutral dynamic is given by

$$\ln(S_t) - \ln(S_{t-\Delta}) = r\Delta - \tfrac{1}{2}\sigma^2\Delta + \sigma\sqrt{\Delta}z_t^* \qquad z_t^*|F_{t-1} \sim N(0,1). \tag{3.26}$$

The continuous-time limit of this risk-neutral process is given by

$$d(\ln(S_t)) = \left(r - \tfrac{1}{2}\sigma^2\right)dt + \sigma dz^*(t),$$

where $z^*(t)$ is a Wiener process under Q. This is the risk-neutral process in the Black-Scholes-Merton (BSM) model. In the diffusion limit the options are thus priced using the BSM formula.

Consider a European option with strike price K and $T - t = M\Delta$ days to maturity. The call price can be written as

$$C_{\Delta,t} = e^{-rM\Delta}S_tE_t^Q\left[e^{R_{t,M}}I[R_{t,M} > \ln(K/S_t)]\right] - e^{-rM\Delta}KP_t^Q[R_{t,M} > \ln(K/S_t)],$$

where $R_{t,M} = \ln(S_{t+M\Delta}) - \ln(S_t)$ and where $I[*]$ is the indicator function. Under the assumption of an i.i.d. normal risk-neutral process in (3.26) we can rewrite the call price as

$$C_{\Delta,t} = e^{-rM\Delta} S_t P_{1,t,\Delta} - e^{-rM\Delta} K P_{2,t,\Delta},$$

where

$$P_{1,t,\Delta} = e^{rM\Delta}\Phi\left(\frac{\ln(S_t/K) + \left(r + \frac{1}{2}\sigma^2\right)\Delta M}{\sigma\sqrt{\Delta M}}\right), \quad P_{2,t,\Delta} = \Phi\left(\frac{\ln(S_t/K) + \left(r - \frac{1}{2}\sigma^2\right)\Delta M}{\sigma\sqrt{\Delta M}}\right),$$

where Φ is the c.d.f. of the standard normal distribution.

Note therefore that for the i.i.d. normal discrete-time process, using the parameterization in (3.25), and given the choice of Radon-Nikodym derivative (and thus EMM), the option value is equal to the BSM price for any Δ.

3.6.2 Homoskedastic inverse Gaussian returns

Consider now a homoskedastic version of the inverse Gaussian (IG) model in (3.23) written for a discrete-time interval Δ,

$$\begin{aligned}
R_t &= r\Delta + \left(\zeta(\Delta) + \eta(\Delta)^{-1}\right)\sigma^2(\Delta) + \varepsilon_t \\
\varepsilon_t &= \eta(\Delta) y_t - \eta(\Delta)^{-1}\sigma^2(\Delta) \\
y_t &\sim IG\left(\frac{\sigma^2(\Delta)}{\eta^2(\Delta)}\right).
\end{aligned}$$

As shown above for the heteroskedastic IG case, the risk neutral return distribution is in the same family as the historical model, and can be written as follows

$$\begin{aligned}
R_t &= r\Delta + \left(\zeta^*(\Delta) + \eta^*(\Delta)^{-1}\right)\sigma^{*2}(\Delta) + \varepsilon_t^* \\
\varepsilon_t^* &= \eta^*(\Delta) y_t^* - \eta^*(\Delta)^{-1}\sigma^{*2}(\Delta) \\
y_t^* &\overset{Q}{\sim} IG\left(\frac{\sigma^{*2}(\Delta)}{\eta^{*2}(\Delta)}\right),
\end{aligned}$$

where

$$\begin{aligned}
\eta^*(\Delta) &= \frac{\eta(\Delta)}{1+2\nu(\Delta)\eta(\Delta)} \\
\sigma^{*2}(\Delta) &= \frac{\sigma^2(\Delta)}{(1+2\nu(\Delta)\eta(\Delta))^{3/2}} \\
\zeta^*(\Delta) &= \frac{1-2\eta^*(\Delta)-\sqrt{1-2\eta^*(\Delta)}}{\eta^*(\Delta)^2},
\end{aligned}$$

and where $\nu(\Delta)$ solves (3.4) and is given by

$$\nu(\Delta) = \frac{1}{2\eta(\Delta)}\left[\frac{\left(2+\zeta(\Delta)^2\eta(\Delta)^3\right)^2}{4\zeta(\Delta)^2\eta(\Delta)^2} - 1\right],$$

Consider a European option with strike price K and $T-t = M\Delta$ days to maturity. The call price can be written as

$$C_{\Delta,t} = e^{-rM\Delta}S_tP_{1,t,\Delta} - e^{-rM\Delta}KP_{2,t,\Delta}. \tag{3.27}$$

The formulas for $P_{1,t,\Delta}$ and $P_{2,t,\Delta}$ can be computed using Fourier inversion of the risk-neutral log MGF of $\Psi^{Q*}_{t,M}(u)$

$$\begin{aligned}
P_{1,t,\Delta} &= \frac{e^{rM\Delta}}{2} + \int_0^{+\infty} \mathscr{R}] \left[\frac{\exp\left(\Psi^{Q*}_{t,M}(-1-iu) - iu\ln\left(\frac{K}{S_t}\right)\right)}{i\pi u}\right] du \\
P_{2,t,\Delta} &= \frac{1}{2} + \int_0^{+\infty} \mathscr{R}] \left[\frac{\exp\left(-iu\ln\left(\frac{K}{S_t}\right) + \Psi^{Q*}_{t,M}(-iu)\right)}{i\pi u}\right] du,
\end{aligned}$$

where

$$\Psi^{Q*}_{t,M}(u) \equiv \ln\left(E^Q_t\left[\exp\left(uR_{t,M}\right)\right]\right) = -\left(r\Delta + \zeta(\Delta)\sigma^2(\Delta)\right)Mu + \frac{\left[\left(1-\sqrt{1+2u\eta^*(\Delta)}\right)\sigma^{*2}(\Delta)M\right]}{\eta^*(\Delta)^2}.$$

Christoffersen, Heston and Jacobs (2006) show that in the heteroskedastic case, the stochastic volatility model in Heston (1993a) with perfectly correlated shocks can be obtained as a limit of the IG-GARCH model when Δ and $\eta(\Delta)$ go to zero. [16] This limit obtains when using a particular parameterization for the IG-GARCH model and the parameterization $\zeta(\Delta) = \lambda - \eta(\Delta)^{-1}$ for the return mean, where λ can be interpreted as the price of equity risk. As the homoskedastic IG model is a special case of the IG-GARCH model it will converge to the homoskedastic Heston (1993a) process which is simply the geometric Brownian motion underlying the Black-Scholes model. The continuous-time limit of the risk-neutral process is thus again given by

$$d(\ln(S_t)) = \left(r - \tfrac{1}{2}\sigma^2\right) dt + \sigma dz^*(t).$$

Figure 2 illustrates the convergence of the homoskedastic IG option price in (3.27) to the BSM price when Δ goes to zero. In the figure we plot the ratio of the homoskedastic IG option price to the Black-Scholes price against the number of trading intervals per day. We use $r = 0$, $K = 100$, $S = 100, M\Delta = 180$. We let $\eta(\Delta) = \eta\Delta$, $\sigma^2(\Delta) = \sigma^2\Delta$, and set $\lambda\sigma^2 = .07$ to match a 7% equity risk premium. Return volatility is set to 10% per year ($\sigma^2 = .01$) in the top row and 20% in the bottom row ($\sigma^2 = .04$). The IG parameter η is set so as to generate a daily skewness of -1 in the left column and -0.5 in the right column. The figure shows that even for these relatively high levels of skewness the convergence of the skewed IG discrete-time option price to the Black-Scholes option price is quite rapid.

3.6.3 Heteroskedastic normal returns

Consider the Heston and Nandi (2000) model

$$\begin{aligned} R_t &= r\Delta + \lambda\sigma_t^2 + \sigma_t z_t \\ \sigma_{t+\Delta}^2 &= \omega + \beta\sigma_t^2 + \alpha(z_t - \rho\sigma_t)^2. \end{aligned} \tag{3.28}$$

16. Christoffersen, Heston and Jacobs (2006) also show that an alternative pure jump limit can be obtained in the inverse Gaussian model.

Defining $v_{t+\Delta} = \sigma^2_{t+\Delta}/\Delta$, we have

$$v_{t+\Delta} = \omega_v + \beta v_t + \alpha_v \left(z_t - \rho_v \sqrt{v_t}\right)^2, \tag{3.29}$$

with $\omega_v = \omega/\Delta$, $\alpha_v = \alpha/\Delta$ and $\rho_v = \rho\sqrt{\Delta}$. The conditional correlation is

$$Corr_{t-\Delta}(v_{t+\Delta}, R_t) = -\frac{sign(\rho_v)\sqrt{2\rho_v^2 v_t}}{\sqrt{1+2\rho_v^2 v_t}},$$

so that the correlation goes to plus or minus one when the interval shrinks to zero. Using the parameterization $\alpha(\Delta) = \frac{1}{4}\varsigma^2\Delta^2$, $\beta(\Delta) = 0$, $\omega(\Delta) = (\kappa\theta - \frac{1}{4}\varsigma^2)\Delta^2$, and $\rho(\Delta) = \frac{2}{\varsigma\Delta} - \frac{\kappa}{\varsigma}$, and following Foster and Nelson (1994), Heston and Nandi derive the diffusion limit for the physical process

$$\begin{aligned} d\ln(S_t) &= (r + \lambda v)dt + \sqrt{v}dz \\ dv &= \kappa(\theta - v)dt + \varsigma\sqrt{v}dz, \end{aligned} \tag{3.30}$$

which corresponds to a special case of the stochastic volatility model in Heston (1993a) with perfectly correlated shocks to stock price and volatility.

The Heston-Nandi discrete-time option price is

$$C_{t,\Delta} = S_t P_{1,t,\Delta} - e^{-rM\Delta} K P_{2,t,\Delta},$$

where the formulas for $P_{1,t,\Delta}$ and $P_{2,t,\Delta}$, which rely on Fourier inversion, are provided in Heston and Nandi (2000).

Note that markets are complete in the limiting case with $\rho = -1$ because there is only one source of uncertainty. Below we analyze the more general case of a discrete-time two-shock stochastic volatility model and its continuous-time limit where $-1 < \rho < 1$, which implies that markets are incomplete even in continuous time.

Figure 3 shows the convergence of the Heston and Nandi (2000) discrete-time GARCH option price to the continuous-time SV option price in Heston (1993a). We plot the ratio of the Heston and Nandi (2000) price to the Heston (1993a) price as the number of trading intervals until maturity gets large. We use $r = 0$, $K = 100$, $S = 100$, $M\Delta = 180$, $\kappa = 2$, and shock correlation $\rho = -1$. Return volatility is set to 10% per year ($v = \theta = .01$) in the top row and 20% in the bottom row ($v = \theta = .04$). The volatility of volatility parameter ς is set to 0.1 in the left column and 0.2 in the right column.

Figure 3 indicates that convergence is very fast, suggesting that the added incompleteness arising from discrete time is minimal. By comparison, convergence is slower in Figure 2 because of the conditional skewness in the discrete-time process. Note that following Heston and Nandi (2000), Figure 3 has trading intervals until maturity (180 days) on the horizontal axis whereas Figure 2 has trading intervals per day on the horizontal axis. Thus convergence is indeed extremely fast in Figure 3.

3.7 Stochastic Volatility Models

In this section, we first develop a discrete-time two-shock stochastic volatility model and derive its continuous-time limit. Subsequently we compare the risk neutralization for this model with the risk neutralization in the continuous-time SV model, and we discuss risk neutralization in the GARCH model as a special case of this approach. We also discuss the issue of market incompleteness and the resulting non-uniqueness of option prices, again by discussing similarities and differences between the continuous- and discrete-time setups.

3.7.1 A discrete-time stochastic volatility model

Popular continuous-time stochastic volatility models such as Heston (1993a) contain two (correlated) innovations, whereas the GARCH processes considered in this chapter contain a single innovation. Nelson (1991) and Duan (1997) derive a continuous-time two-innovation stochastic volatility model as the limit of a GARCH model, but as noted by Corradi (2000) for instance, a given discrete-time model can have several continuous-time limits and vice versa. [17] As shown above, Heston and Nandi (2000) derive a limit

17. See also Nelson and Foster (1994), Foster and Nelson (1996), Nelson (1996) and Ritchken and Trevor (1999) for limit results.

to their proposed GARCH process that contains two perfectly correlated shocks. This limit amounts to a one-shock process, and is therefore intuitively similar to a GARCH process.

With this in mind, we now analyze the limits of a class of discrete-time stochastic volatility processes, which contain two (potentially correlated) shocks.[18] We derive the continuous-time limits for these processes, and then analyze the GARCH limit as a special case.

Consider the return and volatility dynamics

$$\begin{aligned} R_t &= \ln(S_t/S_{t-1}) = \mu_t + \sigma_t z_{1,t} \\ \sigma_{t+1}^2 &= f(\sigma_t^2, z_{2,t}), \end{aligned}$$

where

$$z_t \equiv (z_{1,t}, z_{2,t})' \sim N\left(\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\right).$$

The log MGF is given by

$$\Psi_t(u_1,u_2) = \log[E_{t-1}(\exp(-u_1 z_{1,t} - u_2 z_{2,t}))] = \frac{1}{2}\left[(u_1+\rho u_2)^2 + \left(1-\rho^2\right)u_2^2\right].$$

By analogy with the one-shock linear case (3.2), we define the following Radon-Nikodym derivative

$$\frac{dQ}{dP}|F_t = \exp\left(-\sum_{i=1}^{t}(v_{1,i}z_{1,i} + v_{2,i}z_{2,i} + \Psi_i(v_{1,i}, v_{2,i}))\right). \qquad (3.31)$$

Using an approach similar to the one-shock case, one can show that the probability measure Q defined by the Radon-Nikodym derivative is an EMM if and only if

$$\Psi_t(v_{1,t}-\sigma_t, v_{2,t}) - \Psi_t(v_{1,t}, v_{2,t}) + \mu_t - r = \frac{1}{2}\sigma_t^2 - (v_{1,t}+\rho v_{2,t})\sigma_t + \mu_t - r = 0. \quad (3.32)$$

This is one equation in two unknowns, namely $v_{1,t}$ and $v_{2,t}$. Thus the second shock

18. See Ghysels, Harvey and Renault (1995) for a review of discrete-time stochastic volatility models. See Feunou and Tedongap (2009) for a recent discrete-time multifactor stochastic volatility model.

provides a new source of non-uniqueness to be discussed further below.

The risk neutral log MGF is given by

$$\begin{aligned} E_{t-1}^{Q}[\exp(-u_1 z_{1,t} - u_2 z_{2,t})] &= E_{t-1}^{P}\left[\left(\frac{\frac{dQ}{dP}|F_t}{\frac{dQ}{dP}|F_{t-1}}\right)\exp(-u_1 z_{1,t} - u_2 z_{2,t})\right] \\ &= \exp(\Psi_t(u_1 + \nu_{1,t}, u_2 + \nu_{2,t}) - \Psi_t(\nu_{1,t}, \nu_{2,t})), \end{aligned}$$

where

$$z_t = (z_{1,t}, z_{2,t})' \overset{Q}{\sim} N\left(\left((-\nu_{1,t} - \rho\nu_{2,t}, -\nu_{1,t}\rho - \nu_{2,t})', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\right). \quad (3.33)$$

We now illustrate this risk-neutralization for a specific parametric example

$$\begin{aligned} R_t &= r + \lambda\sigma_t^2 - \frac{1}{2}\sigma_t^2 + \sigma_t z_{1,t} \\ \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha(z_{2,t} - \rho\sigma_t)^2. \end{aligned} \quad (3.34)$$

The dynamic in (3.34) can be thought of as a stochastic volatility (two-shock) generalization of the GARCH dynamic in Heston and Nandi (2000). According to (3.33) the risk-neutral model is given by

$$\begin{aligned} R_t &= r - \frac{1}{2}\sigma_t^2 + \sigma_t z_{1,t}^* \\ \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha(z_{2,t}^* - \nu_{1,t}\rho - \nu_{2,t} - \rho\sigma_t)^2, \end{aligned} \quad (3.35)$$

where

$$z_t^* = \begin{pmatrix} z_{1,t}^* = z_{1,t} + \nu_{1,t} + \rho\nu_{2,t} \\ z_{2,t}^* = z_{2,t} + \nu_{1,t}\rho + \nu_{2,t} \end{pmatrix} \overset{Q}{\sim} N\left(\left((0,0)', \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\right).$$

In the one-shock GARCH case above, we could simply solve (3.4) by choosing the scalar ν_t as a function of the GARCH parameters. Determining $\nu_{1,t}$ and $\nu_{2,t}$ in a model with two innovations is somewhat more complex, but the intuition underlying the procedure is critical to understanding the link with the continuous-time literature. From (3.32) and (3.34) we have $\nu_{1,t} + \nu_{2,t}\rho = \lambda\sigma_t$. We then note that if we want to preserve the affine structure in (3.35) we need $\nu_{2,t} = \nu_2\sigma_t$, which yields the risk neutral dynamic

$$\begin{aligned} R_t &= r - \frac{1}{2}\sigma_t^2 + \sigma_t z_{1,t}^* \\ \sigma_{t+1}^2 &= \omega + \beta\sigma_t^2 + \alpha(z_{2,t}^* - \rho^*\sigma_t)^2, \end{aligned} \tag{3.36}$$

with $\rho^* = \rho + \nu_2(1-\rho^2) + \lambda\rho$. The condition on the price of risk needed to preserve the affine structure is similar to the one usually used in the Heston (1993a) model. Note that conditional on the assumption regarding the price of volatility risk, Proposition 2 can be generalized to address existence and uniqueness of a solution to (3.32).

Note that while λ, which is the price of equity risk, can be estimated from returns, ν_2, which arises from the new separate volatility shock, is not identified from the return on the underlying asset only. It must be estimated using returns as well as option prices. This is of course also the case in continuous-time SV models. The analysis is therefore very similar to the one usually employed in continuous time.

Using an approach similar to that taken in Heston and Nandi (2000), option valuation in this discrete-time SV model can be done via Fourier inversion of the conditional characteristic function.

3.7.2 A diffusion limit of the discrete-time stochastic volatility model

We first write the discrete-time stochastic volatility model as

$$R_t = r\Delta + \lambda\sigma_t^2 - \frac{1}{2}\sigma_t^2 + \sigma_t z_{1,t} \tag{3.37}$$

$$\sigma_{t+\Delta}^2 = \omega + \beta\sigma_t^2 + \alpha(z_{2,t} - \rho\sigma_t)^2. \tag{3.38}$$

Reparameterizing $v_{t+\Delta} = \sigma_{t+\Delta}^2/\Delta$, we have

$$v_{t+\Delta} = \omega_v + \beta v_t + \alpha_v\left(z_{2,t} - \rho_v\sqrt{v_t}\right)^2 \tag{3.39}$$

with $\omega_v = \omega/\Delta$, $\alpha_v = \alpha/\Delta$ and $\rho_v = \rho\sqrt{\Delta}$.

Following Heston and Nandi (2000) we use the parameterization $\alpha(\Delta) = \frac{1}{4}\varsigma^2\Delta^2$, $\beta(\Delta) = 0$, $\omega(\Delta) = (\kappa\theta - \frac{1}{4}\varsigma^2)\Delta^2$, and $\rho(\Delta) = \frac{2}{\varsigma\Delta} - \frac{\kappa}{\varsigma}$. As $\Delta \to 0$ the dynamic in (3.37) and (3.39) converges to

$$\begin{aligned} d\ln(S_t) &= (r + \lambda v_t - \frac{1}{2}v_t)dt + \sqrt{v_t}dz_1 \\ dv_t &= \kappa(\theta - v_t)dt + \varsigma\sqrt{v_t}dz_2, \end{aligned} \tag{3.40}$$

where z_1 and z_2 are two Wiener processes such that $dz_1dz_2 = -\rho dt$. Note that the discrete-time conditional correlation is given by

$$corr_{t-\Delta}(v_{t+\Delta}, R_t) = -\frac{\rho sign(\rho_v)\sqrt{2\rho_v^2 v_t}}{\sqrt{1 + 2\rho_v^2 v_t}}.$$

As $\Delta \to 0$, the variance asymmetry parameter $\rho_v(\Delta)$ approaches positive or negative infinity, and therefore the correlation approaches ρ or $-\rho$ in the limit. Also, as $\Delta \to 0$, the risk neutral discrete-time stochastic volatility model (3.36) converges to the following dynamic

$$\begin{aligned} d\ln(S_t) &= (r - \frac{1}{2}v_t)dt + \sqrt{v_t}dz_1^* \\ dv_t &= [\kappa(\theta - v_t) + \varsigma(v_2(1-\rho^2) + \lambda\rho)v_t]dt + \varsigma\sqrt{v_t}dz_2^*. \end{aligned} \tag{3.41}$$

where z_1^* and z_2^* are two Wiener processes such that $dz_1^*dz_2^* = -\rho dt$.

3.7.3 The relationship with the continuous-time affine SV model

Both (3.40) and (3.41) are square root stochastic volatility models of the type proposed by Heston (1993a). We now link our discrete-time stochastic volatility model and its risk-neutralization to the conventional risk-neutralization in the Heston (1993a) model. Assume for simplicity that the parameterization of the conditional mean dynamic under the physical measure is given by (3.40). Heston (1993a) proposes the following risk neutralization [19]

$$
\begin{aligned}
d\ln(S_t) &= (r-\frac{1}{2}v_t)dt+\sqrt{v_t}dz_1^* \\
dv_t &= [\kappa(\theta-v_t)-\varsigma\chi^* v_t]dt+\varsigma\sqrt{v_t}dz_2^*,
\end{aligned}
\tag{3.42}
$$

where z_1^* and z_2^* are two Wiener process under the risk neutral probability Q and

$$
\begin{aligned}
dz_1^* &= dz_1+\left(\lambda-\frac{1}{2}\right)\sqrt{v_t}dt \\
dz_2^* &= dz_2+\chi^*\sqrt{v_t}dt.
\end{aligned}
\tag{3.43}
$$

In the discrete-time stochastic volatility model, the parameter λ in (3.34) captures the price of equity risk, and v_2 captures the price of volatility risk. In the Heston model, the price of equity risk λ plays the same role as in the discrete-time model, and we have also a price of volatility risk χ^* which ensures the affine structure of the risk-neutral process. Comparing (3.42) and (3.41), we find

$$
\chi^* = v_2(1-\rho^2)+\lambda\rho. \tag{3.44}
$$

which amounts to the assumption on the price of risk used in Pan (2002). Note that for $\rho=0$, the continuous-time price of volatility risk χ^* is not related to λ, but is simply equal to the discrete-time price of volatility risk v_2. Moreover, this mapping between the price of volatility risk in discrete-time and continuous-time stochastic volatility models also provides insight into the relationship between the discrete-time GARCH model and the available continuous-time literature. While the GARCH model contains a single innovation, it can usefully be thought of as a special case of the two-shock discrete-time stochastic volatility model in (3.35), for $\rho=1$ (or $\rho=-1$). In this case, from (3.44), $\chi^*=\lambda$ (or $-\lambda$). Because the GARCH model contains a single shock, the specification of the equity risk premium λ does double duty: it also implicitly defines the price of volatility risk, which is perfectly correlated with the price of equity risk by design. In other

19. Notice that for ease of interpretation, in our notation the price of volatility risk χ^* has been rescaled by $1/\varsigma$ compared to the notation in Heston (1993a).

words, the GARCH return dynamic implicitly makes an assumption about the volatility risk premium. The parameter governing the equity risk premium also determines the volatility risk premium. Strictly speaking therefore, in the case of the GARCH model the only assumption we make in our approach is on the form of the Radon-Nikodym derivative. All other assumptions needed for risk-neutral valuation are implicit in the specification of the return dynamic. Put differently, some important assumptions on the equilibrium supporting the valuation problem are implicitly incorporated in the risk premium assumption for the return dynamic.

3.7.4 Stochastic Volatility and GARCH

The discussion above indicates that while it is useful to distinguish between one-shock and two-shock models, our analysis of discrete-time GARCH option valuation models is very similar to the analysis of continuous-time SV option valuation models. Most existing papers on option pricing in discrete time assume normally distributed returns and, in the words of Rubinstein (1976), "complete" the markets by assuming a representative agent with certain preferences, such as for instance constant relative risk aversion.[20] Our approach, much like the one used in the continuous-time stochastic volatility literature, is to let the researcher specify an empirically realistic return dynamic for the underlying asset, and subsequently provide an equivalent martingale measure that enables option pricing using a no-arbitrage argument. Proposition 3.2.1 provides the form of the EMM and Proposition 3.4.1 provides the no-arbitrage option pricing result. Whereas the assumption on the representative agent's utility function "completes" the market in the standard normal discrete-time setting, the Radon-Nikodym derivative "completes" the market in our setup. Conditional on the choice of Radon-Nikodym derivative which is linear in the return innovation, our approach provides a unique EMM.

The only difference between GARCH option valuation and option valuation with stochastic volatility is that GARCH models can be viewed as special cases of discrete-time stochastic volatility models. In the GARCH model, one parameter determines the volatility risk premium as well as the equity risk premium, and therefore the volatility

20. See for example Rubinstein (1976), Brennan (1979), and Duan (1995).

risk premium is implicitly specified by the GARCH dynamic. This is consistent with the interpretation of the GARCH model as a one-shock model with perfectly correlated equity and volatility innovations.[21]

Section 3.3 illustrates that it is possible to generalize the EMM specification, although in most cases it is not straightforward to obtain analytical results. We therefore limit our discussion to the case of the quadratic EMM with normal innovations in Section 3.3.3, which contains the linear EMM as a special case. This indicates that the uniqueness result obtained for the GARCH model discussed above is due to the choice of the linear EMM. In the more general quadratic case, we obtain an infinite number of valid EMMs, as illustrated in Figure 1.

3.8 Conclusion

This chapter provides valuation results for contingent claims in a discrete-time infinite state space setup. Most of our analysis focuses on a class of Radon-Nikodym derivatives for which the risk neutral return dynamic is the same as the physical dynamic for a wide class of processes, but with a different parameterization which we are able to characterize completely. We also discuss more general choices of Radon-Nikodym derivatives. Our valuation argument applies to a large class of conditionally normal and non-normal stock returns with flexible time-varying mean and volatility, as well as a potentially time-varying price of risk. This setup generalizes the result in Duan (1995) in the sense that we do not restrict the returns to be conditionally normal, nor do we restrict the price of risk to be constant.

Our results apply to some of the most widely used discrete-time processes in finance, such as GARCH processes. We also apply our approach to the analysis of discrete-time processes with multiple innovations, such as discrete-time stochastic volatility processes. To provide intuition for our findings, we extensively discuss the relationship between our

21. While it could be argued that this structure limits the usefulness of the GARCH model, one has to keep in mind that this structure is exactly what makes the GARCH model econometrically tractable. Indeed, the success of the GARCH model in modeling returns, and its growing popularity in modeling options, are precisely due to the fact that despite its simple structure it provides a very good fit.

results and existing results for continuous-time stochastic volatility models, which can be derived as limits of our discrete-time dynamics.

Our results suggest a number of interesting avenues for further research. First, an extensive empirical comparison of option valuation with non-normal and heteroskedastic innovations should prove interesting. Combining non-normality and heteroskedasticity attempts to correct the biases associated with the conditionally normal GARCH model. These biases are similar to those displayed by the Heston (1993a) model, which the continuous-time literature has sought to remedy by adding (potentially correlated) jumps in returns and volatility.[22] A comparison with these models may prove valuable. Second, it is well-known that the risk-neutralization of existing models is not satisfactory from an empirical perspective.[23] The implications of alternative Radon-Nikodym derivatives for the option valuation models' empirical performance therefore ought to be studied. A comparison between linear and quadratic EMMs for normal innovations may provide a valuable starting point. Third, while we advocate separating the valuation issue and the general equilibrium setup that supports it, the general equilibrium foundations of our results are of course very important. It may prove possible to characterize the equilibrium setup that gives rise to the risk neutralization proposed for some of the processes considered in this chapter. However, this is by no means a trivial problem, and it is left for future work.

22. See for example Bakshi, Cao and Chen (1997), Bates (2000), Broadie, Chernov and Johannes (2007), Carr and Wu (2004), Eraker, Johannes and Polson (2003), Eraker (2004), Huang and Wu (2004) and Pan (2002).

23. See for example Broadie, Chernov and Johannes (2007).

3.9 Appendix

3.9.1 Proof of Proposition 3.2.2.

Define $f(v) = \Psi(v) - \Psi(v-1)$. Existence is obtained if $f(v)$ can take any real value. Uniqueness is demonstrated if $f(v) = E[R-r] \equiv \pi$ has a unique solution for any given value of π. By assumption, Ψ tends to infinity at the boundaries of its domain, therefore $\Psi(u_1) = +\infty$ and $\Psi(u_2) = +\infty$. Ψ is also continuous because it is twice differentiable on its domain. The domain of $f(v)$ is (u_1+1, u_2). Since Ψ is continuous $f(.)$ is also continuous. We get

$$\begin{aligned} f(u_1+1) &= \Psi(u_1+1) - \Psi(u_1) = -\Psi(u_1) = -\infty \\ f(u_2) &= \Psi(u_2) - \Psi(u_2-1) = \Psi(u_2) = +\infty, \end{aligned}$$

since $\Psi(u_1) = +\infty$ and $\Psi(u_2) = +\infty$. Hence $f(.)$ is continuous and can attain $-\infty$ or $+\infty$. Thus there exists a value v in the domain of the continuous function $f(.)$ such that $f(v) = \pi$ for any value $\pi \in (-\infty, +\infty)$. Furthermore, we have that $f'(u) = \Psi'(u) - \Psi'(u-1)$. Convexity of Ψ implies that $\Psi'(.)$ is increasing. Thus, if $f'(u) = \Psi'(u) - \Psi'(u-1) > 0$, then $f(.)$ is increasing. Therefore, $f(.)$ is increasing and continuous, which implies that $f(.)$ is a bijection, and uniqueness follows.

3.9.2 Proof of Lemma 4.

For a self financing strategy we have

$$\begin{aligned} G_{t+1} = V_{t+1} &= \eta_{t+1} S_{t+1} + \delta_{t+1} C_{t+1} + \psi_{t+1} B_{t+1} \\ &= \eta_t S_{t+1} + \delta_t C_{t+1} + \psi_t B_{t+1}. \end{aligned}$$

We also have

$$G_t = \sum_{i=0}^{t-1} \eta_i (S_{i+1} - S_i) + \sum_{i=0}^{t-1} \delta_i (C_{i+1} - C_i) + \sum_{i=0}^{t-1} \psi_i (B_{i+1} - B_i).$$

It follows that

$$G_{t+1} - G_t = \eta_t(S_{t+1} - S_t) + \delta_t(C_{t+1} - C_t) + \psi_t(B_{t+1} - B_t).$$

We can trivially also write

$$G^B_{t+1} - G^B_t = G^B_{t+1} - G^B_t + \underbrace{\left(\frac{G_{t+1}}{B_t} - \frac{G_{t+1}}{B_t}\right)}_{=0}.$$

This implies that

$$\begin{aligned} G^B_{t+1} - G^B_t &= (\eta_t S_{t+1} + \delta_t C_{t+1} + \psi_t B_{t+1})\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) \\ &+\frac{1}{B_t}(\eta_t(S_{t+1} - S_t) + \delta_t(C_{t+1} - C_t) + \psi_t(B_{t+1} - B_t)) \\ &= \eta_t\left[S_{t+1}\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) + \frac{1}{B_t}(S_{t+1} - S_t)\right] \\ &+\delta_t\left[C_{t+1}\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) + \frac{1}{B_t}(C_{t+1} - C_t)\right] \\ &+ \underbrace{\psi_t B_{t+1}\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) + \frac{1}{B_t}\psi_t(B_{t+1} - B_t).}_{=0} \end{aligned}$$

Then

$$G^B_{t+1} - G^B_t = \eta_t\left[S_{t+1}\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) + \frac{1}{B_t}(S_{t+1} - S_t)\right] + \delta_t\left[C_{t+1}\left(\frac{1}{B_{t+1}} - \frac{1}{B_t}\right) + \frac{1}{B_t}(C_{t+1} - C_t)\right]$$

$$= \eta_t(S^B_{t+1} - S^B_t) + \delta_t(C^B_{t+1} - C^B_t) + \left(\eta_t\frac{S_{t+1}}{B_t} - \eta_t\frac{S_{t+1}}{B_t}\right) + \left(\delta_t\frac{C_{t+1}}{B_t} - \delta_t\frac{C_{t+1}}{B_t}\right),$$

and therefore

$$G^B_{t+1} - G^B_t = \eta_t(S^B_{t+1} - S^B_t) + \delta_t(C^B_{t+1} - C^B_t). \qquad \forall t = 1, ..., T-1.$$

Because $G_0 = G_0^B = 0$ the discounted gain can be written as the sum of past changes

$$G_t^B = \sum_{i=0}^{t-1}(G_{i+1}^B - G_i^B) \; \forall t = 1,...,T.$$

Therefore the discounted gain can be written

$$G_t^B = \sum_{i=0}^{t-1} \eta_i(S_{i+1}^B - S_i^B) + \sum_{i=0}^{t-1} \delta_i(C_{i+1}^B - C_i^B),$$

and the proof is complete.

3.9.3 Proof of Proposition 3.2.3.

From Lukacs (1970), page 119, we have the Kolmogorov canonical representation of the log-moment generating function of an infinitely divisible distribution function. This result stipulates that a function Ψ is the log-moment generating function of an infinitely divisible distribution with finite second moment if, and only if, it can be written in the form

$$\Psi(u) = -uc + \int_{-\infty}^{+\infty} \left(e^{-ux} - 1 + ux\right) \frac{dK(x)}{x^2},$$

where c is a real constant while $K(u)$ is a nondecreasing and bounded function such that $K(-\infty) = 0$. Applying this theorem gives the following form for $\Psi_t(u)$,

$$\Psi_t(u) = -uc_{t-1} + \int_{-\infty}^{+\infty} \left(e^{-ux} - 1 + ux\right) \frac{dK_{t-1}(x)}{x^2}, \tag{3.45}$$

where c_{t-1} is a random variable known at $t-1$, and $K_{t-1}(x)$ is a function known at $t-1$, which is nondecreasing and bounded so that $K_{t-1}(-\infty) = 0$. Using relation (3.9) and the characterization (3.45) we can write $\Psi_t^{Q*}(u)$ as

$$\Psi_t^{Q*}(u) = \int_{-\infty}^{+\infty} \left(e^{-ux} - 1 + ux\right) \frac{dK_{t-1}^*(x)}{x^2},$$

where

$$K_{t-1}^*(x) = \int_{-\infty}^{x} e^{-\nu_t y} dK_{t-1}(y).$$

This implies that

$$K_{t-1}^{*}(-\infty)=0,$$

$K_{t-1}^{*}(x)$ is obviously non-decreasing since $K_{t-1}(x)$ is non-decreasing, $K_{t-1}^{*}(\infty)<\infty$, because $K_{t-1}(\infty)<\infty$, and $e^{-\nu_t y}$ is a decreasing function of y which converge to 0. Recall that ν_t is the generalized price of risk, which is positive and known at time $t-1$.

In conclusion we have constructed a constant c_{t-1}^{*} $(=0)$ and a non-decreasing bounded function $K_{t-1}^{*}(x)$, with $K_{t-1}^{*}(-\infty)=0$, such that

$$\Psi_t^{Q*}(u)=-uc_{t-1}^{*}+\int_{-\infty}^{+\infty}\left(e^{-ux}-1+ux\right)\frac{dK_{t-1}^{*}(x)}{x^2}.$$

Hence, according to the Kolmogorov canonical representation, the conditional distribution of ε_t^{*} is infinitely divisible.

BIBLIOGRAPHY

[1] Amin, K. and V. Ng (1993), ARCH Processes and Option Valuation, Manuscript, University of Michigan.

[2] Bakshi, C., Cao, C. and Z. Chen (1997), Empirical Performance of Alternative Option Pricing Models, Journal of Finance, 52, 2003-2049.

[3] Barone-Adesi, G., R. Engle, and L. Mancini (2008), A Garch Option Pricing Model in Incomplete Markets, Review of Financial Studies, 21, 1223-1258.

[4] Bates, D. (1996), Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, Review of Financial Studies, 9, 69-107.

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

[6] Bates, D. (2003), Empirical Option Pricing: a Retrospection, Journal of Econometrics, 116, 387-404.

[7] Bernardo, A., and O. Ledoit (2000), Gain, Loss, and Asset Pricing, Journal of Political Economy, 108, 144-172.

[8] Bick, A. (1990), On Viable Diffusion Price Processes of the Market Portfolio, Journal of Finance, 45, 673-689.

[9] Bjork, T. and I. Slinko (2006), Towards a General Theory of Good-Deal Bounds, Review of Finance, 10, 221-260.

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

[11] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327.

[12] Bollerslev, T., Chou, R. and K. Kroner (1992), ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence, Journal of Econometrics, 52, 5-59.

[13] Bollerslev, T., M. Gibbons, and H. Zhou (2005), Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities, Manuscript, Duke University.

[14] Bollerslev, T. and H.O. Mikkelsen (1996), Long-Term Equity AnticiPation Securities and Stock Market Volatility Dynamics, Journal of Econometrics, 92, 75-99.

[15] Bondarenko, O., and I. Longarela (2004), Benchmark Good Deal Bounds: an Application to Stochastic Volatility Models of Option Pricing, Working Paper, Stockholm School of Economics.

[16] Brennan, M. (1979), The Pricing of Contingent Claims in Discrete-Time Models, Journal of Finance, 34, 53-68.

[17] Broadie, M., Chernov, M. and M. Johannes (2007), Model Specification and Risk Premia: Evidence from S&P500 Futures Options Market, Journal of Finance, 62, 1453-1490.

[18] Buhlmann, H., F. Delbaen, P. Embrechts, and A. Shiryaev (1996), No-Arbitrage, Change of Measure and Conditional Esscher Transforms, CWI Quarterly, 9, 291-317.

[19] Buhlmann, H., Delbaen, F., Embrechts, P., and A. Shiryaev (1998), On Esscher Transforms in Discrete Finance Models, ASTIN Bulletin, 28, 171–186.

[20] Camara, A. (2003), A Generalization of the Brennan-Rubinstein Approach for the Pricing of Derivatives, Journal of Finance, 58, 805-819.

[21] Carr, P. and L. Wu (2004), Time-Changed Levy Processes and Option Pricing, Journal of Financial Economics, 17, 113–141.

[22] Carr, P. and L. Wu (2007), Variance Risk Premia, forthcoming in the Review of Financial Studies.

[23] Christoffersen, P., S. Heston, and K. Jacobs (2006), Option Valuation with Conditional Skewness, Journal of Econometrics, 131, 253-284.

[24] Christoffersen, P. and K. Jacobs (2004), Which GARCH Model for Option Valuation? Management Science, 50, 1204-1221.

[25] Cochrane, J., and J. Saa-Requejo (2000), Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets, Journal of Political Economy, 108, 79-119.

[26] Constantinides, G., Jackwerth, J. and S. Perrakis (2009), Mispricing of S&P 500 Index Options, Review of Financial Studies, 22, 1247-1278.

[27] Constantinides, G. M. and S. Perrakis (2002), Stochastic Dominance Bounds on Derivative prices in a Multiperiod Economy with Proportional Transaction Costs, Journal of Economic Dynamics and Control, 26, 1323-1352.

[28] Constantinides, G. M. and S. Perrakis (2007), Stochastic Dominance Bounds on American Option Prices in Markets with Frictions, Review of Finance 11, 71-115.

[29] Corradi, V. (2000), Reconsidering the Continuous-Time Limit of the GARCH(1,1) Process, Journal of Econometrics, 96, 145-153.

[30] Cox, J. and S. Ross (1976), The Valuation of Options for Alternative Stochastic Processes, Journal of Financial Economics, 3, 145-166.

[31] Cox, J., Ross, S. and M. Rubinstein (1979), Option Pricing: A Simplified Approach, Journal of Financial Economics, 7, 229-264.

[32] Dai, Q., Le, A., and K. Singleton (2006), Discrete-Time Dynamic Term Structure Models with Generalized Market Prices of Risk, Manuscript, Stanford University.

[33] Diebold, F. and J. Lopez (1995), Modeling Volatility Dynamics, in Macroeconometrics: Developments, Tensions and Prospects, Kevin Hoover (ed.). Boston: Kluwer Academic Press.

[34] Duan, J.-C. (1995), The GARCH Option Pricing Model, Mathematical Finance, 5, 13-32.

[35] Duan, J.-C. (1997), Augmented GARCH(p,q) Process and Its Diffusion Limit, Journal of Econometrics, 79, 97-127.

[36] Duan, J.-C. (1999), Conditionally Fat-Tailed Distributions and the Volatility Smile in Options, Manuscript, University of Toronto.

[37] Duan, J.-C., Ritchken, P. and Z. Sun (2005), Jump Starting GARCH: Pricing and Hedging Options with Jumps in Returns and Volatilities, Working Paper, Rotman School, University of Toronto.

[38] Engle, R. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation, Econometrica, 50, 987-1008.

[39] Engle, R., D. M. Lilien and R.P. Robins (1987), Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model, Econometrica, 55, 391-407.

[40] Engle, R. and V. Ng (1993), Measuring and Testing the Impact of News on Volatility, Journal of Finance, 48, 1749-1778.

[41] Eraker, B. (2004), Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, Journal of Finance, 59, 1367-1403.

[42] Eraker, B., M. Johannes and N. Polson (2003), The Impact of Jumps in Volatility and Returns, Journal of Finance, 58, 1269-1300.

[43] Esscher, F. (1932), On the Probability Function in the Collective Theory of Risk, Skandinavisk Aktuarietidskrift, 15, 175-195.

[44] Feller, W. (1968), An Introduction to Probability Theory and its Applications. John Wiley & Sons.

[45] Feunou, B., and R. Tedongap, (2009), A Multifactor Stochastic Volatility Model with Time-Varying Conditional Skewness, Working Paper, Department of Economics, Duke University.

[46] Foster, D. P. and D.B. Nelson (1996), Continuous Record Asymptotics for Rolling Sample Variance Estimators, Econometrica, 64, 139-174.

[47] French, K., G. W. Schwert, and R. Stambaugh (1987), Expected Stock Returns and Volatility, Journal of Financial Economics, 19, 3-30.

[48] Garcia, R. and E. Renault (1998), A Note on Hedging in ARCH-type Option Pricing Models, Mathematical Finance, 8, 153-161.

[49] Gerber, H., and E. Shiu (1994), Option Pricing By Esscher Transforms, Transactions of the Society of Actuaries, 46, 99-191.

[50] Glosten, L. R., R. Jagannathan, and D. Runkle (1993), On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks, Journal of Finance, 48, 1779-1801.

[51] Ghysels, E., A. Harvey, and E. Renault (1995), Stochastic Volatility, Handbook of Statistics 14, Statistical Methods in Finance, G.S. Maddala and C.R. Rao (eds.), North Holland. Amsterdam.

[52] Gourieroux, C. and A. Monfort (2007), Econometric Specification of Stochastic Discount Factor Models, Journal of Econometrics, 136, 509-530.

[53] Harrison, M. and D. Kreps (1979), Martingales and Arbitrage in Multi-period Securities Markets, Journal of Economic Theory, 20, 381-408.

[54] Harrison, M., and S. Pliska (1981), Martingales and Stochastic Integrals in the Theory of Continuous Trading, Stochastic Processes and their Applications, 11, 215–260.

[55] He, H. and H. Leland (1993), On Equilibrium Asset Price Processes, Review of Financial Studies, 6, 593-617.

[56] Heston, S. (1993a), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, Review of Financial Studies, 6, 327-343.

[57] Heston, S. (1993b), Invisible Parameters in Option Prices, Journal of Finance, 48, 933-947.

[58] Heston, S. (2004), Option Valuation with Infinitely Divisible Distributions, Quantitative Finance, 4, 515-524.

[59] Heston, S. and S. Nandi (2000), A Closed-Form GARCH Option Pricing Model, Review of Financial Studies, 13, 585-626.

[60] Huang, J.-Z. and L. Wu (2004), Specification Analysis of Option Pricing Models Based on Time-Changed Levy Processes, Journal of Finance, 59, 1405–1439.

[61] Jacod, J. and A. Shiryaev (1998), Local Martingales and the Fundamental Asset Pricing Theorems in the Discrete-time Case, Finance and Stochastics, 259-273.

[62] Karatzas, I., and S. Shreve (1998), Methods of Mathematical Finance, Springer-Verlag.

[63] Levy, H. (1985), Upper and Lower Bounds of Put and Call Option Value: Stochastic Dominance Approach, The Journal of Finance, 40, 1197-1217.

[64] Lukacs, E. (1970), Characteristic Functions, Griffin. London

[65] Madan, D. B., and E. Seneta (1990), The Variance Gamma Model from Share Market Returns, Journal of Business, 63, 511-524.

[66] Maheu, J., and T. McCurdy (2004), News Arrival, Jump Dynamics and Volatility Components for Individual Stock Returns, Journal of Finance, 59, 755–793.

[67] Nelson, D. B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59, 347-370.

[68] Nelson, D. B. (1996), Asymptotically Optimal Smoothing with Arch Models, Econometrica, 64, 561-573.

[69] Nelson, D. B. and D. P. Foster (1994), Asymptotic Filtering Theory for Univariate Arch Models, Econometrica, 62, 1-41.

[70] Oancea, I. and S. Perrakis (2007), Stochastic Dominance and Option Pricing in Discrete and Continuous Time: an Alternative Paradigm, Working Paper, Concordia University.

[71] Pan, J. (2002), The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, Journal of Financial Economics, 63, 3–50.

[72] Perrakis, S. (1986), Option Bounds in Discrete Time: Extensions and the Pricing of the American Put, Journal of Business, 59, 119-141.

[73] Perrakis, S. and P. Ryan (1984), Option Bounds in Discrete Time, Journal of Finance, 39, 519-525.

[74] Ritchken, P.(1985), On Option Pricing Bounds, The Journal of Finance, 40, 1219-1233.

[75] Ritchken, P. and S. Kuo (1988), Option Bounds with Finite Revision Opportunities, Journal of Finance, 43, 301-308.

[76] Ritchken, P. and R. Trevor (1999), Option Pricing Under GARCH and Stochastic Volatility, Journal of Finance, 54, 377-402.

[77] Rubinstein, M. (1976), The Valuation of Uncertain Income Streams and the Pricing of Options, Bell Journal of Economics, 7, 407-425.

[78] Satchell, S., and A. Timmermann (1996), Option Pricing with GARCH and Systematic Consumption Risk, Derivatives Use, Trading & Regulation, Part II, 1, 353-67.

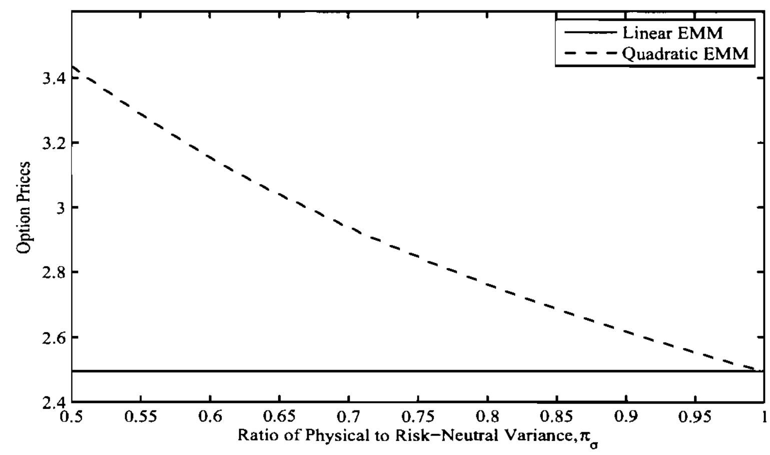
[79] Schroder, M. (2004), Risk Neutral Parameter Shifts and Derivatives Pricing in Discrete Time, Journal of Finance, 59, 2375-2402.

[80] Schwert, G.W. (1989), Why Does Stock Market Volatility Change over Time? Journal of Finance, 44, 1115-1153.

[81] Shiryaev, A. (1999), Essentials of Stochastic Finance. Facts, Models, Theory, World Scientific.

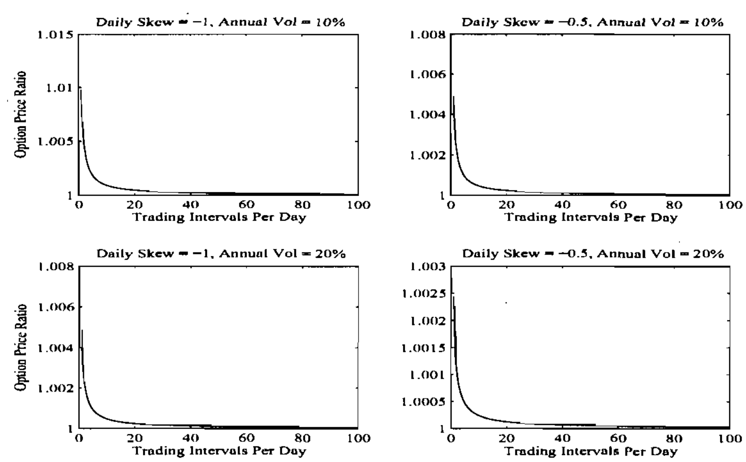
[82] Siu, T.K., H. Tong, and H. Yang (2004), On Pricing Derivatives under GARCH Models: A Dynamic Gerber-Shiu Approach, North American Actuarial Journal, 8, 17-31.

Figure 3.1: **Option Prices from Linear and Quadratic EMMs.**



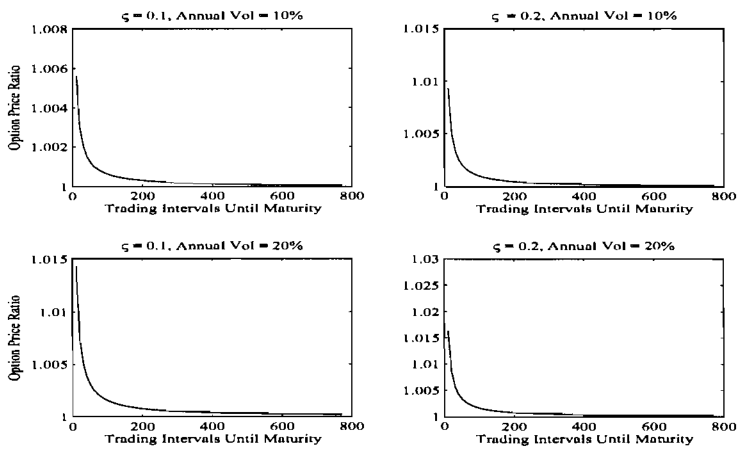
We use the linear and quadratic EMMs to compute the price of a one-month to maturity, at-the-money call option with an underlying asset price of 100. We assume a risk-free rate of 5%, an underlying asset mean return of 10% and a physical asset volatility of 20% per year. In the quadratic EMM we let the ratio of the physical to risk-neutral variance, π_σ, vary from 0.5 to 1.

Figure 3.2: **Convergence of Homoskedastic Inverse Gaussian to Black-Scholes Option Price**



We plot the ratio of the homoskedastic IG option price to the Black-Scholes price as the number of trading intervals per day gets large. We use $r = 0$, $K = 100$, $S = 100, M\Delta = 180$. We let $\eta(\Delta) = \eta\Delta$, $\sigma^2(\Delta) = \sigma^2\Delta$, and set $\lambda\sigma^2 = .07$ to match a 7% equity risk premium. Return volatility is set to 10% per year ($\sigma^2 = .01$) in the top row and 20% in the bottom row ($\sigma^2 = .04$). The IG parameter η is set so as to generate a daily skewness of -1 in the left column and -0.5 in the right column.

Figure 3.3: **Convergence of GARCH to Stochastic Volatility Option Price**



We plot the ratio of the Heston and Nandi (2000) discrete-time GARCH option price to the continuous-time SV option price in Heston (1993a) as the number of trading intervals until maturity gets large. We use $r = 0$, $K = 100$, $S = 100$, $M\Delta = 180$, $\kappa = 2$, and shock correlation $\rho = -1$. Return volatility is set to 10% per year ($v = \theta = .01$) in the top row and 20% in the bottom row ($v = \theta = .04$). The volatility of volatility parameter ς is set to 0.1 in the left column and 0.2 in the right column.

CHAPTER 4

AFFINE STOCHASTIC SKEWNESS MODELS

Abstract

We develop a conditional arbitrage pricing theory (APT) model where factors and idiosyncratic noises are both heteroscedastic and asymmetric. The model features both stochastic volatility and conditional skewness (SVS model), as well as conditional leverage effects. We explicitly allow asset prices to be asymmetric conditional on current factors and past information, termed contemporaneous asymmetry. Conditional skewness is driven by conditional leverage effects (through factor loadings) and contemporaneous asymmetry (through idiosyncratic skewness). We estimate and test three versions of the SVS model using several equities and indexes daily returns, as well as daily index options data. Results suggest that contemporaneous asymmetry is particularly important in several dimensions. It helps to match sample return skewness, negative and significant cross-correlations between returns and squared returns, as well as positive and significant cross-correlations between returns and cubed returns. Further diagnostics suggest that SVS models with contemporaneous asymmetry show a better option pricing performance compared to contemporaneous normality and existing affine GARCH models, especially for in-the-money call options and short-maturity contracts.

4.1 Introduction

Three relevant stylized facts have emerged from the analysis of financial time series, namely, time-varying conditional variance (or heteroscedasticity), time-varying conditional leverage effect, and time-varying conditional skewness. Since these time series characteristics are common to many financial assets, and given that these assets are likely to be affected by the same economic risk factors, time series properties of the factors combined with asset's systematic risk and idiosyncratic characteristics, will have important implications for the time series of asset returns. This article develops a conditional arbitrage pricing theory (APT) model where factors and idiosyncratic noises are both heteroscedastic and asymmetric. Heteroscedasticity in the factors implies heteroscedasticity in asset returns, as well as time-varying conditional skewness and leverage effect. Our approach does not tackle independently time series and cross-sectional characteristics of asset returns. In fact, leverage effect arises from asset systematic risk (asset's factor loading or beta), heteroscedasticity results from asset's beta and idiosyncratic volatility, and conditional skewness relates to both asset's beta, idiosyncratic volatility and idiosyncratic skewness.

The autoregressive conditional heteroscedasticity (ARCH, Engle (1982)) and its generalization (GARCH, Bollerslev (1986)) have been widely used in modeling time-series variation in conditional variance. While return volatility is completely determined as a function of past observed returns in ARCH and GARCH models, an alternative approach, which has become more popular recently, is the stochastic volatility (SV) model, where return volatility is an unobserved component which undergoes shocks from a different source other than return shocks. Most empirical applications of SV and GARCH models assume that the conditional distribution of returns is symmetric. Even if these models help matching the observed unconditional kurtosis in actual data, they fail to match unconditional asymmetries (skewness and leverage effects). Allowing for conditional leverage effect in GARCH models (Nelson (1991) and Engle and Ng (1993)) helps to match these unconditional asymmetries. Heston and Nandi (2000, hereinafter HN), Christoffersen et al. (2008) and Christoffersen, Heston and Jacobs (2006, hereinafter CHJ) are examples of GARCH models which belong to the discrete-time affine class, and feature conditional leverage effect, while only CHJ studies also conditional skewness. Conditionally nonsymmetric return innovations are important as in option pricing for example, where heteroscedasticity and leverage effect alone do not suffice to explain the "option smirk". However, skewness in the CHJ's model is still deterministically related to volatility and both skewness and volatility undergo return shocks.

Existing affine GARCH and SV models are univariate and do not have a straightforward generalization to multiple returns and multiple volatility components without loosing their main advantage. They also focus on explaining time-series characteristics of returns and loose interest on the cross-sectional dimension. As argued in the beginning, financial assets are likely to be affected by the same economic risk factors. Then, time series properties of the factors combined with asset's systematic risk and idiosyncratic characteristics, will have important implications for the time series of asset returns. Our model belongs to the discrete-time affine class, features both stochastic volatility and skewed return innovations (SVS model), and appropriately takes part into multiple assets and multiple factors. The affine property of the model allows for a closed-form derivation of asset's risk premium and option prices under no arbitrage. We derive the risk-neutral version of our conditional APT model and show that asset's risk premium and option prices are also a function of asset's beta, idiosyncratic volatility and idiosyncratic skewness. The model then allows for a direct analysis of the sensitivity of an individual asset's option prices to asset's beta, idiosyncratic volatility and idiosyncratic skewness. The affine property of the model also leads to a GMM estimation based on exact moment conditions (see Jiang and Knight (2002) for the case of continuous-time processes, and Feunou and Tédongap (2008) for the discrete-time setting). We distinguish agent and econometrician information sets in our SV setting and provide explicit GARCH counterparts of volatility, conditional skewness and leverage effects.

Harvey and Siddique (1999, hereinafter HS) also consider a nonsymmetric conditional distribution of return with volatility and skewness as two separate factors which follow GARCH-type processes. Their autoregressive conditional skewness is a simple way to model conditional asymmetry and provides an easy methodology to estimate time-varying conditional skewness because of the availability of the likelihood function. However, the non-affinity of their model is a practical limitation, for example for solving option pricing models. The price of a european call option does not exist in closed-form, as opposed to affine GARCH models previously cited. Therefore, solving such a price would involve numerical methods or simulation techniques which are time-consuming. Our model is affine with skewness and volatility being affine combinations of the same factors. We assume that factors follow a multivariate autoregressive gamma process and that idiosyncratic noises are combinations of inverse Gaussian shocks whose variances and skewness are functions of the factors. In consequence, all conditional moments of returns are affine combinations of the factors, with coefficients given by cross-sectional characteristics of the asset. Interestingly, our discrete-time conditional APT model has several continuous-time limits, including affine jump-diffusion models with stochastic intensities.

We apply the GMM procedure suggested by Feunou and Tédongap (2008) to estimate a single factor univariate SVS model using several equities and indexes daily returns. Because we only use asset returns at this stage, this corresponds to the historical dynamics. This estimation technique permits a direct evaluation of the model performance in replicating well-known stylized facts as the persistence of volatility through the autocorrelation of squared returns as shown in Figure 4.2, the negative correlation between returns and future squared returns as shown in Figure 4.3, and the positive correlation between returns and cubed returns, especially for small stocks, as shown in Figure 4.4. We apply the unscented Kalman filter to estimate cumulants of the factors conditional on observable returns, as they are necessary to evaluate GARCH counterparts of volatility and conditional skewness. We further estimate the single factor and the two-factor SVS models using index daily option data. This corresponds to the risk-neutral dynamics. We test for a specification that allows for contemporaneous asymmetry, and also for a specification with contemporaneous normality. We compare the SVS model performance to the GARCH(1,1) model of Heston and Nandi (2000) and the IG GARCH model of Christoffersen, Heston and Jacobs (2006).

By estimating the historical dynamics, the model's parameters are significantly estimated and the model's implications are striking. We find that contemporaneous asymmetry is positive, and this result is robust across all assets under consideration. Contemporaneous asymmetry is particularly important to match sample return skewness, as well as negative cross-correlations between returns and squared returns. When contemporaneous normality is imposed, unconditional skewness is not matched. We also find that the HN GARCH and the IG GARCH models have the same performance as the SVS with contemporaneous asymmetry in matching significant return moments, but only when cross-correlations between returns and cubed returns are not important. The SVS model with contemporaneous asymmetry performs better in matching significant cross-correlations between returns and cubed returns in addition to other relevant moments of returns. The positive contemporaneous asymmetry in the SVS model dominates negative components of the conditional skewness, and leads to a positive historical conditional skewness, although unconditional skewness is negative and well matched. However, when contemporaneous normality is imposed, conditional skewness becomes negative, consistent with the CHJ's model. However, the model does not match unconditional skewness and short-term leverage effects, and tends to be rejected at conventional level of significance.

Fitting the risk-neutral dynamics using option data, we find that, explicitly allowing for contemporaneous asymmetry leads to substantial gains in option pricing, compared to existing GARCH models with equal or superior number of parameters. The single factor SVS model with contemporaneous asymmetry performs well in-sample, compared to the HN and CHJ's models. The two-factor SVS model has the best in-sample performance, which is not surprising since it nests the single factor SVS model and provides more flexibility in conditional skewness modeling. Contemporaneous asymmetry is negative and this also is not surprising since a more negative risk-neutral conditional skewness is needed to capture strong biases in short-term options. Empirical evidence shows that in-the-money call prices are relatively high compared to the Black-Scholes price, a stylized fact often represented by the well-known "volatility smirk". Our results suggest that all SVS models outperform the HN and CHJ in fitting the actual Black-Scholes implied volatility for in-the-money and deep-in-the-money calls, when one considers short-maturity contracts (less than three months).

The rest of the chapter is organized as follows. Section 4.2 presents the general affine multivariate latent factor model of asset returns. Section 4.3 introduces our discrete-time SVS model, discusses continuous-time limits, derives GARCH counterparts of volatility and skewness, and discusses the filtering method. Section 4.4 presents assets risk-neutral valuation and derives the closed-form option pricing formula consistent with SVS model. Section 4.5 estimates univariate SVS, SV and GARCH models using several equities and indexes daily returns and provides comparisons and diagnostics. Section 4.6 estimates univariate SVS, SV and GARCH models using index daily option data and provides comparisons and diagnostics. Section 4.7 concludes. The appendix contains technical material and proofs.

4.2 Discrete-Time Affine Models

4.2.1 Definition and Overview

We consider a discrete-time affine multivariate latent factor model of returns with time-varying conditional moments, characterized by its conditional cumulant generating function:

$$\Psi_t(x,y;\theta) = \ln E_t\left[\exp\left(x^\top r_{t+1} + y^\top l_{t+1}\right)\right] = A(x,y;\theta) + B(x,y;\theta)^\top l_t, \tag{4.1}$$

where $E_t[\cdot] \equiv E[\cdot \mid I_t]$ denotes the expectation conditional to the $\sigma - algebra$ generated by $(r_s, l_s), s \leq t$ (we denote this information set by I_t). $r_t = (r_{1t}, .., r_{Nt})^\top$ is the vector of observable returns, $l_t = (l_{1t}, .., l_{Kt})^\top$ is the vector of latent factors and θ is the vector of parameters.[1] Notice that the conditional moment generating is exponentially linear in the latent variables l_t. The vector process $\left(r_t^\top, l_t^\top\right)^\top$ is then semi-affine in the sense of Bates (2006). The conditional cumulant generating function of a fully affine process would be also linear in r_t. In all what follows, the parameter θ is withdrawn from functions A and B for expository purposes.

In practice, such processes are specified through the joint dynamics of observable returns r and latent factors l, from which the cumulant generating function (4.1) is obtained. All conditional cumulants of returns are affine functions of the latent factors. In particular, a latent factor l_i itself can be a specific conditional cumulant of returns, which implies some restrictions on the first derivatives of the functions $A(x,y)$ and $B_i(x,y)$. Proposition 4.2.1 below gives necessary and sufficient conditions under which the latent factor l_i is the conditional variance or the conditional asymmetry of the return r_j.

1. Darolles, Gourieroux and Jasiak (2006) study in details conditions for the stationarity in distribution of vector affine processes. The vector process $\left(r_t^\top, l_t^\top\right)^\top$ is stationary in distribution if the conditional moment-generating function $E_t\left[\exp\left(x^\top r_{t+\tau} + y^\top l_{t+\tau}\right)\right]$ converges to the unconditional moment-generating function $E\left[\exp\left(x^\top r_t + y^\top l_t\right)\right]$ as τ approaches infinity.

Proposition 4.2.1. *The factor l_i is the conditional variance of returns r_j if and only if*

$$\left.\frac{\partial^2 A(x,y)}{\partial x_j^2}\right|_{x=0,y=0} = 0 \text{ and } \left.\frac{\partial^2 B_k(x,y)}{\partial x_j^2}\right|_{x=0,y=0} = \mathbf{1}_{\{k=i\}}. \tag{4.2}$$

The factor l_i is the central conditional third moment of returns r_j if and only if

$$\left.\frac{\partial^3 A(x,y)}{\partial x_j^3}\right|_{x=0,y=0} = 0 \text{ and } \left.\frac{\partial^3 B_k(x,y)}{\partial x_j^3}\right|_{x=0,y=0} = \mathbf{1}_{\{k=i\}}. \tag{4.3}$$

Affine models of the form (4.1) with a single return and a single latent factor corresponding to the conditional variance have been widely studied in the literature as GARCH and stochastic volatility models. An extensive review of this literature is given in Shephard (2005). Example 1 below lists most common affine GARCH and SV models that have great success in the literature.

Example 1. *Stochastic Volatility.*
Discrete-time semi-affine univariate latent factor models of returns considered in several empirical studies, are the following stochastic volatility models:

$$r_{t+1} = \mu_r - \lambda_h \mu_h + \lambda_h h_t + \sqrt{h_t} u, \tag{4.4}$$

the volatility process (h_t) satisfies one of the following dynamics:

$$h_{t+1} = (1-\phi_h)\mu_h - \alpha_h + \left(\phi_h - \alpha_h \beta_h^2\right) h_t + \alpha_h \left(\varepsilon_{t+1} - \beta_h \sqrt{h_t}\right)^2, \tag{4.5}$$

$$h_{t+1} = (1-\phi_h)\mu_h + \phi_h h_t + \sigma_h \varepsilon_{t+1}, \tag{4.6}$$

$$h_{t+1} = (1-\phi_h)\mu_h + \phi_h h_t + \sigma_h \sqrt{h_t} \varepsilon_{t+1}, \tag{4.7}$$

u_{t+1} and ε_{t+1} are two i.i.d standard normal shocks, in some cases u_{t+1} and ε_{t+1} are correlated. ρ_{rh} denotes the correlation between u_{t+1} and ε_{t+1}.

The HN model corresponds to equations (4.4) and (4.5) with $\rho_{rh} = 1$. We provide also a generalization of the HN model denoted by HN-S. HN-S combines equations (4.4) and

(4.5) where ρ_{rh} is not necessarily restricted to 1. The autoregressive Gaussian volatility model corresponds to equations (4.4) and (4.6) with $\rho_{rh} = 0$. Finally the square-root volatility model corresponds to equations (4.4) and (4.7). Christoffersen, Heston and Jacobs (2006) also study an affine GARCH model specified by:

$$r_{t+1} = \alpha_h + \lambda_h h_t + \eta_h y_{t+1} \tag{4.8}$$

$$h_{t+1} = w_h + b_h h_t + c_h y_{t+1} + a_h \frac{h_t^2}{y_{t+1}} \tag{4.9}$$

where, given the available information at time t, y_{t+1} has an inverse Gaussian conditional distribution with degrees of freedom parameter h_t / η_h^2. As in the original chapter, we refer to this specification as IGGARCH.

The functions A and B characterizing the cumulant generating functions for these GARCH and SV models are explicitly given in Appendix 4.8.1. One should notice that the volatility processes (4.6) and (4.7) are not well defined since h_t can take negative values. In simulations, one should be careful when using a reflecting barrier at a small positive number to ensure positivity of simulated volatility samples.[2] This can also arise with the process (4.5) unless parameters satisfy a couple of constraints. Note also that if the volatility shock ε_{t+1} in (4.6) is allowed to be correlated to the return shock u_{t+1} in (4.4), then the model becomes non-affine. The HN-S and the IG GARCH specifications will be examined in more details in the empirical part.

A known case of a well-defined affine stochastic volatility model assumes that h_t follows an autoregressive gamma process (see Gourieroux and Jasiak (2001) for more details). However, when combined with the return process (4.4), the model presumes that within a period, return and volatility shocks are mutually independent, what appears to be a counterfactual assumption against the well-documented conditional leverage effect (Black(1976) and Christie (1982)). As discussed above, the autoregressive Gaussian dynamics (4.6), coupled with the return equation (4.4), cannot allow for leverage effect

2. Because of this limitation, autoregressive Gaussian and squared-root stochastic volatility models have been mainly explored in continuous time. To avoid negative values of h_t in simulations for examples, one relies on the true dynamics of $\ln h_t$ using the Itô lemma and works through the logarithmic model.

without the model losing its affine property. This counterfactual assumption is not required for classical SV models (Taylor (1986), Andersen (1994)) and GARCH models (Bollerslev (1986), Nelson (1991), Engle and Ng (1993)). However, these latter models are less tractable in empirical studies because of their non affine property. Then, there has been a trade-off between tractable affine models with counterfactual assumptions and non-tractable non-affine models that do not require these assumptions. In this chapter, we aim at combining both the affine model and the ability of a SV model to take into account important features of the data (fat-tailedness, asymmetry and leverage effect) in a coherent way.

4.2.2 Modeling Conditional Skewness and Leverage Effect in Affine SV Models

While return models of Example 1 are such that the vector $(r_{t+1}, h_{t+1})^\top$ of returns and volatility is affine, the conditional skewness of returns in these models is zero, the exception being the IGGARCH studied below. The literature on asset return models has evolved so far and empirical evidence upon path dependence of conditional skewness as well as its importance and contribution to risk management and asset pricing rose in recent studies. Higher moments, and especially skewness, are implicitly priced in nonlinear asset pricing models (Bansal and Viswanathan (1993), Bansal, Hsieh and Viswanathan (1993), Harvey and Siddique (2000)). Harvey and Siddique (2000) reject the constant conditional skewness assumption and argue that time-varying conditional skewness is relevant in asset pricing.

In their original paper, CHJ argue that while specification (4.4) combined with (4.7) or (4.5) generated multi-step ahead conditional skewness, single-period innovations remain Gaussian in these models which explains their failure to fit short-term options. The necessity to model return skewness has thus become of first order importance.

Harvey and Siddique (2000) model conditional skewness as a GARCH process and the IGGARCH process (described in Example 1) restricts conditional skewness to be deterministically related to volatility ($s_t = 3\eta_h/\sqrt{h_t}$). Liesenfeld and Jung (2000) introduce SV models with conditional heavy tails. However, SV models with conditional asymmetry have received less attention so far. We depart from previous literature by allowing skewness, as well as other higher order moments, to undergo unobservable shocks, which in general can be uncorrelated or linearly independent from returns and volatility shocks. Most importantly, we keep the affine property of the overall system, with a straightforward generalization to a cross-section of returns.

In this section, we explain our approach for accounting for both conditional skewness and leverage effect in a general affine univariate SV -type model. Existing affine SV models basically lead to a couple of equations of the form:

$$r_{t+1} = e(h_t) + \sqrt{h_t} u_{t+1} \tag{4.10}$$

$$h_{t+1} = m(h_t) + \sqrt{v(h_t)} \varepsilon_{t+1} \tag{4.11}$$

where u_{t+1} and ε_{t+1} are two errors with zero means and unit variances. Written in this form, the conditional skewness of returns is zero unless u_{t+1} is conditionally asymmetric. These models do not allow for the leverage effect unless the shocks u_{t+1} and ε_{t+1} are correlated. However, it is generally assumed that u_{t+1} is Gaussian and therefore unusual to assume a conditional correlation when at least one of the shocks is non-Gaussian. This is a potential limitation that typically arises when u_{t+1} is Gaussian and equation (4.11) is such that h_t is an autoregressive gamma process. Since the leverage effect is the nonzero conditional covariance between returns and volatility, projecting r_{t+1} onto h_{t+1} should lead to a nonzero slope coefficient. Therefore, we suggest to account for skewness and leverage effect in asset returns by projecting returns r_{t+1} onto volatility h_{t+1} and characterizing the projection error. This will basically lead to a return equation of the form:

$$r_{t+1} = g(h_t) + \beta h_{t+1} + \sqrt{h_t - \beta^2 v(h_t)} u_{t+1} \tag{4.12}$$

where u_{t+1} is an error with mean zero and unit variance. One could still endow u_{t+1} with a suitable distribution conditional on $\langle h_{t+1}, I_t \rangle$ such that combining (4.11) with (4.12) leads to an affine stochastic volatility model of asset returns. The model will now account for the leverage effect through β. The conditional skewness will also depend on β as well as on the asymmetry of the shock u_{t+1} conditional on $\langle h_{t+1}, I_t \rangle$, if any. We refer to the asymmetry of observable returns conditional on current factors and past information as the contemporaneous asymmetry.

It is easier to conceive a semi-affine one-factor SV model as in Example 1, with a directly specified equation for volatility dynamics, precisely because of tractable properties of the standard normal distribution appearing in both return and volatility shocks. However, it is more challenging to think of a semi-affine two-factor model with stochastic skewness as additional factor, such that both equations for volatility and skewness dynamics are directly specified. The reason is that, while conditional asymmetry of returns appears to be a necessary and sufficient condition to generate time-variation in conditional skewness, asymmetric distributions are not as tractable as the normal distribution. A strategy to get equations which explicitly characterize the joint dynamics of returns, volatility and skewness would be to first specify a semi-affine two-factor model with arbitrary linearly independent latent factors, and:

- find volatility and conditional skewness in terms of the two factors,
- then, invert the previous relationship to determine the two factors in terms of volatility and skewness,
- and finally, replace the factors in the initial return model to get the joint dynamics of returns, volatility and skewness.

In the next section, we develop a semi-affine multivariate latent factor model of returns such that both conditional variance h_t and conditional skewness s_t are stochastic. Moreover, the vector $\left(r_{t+1}, h_{t+1}, s_{t+1} h_{t+1}^{3/2}\right)^\top$ is affine in the case of a single return and two linearly independent latent factors.

4.3 An Affine Multivariate Latent Factor Model with Stochastic Skewness

4.3.1 General Setup

The dynamics of returns in our model are built upon shocks drawn from a standardized inverse Gaussian distribution. The cumulant generating function of a discrete random variable which follows a standardized inverse Gaussian distribution of parameter

s, denoted $SIG(s)$, is given by:

$$\psi(u;s) = \ln E[\exp(uX)] = -3s^{-1}u + 9s^{-2}\left(1 - \sqrt{1 - \frac{2}{3}su}\right). \quad (4.13)$$

For such a random variable, one has $E[X] = 0$, $E\left[X^2\right] = 1$ and $E\left[X^3\right] = s$, meaning that s is the skewness of X. In addition to the fact that the SIG distribution is directly parameterized by its skewness, the limiting distribution when the skewness s tends to zero is the standard normal distribution, that is $SIG(0) \equiv \mathcal{N}(0,1)$. This particularity makes the SIG an ideal building block for studying departures from normality.

For each variable in all what follows, the time subscript denotes the date from which the value of the variable is observed by the economic agent. We assume that components of the vector r_t of N returns on financial assets follow the dynamics:

$$r_{j,t+1} = \ln\frac{S_{j,t+1}}{S_{j,t}} = \mu_{j0} + \sum_{i=1}^{K}\lambda_{ji}\left(\sigma_{it}^2 - \mu_i\right) + \sum_{i=1}^{K}\beta_{ji}\left(\sigma_{i,t+1}^2 - \mu_i\right) + \sum_{i=1}^{K}\gamma_{ji}\sigma_{i,t+1}u_{ji,t+1} \quad (4.14)$$

where S_{jt} is the price of the j^{th} asset and $u_{ji,t+1} \mid \left\langle\sigma_{t+1}^2, I_t\right\rangle \sim SIG\left(\eta_{ji}\left(\gamma_{ji}\sigma_{i,t+1}\right)^{-1}\right)$. The components of the latent vector σ_t^2 are K linearly independent positive factors driving all returns' dynamics. For identification, we impose $\gamma_{1i} = 1$, $\forall i$. The NK return shocks $u_{ji,t+1}$ are mutually independent conditionally on $\left\langle\sigma_{t+1}^2, I_t\right\rangle$. If $\eta_{ji} = 0$, then $u_{ji,t+1}$ is a standard normal shock. The time t information set I_t contains past realizations of returns $\underline{r_t} = \{r_t, r_{t-1}, \ldots\}$ and latent factors $\underline{\sigma_t^2} = \{\sigma_t^2, \sigma_{t-1}^2, \ldots\}$. The return dynamics (4.14) can also be written in vector forms:

$$r_{j,t+1} = \delta_{jt} + \beta_j^{\top}\sigma_{t+1}^2 + \sigma_{t+1}^{\top}\left(\gamma_j u_{j,t+1}\right) \text{ or } r_{t+1} = \delta_t + \beta^{\top}\sigma_{t+1}^2 + \left(\gamma u_{t+1}\right)^{\top}\sigma_{t+1} \quad (4.15)$$

where $\delta_{jt} = \mu_{j0} - (\lambda_j + \beta_j)^{\top}\mu + \lambda_j^{\top}\sigma_t^2$ and $\delta_t = \mu_0 - (\lambda + \beta)^{\top}\mu + \lambda^{\top}\sigma_t^2$. The vector μ is the unconditional mean of the stationary process σ_t^2. In consequence μ_0 is the vector of unconditional expected returns. λ, β and η are $K \times N$ matrices such that

$\lambda^\top = [\lambda_{ji}]$, $\beta^\top = [\beta_{ji}]$ and $\eta^\top = [\eta_{ji}]$, and λ_j, β_j and η_j are the j^th column of the matrices λ, β and η respectively. Similarly, γu_{t+1} is the $K \times N$ matrix process such that $(\gamma u_{t+1})^\top = [\gamma_{ji} u_{ji,t+1}]$ and $\gamma_j u_{j,t+1}$ represents the j^{th} column of γu_{t+1}.

Under previous assumptions on u_{t+1}, the cumulant generating function of returns conditional to $\langle \sigma_{t+1}^2, I_t \rangle$ is given by:

$$\ln E\left[\exp\left(x^\top r_{t+1}\right) \mid \sigma_{t+1}^2, I_t\right] = x^\top \delta_t + \sum_{i=1}^{K} \sum_{j=1}^{N} \left(\beta_{ji} x_j + \psi\left(x_j; \eta_{ji}\right) \gamma_{ji}^2\right) \sigma_{i,t+1}^2. \quad (4.16)$$

The process σ_t^2 is assumed to be affine with the conditional cumulant generating function

$$\Psi_t^\sigma(y) = \ln E\left[\exp\left(y^\top \sigma_{t+1}^2\right) \mid I_t\right] = a(y) + b(y)^\top \sigma_t^2. \quad (4.17)$$

In this case, the vector $\left(r_{t+1}^\top, \left(\sigma_{t+1}^2\right)^\top\right)^\top$ is semi-affine in the sense of Bates (2006). Its conditional cumulant generating function is given by:

$$\Psi_t(x,y) = \ln E\left[\exp\left(x^\top r_{t+1} + y^\top \sigma_{t+1}^2\right) \mid I_t\right] = A(x,y) + B(x,y)^\top \sigma_t^2,$$

with

$$A(x,y) = \left(\mu_0 - (\lambda + \beta)^\top \mu\right)^\top x + a(f(x,y)) \quad (4.18)$$

$$B(x,y) = \lambda x + b(f(x,y)), \quad (4.19)$$

where $f(x,y) = (f_1(x,y_1), .., f_K(x,y_K))^\top$ with $f_i(x,y_i) = y_i + \sum_{j=1}^{N} \left(\beta_{ji} x_j + \psi\left(x_j; \eta_{ji}\right) \gamma_{ji}^2\right)$. Since the factors are positive, we assume that the vector σ_t^2 follows a multivariate autoregressive gamma process. This process also represents the discrete-time counterpart to continuous-time multivariate square root processes that have previously been examined in the literature.[3] Its log conditional Laplace transform has the form (4.17) with:

$$a(y) = -\sum_{i=1}^{K} v_i \ln(1-\alpha_i y_i) \text{ and } b_i(y) = \sum_{j=1}^{K} \frac{\phi_{ij} y_j}{1-\alpha_j y_j}. \tag{4.20}$$

The $K \times K$ matrix $\Phi = [\phi_{ij}]$ represents the persistence matrix of the vector σ_t^2 and the autoregressive gamma processes σ_{it}^2 are mutually correlated if the off-diagonal elements of Φ are nonzero. The factors are mutually independent conditional on I_t if the off-diagonal elements are zero. In this latter case we note $\phi_i = \phi_{ii}$. In the single factor case, the factor σ_{1t}^2 has the conditional cumulant generating function $\psi_{1t}^{\sigma}(y_1) = a(y_1) + b_1(y_1)\sigma_{1t}^2$, where $a(y_1) = -v_1 \ln(1-\alpha_1 y_1)$ and $b_1(y_1) = \phi_1 y_1 / (1-\alpha_1 y_1)$. The parameter ϕ_1 is the persistence of the factor and the parameters v_1 and α_1 are related to persistence and unconditional mean μ_1 and variance ω_1 as $v_1 = \mu_1^2/\omega_1$ and $\alpha_1 = (1-\phi_1)\omega_1/\mu_1$.

Although our empirical focus will be on the time series dynamics of a single return, it is important to notice that equation (4.15) is a multifactor conditional arbitrage-pricing model. In fact, we assume that a true conditional multifactor representation of expected returns in the cross-section is such that log returns are linear in the factors and the idiosyncratic noise. The vector β_j represents the loadings of asset j on the factors, and this asset's conditional beta is time-invariant. The factors are heteroscedastic and the idiosyncratic noise is a combination of independent heteroscedastic and asymmetric shocks. This constitutes a substantial depart from previous literature, as the true data generating process in existing APT models is, in general, specified such that factors as well as idiosyncratic shocks are implicitly or explicitly homoscedastic and normally distributed. Considering latent factors is also appealing as, in the original APT model of Ross (1976), factors are unknown. Also, focusing on positive factors in not restrictive as any arbitrary economic factor, say F_t, can be written as a difference of two nonnegative factors, say $\sigma_{1t}^2 - \sigma_{2t}^2$, where $\sigma_{1t}^2 = \max(F_t, 0)$ and $\sigma_{2t}^2 = \max(-F_t, 0)$.

3. See for example Singleton (2001).

4.3.2 Expected Returns, Conditional Variance, Skewness and Leverage Effects

In the previous section, we do not model directly volatility and conditional skewness as well as other higher moments of returns. Instead, we relate returns to a finite number of stochastic linearly independent positive factors. In this section, we relate expected returns, conditional variance, skewness and leverage effects to these factors and discuss important features of the model.

Proposition 4.3.1. *Conditional on I_t, the mean μ^r_{jt}, the variance h_{jt} and the skewness s_{jt} of returns r_j are expressed as follows:*

$$\mu^r_{jt} = \mu_{j0} - (\lambda_j + \beta_j)^\top \mu + \lambda_j^\top \sigma_t^2 + \beta_j^\top m_t^\sigma = c_{j0,\mu} + \sum_{i=1}^{K} c_{ji,\mu}\sigma_{it}^2, \tag{4.21}$$

$$h_{jt} = \beta_j^\top V_t^\sigma \beta_j + \left(\gamma_j^2\right)^\top m_t^\sigma = c_{j0,h} + \sum_{i=1}^{K} c_{ji,h}\sigma_{it}^2, \tag{4.22}$$

$$s_{jt} h_{jt}^{3/2} = \left(\beta_j \otimes \beta_j\right)^\top S_t^\sigma \beta_j + 3\left(\gamma_j^2\right)^\top V_t^\sigma \beta_j + \left(\gamma_j^2 \eta_j\right)^\top m_t^\sigma = c_{j0,s} + \sum_{i=1}^{K} c_{ji,s}\sigma_{it}^2, \tag{4.23}$$

where the coefficients $c_{jn,l}$ depend on model's parameters,

$$m_t^\sigma = E\left[\sigma_{t+1}^2 \mid I_t\right],\ V_t^\sigma = E\left[\left(\sigma_{t+1}^2 - m_t^\sigma\right)\left(\sigma_{t+1}^2 - m_t^\sigma\right)^\top \mid I_t\right]$$

and

$$S_t^\sigma = E\left[\left(\left(\sigma_{t+1}^2 - m_t^\sigma\right) \otimes \left(\sigma_{t+1}^2 - m_t^\sigma\right)\right)\left(\sigma_{t+1}^2 - m_t^\sigma\right)^\top \mid I_t\right].$$

This proposition is established by taking the first, second and third derivative of the conditional cumulant generating function of returns r_{jt} given by equations 4.18 and 4.19.

The linearity of expected returns, volatility and conditional asymmetry of returns in terms of the factors results from the fact that components of the vector m_t^σ, and of the matrices V_t^σ and S_t^σ, are also linear in the σ_{it}^2's. This is a consequence of the affine structure of the process σ_t^2. Also, note that the bivariate vector $\left(h_{jt}, s_{jt} h_{jt}^{3/2}\right)^\top$ is not deterministically related to contemporaneous and past returns as for GARCH-type processes (Harvey and Siddique (1999) and Feunou and Tédongap (2009)), as well as many other authors.[4] For this reason, we label the present model stochastic volatility and skewness (SVS).

Proposition 4.3.2. *Conditional on I_t, the covariance between returns r_j and volatility h_j (leverage effect) and the covariance between returns r_j and skewness $s_j h_j^{3/2}$ are given by:*

$$Cov\left(r_{j,t+1}, h_{j,t+1} \mid I_t\right) = c_{j,h}^{\top} V_t^{\sigma} \beta_j = c_{j0,rh} + \sum_{i=1}^{K} c_{ji,rh} \sigma_{it}^2 = c_{j0,rh} + c_{j,rh}^{\top} \sigma_t^2 \tag{4.24}$$

$$Cov\left(r_{j,t+1}, s_{j,t+1} h_{j,t+1}^{3/2} \mid I_t\right) = c_{j,s}^{\top} V_t^{\sigma} \beta_j = c_{j0,rs} + \sum_{i=1}^{K} c_{ji,rs} \sigma_{it}^2 = c_{j0,rs} + c_{j,rs}^{\top} \sigma_t^2 \tag{4.25}$$

where the coefficients $c_{jn,rl}$ depend on model's parameters.

It is not surprising that the parameter β_j governs conditional leverage effect as it represents the slope of linear projection of returns on current factors. For a negative correlation between spot returns and variance, and consistently with the postulate of Black (1976) and the leverage effect documented by Christie (1982) and others, the parameter β_j may be expected to be negative, in particular for the single-factor case.

It should be noted that, in our SVS model, although the parameter η_j dictates contemporaneous asymmetry of returns (that is, the asymmetry of returns conditional on current factors and past information), it is not the only parameter determining conditional skewness as shown in equation (4.23). The parameter β_j, which alone characterizes leverage effect, also plays a central role in generating conditional asymmetry in returns, even if returns are contemporaneously normally distributed, that is when $\eta_j = 0$. In contrast to existing SV models with leverage effect as discussed in Example 1, where leverage effect generates skewness only in multiple-period returns, in our setting, leverage effect invokes skewness in single-period returns as well. If $\beta_j = 0$, there is no leverage effect. In addition, there is also no skewness unless $\eta_j \neq 0$. Then, contemporaneous asymmetry

4. Hansen (1994), Jondeau and Rockinger (2003), and Leon, Rubio and Serna (2005), do not explicitly model conditional skewness, but related shape parameters of the conditional return distribution using GARCH-type dynamics.

in this model reinforces the effect of the leverage parameter β_j in generating conditional skewness. In other words, time-varying conditional skewness in this model is a combination of conditional leverage effect (through β_j) and contemporaneous asymmetry (through η_j).

To better understand the flexibility of the SVS model in generating conditional skewness, we refer to the single-factor SVS. Equation (4.23) shows that conditional skewness is the sum of three terms. The first two terms have the same sign, which is the sign of β_j as components of the matrices V_t^{σ} and S_t^{σ} are positive. The last term has the sign of η_j as m_t^{σ} is positive. As discussed previously, a negative value of β_j is necessary to generate the documented negative leverage effect. If so, the first two terms in (4.23) are negative. The sign of conditional skewness will depend on contemporaneous asymmetry η_j. If η_j is zero or negative, then conditional skewness is negative over time as in the IGGARCH model. This also arises if η_j is positive, but not enough that the third term dominates the first two. If it does, then conditional skewness is positive over time. Also remark that skewness of the j^{th} financial asset may change sign over time if η_j is positive and such that $c_{j0,s}c_{j1,s} < 0$. This will then be consistent with the empirical evidence in Harvey and Siddique (1999) that conditional skewness changes sign over time. The findings in Feunou and Tédongap (2009) also suggest that returns's innovations are conditionally normal or weakly positively skewed most of the time, but undergo unfrequent and large drops in conditional skewness. However, it is recognized in the literature that a negative conditional skewness is particularly important for explaining strong biases in option prices.

While $\sigma_{1t}^2, .., \sigma_{Kt}^2$ are the primitive predictive variables in our SVS model, predictability when $K \geq 2$ can also be related to conditional variance and skewness which are economically interpretable. For example, empirical facts tend to support that an increase in volatility drives up expected returns, as the economic agent requires more premium when it becomes more riskier to invest in stocks. The economic agent dislikes high return volatility, as he prefers positive skewness, i.e; extreme positive returns are more likely to realize than extreme negative returns. Therefore, economic agents would pay a premium in exchange of positive skewness, and require a premium to compensate for negative skewness. In the two-factor case, $K=2$, and if $c_{j1,h}c_{j2,s} \neq c_{j1,s}c_{j2,h}$, one can invert relations (4.22) and (4.23) to obtain σ_{1t}^2 and σ_{2t}^2 in terms of h_{jt} and $s_{jt}h_{jt}^{3/2}$. By plugging these relations in (4.14), one gets returns in terms of volatility and skewness, instead of initial factors. The IGGARCH does not disentangle skewness from volatility whereas the two-factor SVS disentangles these two measures while maintaining a semi-affine structure of the model. This separation results from the decomposition of return shocks into two linearly independent IG components with individual conditional variances having specific affine dynamics.

4.3.3 Continuous-Time Limits

Although the present SVS model is written in discrete time, we are interested in its continuous-time limits. These limits are useful to make the link with the huge literature on stock's returns modeling which have been casted in continuous time. Following several papers which derive continuous-time limits of discrete-time processes (among which we can cite Nelson (1991), Foster and Nelson (1994)), we write the model for a small time interval, and let the time interval shrink to zero. For a small time interval Δ, the return's equation (4.14) becomes:

$$\ln\frac{S_{j,t+\Delta}}{S_{j,t}} = \mu_{j0} + \sum_{i=1}^{K}\lambda_{ji}\left(\sigma_{i,t}^2 - \mu_i\right) + \sum_{i=1}^{K}\beta_{ji}\left(\sigma_{i,t+\Delta}^2 - \mu_i\right) + \sum_{i=1}^{K}\gamma_{ji}\sigma_{i,t+\Delta}u_{ji,t+\Delta}. \quad (4.26)$$

For simplicity we assume that the factors are independent. Let us consider the following parameterizations

$$\mu_{j0}(\Delta) = \mu_{j0}\Delta,\ \beta_{ji}(\Delta) = \frac{\beta_{ji}}{\Delta},\ \lambda_{ji}(\Delta) = \lambda_{ji} - \beta_{ji}\frac{\exp(-\kappa_i\Delta)}{\Delta}$$
$$\phi_i(\Delta) = \exp(-\kappa_i\Delta),\ \alpha_i(\Delta) = \frac{\omega_i\Delta^2}{2\exp(-\kappa_i\Delta)},\ \nu_i(\Delta) = \theta_i\left(1-\exp(-\kappa_i\Delta)\right)\frac{2\exp(-\kappa_i\Delta)}{\omega_i\Delta}.$$

Letting $v_{it} = \sigma_{i,t}^2 / \Delta$ represents factor per unit time, by taking the limit as Δ approaches zero, it follows that v_{it} converges weakly to the following square-root process:

$$dv_{it} = \kappa_i (\theta_i - v_{it})\, dt + \sqrt{\omega_i}\sqrt{v_{it}} dB_{it}, \tag{4.27}$$

where dB_{it} is a Wiener process. Similar arguments can be found in HN and CHJ. We also refer to Gourieroux and Jasiak (2006) where it is shown that the discrete time univariate autoregressive gamma process converge weakly to a square-root process. $\sigma_{i,t+\Delta}\sqrt{\Delta} u_{ji,t+\Delta}$ has two different continuous time limits depending on the value of the parameter η_{ji}, see CHJ for more details.

If $\eta_{ji} = 0$ then $\sigma_{i,t+\Delta}\sqrt{\Delta} u_{ji,t+\Delta}$ converges to $\sqrt{v_{it}} dW_{ji,t}$ as Δ shrinks to zero, where $W_{ji,t}$ is a Wiener process. In contrast if $\eta_{ji} \neq 0$, then $\sigma_{i,t+\Delta}\sqrt{\Delta} u_{ji,t+\Delta}$ converges to $-\left(3\gamma_{ji} v_{it} / \eta_{ji}\right) dt + \left(\eta_{ji} / 3\gamma_{ji}\right) dJ_{ji,t}$ as Δ shrinks to zero, where $J_{ji,t}$ is a pure jump inverse Gaussian process with degree of freedom $9\gamma_{ji}^2 v_{it} \Big/ \eta_{ji}^2$ on interval $[t, t+dt]$. We then show that the limiting distribution of the SVS model in continuous time is a stochastic volatility process where the return is a sum of diffusion and pure jump inverse Gaussian processes:

$$\begin{aligned} d\ln S_{jt} = & \left[\mu_{j0} + \sum_{i=1}^{K} \lambda_{ji} (v_{it} - \theta_i) - \sum_{i:\eta_{ji} \neq 0} \frac{3\gamma_{ji}^2 v_{it}}{\eta_{ji}} \right] dt + \sum_{i=1}^{K} \beta_{ji} \sqrt{\omega_i}\sqrt{v_{it}} dB_{it} \\ & + \sum_{i:\eta_{ji}=0} \gamma_{ji} \sqrt{v_{it}} dW_{ji,t} + \sum_{i:\eta_{ji}\neq 0} \frac{\eta_{ji}}{3} dJ_{ji,t}. \end{aligned} \tag{4.28}$$

4.3.4 GARCH versus SVS: Filtering the Unobservable Factors

In GARCH models, the information set I_t is exactly the sigma algebra generated by $(r_s, s \leq t)$ ($\underline{r_t}$ hereafter), hence both the economic agent and the econometrician view the same information. This is a strong assumption that is implicit in GARCH models. In SV models in general, and the present SVS model in particular, the econometrician does not observe $\underline{\sigma_t^2}$ (the sigma algebra generated by $(\sigma_s^2, s \leq t)$), only known by the economic agent. While the moments in Proposition 4.3.1 are conditional on information $I_t = \underline{r_t} \bigcup \underline{\sigma_t^2}$, one can also derive their GARCH counterparts, meaning same return moments now conditional on econometrician's information, $\underline{r_t}$ only. Without loss of generality, we derive these conditional moments for the case of a single return ($N = 1$). However, the formulas can be generalized to multiple returns as well. Let $\mu_t^{r,G}$, h_t^G and s_t^G respectively denote the mean, the variance and the skewness of r_{t+1} conditional on $\underline{r_t}$. One has:

$$\mu_t^{r,G} = c_{0\mu} + c_\mu^\top G_{\mu t} \text{ and } h_t^G = c_{0h} + c_h^\top G_{\mu t} + c_\mu^\top G_{ht} c_\mu, \tag{4.29}$$

$$s_t^G \left(h_t^G\right)^{3/2} = c_{0s} + c_s^\top G_{\mu t} + c_\mu^\top G_{ht} c_h + \left(c_\mu \otimes c_\mu\right)^\top G_{st} c_\mu \tag{4.30}$$

where

$$G_{\mu t} = E\left[\sigma_t^2 \mid \underline{r_t}\right] \text{ and } G_{ht} = E\left[\sigma_t^2 \left(\sigma_t^2\right)^\top \mid \underline{r_t}\right] - E\left[\sigma_t^2 \mid \underline{r_t}\right] E\left[\sigma_t^2 \mid \underline{r_t}\right]^\top, \tag{4.31}$$

$$\begin{aligned} G_{st} = E&\left[\left(\sigma_t^2 \otimes \sigma_t^2\right)\left(\sigma_t^2\right)^\top \mid \underline{r_t}\right] - 3E\left[\left(\sigma_t^2 \otimes \sigma_t^2\right) \mid \underline{r_t}\right] E\left[\sigma_t^2 \mid \underline{r_t}\right]^\top \\ &+ 2\left(E\left[\sigma_t^2 \mid \underline{r_t}\right] \otimes E\left[\sigma_t^2 \mid \underline{r_t}\right]\right) E\left[\sigma_t^2 \mid \underline{r_t}\right]^\top, \end{aligned} \tag{4.32}$$

are mean, variance and third central moment of the latent vector σ_t^2 conditional upon observed returns $\underline{r_t}$.

Disentangling the agents and the econometricians information sets in return modeling can be crucial. In the single-factor SVS model, return's conditional variance and third central moment are perfectly correlated to the agent, whereas it is the contrary to the econometrician, unless returns are unpredictable by the factor ($c_\mu = 0$). When $c_\mu \neq 0$, the SVS model generates, conditional to observed returns, an asymmetry that is not perfectly correlated to the variance, although this correlation remains high for a persistent factor. In contrast, conditional variance and third central moment are perfectly correlated in the IGGARCH, given past observed returns.

GARCH counterparts of leverage effect and conditional covariance between returns and skewness are defined by:

$$Cov\left(r_{t+1}, h_{t+1}^G \mid \underline{r_t}\right) \text{ and } Cov\left(r_{t+1}, s_{t+1}^G \left(h_{t+1}^G\right)^{3/2} \mid \underline{r_t}\right).$$

These two quantities are difficult to express in terms of the moments of the latent vector σ_t^2 conditional on observed returns $\underline{r_t}$. Instead, we consider the following two quantities:

$$Cov\left(r_{t+1}, h_{t+1} \mid \underline{r_t}\right) = c_{0,rh} + c_{rh}^{\top} G_{\mu t} + c_{\mu}^{\top} G_{ht} \Phi^{\top} c_h \tag{4.33}$$

$$\text{and, } Cov\left(r_{t+1}, s_{t+1} h_{t+1}^{3/2} \mid \underline{r_t}\right) = c_{0,rs} + c_{rs}^{\top} G_{\mu t} + c_{\mu}^{\top} G_{ht} \Phi^{\top} c_s, \tag{4.34}$$

where Φ represents the persistence matrix of the latent vector.

We now describe how to compute expectations in (4.31) and (4.32). Various strategies to deal with non-linear state-space systems have been proposed in the filtering literature: the Extended Kalman Filter, the Particle Filter and more recently the Unscented Kalman Filter that we apply in this chapter.[5] Since our SVS model has the standard state space representation, on can use Kalman Filter-based techniques to compute $G_{\mu t}$, G_{ht} and G_{st}. As these methods will not guarantee that $E\left[\sigma_{it}^2 \mid \underline{r_t}\right]$ is positive, it would be more convenient to filter $\omega_{it} = \ln \sigma_{it}^2$. Let $\omega_t = (\omega_{1t}, .., \omega_{kt})^{\top}$.

The basic framework of Kalman filter techniques involves estimation of the state of a discrete-time nonlinear dynamic system of the form:

$$r_{t+1} = H\left(\omega_{t+1}, u_{t+1}^*\right) \tag{4.35}$$

$$\omega_{t+1} = F\left(\omega_t, \varepsilon_{t+1}^*\right), \tag{4.36}$$

where u_{t+1}^* and ε_{t+1}^* are not necessarily but conventionally two Gaussian noises. For this reason, we log-normally approximate our model, which in the one-factor case leads to:

$$H\left(\omega_{1,t+1}, u_{1,t+1}^*\right) = \mu_0 + \beta_1 \exp\left(\omega_{1,t+1}\right) + \exp\left(\frac{\omega_{1,t+1}}{2}\right) \left[\exp\left(\ln\left(\frac{9}{s\left(\omega_{1,t+1}\right)\sqrt{s\left(\omega_{1,t+1}\right)^2 + 9}}\right) + \sqrt{\ln\left(\frac{s\left(\omega_{1,t+1}\right)^2 + 9}{9}\right)} u_{1,t+1}^*\right) - \frac{3}{s\left(\omega_{1,t+1}\right)}\right] \tag{4.37}$$

5. See Leippold and Wu (2003) and Bakshi, Carr and Wu (2005) for application in finance, Julier et al. (1995) and Julier and Uhlmann (1996) for details and Wan and van der Merwe (2001) for textbook treatment.

and

$$F\left(\omega_{1t},\varepsilon_{1,t+1}^{*}\right)=\ln\left(\frac{m\left(\omega_{1t}\right)^{2}}{\sqrt{m\left(\omega_{1t}\right)^{2}+v\left(\omega_{1t}\right)}}\right)+\sqrt{\ln\left(\frac{m\left(\omega_{1t}\right)^{2}+v\left(\omega_{1t}\right)}{m\left(\omega_{1t}\right)^{2}}\right)}\varepsilon_{1,t+1}^{*},\tag{4.38}$$

where

$$\begin{aligned}
s\left(\omega_{1,t+1}\right)&=\eta_{1}\exp\left(-\frac{\omega_{1,t+1}}{2}\right)\\
m\left(\omega_{1t}\right)&=\left(1-\phi_{1}\right)\mu_{1}+\phi_{1}\exp\left(\omega_{1t}\right)\\
v\left(\omega_{1t}\right)&=\left(1-\phi_{1}\right)^{2}\sigma_{1}^{2}+\frac{2\left(1-\phi_{1}\right)\phi_{1}\sigma_{1}^{2}}{\mu_{1}}\exp\left(\omega_{1t}\right).
\end{aligned}$$

Details on this log-normal approximation of the one-factor SVS model are provided in appendix 4.8.3.

Let $\omega_{t|\tau}$ be the estimate of ω_t using returns up to and including time τ, $\underline{r_\tau}$, and let $P_{t|\tau}^{\omega\omega}$ be its covariance. Given the joint distribution of $\left(\omega_t^{\top},u_{t+1}^{*\top},\varepsilon_{t+1}^{*\top}\right)^{\top}$ conditionally to $\underline{r_t}$, the filter predicts what future state and returns will be using process models. Optimal predictions and associated mean squared errors are given by:

$$\omega_{t+1|t}=E\left[\omega_{t+1}\mid\underline{r_t}\right]=E\left[F\left(\omega_t,\varepsilon_{t+1}^{*}\right)\mid\underline{r_t}\right]\tag{4.39}$$

$$r_{t+1|t}=E\left[r_{t+1}\mid\underline{r_t}\right]=E\left[H\left(\omega_{t+1},\varepsilon_{t+1}^{*}\right)\mid\underline{r_t}\right]\tag{4.40}$$

$$P_{t+1|t}^{\omega\omega}=E\left[\left(\omega_{t+1}-\omega_{t+1|t}\right)\left(\omega_{t+1}-\omega_{t+1|t}\right)^{\top}\mid\underline{r_t}\right]\tag{4.41}$$

$$P_{t+1|t}^{rr}=E\left[\left(r_{t+1}-r_{t+1|t}\right)\left(r_{t+1}-r_{t+1|t}\right)^{\top}\mid\underline{r_t}\right]\tag{4.42}$$

$$P_{t+1|t}^{\omega r}=E\left[\left(\omega_{t+1}-\omega_{t+1|t}\right)\left(r_{t+1}-r_{t+1|t}\right)^{\top}\mid\underline{r_t}\right].\tag{4.43}$$

The join distribution of $\left(\omega_t^{\top},u_{t+1}^{*\top},\varepsilon_{t+1}^{*\top}\right)^{\top}$ conditionally to $\underline{r_t}$ is conventionally assumed Gaussian. To the contrary of the standard Kalman filter where the functions H and F are linear, the precise values of the conditional moments (4.39) to (4.43) can not be determined analytically in our model because the functions H and F are strongly nonlinear. Alternative methods produce approximations of these conditional moments.

The Extended Kalman Filter linearizes the functionals H and F in the state-space system to determine the conditional moments analytically. While this simple linearization maintains a first-order accuracy, it can introduce large errors in the true posterior mean and covariance of the transformed random variable which may lead to sub-optimal performance and sometimes to divergence of the filter. The Particle Filter uses Monte-Carlo simulations of the relevant distributions to get estimates of moments. In contrast, the Uncented Kalman Filter adresses the approximation issues of the Extended Kalman filter and the computational issues of the Particle Filter. It represents the distribution of $\left(\omega_t^\top, u_{t+1}^{*\top}, \varepsilon_{t+1}^{*\top}\right)^\top$ conditional on $\underline{r_t}$ by a minimal set of carefully chosen points. This reduces the computational burden but maintain second-order accuracy. Details on the Unscented Kalman Filter are provided in appendix 4.8.4.

The next step is to use current returns to update estimate (4.39) of the state. In the Kalman filter, a linear update rule is specified, where the weights are chosen to minimize the mean squared error of the estimate. This rule is given by:

$$\omega_{t+1|t+1} = \omega_{t+1|t} + K_{t+1}\left(r_{t+1} - r_{t+1|t}\right) \tag{4.44}$$

$$P^{\omega\omega}_{t+1|t+1} = P^{\omega\omega}_{t+1|t} - K_{t+1} P^{rr}_{t+1|t} K^\top_{t+1} \tag{4.45}$$

$$K_{t+1} = P^{\omega r}_{t+1|t}\left(P^{rr}_{t+1|t}\right)^{-1}. \tag{4.46}$$

Once the Kalman recursion outlined above delivers the estimates $\omega_{t|t}$ and $P^{\omega\omega}_{t|t}$ for the whole sample, the statistics $G_{\mu t}$, G_{ht} and G_{st} can be computed using approximations of moments of a nonlinear function of a Gaussian random variable. Without loss of generality, appendix 4.8.5 derives corresponding formulas in the univariate case.

4.4 Asset Pricing with Stochastic Skewness

In the context of asset and derivative pricing, one would like to characterize a probability measure under which the expected gross return on any risky security equals the gross return on a safe security. It is sufficient to define a change of measure $Z_{t,t+1}$ from historical to risk-neutral, or equivalently to specify a stochastic discount factor $M_{t,t+1}$ from which investors value financial payoffs (see Gourieroux and Monfort (2006) and Christoffersen et al. (2009)). The change of measure $Z_{t,t+1}$ should satisfy the following conditions:

$$E\left[Z_{t,t+1} \mid I_t\right] = 1 \text{ and } E^*\left[\exp\left(r_{j,t+1}\right) \mid I_t\right] \equiv E\left[Z_{t,t+1}\exp\left(r_{j,t+1}\right) \mid I_t\right] = \exp\left(r_{f,t+1}\right), \tag{4.47}$$

where $r_{j,t+1}$ and $r_{f,t+1}$ refer to the j^{th} risky return and the risk-free rate from date t to date $t+1$, respectively, and where $E^*\left[\cdot \mid I_t\right] \equiv E\left[Z_{t,t+1}\left(\cdot\right) \mid I_t\right]$ denotes the risk-neutral expectation associated with the density $Z_{t,t+1}$.

Given the historical return dynamics (4.14), we would like to find a change of measure such that risk-neutral return dynamics is also an affine SVS model similar to (4.14). Exploiting the affine property, we assume that the change of measure $Z_{t,t+1}$ is given by:

$$Z_{t,t+1} = \exp\left(-A\left(\kappa,\pi\right) - B\left(\kappa,\pi\right)^{\top}\sigma_t^2 + \kappa^{\top}r_{t+1} + \pi^{\top}\sigma_{t+1}^2\right), \tag{4.48}$$

which, by definition and specification, satisfies $E\left[Z_{t,t+1} \mid I_t\right] = 1$. In appendix 4.8.2 we show that the necessary and sufficient condition (on the change of measure) such that the SVS model is preserved under both measures is $f\left(\kappa,\pi\right) = 0$ where $f(\cdot)$ is defined in equations 4.18 and 4.19. It implies that

$$Z_{t,t+1} = \exp\left(-\kappa^{\top}\delta_t + \kappa^{\top}r_{t+1} + \pi^{\top}\sigma_{t+1}^2\right) \text{ and } E\left[Z_{t,t+1} \mid \left\langle\sigma_{t+1}^2, I_t\right\rangle\right] = 1. \tag{4.49}$$

In particular, an implication of the second equation is that the moment generating function of σ_{t+1}^2, conditional to I_t, does not change from the physical to the risk-neutral measure. Thus, the factors still follow the same multivariate autoregressive gamma under the risk-neutral dynamics. Appendix 4.8.2 finally shows that the risk-neutral dynamics of returns is given by:

$$r_{j,t+1} = r_f - a\left(q_j^2\theta_j^*\right) - \sum_{i=1}^{K} b_i\left(q_j^2\theta_j^*\right)\sigma_{it}^2 + \sum_{i=1}^{K}\left(\beta_{ji}^* q_{ji}^2\right)\sigma_{i,t+1}^2 + \sum_{i=1}^{K}\gamma_{ji}^*\sigma_{i,t+1}u_{ji,t+1}^* \tag{4.50}$$

with $u_{ji,t+1}^* \mid \langle\sigma_{t+1}^2, I_t\rangle \sim SIG\left(\eta_{ji}^*\left(\gamma_{ji}^*\sigma_{i,t+1}\right)^{-1}\right)$, and where $q_j^2\theta_j^*$ denotes the $K\times 1$ vector with components $q_{ji}^2\theta_{ji}^*$. The risk-neutral parameters are defined by:

$$q_{ji}^2 = \left(1 - (2/3)\,\eta_{ji}\kappa_j\right)^{-3/2} \text{ and } \theta_{ji}^* = \beta_{ji}^* + \psi\left(1;\eta_{ji}^*\right)\gamma_{ji}^2$$
$$\beta_{ji}^* = \left(\beta_{ji} + \psi'\left(\kappa_j;\eta_{ji}\right)\gamma_{ji}^2\right)/q_{ji}^2,\quad \eta_{ji}^* = \eta_{ji}/\left(1-(2/3)\,\eta_{ji}\kappa_j\right) \text{ and } \gamma_{ji}^* = q_{ji}\gamma_{ji}.$$

The return dynamics (4.15) and the no-arbitrage restrictions (4.47) lead to the characterization of the asset's risk premium, which in our model is given by:

$$\mu_{0,j} - r_f = -a\left(q_j^2\theta_j^*\right) + \left(\beta_j - b\left(q_j^2\theta_j^*\right)\right)^{\top}\mu. \tag{4.51}$$

In most empirical studies, ingredients of the return dynamics that are important for explaining actual time series properties of returns, and which turn out to be relevant also for explaining characteristics of observed option prices (for example leverage effects and conditional skewness), are studied separately from features that relate to actual cross-sectional properties of asset returns. We argue that time series and cross-sectional properties of returns result from the same features, and that these features should not be modeled independently.[6] We see for example that, if the factors are heteroscedastic and idiosyncratic shocks are heteroscedastic and asymmetric, as in our model, leverage effects are determined by asset's factor loadings (β_j), and conditional skewness is determined by both factor loadings and idiosyncratic skewness (η_j). In addition, no-arbitrage equilibrium restrictions imply that asset's risk premium depends both on factor

6. A similar argument can be found in Santos and Veronezi (2008). The authors argue that the equity premium puzzle and the value premium puzzle cannot be tackled independently, as any economic mechanism proposed to address one of them immediately has general equilibrium implications for the other.

loadings, idiosyncratic volatility (γ_j) and idiosyncratic skewness. Our model offers a tractable framework to address simultaneously time series and cross-sectional properties of asset returns as well as of asset's option prices.

When factors are independent ($\phi_{ij} = 0$ for $i \neq j$), it is convenient to write the risk-neutral return dynamics as:

$$\begin{aligned} r_{1,t+1} &= r_f - a^*(\theta_1^*) - \sum_{i=1}^{K} b_i^*(\theta_1^*)\sigma_{it}^{*2} + \sum_{i=1}^{K} \beta_{1i}^*\sigma_{i,t+1}^{*2} + \sum_{i=1}^{K} \sigma_{i,t+1}^* u_{1i,t+1}^*, \\ r_{j,t+1} &= r_f - a^*(\theta_j^{**}) - \sum_{i=1}^{K} b_i^*(\theta_j^{**})\sigma_{it}^{*2} + \sum_{i=1}^{K} \beta_{ji}^{**}\sigma_{i,t+1}^{*2} + \sum_{i=1}^{K} \gamma_{ji}^{**}\sigma_{i,t+1}^* u_{ji,t+1}^*, \; 2 \leq j \leq N, \end{aligned} \tag{4.52}$$

where $\sigma_{it}^* = q_{1i}\sigma_{it}$ with $q_{1i}^2 = \left(1 - (2/3)\,\eta_{1i}\kappa_1\right)^{-3/2}$, $u_{1i,t+1}^* \mid \langle \sigma_{t+1}^{*2}, I_t \rangle \sim SIG\left(\eta_{1i}^* \sigma_{i,t+1}^{*-1}\right)$ and $u_{ji,t+1}^* \mid \langle \sigma_{t+1}^{*2}, I_t \rangle \sim SIG\left(\eta_{ji}^*\left(\gamma_{ji}^{**}\sigma_{i,t+1}^*\right)^{-1}\right)$. The vector process σ_t^{*2} is a multivariate autoregressive gamma under the risk-neutral measure, with parameters of the i^{th} factor, σ_{it}^{*2}, given by:

$$\alpha_i^* = q_{1i}^2\alpha_i, \;\; \nu_i^* = \nu_i \;\text{ and }\; \phi_i^* = \phi_i.$$

Parameters in the first risk-neutral return equation ($j = 1$) are given by:

$$\beta_{1i}^* = \left(\beta_{1i} + \psi'(\kappa_1;\eta_{1i})\right)/q_{1i}^2, \;\; \eta_{1i}^* = \eta_{1i}/\left(1 - (2/3)\,\eta_{1i}\kappa_1\right) \;\text{ and }\; \theta_{1i}^* = \beta_{1i}^* + \psi(1;\eta_{1i}^*).$$

The functions $a^*(\cdot)$ and $b^*(\cdot)$ are analogue to the functions $a(\cdot)$ and $b(\cdot)$ in (4.17), and similarly characterize the cumulant generating function of the multivariate autoregressive gamma process σ_t^{*2} under the risk-neutral dynamics. Parameters in the second risk-neutral return equation ($2 \leq j \leq N$) are given by:

$$\beta_{ji}^{**} = \beta_{ji}^* q_{ji}^2/q_{1i}^2, \;\; \eta_{ji}^* = \eta_{ji}/\left(1 - (2/3)\,\eta_{ji}\kappa_1\right) \;\text{ and }\; \theta_{ji}^{**} = \beta_{ji}^{**} + \psi(1;\eta_{ji}^*)\,\gamma_{ji}^{**2}.$$

Because θ_1^* is related to β_1^* and η_1^*, and θ_j^{**} is related to β_j^{**}, γ_j^{**} and η_j^* for $2 \leq j \leq N$, the risk-neutral dynamics of every asset has K parameters less compared to its historical dynamics. This is analogous to the IGGARCH risk-neutral model of CHJ (which is a single-factor model) where the parameter governing conditional skewness is a function of other parameters. This result is in fact more general, and, as we show in this article, for a K factor model, the risk-neutral dynamics of an individual asset return has K independent parameters less compared to the physical model.

In the single return case ($N = 1$), the joint dynamics of returns and factors under the risk-neutral distribution is characterized by the following cumulant generating function:

$$\Psi_t^*(x,y) = \ln E^*\left[\exp\left(xr_{t+1} + y^\top \sigma_{t+1}^2\right) \mid I_t\right] = A^*(x,y) + B^*(x,y)^\top \sigma_t^2$$

where the functions $A^*(\cdot,\cdot)$ and $B^*(\cdot,\cdot)$ are analogue to the functions $A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ in (4.18) and (4.19) respectively. Let $\Psi_{t,t+\tau}^{*r}(x)$ denote the conditional log moment generating function of aggregate returns $\sum_{i=1}^{\tau} r_{t+i}$, under the risk-neutral measure. One has

$$E^*\left[\exp\left(x\sum_{i=1}^{\tau} r_{t+i}\right) \mid I_t\right] = \exp\left(\Psi_{t,t+\tau}^{*r}(x)\right) = \exp\left(A_r^*(x;\tau) + B_r^*(x;\tau)^\top \sigma_t^2\right),$$

where the sequence of functions $A_r^*(x;\tau)$ and $B_r^*(x;\tau)$ satisfy the following recursion:

$$A_r^*(x;\tau) = A_r^*(x;\tau-1) + A^*(x, B_r^*(x;\tau-1)) \text{ and } B_r^*(x;\tau) = B^*(x, B_r^*(x;\tau-1)),$$

with $A_r^*(x;1) = A^*(x,0)$ and $B_r^*(x;1) = B^*(x,0)$.

The price at date t of a european call option with strike price X and maturity τ, is given by

$$C\left(\tau,\sigma_t^2,\frac{X}{S_t}\right) = \exp(-r\tau)\left[S_t\left(\frac{1}{2}\exp(r\tau) + C_1\left(\tau,\sigma_t^2,\frac{X}{S_t}\right)\right) - X\left(\frac{1}{2} + C_2\left(\tau,\sigma_t^2,\frac{X}{S_t}\right)\right)\right],$$

where

$$C_1\left(\tau,\sigma_t^2,\frac{X}{S_t}\right)=\int_0^{+\infty}\frac{1}{\pi u}Im\left(\exp\left(A^*\left(1+iu;\tau\right)+B^*\left(1+iu;\tau\right)^\top\sigma_t^2-iu\ln\left(\frac{X}{S_t}\right)\right)\right)du.$$
$$C_2\left(\tau,\sigma_t^2,\frac{X}{S_t}\right)=\int_0^{+\infty}\frac{1}{\pi u}Im\left[\exp\left(A^*\left(iu;\tau\right)+B^*\left(iu;\tau\right)^\top\sigma_t^2-iu\ln\left(\frac{X}{S_t}\right)\right)\right]du.$$

4.5 Estimation and Comparison of Affine SVS, SV and GARCH Models Using Equity and Index Daily Returns

4.5.1 Estimation Methodology and Data

Return unconditional moments can be computed in closed-form in a discrete-time affine multivariate latent factor model, as shown in Feunou and Tédongap (2008). All these moments are functions of the parameter vector θ that governs both returns and factors dynamics. We can then choose M informative moments to perform GMM estimation of the model. Assuming a single return, we choose M moments of the form $\mu_{r,j}(n,m)=E\left[r_t^n r_{t+j}^m\right]$ such that $1\leq j\leq J$, $0\leq n\leq Q$ and $0<m\leq Q-n$, that means M among $Q+JQ(Q-1)/2$ moments of order less than or equal to Q. Since moments of observed returns implied by a given model can directly be compared to their sample equivalent, our estimation setup is more likely to evaluate the performance of a given model in replicating well-known stylized facts like autocorrelation of squared returns, absence of autocorrelation of returns, leverage effect which can be captured via autocoskewness, unconditional fat-tailedness and asymmetries of returns. Model performance in replicating these empirical facts is assessed by including corresponding moments.

Let $g_t(\theta)=\left[r_t^{n_i}r_{t+j_i}^{m_i}-\mu_{r,j_i}(n_i,m_i)\right]_{1\leq i\leq M}$ denote the $M\times 1$ vector of retained moments. We have $E\left[g_t(\theta)\right]=0$ and we define the sample counterpart of this moment condition as follows:

$$\widehat{g}(\theta)=\begin{pmatrix}\widehat{E}\left[r_t^{n_1}r_{t+j_1}^{m_1}\right]-\mu_{r,j_1}(n_1,m_1)\\ \cdot\\ \cdot\\ \widehat{E}\left[r_t^{n_M}r_{t+j_M}^{m_M}\right]-\mu_{r,j_M}(n_M,m_M)\end{pmatrix}. \tag{4.53}$$

Given the $M\times M$ weighting matrix $\widehat{W}$, the GMM estimator $\widehat{\theta}$ of the parameter vector is given by:

$$\widehat{\theta}=\arg\min_{\theta}\widehat{g}(\theta)^{\top}\widehat{W}\widehat{g}(\theta). \tag{4.54}$$

Interestingly, the variance-covariance matrix of $g_t(\theta)$ does not depend on the vector of parameter θ. This is an advantage since with a nonparametric empirical variance-covariance matrix of moment conditions, the optimal GMM procedure is readily implemented in one step. In addition, for two different models estimated via the same moment conditions and weighting matrix, the minimum value of the GMM objective function itself is a criterion for comparison of alternative models.

In some cases, this GMM procedure also has a huge numerical advantage compared to the maximum likelihood estimation even when the likelihood function can be derived. Maximum likelihood estimation becomes difficult to perform numerically especially when the support of the likelihood function is parameter-dependent. This is the case in the IG-GARCH model of Christoffersen, Heston and Jacobs (2006) which can also be estimated through this GMM method. On the other hand, the maximum likelihood estimation of semi-affine latent variable models of Bates (2006) and the quasi-maximum likelihood estimation based on the Kalman recursion have the limitation that critical unconditional higher moments (skewness and kurtosis) of returns can be poorly estimated due to the second order approximation of the distribution of the latent variable conditional on observable returns. Moreover, in single-stage estimation and filtering methods like the Unscented Kalman Filter and the Bates (2006)'s algortihm, one can argue that approximations affect both parameter and state estimations. Instead, our GMM procedure matches critical higher moments and requires no approximation for parameter estimation. Provided with the GMM estimates of model parameters, Bates (2006)'s procedure or any other filtering procedure like the Unscented Kalman Filter can be followed for the state estimation. In this sense, approximations required by these techniques will only affect state estimation.

We estimate the single-factor SVS, the HN-S volatility and the IG GARCH models using daily returns on S&P500 and CRSP indexes, as well as daily returns on the six Fama and French size and book-to-market sorted portfolios. As explained in Fama and French (1993), the six portfolios are the outcome of the intersection of two independent sorts. Stocks are sorted into two size groups— **S** (small; that is, market capitalization below the NYSE median) and **B** (big; that is, market capitalization above the NYSE median)—and into three book-to-market groups—**G** (growth; that is, in the bottom 30 percent of the NYSE book-to-market), **N** (neutral; that is, in the middle 40 percent of the NYSE book-to-market) and **V** (value; that is, in the top 30 percent of the NYSE book-to-market). The six portfolios are commonly labelled SG, SN, SV, BG, BN and BV. Table 4.1 summarizes basic descriptive statistics of these returns. It shows the well-documented facts that asset returns are negatively skewed and fat-tailed. Small stocks are generally more negatively skewed than big stocks and a growth portfolio has lower average returns and higher negative skewness compared to a value portfolio of the same size.

4.5.2 Parameter Estimation

To perform the GMM procedure for each series, we need to decide which moments to match. To achieve this task, we refer to the relative importance of return moments. We consider the moments

$$\left\{E\left[r_t^j\right]\right\}_{j\geq 1},$$

in order to match the critical first moments of asset returns. Indeed, we do not estimate the unconditional mean of returns μ_0, which we set to its sample counterpart. Figure 4.2 displays autocorrelations of squared returns which are significant up to the twentieth lag. Given the positive and significant autocorrelation of squared returns, we consider the moments

$$\left\{E\left[r_t^2 r_{t+j}^2\right]\right\}_{j\geq 1},$$

in order to match these autocorrelations. For this set, we choose up to five leads for all stocks under consideration. To assess the ability of our SVS models to match significant autocoskewness and autocokurtosis, we add the following moments

$$\left\{E\left[r_t r_{t+j}^2\right], E\left[r_t r_{t+j}^3\right]\right\}_{j\geq 1}.$$

The negative and significant cross-correlation between returns and squared returns for various leads as shown in Panel A of Figure 4.3 is an empirical fact characterizing the well-known leverage effect. We choose up to five leads for small stocks and up to three leads for big stocks and market indexes. As shown in Panel B of Figure 4.3, similar cross-correlations for various lags are not significant. The cross-correlation between returns and cube returns is positive and significant, at least for the first three leads as shown in Panel A of Figure 4.4, especially for small stocks. Panel B of Figure 4.4 shows that similar cross-correlations for various lags are not significant.

The moments are weighted by the diagonal of the inverse of their covariance matrix:

$$\widehat{W} = Diag\left\{\left(\widehat{Var}\left[g_t\right]\right)^{-1}\right\}.$$

This matrix is nonparametric and puts more weight on moments with low variability. Estimation results for one-factor SVS models are shown in Table 4.2 for small stocks, in Table 4.3 for big stocks and in Table 4.4 for market indexes, both for contemporaneous asymmetry, contemporaneous normality, as well as alternative SV and GARCH models. For single factor SVS models, the parameter λ_1 is not estimated. The reason is that, due to the high expected persistence of the factor, it would be difficult in the return equation (4.14) to identify λ_1 and β_1 separately. To avoid this identification problem, we set $\lambda_1 = 0$.

We first focus on the first panel of Tables 4.2, 4.3 and 4.4, for estimation results in the context of contemporaneous asymmetry of returns, that is, η_1 is estimated. Starting with the measure equation (4.14), estimation output confirms that projecting returns onto the latent factor results in a significant and negative coefficient β_1, then corroborating the story that an increase in contemporaneous volatility lowers asset payoffs. The coefficient η_1 in the return equation is significant and positive and this results is robust across all stocks under consideration. This suggests that the distribution of daily returns conditional upon contemporaneous volatility is asymmetric. This result contrasts with the findings of Forsberg and Bollerslev (2002) that daily returns are normal conditional upon current realized volatility. Coming to the state dynamics, estimation results show that the factor governing daily return dynamics is highly persistent, with significant estimates of the coefficient of persistence, 0.963 and 0.948 for S&P500 and CRSP indexes respectively. This also means that daily return volatility and conditional asymmetry as perceived by agents are highly persistent as well, since they are linear in the factor. All estimates for the single factor SVS with contemporaneous asymmetry are significant. In addition, the J-test of over-identifying restrictions does not reject the model, but on small value stocks.

We now assess how important is contemporaneous asymmetry for asset return modeling. The second panel of Tables 4.2, 4.3 and 4.4 shows estimation results in the context of contemporaneous normality of returns, that is, with the constraint $\eta_1 = 0$. As for contemporaneous asymmetry, all parameters are significantly estimated and, in comparison, there is a decrease in the magnitude of the leverage parameter and an increase in the persistence of the factors— estimate of the persistence even becomes unrealistically greater or equal to 1 for some assets under consideration as we do not explicitly impose a restriction on this parameter in our estimation. For estimation results with a realistic persistence of factor, models are or tend to be rejected in the data. The maximum p-value for the J-test of over-identifying restrictions is 0.05. The sharp decrease in GMM criterion from contemporaneous normality to contemporaneous asymmetry also suggests that tests favor the latter compared to the former. The GMM criterion falls from 94.72 to 34.12 for CRSP index, and from 101.87 to 24.06 for S&P500 index.

The third panel of Tables 4.2, 4.3 and 4.4 shows estimation results for the HN-S volatility specification discussed in Section 4.2.1, and the fourth panel of Table 4.4 shows estimation results for the IG GARCH specification also discussed in Section 4.2.1. For big stocks and market indexes, HN-S volatility and IG GARCH specifications are comparable to the single factor SVS with contemporaneous asymmetry. Notice that, cokurtosis moments are not significant for these stocks and therefore not included for the GMM estimation. Instead, cokurtosis moments are significant for small stocks and, when included for the GMM estimation, results show that the single factor SVS model is preferred to the HN-S volatility specification.

For small stocks, Table 4.6 compares model unconditional moments to their sample counterparts across different models. Tables 4.7 and 4.8 show similar comparisons for big stocks and market indexes. Tables show ratios of model unconditional moments to their sample counterparts. The closer to one is the ratio, the better the model matches the moment. Mean, variance and kurtosis are perfectly matched by all models and this is robust across all stocks. It is also the case for autocorrelations of squared returns. A straightforward remark is how accurate the model with $\eta_1 \neq 0$ matches selected moments better than the model with $\eta_1 = 0$. In particular, Table 4.6 shows that skewness (moment 3) is not well matched by the model with contemporaneous normality, and, as shown in Tables 4.7 and 4.8, this matching is the worst when autocokurtosis is not significant. Contemporaneous normality matches autocoskewness better at long horizons ($j > 2$) than at short horizons ($j \leq 2$), while it is the contrary for contemporaneous asymmetry.

Finally, as we mentioned previously, the choice of the moments used in the GMM procedure is crucial when intended to reproduce important empirical facts. While the cross-correlation between returns and cubed returns is in general not significant for big stocks and market indexes, these moments are not matched by the GMM estimates when selected for estimation, except for the first lead where it appears weakly significant for some of these stocks. However, for small stocks, this moment is significant empirically as shown in Panel A of Figure 4.4 for the three first leads, and Table 4.6 shows that the GMM estimates reproduce the moments as well. Next, we filter the latent factors using the GMM estimates of parameters.

4.5.3 State Estimation

We use the Unscented Kalman Filter algorithm with our GMM estimates to filter the latent factor σ_{1t}^2 that we use to compute GARCH counterparts of volatility and conditional skewness, i.e h_t^G and s_t^G. Figure 4.5 displays the time series of GARCH counterparts of volatility and conditional skewness for the CRSP and the S&P500 indexes, for contemporaneous asymmetry ($\eta_1 \neq 0$) as well as for contemporaneous normality ($\eta_1 = 0$). Asset returns in our sample as plotted in Figure 4.1, are characterized by moderately high volatility at the beginning of the sample (1990-1992), followed by low volatility (1993-1996), then high volatility (1997-2003) and low or moderately high volatility at the end of the sample (2004-2005). This volatility pattern is well-matched by the volatility time series plotted in the first and the second rows of Figure 4.5. Also notice the slightly difference between volatility time series in different columns of the figure, due to the effect of contemporaneous asymmetry. Volatility is more persistent for contemporaneous normality.

The third and the fourth rows of the figure show the pattern of the GARCH counterpart of conditional skewness. Overall results are striking. As shown in the figure, conditional skewness is negative for contemporaneous normality, and this is consistent with the IG-GARCH model of Christoffersen, Heston and Jacobs (2006). We also recall that critical unconditional third order moments of returns, skewness and leverage effects, are not well-matched by GMM estimates under contemporaneous normality. In contrast, if contemporaneous asymmetry is allowed, we find that GMM estimates match unconditional skewness and leverage effects very well and, in this case, Figure 4.5 shows that conditional skewness is positive, and its mean has a larger magnitude compared to the contemporaneous normality case. Figure 4.6 confirms that these results hold for individual portfolios as well.

4.6 Estimation and Comparison of Affine SVS, SV and GARCH Models Using Index Option Prices

We conduct our empirical analysis using 9 years of data on S&P500 index call options. We use option data for Wednesday only in the the period from 1996 to 2004. If Wednesday is a holiday, we use the next trading day. Using only Wednesday data allows us to study a long time series, which is useful considering the highly persistent factors. Besides, using Wednesday is common practice in the literature to limit the impact of holidays and day-of-the week effects (see Heston and Nandi (2000), Christoffersen and Jacobs (2004, 2006)).

Table 4.9 presents the number of contracts used by moneyness (Panel (a)) and maturity (Panel (b)), and provides a cross-tabulation across moneyness and maturity (Panel (c)). From panel (a), we can observe a skewed Black-Scholes IV curve which is materialized by a high in-the-money call option implied volatility, compared to out-of-the-money. This suggests a necessity to model risk-neutral skewness in a flexible way. This pattern differs substantially across maturity groups. Short maturity options display an asymmetric smile pattern with high deep out-of-the-money implied volatility. This pattern reverses gradually when maturity increases to finally yield a smirked curve with high deep in-the-money implied volatilities.

Implied volatility show little variation across maturities, but a variable pattern across moneyness classes. For deep out-of-the-money call, implied volatility decreases with maturity, while it shows a smile-shape for out-of-the-money. For in-the-money, implied volatility increases with maturity while it shows an asymmetric smile pattern for deep in-the-money. Later, we evaluate empirically the ability of different models to replicate these observed patterns. Smirked implied volatility patterns for short maturities suggest a skewed one-step ahead conditional return distribution, while the reversion of this pattern and its persistence for longer maturities suggest more than one factor in risk-neutral conditional return distribution. Conditional skewness controls short-term properties while multiple factors control the long-term.

In this section, we estimate risk-neutral versions of the following models: single factor SVS both with contemporaneous asymmetry (SVS1f) and contemporaneous normality (SVS1f, $\eta = 0$), two-factor SVS with contemporaneous asymmetry (SVS2f), HN GARCH and IG GARCH. One challenge facing with unobservable factors is the joint estimation of risk-neutral parameters and latent factors. Several methods have been used in the literature and can be divided into two categories. The first approach considers latent factors as parameters (Bakshi, Cao and Chen (1997), Bates (2000) Huang and Wu (2004) and Christoffersen Heston and Jacobs (2007)), and the second approach filters latent factors using time series of underlying returns in a Bayesian framework (see Jones (2003) and Eraker (2004)).

4.6.1 Estimation Methodology

We follow the first approach here described. Without loss of generality we describe the method for an SVS model, as the same approach is applied to others. Consider a sample of T Wednesdays of option data ($T = 463$ corresponds to the number of Wednesdays in our sample). Given starting values for the structural parameter vector θ^* and the vector σ_t^{*2} of latent factors under the risk-neutral model, the iterative procedure proceeds as follows:

Step 1: For a given set of structural parameters, θ, solve T sums of squared pricing errors optimization problems of the form:

$$\hat{\sigma}_t^{*2} = \arg\min \sum_{n=1}^{N_t} \left(C_{nt} - C_n\left(\theta^*, \sigma_t^{*2}\right)\right)^2, \quad t = 1, 2, ..., T, \tag{4.55}$$

where C_{nt} is the observed price of contract n on day t and $C_n\left(\theta^*, \sigma_t^{*2}\right)$ is the corresponding model price. N_t is the number of contracts available on day t.

Step 2: For a given estimated factor $\hat{\sigma}_t^{*2}$ obtained from Step 1, solve one aggregate sum of squared pricing errors optimization problem of the form :

$$\hat{\theta}^* = \arg\min \sum_{t=1}^{T} \sum_{n=1}^{N_t} \left(C_{nt} - C_n\left(\theta^*, \hat{\sigma}_t^{*2}\right)\right)^2. \tag{4.56}$$

The procedure iterates between Step 1 and Step 2 until no further significant decreases in the overall objective function in Step 2 are obtained.

4.6.2 Risk-Neutral Volatility and Conditional Skewness

Table 4.10 shows parameter estimates of risk-neutral models. As shown in the table, all models deliver persistent factors. Even for the two-factor SVS model, the second factor is still very persistent. For the two-factor model, estimates of the two parameters driving contemporaneous asymmetry of returns (η_1^* and η_2^*) are negative, but lower in absolute value, compared to the corresponding parameter in the single-factor SVS model. This negative contemporaneous asymmetry contrasts with the positive contemporaneous asymmetry found when estimating the single-factor SVS using return data. Because the conditional risk neutral distribution is highly negatively skewed, a negative contemporaneous asymmetry is needed. Also in the two-factor risk-neutral SVS, the factor associated with the lowest (in absolute value) negative contemporaneous skewness is the most persistent.

Figure 4.7 shows annualized time series of risk-neutral volatility and conditional skewness for all models. The figure shows high co-movements of volatility and skewness across different risk-neutral models. Meanwhile, the level of volatility increases with the flexibility in conditional skewness modeling. The two-factor SVS model generates the highest level of volatility, then follows the SVS1F, the IG GARCH, the HN GARCH and finally the SVS1F with $\eta = 0$. The risk-neutral one-day ahead conditional skewness is the highest with the IG GARCH, while still comparable to the SVS1F, and the lowest in the SVS1F with $\eta = 0$. By construction, it is zero in the HN GARCH model. Conditional skewness in the IG GARCH increases (in absolute value) as volatility lowers. Relaxing this link as in other specifications reduces the level of conditional skewness.

4.6.3 Model Diagnostics

For all models, Table 4.11 shows the relative root mean squared error (RRMSE), that is the root mean squared error (RMSE) divided by the sample mean of call price, across different moneyness and maturity classes. As expected, the SVS2F model has the best in-sample fit in every moneyness class as well as every maturity class. It provides more flexibility than other models as it has a superior number of parameters/factors, twice the number of parameters/factors in other specifications. Using all available option data, the RRMSE for the SVS2F is 5.8%, followed by 7.8% for the SVS1F, 8.5% for the IG GARCH, 9.2% for the HN GARCH and finally 10.3% for the SVS1F with $\eta = 0$. In this order, we argue that flexibility in conditional asymmetries modeling also reduces option pricing errors. For short it should be noted that, even if the one-day ahead conditional skewness is zero in the HN GARCH model, it is not the case for the multi-day ahead conditional skewness, due to the leverage effect. A negative one-day ahead conditional skewness reduces pricing errors for short-term option contracts. This intuition is confirmed in CHJ (2006), where the authors find an improvement of the IG GARCH over the HN GARCH on short-maturity options. Our results shed light once more on this fact, and we argue that flexibility in one-day ahead conditional skewness modeling decreases the RRMSE for short maturities. As shown in Panel (b) of Table 4.11, for maturities less than one month, the best performance measured by the RRMSE is attributable to the SVS2F (6.2%), then SVS1F models (10.6% and 10.5%), the IG GARCH (11.6%) and the HN model (11.8%). The two-factor model has the best performance along all dimensions as also shown in Panel (a) of Table 4.11.

We summarize the model relative bias (the bias divided by the average price) in Table 4.12. Although the bias is generally low for all models under consideration, for maturities less than one month, it is the highest for the IG GARCH, which is comparable to that of the SVS1F, $\eta = 0$, and more than twice the bias for the SVS1F. The HN GARCH does better among single factor models for these maturities, and is comparable to the SVS2F. For deep in-the-money call options, the SVS1F bias the least, followed by the IG GARCH, then the SVS2F, the HN GARCH and the SVS1F, $\eta = 0$.

We finally represent, the observed and model's Black-Scholes implied volatilities along different dimensions. We retain our analysis to in-the-money call options and short-maturity contracts. In Figure 4.8 we fix the maturity class and represent implied volatility as function of moneyness. Using all available option data, the first panel of Figure 4.8 shows that all SVS models outperform HN and IG GARCH models in fitting deep in-the-money implied volatility, and have comparable performances for in-the-money implied volatility. For same options but maturities less than one month, the SVS1F and the HN GARCH fits perfectly in-the-money are outperform the IG GARCH deep in-the-money. For maturities between one and two months, the SVS1F, $\eta = 0$ perfectly fits deep in-the-money, while all SVS models outperform GARCH models and have comparable fits in-the-money. The SVS1F maintains its lead over GARCH models for maturities between two and three months, both in-the-money and deep in-the-money.

In Figure 4.9 we fix moneyness class and represent implied volatility as function of maturity. For short-maturity contracts as shown in the first panel of the figure, the SVS2F outperforms all models at every maturity. The SVS1F and the HN GARCH on one hand, and the SVS1F, $\eta = 0$ and the IG GARCH on the other hand, have comparable fits for maturities less than one month. Coming to in-the-money call options in the fifth panel, the SVS1F and the HN GARCH have comparable fits of the IV curve for maturities less than three months and outperform the IG GARCH. For deep in-the-money call options of less than three months of maturity, the SVS1F and the SVS2F have the lead over GARCH models in fitting observed Black-Scholes implied volatilities. Finally, these implied volatility curves confirm main analyses and model rankings resulting from RRMSE. The SVS1F, $\eta = 0$ even seems to do better in terms of implied volatility fit.

Two main findings can summarize this section. First we confirm CHJ (2006)'s result, which is the one-step ahead conditional skewness is particularly useful to price short maturities options. It also relaxes the link between the multi-step ahead skewness and the asymmetry in the volatility's motion imposed by the conditional normal GARCH or SV's models. Second it is important to include more than one factor in the returns' dynamic, this allows to disentangle the one-step ahead conditional volatility and skewness, and it helps to match the whole term structure of the risk neutral conditional volatility, skewness and kurtosis (the two factor SVS model outperforms all the others in the maturity dimension).

4.7 Conclusion and Future Work

This chapter presents a new approach for modeling conditional skewness in a discrete-time affine multivariate latent factor model with volatility. The model explicitly allows returns to be asymmetric conditional on current factors and past information. This contemporaneous asymmetry is shown to be particularly important for the model to fit both return and option data. An empirical investigation suggests that the flexibility that the model offers for conditional skewness, increases its option pricing performance relative to existing affine GARCH and SV models. In particular, SVS models with contemporaneous asymmetry outperform existing affine GARCH and SV models especially, for in-the-money calls and short-maturity contracts.

Although the model is flexible enough to accommodate both multiple returns and multiple factors, our analysis focuses on the single return case. In a future research, it would be interesting to study the implications of the model for a parsimonious multiple returns setting as well.

4.8 Appendix

4.8.1 Cumulant Generating Functions of Affine SV and GARCH Models

The functions A and B characterizing the cumulant-generating functions for GARCH and SV models in Example 1 are given by:

$$A(x,y) = (\mu_r - \lambda_h\mu_h)x + ((1-\phi_h)\mu_h - \alpha_h)y - \frac{1}{2}\ln(1-2\alpha_h y) \tag{4.57}$$

$$B(x,y) = \lambda_h x + (\phi_h - \alpha_h\beta_h^2)y + \frac{1}{2}x^2 + \frac{\alpha_h y}{1-2\alpha_h y}(\beta_h - \rho_{rh}x)^2, \tag{4.58}$$

for the HN-S specification, by

$$A(x,y) = (\mu_r - \lambda_h\mu_h)x + (1-\phi_h)\mu_h y + \frac{1}{2}\sigma_h^2 y^2 \tag{4.59}$$

$$B(x,y) = \lambda_h x + \phi_h y + \frac{1}{2}x^2, \tag{4.60}$$

for the autoregressive Gaussian specification, by

$$A(x,y) = (\mu_r - \lambda_h\mu_h)x + (1-\phi_h)\mu_h y \tag{4.61}$$

$$B(x,y) = \lambda_h x + \phi_h y + \frac{1}{2}(x^2 + 2\rho_{rh}\sigma_h xy + \sigma_h^2 y^2), \tag{4.62}$$

for the square-root specification and finally, by

$$A(x,y) = \alpha_h x + w_h y - \frac{1}{2}\ln(1-2a_h\eta_h^4 y) \tag{4.63}$$

$$B(x,y) = \lambda_h x + b_h y + \frac{1-\sqrt{(1-2a_h\eta_h^4 y)(1-2\eta_h x - 2c_h y)}}{\eta_h^2}, \tag{4.64}$$

for the IG GARCH specification.

4.8.2 Change of Measure, Risk-Neutral Dynamics of Returns and Option Pricing

For the multifactor SVS model, we assume a change of measure $Z_{t,t+1}$ given by:

$$Z_{t,t+1} = \exp\left(-A(\kappa,\pi) - B(\kappa,\pi)^{\top}\sigma_t^2 + \kappa^{\top}r_{t+1} + \pi^{\top}\sigma_{t+1}^2\right), \tag{4.65}$$

and which, by definition and specification, satisfies $E\left[Z_{t,t+1} \mid I_t\right] = 1$. We are interested in deriving return dynamics under the risk-neutral measure. We first look at expected returns conditional to $\left\langle \sigma_{t+1}^2, I_t \right\rangle$. For the return r_j we show that:

$$\begin{aligned} E^*\left[r_{j,t+1} \mid \left\langle \sigma_{t+1}^2, I_t \right\rangle\right] &= E^*\left[Z_{t,t+1} r_{j,t+1} \mid \left\langle \sigma_{t+1}^2, I_t \right\rangle\right] \\ &= \left[\delta_{jt} + \sum_{i=1}^{K} \left(\beta_{ji} + \psi'\left(\kappa_j; \eta_{ji}\right) \gamma_{ji}^2\right) \sigma_{i,t+1}^2\right] \exp\left(-\Psi_t^{\sigma}\left(f\left(\kappa, \pi\right)\right) + f\left(\kappa, \pi\right)^{\top} \sigma_{t+1}^2\right), \quad (4.66) \end{aligned}$$

where $\psi'\left(\cdot;\cdot\right)$ is the first derivative of $\psi\left(\cdot;\cdot\right)$ with respect to its first argument.

Expected returns are linear on factors under the physical measure. They will also be linear on factors under the risk-neutral measure if and only if $f\left(\kappa, \pi\right) = 0$. We assume that the parameters κ and π in our change of measure specification satisfy this condition. Also notice that this condition implies the followings:

$$Z_{t,t+1} = \exp\left(-\kappa^{\top} \delta_t + \kappa^{\top} r_{t+1} + \pi^{\top} \sigma_{t+1}^2\right) \text{ and } E\left[Z_{t,t+1} \mid \left\langle \sigma_{t+1}^2, I_t \right\rangle\right] = 1. \quad (4.67)$$

In particular, an implication of the second equation is that the moment generating function of σ_{t+1}^2, conditional to I_t, does not change from the physical to the risk-neutral measure. Thus, the factors still follow a multivariate autoregressive gamma under the risk-neutral measure. Return innovations under the risk-neutral measure, and conditional to $\left\langle \sigma_{t+1}^2, I_t \right\rangle$, are given by:

$$r_{j,t+1} - E^*\left[r_{j,t+1} \mid \left\langle \sigma_{t+1}^2, I_t \right\rangle\right] = \sum_{i=1}^{K} \gamma_{ji} \sigma_{i,t+1}\left(u_{ji,t+1} - \psi'\left(\kappa_j; \eta_{ji}\right) \gamma_{ji} \sigma_{i,t+1}\right). \quad (4.68)$$

Finding the distribution of the terms $\gamma_{ji} \sigma_{i,t+1}\left(u_{ji,t+1} - \psi'\left(\kappa_j; \eta_{ji}\right) \gamma_{ji} \sigma_{i,t+1}\right)$ will achieve the return dynamics under the risk-neutral measure. This distribution can be detected through their moment generating function. We show that:

$$\begin{aligned} &E^*\left[\exp\left(x_j \gamma_{ji} \sigma_{i,t+1}\left(u_{ji,t+1} - \psi'\left(\kappa_j; \eta_{ji}\right) \gamma_{ji} \sigma_{i,t+1}\right)\right) \mid \left\langle \sigma_{t+1}^2, I_t \right\rangle\right] \\ &\qquad = \exp\left(\left(\psi\left(x_j + \kappa_j; \eta_{ji}\right) - \psi\left(\kappa_j; \eta_{ji}\right) - x_j \psi'\left(\kappa_j; \eta_{ji}\right)\right) \gamma_{ji}^2 \sigma_{i,t+1}^2\right) \\ &\qquad = \exp\left(\psi\left(x_j; \eta_{ji}^*\right) q_{ji}^2 \gamma_{ji}^2 \sigma_{i,t+1}^2\right) \quad (4.69) \end{aligned}$$

where we have also shown that

$$\psi(x_j+\kappa_j;\eta_{ji})-\psi(\kappa_j;\eta_{ji})-x_j\psi'(\kappa_j;\eta_{ji})=q_{ji}^2\psi\left(x_j;\eta_{ji}^*\right) \tag{4.70}$$

with

$$q_{ji}^2=\left(1-\frac{2}{3}\eta_{ji}\kappa_j\right)^{-3/2} \text{ and } \eta_{ji}^*=\eta_{ji}\left(1-\frac{2}{3}\eta_{ji}\kappa_j\right)^{-1}. \tag{4.71}$$

Equation (4.69) means that the term $\gamma_{ji}\sigma_{i,t+1}\left(u_{ji,t+1}-\psi'(\kappa_j;\eta_{ji})\gamma_{ji}^2\sigma_{i,t+1}\right)$ can also be written

$$\gamma_{ji}\sigma_{i,t+1}\left(u_{ji,t+1}-\psi'(\kappa_j;\eta_{ji})\gamma_{ji}\sigma_{i,t+1}\right)=\gamma_{ji}^*\sigma_{i,t+1}u_{ji,t+1}^* \tag{4.72}$$

where

$$\gamma_{ji}^*=q_{ji}\gamma_{ji} \text{ and } u_{ji,t+1}^*\mid\left\langle\sigma_{t+1}^2,I_t\right\rangle\sim SIG\left(\eta_{ji}^*\left(\gamma_{ji}^*\sigma_{i,t+1}\right)^{-1}\right). \tag{4.73}$$

The return r_j should also satisfies the no-arbitrage condition:

$$E^*\left[\exp\left(r_{j,t+1}\right)\mid I_t\right]=\exp\left(r_{f,t+1}\right) \tag{4.74}$$

where $r_{f,t+1}$ is the risk-free rate. We show that:

$$E^*\left[\exp\left(r_{j,t+1}\right)\mid I_t\right]=\exp\left(\delta_{jt}+\Psi_t^{\sigma}\left(f\left(e_j+\kappa,\pi\right)\right)\right)=\exp\left(\delta_{jt}+\Psi_t^{\sigma}\left(\theta_j\right)\right) \tag{4.75}$$

where e_j is the $N\times 1$ vector with all components equal to zero but the j^{th} component equals one, and $\theta_j=f(e_j+\kappa,\pi)$ is the $K\times 1$ vector which components are given by:

$$\theta_{ji}=\beta_{ji}+\left(\psi\left(1+\kappa_j;\eta_{ji}\right)-\psi\left(\kappa_j;\eta_{ji}\right)\right)\gamma_{ji}^2. \tag{4.76}$$

We further show that

$$\theta_{ji}=q_{ji}^2\theta_{ji}^* \text{ where } \theta_{ji}^*=\beta_{ji}^*+\psi\left(1;\eta_{ji}^*\right)\gamma_{ji}^2 \text{ and } \beta_{ji}^*=\left(\beta_{ji}+\psi'\left(\kappa_j;\eta_{ji}\right)\gamma_{ji}^2\right)/q_{ji}^2. \tag{4.77}$$

Assuming that the risk-free rate is constant and equal to r_f, equation (4.75) implies

$$\lambda_j = -b(\theta_j) \quad \text{and} \quad \mu_{0,j} = r_f + (\lambda_j + \beta_j)^\top \mu - a(\theta_j). \tag{4.78}$$

Finally, the dynamics of returns under the risk-neutral measure are given by:

$$r_{j,t+1} = r_f - a\left(q_j^2 \theta_j^*\right) - \sum_{i=1}^{K} b_i\left(q_j^2 \theta_j^*\right)\sigma_{it}^2 + \sum_{i=1}^{K} \beta_{ji}^* q_{ji}^2 \sigma_{i,t+1}^2 + \sum_{i=1}^{K} \gamma_{ji}\sigma_{i,t+1} u_{ji,t+1}^*. \tag{4.79}$$

where $q_j^2\theta_j^*$ denotes the $K \times 1$ vector with components $q_{ji}^2\theta_{ji}^*$.

4.8.3 Second Order Log-normal Approximation of Positive Random Variables

As mentioned in section 4.3.4 we now provide more detail on the log-normal approximation of the one factor version of our model. This approximation is required in order to implement the Unscented Kalman Filter. The second order lognormal approximation of a positive random variable X with mean μ_x and variance σ_x^2 is given by:

$$X \approx \exp\left(\ln\left(\frac{\mu_x^2}{\sqrt{\mu_x^2 + \sigma_x^2}}\right) + \sqrt{\ln\left(\frac{\mu_x^2 + \sigma_x^2}{\mu_x^2}\right)}\varepsilon_X\right) \tag{4.80}$$

where ε_X is a standard normal random variable.

Given (4.80), the second order lognormal approximation of a standardized inverse Gaussian random variable u with positive skewness s is given by:

$$u \approx \exp\left(\ln\left(\frac{9}{s\sqrt{s^2+9}}\right) + \sqrt{\ln\left(\frac{s^2+9}{9}\right)}\varepsilon\right) - \frac{3}{s} \tag{4.81}$$

where ε is a standard normal random variable.

Given (4.80), the second order lognormal approximation for the dynamics of a stationary univariate autoregressive gamma process X_{t+1} with mean μ_x, variance σ_x^2 and persistence ϕ_x is given by:

$$X_{t+1} \approx \exp\left(\ln\left(\frac{m(X_t)^2}{\sqrt{m(X_t)^2 + v(X_t)}}\right) + \sqrt{\ln\left(\frac{m(X_t)^2 + v(X_t)}{m(X_t)^2}\right)}\varepsilon_{X,t+1}\right) \tag{4.82}$$

where

$$m(X_t) = (1-\phi_x)\mu_x + \phi_x X_t \tag{4.83}$$

$$v(X_t) = (1-\phi_x)^2\sigma_x^2 + \frac{2(1-\phi_x)\phi_x\sigma_x^2}{\mu_x}X_t \tag{4.84}$$

and $\varepsilon_{X,t+1}$ is a i.i.d. standard normal shock.

4.8.4 The Unscented Kalman Filter

As mentioned in section 4.3.4, we now provide more details on the Unscented Kalman Filter. The Unscented Kalman Filter is essentially an approximation of a nonlinear transformation of probability distribution coupled with the Kalman Filter. It has been introduced in the engineering literature by Julier et al. (1995) and Jullier and Uhlmann (1996). (See also Wan and van der Merwe (2001) for general introduction) and, to our knowledge, was first imported in Finance by Leippold and Wu (2003).

The Unscented Filter selects a set of sigma points in the distribution of $\left(\omega_t^\top, u_{t+1}^{*\top}, \varepsilon_{t+1}^{*\top}\right)^\top$ conditional on $\underline{r}_t$. This distribution is assumed Gaussian with mean

$$\bar{\chi} = \left(\omega_{t|t}^\top, \bar{u}^\top, \bar{\varepsilon}^\top\right)^\top$$

and variance

$$P^{\chi\chi} = \begin{pmatrix} P_{t|t}^{\omega\omega} & P^{\omega u} & P^{\omega\varepsilon} \\ P^{u\omega} & P^{uu} & P^{u\varepsilon} \\ P^{\varepsilon\omega} & P^{\varepsilon u} & P^{\varepsilon\varepsilon} \end{pmatrix}.$$

Following Julier et al. (1995) we consider the $2n+1$ sigma points $\chi_i = \left(\omega_{i,t|t}^\top, u_i^\top, \varepsilon_i^\top\right)^\top$ with associated weights W_i defined by:

$$\begin{aligned} \chi_0 &= \bar{\chi}, & W_0 &= \kappa/(n+\kappa) \\ \chi_i &= \bar{\chi} + \left(\sqrt{(n+\kappa)P^{\chi\chi}}\right)_i, & W_i &= 1/2(n+\kappa) \\ \chi_{i+n} &= \bar{\chi} - \left(\sqrt{(n+\kappa)P^{\chi\chi}}\right)_i, & W_i &= 1/2(n+\kappa), \end{aligned} \tag{4.85}$$

where n is the dimension of the vector $\left(\omega_t^\top, u_{t+1}^{*\top}, \varepsilon_{t+1}^{*\top}\right)^\top$, κ is an appropriately chosen real number and $\left(\sqrt{(n+\kappa)P^{\chi\chi}}\right)_i$ is the ith column of the matrix $(n+\kappa)P^{\chi\chi}$.

These sigma points are transformed through state and observation functions to obtain:

$$\omega_{i,t+1|t} = F\left(\omega_{i,t|t}, u_i\right) \text{ and } r_{i,t+1|t} = H\left(\omega_{i,t+1|t}, \varepsilon_i\right)$$

from which approximations of predicted means and covariances are computed as:

$$\widehat{\omega_{t+1|t}} = \sum_{i=0}^{2n} W_i \omega_{i,t+1|t} \text{ and } \widehat{r_{t+1|t}} = \sum_{i=0}^{2n} W_i r_{i,t+1|t} \tag{4.86}$$

$$\widehat{P^{\omega\omega}_{t+1|t}} = \sum_{i=0}^{2n} W_i \left(\omega_{i,t+1|t} - \widehat{\omega_{t+1|t}}\right)\left(\omega_{i,t+1|t} - \widehat{\omega_{t+1|t}}\right)^{\top} \tag{4.87}$$

$$\widehat{P^{rr}_{t+1|t}} = \sum_{i=0}^{2n} W_i \left(r_{i,t+1|t} - \widehat{r_{t+1|t}}\right)\left(r_{i,t+1|t} - \widehat{r_{t+1|t}}\right)^{\top} \tag{4.88}$$

$$\widehat{P^{\omega r}_{t+1|t}} = \sum_{i=0}^{2n} W_i \left(\omega_{i,t+1|t} - \widehat{\omega_{t+1|t}}\right)\left(r_{i,t+1|t} - \widehat{r_{t+1|t}}\right)^{\top}. \tag{4.89}$$

4.8.5 Approximated Moments of a Function of a Normal Random Variable

As mentioned in section 4.3.4, we choose to filter the distribution of $\omega_t = \ln(\sigma_t^2)$ in order to guaranty the positivity of σ_t^2. We now described in detail a procedure to recover the distribution of σ_t^2. Consider a normal random variable X with mean μ_x and variance σ_x^2. Let $Y = f(X)$, where f is a twice differentiable real function. The variable Y admits the second order Taylor approximation

$$Y = f(\mu_x) + f'(\mu_x)(X-\mu_x) + \frac{1}{2}f''(\mu_x)(X-\mu_x)^2, \tag{4.90}$$

which implies that the mean of Y can be approximated by:

$$\mu_y = E[Y] = f(\mu_x) + \frac{1}{2}f''(\mu_x)\sigma_x^2. \tag{4.91}$$

It follows that:

$$Y-\mu_y = f'(\mu_x)(X-\mu_x)+\frac{1}{2}f''(\mu_x)\left[(X-\mu_x)^2-\sigma_x^2\right] \tag{4.92}$$

$$\begin{aligned}(Y-\mu_y)^2 &= f'(\mu_x)^2(X-\mu_x)^2+f'(\mu_x)f''(\mu_x)\left[(X-\mu_x)^3-\sigma_x^2(X-\mu_x)\right]\\ &\quad+\frac{1}{4}f''(\mu_x)^2\left[(X-\mu_x)^4-2\sigma_x^2(X-\mu_x)^2+\sigma_x^4\right],\end{aligned} \tag{4.93}$$

$$\begin{aligned}(Y-\mu_y)^3 &= f'(\mu_x)^3(X-\mu_x)^3+\frac{3}{2}f'(\mu_x)^2f''(\mu_x)\left[(X-\mu_x)^4-\sigma_x^2(X-\mu_x)^2\right]\\ &\quad+\frac{3}{4}f'(\mu_x)f''(\mu_x)^2\left[(X-\mu_x)^5-2\sigma_x^2(X-\mu_x)^3+\sigma_x^4(X-\mu_x)\right]\\ &\quad+\frac{1}{8}f''(\mu_x)^3\left[(X-\mu_x)^6-3\sigma_x^2(X-\mu_x)^4+3\sigma_x^4(X-\mu_x)^2-\sigma_x^6\right].\end{aligned} \tag{4.94}$$

The third and fifth central moments of X are zero whereas the fourth and sixth central moments of X are respectively $3\sigma_x^4$ and $15\sigma_x^6$. Based on that, taking expectations of (4.93) and (4.94) gives the following approximations for the variance and the third moment of Y:

$$\sigma_y^2 = Var[Y] = f'(\mu_x)^2\sigma_x^2+\frac{1}{2}f''(\mu_x)^2\sigma_x^4, \tag{4.95}$$

$$E\left[(Y-\mu_y)^3\right] = 3f'(\mu_x)^2f''(\mu_x)\sigma_x^4+f''(\mu_x)^3\sigma_x^6. \tag{4.96}$$

BIBLIOGRAPHY

[1] Ait-Sahalia, Y., and R. Kimmel (2006), Maximum Likelihood Estimation of Stochastic Volatility Models, Journal of Financial Economics, 83, 413-452.

[2] Andersen, T.-G. (1994), Bayesian Analysis of Stochastic Volatility Models: Comment, Journal of Business and Economic Statistics, American Statistical Association, 12, 389-392.

[3] Andersen, T.-G., H.-G. Chung, and B.-E. Sorensen (1999), Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study, Journal of Econometrics, 91, 61-87.

[4] Bakshi, G., P. Carr, and L., Wu (2008), Stochastic Risk Premiums, Stochastic Skewness in Currency Options, and Stochastic Discount Factors in International Economies, Journal of Financial Economics, 87, 132-156.

[5] Bansal, R., D. Hsieh, and S. Viswanathan (1993), A New Approach to International Arbitrage Pricing, Journal of Finance 48, 1719-1747.

[6] Bansal, R., and S. Viswanathan (1993), No Arbitrage and Arbitrage Pricing: A New Approach, Journal of Finance 48, 1231-1262.

[7] Barndorff-Nielsen, O-E., and S.-Z. Levendorskii (2000), Feller Processes of Negative Inverse Gaussian Type, Quantitative Finance, 1, 318-331.

[8] Barndorff-Nielsen, O.-E., and N. Shephard (2001), Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics, Journal of the Royal Statistical Society, 63, 167-241.

[9] Bates, D.-S. (2006), Maximum Likelihood Estimation of Latent Affine Processes, The Review of Financial Studies, 19, 909-965.

[10] Black, F. (1976), Studies of Stock Price Volatility Changes, Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association (1976), 177-181.

[11] Bolic, M. (2004), Architectures for Efficient Implementation of Particle Filters, PhD dissertation, Stony Brook University.

[12] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327.

[13] Bollerslev, T., R.-F. Engle and D.-B. Nelson (1994), ARCH Models, In R. F. Engle & D. McFadden, eds, .The Handbook of Econometrics, Volume 4, North-Holland, Amsterdam, 2959-3038.

[14] Brooks, C., S.-P. Burke, S. Heravi, and G. Persand (2005), Autoregressive Conditional Kurtosis, Journal of financial Econometric, 3, 399-421.

[15] Christie, A. (1982), The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects, Journal of Financial Economics, 10, 407-432.

[16] Christoffersen, P., C. Dorion, K. Jacobs, and Y. Wang (2008), Volatility Components: Affine Restrictions and Non-Normal Innovations, Forthcoming in the Journal of Business and Economic Statistics.

[17] Christoffersen, P., R. Elkamhi, B. Feunou, and K. Jacobs (2009), Option Valuation with Conditional Heteroskedasticity and Non-Normality, Review of Financial Studies, Forthcoming.

[18] Christoffersen, P., S. Heston, and K. Jacobs (2006), Option Valuation with Conditional Skewness, Journal of Econometrics, 131, 253-284.

[19] Christoffersen, P., S. Heston, and K. Jacobs (2009), The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well, Working Paper, McGill University.

[20] Christoffersen, P., and K. Jacobs (2004), Which GARCH Model for Option Valuation?, Management Science 50, 1204-1221.

[21] Christoffersen, P., K. Jacobs and K. Mimouni (2009), Models for S&P500 Dynamics: Evidence from Realized Volatility, Daily Returns, and Option Prices, Working Paper, McGill University.

[22] Danielsson, J. (2002), Stochastic Volatility in Asset Prices Estimation with Simulated Maximum Likelihood, Journal of Econometrics, 64, 375-400.

[23] Darolles, S., C. Gourieroux, and J. Jasiak (2006), Structural Laplace Transform and Compound Autoregressive Models, Journal of Time Series Analysis, 27, 477-503.

[24] Duan, J., P. Ritchken, and Z. Sun (2002), Option Valuation with Jumps in Returns and Volatility, Working paper, University of Toronto.

[25] Engle, R.-F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, Econometrica, 50, 987-1008.

[26] Engle, R., D. Lilien, and R. Robins (1987), Estimating Time-Varying Risk Premia in the Term Structure: the ARCH-M model. Econometrica, 55, 391-407.

[27] Engle, R., and V. Ng (1993), Measuring and Testing the Impact of News on Volatility, Journal of finance, 48, 1749-1778.

[28] Feunou, B., and R. Tédongap (2009), Regime-Switching in Conditional Skewness, Working Paper, Duke University and Stockholm School of Economics.

[29] Feunou, B., and R. Tédongap (2008), Affine Stochastic Skewness, Working Paper, Duke University and Stockholm School of Economics.

[30] Forsberg, L., and T. Bollerslev (2002), Bridging the Gap Between the Distribution of Realized (ECU) and ARCH Modelling (of the euro): The GARCH-NIG Model, Journal of Applied Econometrics, 17, 535-548.

[31] Gallant, A.-R., D. Hsieh, and G. Tauchen (1998), Estimation of Stochastic Volatility Model with Diagnostics. Journal of Econometrics, 81, 159-192.

[32] Ghysels, E., A. Harvey, and E. Renault (1996), Stochastic Volatility, In Maddala, G.-S., and C.-R. Rao, eds, Statistical methods of finance, Handbook of Statistics series, vol. 14. Amsterdam, New York and Oxford, Elsevier, North Holland, 119-91.

[33] Gourieroux, C., J. Jasiak (2006), Autoregressive Gamma Processes, Journal of Forecasting, 25, 129-152.

[34] Gourieroux, C., A. Monfort, and E. Renault (1993), Indirect Inference, Journal of Applied Econometrics, 8, 85-118.

[35] Hansen, B.-E. (1994), Autoregressive Conditional Density Estimation, International Economic Review, 35, 705-73.

[36] Harvey, C.-R., and A. Siddique (1999), Autoregressive Conditional Skewness, Journal of Financial and Quantitative Analysis, 34, 465-487.

[37] Heston, S.-L. (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, Review of Financial Studies, 6, 327-343.

[38] Heston, S.-L., and S. Nandi (2000), A Closed-Form GARCH Option Pricing Model, Review of Financial Studies, 13, 585-626.

[39] Jacquier, E., N.-G. Polson, and P.-E. Rossi (1994), Bayesian Analysis of Stochastic Volatility Models, Journal of Business and Economic Statistics, 12, 371-389.

[40] Jensen, M.-B., and A. Lunde (2001), The NIG-S ARCH Model: A Fat Tailed, Stochastic, and Autoregressive Conditional Heteroskedastic Volatility Model, Econometrics Journal, 4, 319-342.

[41] Jiang, G.-J., and J.-L. Knight (2002), Estimation of Continuous-Time Processes via the Empirical Characteristic Function, Journal of Business and Economic Statistics, 20, 198-212.

[42] Jondeau, E., and M. Rockinger (2003), Conditional Volatility, Skewness and Kurtosis: Existence, Persistence and Comovements, Journal of Economic Dynamics and Control, 27, 1699-1737.

[43] Julier, S.-J., and J.-K. Uhlmann (1996), A General Method for Approximating Nonlinear Transformations of Probability Distributions, Working Paper, University of Oxford.

[44] Julier, S.-J., J.-K. Uhlmann, and H.-F. Durrant-Whyte (1995), A New Approach for Filtering Nonlinear Systems, In The Proceedings of the American Control Conference, Seattle, 1628-1632.

[45] Kim, S., N. Shephard, and C. Siddhartha (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, Review of Economic studies, 65, 361-393.

[46] Kim, T.-H., and H. White (2004), On More Robust Estimation of Skewness and Kurtosis, Finance Research Letters, 1, 57-73.

[47] Koopman, S.-J., and J. Sandmann (1998), Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, Journal of Econometrics, 87, 271-301.

[48] Koopman, S.-J., and E.-H. Uspensky (2002), The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets, Journal of applied Econometrics, 17, 667-689.

[49] Leon, A., G. Rubio, and G. Serna (2005), Autoregressive conditional volatility, skewness and kurtosis, The quatrely review of Economics and Finance, 45, 599-618.

[50] Leippold, M., and L. Wu (2003), Design and Estimation of Multi-currency Quadratic Models, European Finance Review, 7, 47-73.

[51] Liesenfeld, R. and , R.-C. Jung (2000), Stochastic Volatility Models: Conditional Normality Versus Heavy-Tailed Distribution, Journal of applied Econometrics, 15, 137-160.

[52] Nandi, S. (1998), How Important is the Correlation Between Returns and Volatility in a Stochastic Volatility Model? Empirical Evidence from Pricing and Hedging in the S&P500 Index Options Market, Journal of Banking & Finance, 22, 589-610.

[53] Nelson, D. B. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Econometrica, 59, 347-370.

[54] Shephard, N. (2005), Stochastic Volatility: Selected Readings, Oxford: Oxford University Press.

[55] Singleton, K.-J. (2001), Empirical Dynamic Asset Pricing, Princeton, NJ: Princeton University Press.

[56] Taylor, S.-J. (1987), Forecasting the Volatility of Currency Exchange Rates, International Journal of Forecasting, 3, 159-170.

Table 4.1: **Summary Statistics of Daily Stocks Returns for the Period 1990-2005.** The entries of the table are summary statistics of daily stock returns. Means and standard deviations are annualized values using 252 trading days per year, whereas skewness and kurtosis are daily values. Return data span the period January 2, 1990 to December 30, 2005.

r	Mean	Std. Dev.	Skewness	Kurtosis
SG	6.11	19.12	-0.46	6.61
SN	13.84	13.45	-0.46	6.36
SV	15.45	12.50	-0.61	6.99
BG	10.05	17.24	-0.06	6.78
BN	11.30	14.34	-0.19	7.04
BV	10.88	14.09	-0.30	6.88
CRSP	10.10	15.54	-0.21	7.21
S&P500	9.17	16.09	-0.01	6.69

Table 4.2: **Estimation Results on Small Stocks.**

The entries of the table are GMM parameter estimates of the single factor SVS, the HN-S volatility and the HN GARCH(1,1) models on small stocks. To perform the GMM estimation for all the three models, we use the same sixteen moment conditions that are all significant for small stocks as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are $\left\{E\left[r_t^j\right]\right\}_{j=2}^4$, $\left\{E\left[r_t^2 r_{t+j}^2\right], E\left[r_t r_{t+j}^2\right]\right\}_{j=1}^5$ and $\left\{E\left[r_t r_{t+j}^3\right]\right\}_{j=1}^3$. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations. Standard errors (in parenthesis) are given below the estimates.

SVS1f, $\eta \neq 0$	SG	SN	SV
β_1	-24.25	-36.51	-52.77
	(3.73)	(7.01)	(8.61)
η_1	5.25E-03	3.34E-03	3.79E-03
	(1.16E-03)	(1.03E-03)	(9.86E-04)
μ_1	1.35E-04	6.82E-05	5.76E-05
	(1.43E-05)	(7.12E-06)	(5.37E-06)
ϕ_1	0.979	0.979	0.930
	(0.027)	(0.023)	(0.025)
$\sqrt{\omega_1}$	1.42E-04	6.75E-05	5.54E-05
	(1.86E-05)	(7.64E-06)	(5.87E-06)
Criterion	125.04	128.66	126.01
J-Stat	17.77	15.50	26.48
p-value	0.06	0.11	0.00

SVS1f, $\eta = 0$	SG	SN	SV
β_1	-20.47	-29.02	-40.62
	(3.31)	(5.95)	(6.95)
η_1			
μ_1	1.38E-04	7.00E-05	6.00E-05
	(1.50E-05)	(7.48E-06)	(5.71E-06)
ϕ_1	1.016	1.019	0.979
	(0.028)	(0.024)	(0.026)
$\sqrt{\omega_1}$	1.42E-04	6.83E-05	5.76E-05
	(1.92E-05)	(8.00E-06)	(6.23E-06)
Criterion	168.00	171.50	183.80
J-Stat	22.49	13.54	22.29
p-value	0.02	0.26	0.02

HN-S Volatility	SG	SN	SV
λ_h	-14.12	-22.36	-33.68
	(2.45)	(6.36)	(6.75)
μ_h	1.42E-04	7.16E-05	6.29E-05
	(1.60E-05)	(7.94E-06)	(6.31E-06)
ϕ_h	0.983	0.988	0.941
	(0.029)	(0.022)	(0.027)
α_h	2.46E-06	8.40E-07	3.69E-06
	(3.87E-06)	(1.59E-06)	(1.68E-06)
ρ_{rh}	0.607	0.587	0.324
	(0.521)	(0.546)	(0.132)
β_h	476.36	767.80	333.12
	(367.61)	(729.89)	(88.73)
Criterion	142.46	154.57	162.02
J-Stat	22.58	13.59	25.04
p-value	0.01	0.14	0.00

Table 4.3: **Estimation Results on Big Stocks.**

The entries of the table are GMM parameter estimates of the single factor SVS and the HN-S volatility models on big stocks. To perform the GMM estimation for all the three models, we use the same eleven moment conditions that are all significant for big stocks and market indexes as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are $\left\{E\left[r_t^j\right]\right\}_{j=2}^{4}$, $\left\{E\left[r_t^2 r_{t+j}^2\right]\right\}_{j=1}^{5}$ and $\left\{E\left[r_t r_{t+j}^2\right]\right\}_{j=1}^{3}$. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations. Standard errors (in parenthesis) are given below the estimates.

SVS1f, $\eta \neq 0$				***SVS1f, $\eta = 0$***				***HN-S Volatility***			
	BG	BN	BV		BG	BN	BV		BG	BN	BV
β_l	-26.45	-21.95	-19.74	β_l	-12.00	-13.26	-14.40	λ_h	-1.44	-5.38	-8.76
	(5.37)	(6.11)	(5.98)		(2.73)	(4.66)	(5.31)		(2.96)	(3.82)	(4.88)
η_l	8.83E-03	4.77E-03	2.97E-03	η_l				μ_h	1.18E-04	8.15E-05	7.83E-05
	(1.85E-03)	(1.30E-03)	(1.36E-03)						(1.20E-05)	(8.73E-06)	(8.80E-06)
μ_l	1.09E-04	7.81E-05	7.61E-05	μ_l	1.15E-04	8.03E-05	7.75E-05	ϕ_h	0.930	0.969	0.995
	(1.04E-05)	(8.26E-06)	(8.12E-06)		(1.14E-05)	(8.54E-06)	(8.54E-06)		(0.035)	(0.036)	(0.004)
ϕ_l	0.946	0.973	1.005	ϕ_l	0.961	0.995	1.021	α_h	5.24E-06	1.65E-06	4.05E-07
	(0.035)	(0.041)	(0.026)		(0.031)	(0.040)	(0.026)		(3.52E-06)	(2.07E-06)	(3.51E-07)
$\sqrt{\omega_l}$	1.12E-04	8.69E-05	8.48E-05	$\sqrt{\omega_l}$	1.26E-04	8.90E-05	8.54E-05	ρ_{rh}	0.662	0.620	1.000
	(1.45E-05)	(1.43E-05)	(1.58E-05)		(1.73E-05)	(1.52E-05)	(1.65E-05)		(0.162)	(0.423)	(0.000)
								β_h	421.08	765.96	1267.48
Criterion	33.12	16.43	10.45	Criterion	117.18	50.95	30.03		(152.61)	(513.86)	(520.29)
J-Stat	8.93	6.55	3.33	*J*-Stat	18.72	12.36	7.34	Criterion	32.56	16.52	10.81
p-value	0.11	0.26	0.65	p-value	0.00	0.05	0.29				
								J-Stat	11.56	7.02	4.45
								p-value	0.02	0.13	0.49

Table 4.4: **Estimation Results on Market Indexes.**

The entries of the table are GMM parameter estimates of the single factor SVS, the HN-S volatility and the IG GARCH models on market indexes. To perform the GMM estimation for all the three models, we use the same eleven moment conditions that are all significant for big stocks and market indexes as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are $\left\{E\left[r_t^j\right]\right\}_{j=2}^{4}$, $\left\{E\left[r_t^2 r_{t+j}^2\right]\right\}_{j=1}^{5}$ and $\left\{E\left[r_t r_{t+j}^2\right]\right\}_{j=1}^{3}$. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations. Standard errors (in parenthesis) are given below the estimates.

SVSIf, $\eta \neq 0$			**SVSIf, $\eta = 0$**			**HN-S Volatility**			**IG GARCH**		
	CRSP	S&P500		CRSP	S&P500		CRSP	S&P500		CRSP	S&P500
β_1	-30.09	-25.29	β_1	-17.19	-10.46	λ_h	-5.29	-0.35	λ_h	8.20E+02	9.74E+02
	(6.44)	(5.84)		(3.76)	(3.11)		(3.03)	(3.03)		(9.29E+02)	(1.17E+03)
η_1	7.65E-03	7.76E-03	η_1			μ_h	9.58E-05	1.03E-04	η_h	-1.23E-03	-1.03E-03
	(1.76E-03)	(1.64E-03)					(1.04E-05)	(1.05E-05)		(1.41E-03)	(1.26E-03)
μ_1	8.79E-05	9.67E-05	μ_1	9.28E-05	1.01E-04	ϕ_h	0.939	0.945	w_h	5.24E-06	5.42E-06
	(9.05E-06)	(9.72E-06)		(9.81E-06)	(1.02E-05)		(0.049)	(0.042)		(4.22E-06)	(3.62E-06)
ϕ_1	0.948	0.963	ϕ_1	0.990	0.968	α_h	3.83E-06	3.48E-06	b_h	-2.026	-2.533
	(0.043)	(0.039)		(0.040)	(0.035)		(4.05E-06)	(3.59E-06)		(2.471)	(3.338)
$\sqrt{\omega_1}$	9.43E-05	9.89E-05	$\sqrt{\omega_1}$	1.02E-04	1.10E-04	ρ_{rh}	0.645	0.712	c_h	4.44E-06	3.69E-06
	(1.16E-05)	(1.34E-05)		(1.38E-05)	(1.57E-05)		(0.244)	(0.315)		(6.67E-06)	(5.53E-06)
						β_h	495.16	521.14	a_h	1.50E+04	2.88E+04
Criterion	34.12	24.06	Criterion	94.72	101.87		(304.31)	(303.81)		(1.21E+05)	(1.26E+05)
J-Stat	8.24	7.19	J-Stat	13.49	17.50	Criterion	35.27	23.96	Criterion	35.14	23.99
p-value	0.14	0.21	p-value	0.04	0.01						
						J-Stat	10.02	12.46	J-Stat	10.19	11.27
						p-value	0.04	0.01	p-value	0.04	0.02

Table 4.5: **Estimation Results. c's coefficients.**

The entries of the table are loadings of expected returns, volatility, asymmetry and leverage effects on factors, using GMM parameter estimates of the single factor SVS, the HN-S volatility and the IG GARCH models on small stocks and market indexes. To perform the GMM estimation for all three models on small stocks, we use the same sixteen moment conditions that are all significant for small stocks as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are $\left\{E\left[r_t^j\right]\right\}_{j=2}^{4}$, $\left\{E\left[r_t^2 r_{t+j}^2\right], E\left[r_t r_{t+j}^2\right]\right\}_{j=1}^{5}$ and $\left\{E\left[r_t r_{t+j}^3\right]\right\}_{j=1}^{3}$. For all four models on market indexes, we use the same eleven moment conditions that are all significant for big stocks and market indexes as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are $\left\{E\left[r_t^j\right]\right\}_{j=2}^{4}$, $\left\{E\left[r_t^2 r_{t+j}^2\right]\right\}_{j=1}^{5}$ and $\left\{E\left[r_t r_{t+j}^2\right]\right\}_{j=1}^{3}$. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations.

SVS1f, $\eta \neq 0$				**SVS1f, $\eta = 0$**				**HN-S Volatility**			
	SG	SN	SV		SG	SN	SV		SG	SN	SV
$c_{0\mu}$	3.44E-03	2.99E-03	3.44E-03	$c_{0\mu}$	3.12E-03	2.62E-03	3.00E-03	$c_{0\mu}$	2.25E-03	2.15E-03	2.73E-03
$c_{1\mu}$	-23.75	-35.76	-49.07	$c_{1\mu}$	-20.80	-29.57	-39.78	$c_{1\mu}$	-14.123	-22.362	-33.677
c_{0h}	2.80E-06	1.41E-06	4.08E-06	c_{0h}	-2.21E-06	-1.32E-06	1.23E-06	c_{0h}	0	0	0
c_{1h}	0.983	0.983	0.949	c_{1h}	1.014	1.017	0.983	c_{1h}	1	1	1
c_{0s}	1.41E-08	4.49E-09	1.29E-08	c_{0s}	-3.15E-10	-1.45E-10	-1.71E-10	c_{0s}	0	0	0
c_{1s}	4.70E-03	2.98E-03	2.41E-03	c_{1s}	2.89E-04	2.23E-04	-2.72E-04	c_{1s}	0	0	0
c_{0r}	-2.10E-10	-7.09E-11	-7.97E-10	c_{0r}	-1.05E-10	-4.84E-11	-5.68E-11	c_{0r}	0	0	0
c_{1r}	-1.47E-04	-9.87E-05	-3.67E-04	c_{1r}	9.65E-05	7.45E-05	-9.04E-05	c_{1r}	-1.42E-03	-7.58E-04	-7.95E-04

SVS1f, $\eta \neq 0$			**SVS1f, $\eta = 0$**			**HN-S Volatility**			**IG GARCH**		
	CRSP	S&P500		CRSP	S&P500		CRSP	S&P500		CRSP	S&P500
$c_{0\mu}$	2.91E-03	2.72E-03	$c_{0\mu}$	1.98E-03	1.39E-03	$c_{0\mu}$	9.07E-04	4.00E-04	$c_{0\mu}$	-4.94E-05	-4.79E-04
$c_{1\mu}$	-28.52	-24.34	$c_{1\mu}$	-17.02	-10.13	$c_{1\mu}$	-5.29	-0.35	$c_{1\mu}$	4.70	8.26
c_{0h}	4.61E-06	3.63E-06	c_{0h}	9.48E-07	3.24E-06	c_{0h}	0	0	c_{0h}	0	0
c_{1h}	0.957	0.967	c_{1h}	0.990	0.969	c_{1h}	1	1	c_{1h}	1	1
c_{0s}	3.29E-08	2.70E-08	c_{0s}	-5.57E-11	-3.91E-10	c_{0s}	0	0	c_{0s}	0	0
c_{1s}	6.35E-03	6.91E-03	c_{1s}	-1.16E-04	-2.34E-04	c_{1s}	0	0	c_{1s}	-3.68E-03	-3.10E-03
c_{0r}	-7.30E-10	-3.48E-10	c_{0r}	-1.86E-11	-1.30E-10	c_{0r}	0	0	c_{0r}	0	0
c_{1r}	-3.01E-04	-1.85E-04	c_{1r}	-3.88E-05	-7.80E-05	c_{1r}	-2.45E-03	-2.58E-03	c_{1r}	-3.59E-03	-3.54E-03

Table 4.6: **Moment Matching for Small Stocks.**

The entries of the table are ratios of model unconditional moments to their empirical counterparts, based on parameter estimates of the single factor SVS and the HN-S volatility models on small stocks. To perform the GMM estimation for all the three models, we use the same sixteen moment conditions that are all significant for small stocks as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are identified by a 1 in the third column. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations.

			SG			SN			SV		
			$\eta \neq 0$	$\eta = 0$	HN	$\eta \neq 0$	$\eta = 0$	HN	$\eta \neq 0$	$\eta = 0$	HN
$E[r_t]$	1	0	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
$E[r_t^2]$	2	1	1.01	1.01	1.01	1.03	1.03	1.03	1.07	1.05	1.08
$E[r_t^3]$	3	1	1.04	1.67	1.27	1.10	1.84	1.44	1.06	1.69	1.39
$E[r_t^4]$	4	1	1.03	1.00	1.03	1.05	1.02	1.04	1.01	0.98	1.00
$E[r_t r_{t+5}]$	5	0	-7.90	-6.75	-5.13	14.12	11.28	9.53	4.59	3.87	3.41
$E[r_t^2 r_{t-5}]$	6	0	-10.96	-10.55	-6.82	-5.16	-4.76	-3.05	-3.22	-3.20	-1.99
$E[r_t r_{t+5}^2]$	7	1	1.29	1.24	1.31	1.84	1.70	1.81	1.54	1.53	1.55
$E[r_t^3 r_{t-5}]$	8	0	-1.95	-1.87	-1.32	-1.08	-0.93	-0.74	-1.00	-0.90	-0.75
$E[r_t^2 r_{t-5}^2]$	9	1	0.99	1.04	0.99	1.00	1.05	1.01	0.96	1.02	0.98
$E[r_t r_{t+5}^3]$	10	0	-1.43	-1.38	-1.11	-4.25	-3.65	-3.14	-25.25	-22.87	-20.35
$E[r_t r_{t+4}]$	11	0	1.60	1.32	1.03	2.37	1.82	1.59	1.92	1.54	1.41
$E[r_t^2 r_{t-4}]$	12	0	-5.63	-5.21	-3.49	-1.65	-1.44	-0.96	-1.53	-1.41	-0.94
$E[r_t r_{t+4}^2]$	13	1	1.24	1.15	1.25	1.67	1.47	1.63	1.94	1.80	1.92
$E[r_t^3 r_{t-4}]$	14	0	1.48	1.37	1.00	-3.01	-2.48	-2.03	5.38	4.62	3.98
$E[r_t^2 r_{t-4}^2]$	15	1	0.92	0.94	0.92	0.91	0.94	0.92	0.93	0.97	0.94
$E[r_t r_{t+4}^3]$	16	0	1.10	1.02	0.85	4.70	3.88	3.45	2.66	2.28	2.13
$E[r_t r_{t+3}]$	17	0	1.39	1.10	0.89	1.42	1.05	0.94	1.48	1.13	1.07
$E[r_t^2 r_{t-3}]$	18	0	-8.00	-7.11	-4.95	-2.46	-2.04	-1.43	-4.59	-3.96	-2.82
$E[r_t r_{t+3}^2]$	19	1	1.39	1.24	1.40	1.77	1.47	1.70	1.53	1.32	1.49
$E[r_t^3 r_{t-3}]$	20	0	2.31	2.06	1.56	1.18	0.93	0.79	0.98	0.80	0.72
$E[r_t^2 r_{t-3}^2]$	21	1	1.02	1.02	1.02	1.00	1.01	1.00	0.94	0.95	0.94
$E[r_t r_{t+3}^3]$	22	1	0.76	0.68	0.59	0.59	0.47	0.43	0.65	0.53	0.52
$E[r_t r_{t+2}]$	23	0	49.92	38.20	32.08	2.65	1.90	1.74	1.89	1.37	1.35
$E[r_t^2 r_{t+2}]$	24	0	-2.95	-2.52	-1.82	-1.94	-1.53	-1.11	-1.91	-1.54	-1.17
$E[r_t r_{t+2}^2]$	25	1	0.67	0.58	0.67	0.68	0.53	0.64	0.73	0.59	0.70
$E[r_t^3 r_{t+2}]$	26	0	-0.73	-0.63	-0.49	-1.75	-1.33	-1.16	-7.19	-5.55	-5.20
$E[r_t^2 r_{t+2}^2]$	27	1	0.75	0.73	0.74	0.79	0.79	0.79	0.79	0.78	0.79
$E[r_t r_{t+2}^3]$	28	1	2.17	1.86	1.66	0.70	0.53	0.50	0.75	0.58	0.59
$E[r_t r_{t+1}]$	29	0	0.53	0.39	0.34	0.85	0.59	0.56	0.93	0.64	0.66
$E[r_t^2 r_{t+1}]$	30	0	986.72	810.83	606.84	-4.68	-3.50	-2.67	-18.79	-14.18	-11.46
$E[r_t r_{t+1}^2]$	31	1	0.94	0.77	0.94	1.00	0.75	0.94	1.09	0.83	1.04
$E[r_t^3 r_{t+1}]$	32	0	2.48	2.05	1.66	-4.31	-3.14	-2.82	3.31	2.42	2.37
$E[r_t^2 r_{t+1}^2]$	33	1	1.12	1.08	1.11	1.10	1.06	1.09	1.13	1.09	1.12
$E[r_t r_{t+1}^3]$	34	1	0.81	0.67	0.62	1.01	0.73	0.72	0.86	0.63	0.67

Table 4.7: **Moment Matching for Big Stocks.**
The entries of the table are ratios of model unconditional moments to their empirical counterparts, based on parameter estimates of the single factor SVS and the HN-S volatility models on big stocks. To perform the GMM estimation for all the three models, we use the same eleven moment conditions that are all significant for big stocks and market indexes as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are identified by a 1 in the third column. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations.

			BG			BN			BV		
			$\eta \neq 0$	$\eta = 0$	HN	$\eta \neq 0$	$\eta = 0$	HN	$\eta \neq 0$	$\eta = 0$	HN
$E[r_t]$	1	0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
$E[r_t^2]$	2	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$E[r_t^3]$	3	1	0.75	-6.71	1.01	0.98	6.77	0.99	1.00	2.01	1.02
$E[r_t^4]$	4	1	1.01	0.98	1.00	1.00	0.96	0.99	1.00	0.99	1.02
$E[r_t r_{t+5}]$	5	0	-1.52	-0.45	-0.12	-2.15	-0.99	-0.65	130.85	79.52	63.02
$E[r_t^2 r_{t+5}]$	6	0	-0.99	-0.47	0.12	-1.11	-0.64	-0.04	-1.38	-0.98	-0.42
$E[r_t r_{t+5}^2]$	7	0	1.80	0.85	1.69	0.95	0.54	0.93	1.19	0.85	1.14
$E[r_t^3 r_{t+5}]$	8	0	-0.49	-0.19	-0.03	-0.33	-0.17	-0.09	-0.52	-0.34	-0.23
$E[r_t^2 r_{t+5}^2]$	9	1	0.91	0.94	0.89	0.96	0.98	0.95	1.06	1.08	1.04
$E[r_t r_{t+5}^3]$	10	0	-0.54	-0.21	-0.02	-0.46	-0.24	-0.13	-1.37	-0.90	-0.70
$E[r_t r_{t+4}]$	11	0	3.77	1.10	0.30	-5.34	-2.42	-1.62	-42.12	-25.26	-20.48
$E[r_t^2 r_{t+4}]$	12	0	-1.64	-0.77	0.18	-0.93	-0.52	-0.04	-1.39	-0.96	-0.42
$E[r_t r_{t+4}^2]$	13	0	-13.38	-6.28	-12.85	7.61	4.25	7.54	6.08	4.23	5.88
$E[r_t^3 r_{t+4}]$	14	0	0.97	0.37	0.07	-1.30	-0.65	-0.34	-3.30	-2.13	-1.49
$E[r_t^2 r_{t+4}^2]$	15	1	1.11	1.15	1.10	1.02	1.04	1.02	0.98	0.99	0.97
$E[r_t r_{t+4}^3]$	16	0	44.70	17.24	1.75	-0.61	-0.31	-0.18	-0.57	-0.37	-0.29
$E[r_t r_{t+3}]$	17	0	-2.00	-0.58	-0.16	11.09	4.91	3.36	-59.56	-35.24	-29.23
$E[r_t^2 r_{t+3}]$	18	0	-2.35	-1.09	0.23	-2.09	-1.14	-0.11	-2.30	-1.56	-0.72
$E[r_t r_{t+3}^2]$	19	1	2.09	0.97	2.05	2.37	1.29	2.35	1.34	0.91	1.31
$E[r_t^3 r_{t+3}]$	20	0	1.12	0.43	0.08	0.50	0.24	0.13	0.51	0.33	0.23
$E[r_t^2 r_{t+3}^2]$	21	1	1.03	1.06	1.03	0.94	0.95	0.94	0.95	0.95	0.95
$E[r_t r_{t+3}^3]$	22	0	-2.66	-1.01	-0.10	0.55	0.27	0.16	1.04	0.66	0.54
$E[r_t r_{t+2}]$	23	0	-2.11	-0.60	-0.17	-7.14	-3.10	-2.17	5.39	3.15	2.67
$E[r_t^2 r_{t+2}]$	24	0	-5.71	-2.62	0.49	-1.64	-0.86	-0.10	-2.30	-1.53	-0.73
$E[r_t r_{t+2}^2]$	25	1	0.64	0.29	0.64	0.68	0.36	0.68	0.74	0.49	0.73
$E[r_t^3 r_{t+2}]$	26	0	-1.29	-0.49	-0.09	1.78	0.85	0.47	0.77	0.48	0.35
$E[r_t^2 r_{t+2}^2]$	27	1	0.96	0.98	0.97	1.02	1.01	1.02	1.03	1.02	1.04
$E[r_t r_{t+2}^3]$	28	0	114.07	42.94	4.52	0.62	0.30	0.18	0.46	0.29	0.24
$E[r_t r_{t+1}]$	29	0	3.99	1.11	0.32	1.01	0.43	0.31	0.50	0.29	0.25
$E[r_t^2 r_{t+1}]$	30	0	-1.29	-0.59	0.10	-1.01	-0.52	-0.07	-0.85	-0.56	-0.27
$E[r_t r_{t+1}^2]$	31	1	1.01	0.46	1.03	0.95	0.49	0.95	1.12	0.73	1.13
$E[r_t^3 r_{t+1}]$	32	0	-0.42	-0.16	-0.03	-0.58	-0.27	-0.15	-0.71	-0.43	-0.33
$E[r_t^2 r_{t+1}^2]$	33	1	0.99	1.00	1.01	1.01	1.00	1.02	1.00	0.99	1.02
$E[r_t r_{t+1}^3]$	34	0	3.35	1.24	0.13	0.71	0.33	0.21	0.64	0.40	0.34

Table 4.8: **Moment Matching for Market Indexes.**
The entries of the table are ratios of model unconditional moments to their empirical counterparts, based on parameter estimates of the single factor SVS and the HN-S volatility models on market indexes. To perform the GMM estimation for all the three models, we use the same eleven moment conditions that are all significant for big stocks and market indexes as depicted in Figures 4.2, 4.3 and 4.4. The corresponding moments are identified by a 1 in the third column. Return data are daily and span the period from January 2, 1990 to December 30, 2005, for a total of 4036 observations.

			CRSP				S&P500			
			$\eta \neq 0$	$\eta = 0$	HN	IG	$\eta \neq 0$	$\eta = 0$	HN	IG
$E[r_t]$	1	0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
$E[r_t^2]$	2	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$E[r_t^3]$	3	1	0.80	5.43	0.95	0.94	0.90	-2.81	1.02	1.04
$E[r_t^4]$	4	1	0.98	0.94	0.97	0.97	1.00	0.98	1.00	0.99
$E[r_t r_{t+5}]$	5	0	-2.03	-0.98	-0.44	0.30	-1.63	-0.39	-0.06	0.50
$E[r_t^2 r_{t+5}]$	6	0	-1.40	-0.98	-0.07	0.58	-1.02	-0.39	0.18	0.63
$E[r_t r_{t+5}^2]$	7	0	1.57	1.10	1.51	1.56	1.70	0.65	1.60	1.62
$E[r_t^3 r_{t+5}]$	8	0	-0.58	-0.36	-0.12	0.12	-0.43	-0.13	-0.01	0.15
$E[r_t^2 r_{t+5}^2]$	9	1	0.89	0.97	0.89	0.90	0.91	0.94	0.90	0.90
$E[r_t r_{t+5}^3]$	10	0	-0.60	-0.37	-0.15	0.09	-0.52	-0.16	0.02	0.22
$E[r_t r_{t+4}]$	11	0	7.39	3.44	1.61	-1.11	8.40	1.98	0.32	-2.63
$E[r_t^2 r_{t+4}]$	12	0	-1.62	-1.08	-0.10	0.65	-1.35	-0.52	0.23	0.83
$E[r_t r_{t+4}^2]$	13	0	12.88	8.60	12.52	12.81	-8.19	-3.15	-7.85	-7.93
$E[r_t^3 r_{t+4}]$	14	0	1.85	1.09	0.38	-0.37	1.05	0.32	0.03	-0.37
$E[r_t^2 r_{t+4}^2]$	15	1	1.03	1.10	1.03	1.04	1.08	1.11	1.08	1.08
$E[r_t r_{t+4}^3]$	16	0	42.62	25.28	10.39	-6.07	-2.41	-0.73	0.10	1.06
$E[r_t r_{t+3}]$	17	0	-4.46	-1.99	-0.97	0.68	-1.77	-0.41	-0.07	0.57
$E[r_t^2 r_{t+3}]$	18	0	-2.62	-1.66	-0.19	1.01	-2.22	-0.86	0.36	1.35
$E[r_t r_{t+3}^2]$	19	1	2.11	1.34	2.07	2.11	2.11	0.81	2.06	2.08
$E[r_t^3 r_{t+3}]$	20	0	1.07	0.61	0.22	-0.21	0.79	0.24	0.02	-0.28
$E[r_t^2 r_{t+3}^2]$	21	1	0.98	1.02	0.98	0.98	0.96	0.98	0.97	0.97
$E[r_t r_{t+3}^3]$	22	0	2.69	1.53	0.66	-0.38	-7.00	-2.12	0.29	3.13
$E[r_t r_{t+2}]$	23	0	-3.66	-1.56	-0.80	0.56	-2.42	-0.56	-0.09	0.79
$E[r_t^2 r_{t+2}]$	24	0	-4.06	-2.45	-0.33	1.52	-3.27	-1.27	0.50	1.98
$E[r_t r_{t+2}^2]$	25	1	0.67	0.41	0.67	0.67	0.66	0.26	0.66	0.67
$E[r_t^3 r_{t+2}]$	26	0	-1.62	-0.88	-0.34	0.32	-3.97	-1.20	-0.10	1.45
$E[r_t^2 r_{t+2}^2]$	27	1	0.89	0.90	0.89	0.89	0.98	1.00	1.00	1.00
$E[r_t r_{t+2}^3]$	28	0	1.77	0.96	0.44	-0.25	3.70	1.12	-0.16	-1.68
$E[r_t r_{t+1}]$	29	0	1.89	0.77	0.42	-0.29	-24.81	-5.73	-0.91	8.29
$E[r_t^2 r_{t+1}]$	30	0	-1.57	-0.91	-0.14	0.57	-1.17	-0.45	0.17	0.71
$E[r_t r_{t+1}^2]$	31	1	1.05	0.61	1.06	1.06	0.99	0.38	1.01	1.02
$E[r_t^3 r_{t+1}]$	32	0	-0.58	-0.30	-0.12	0.12	-0.36	-0.11	-0.01	0.14
$E[r_t^2 r_{t+1}^2]$	33	1	1.05	1.04	1.06	1.06	0.99	1.01	1.01	1.01
$E[r_t r_{t+1}^3]$	34	0	1.58	0.83	0.40	-0.22	6.86	2.07	-0.30	-3.17

Figure 4.1: **Return Series.**

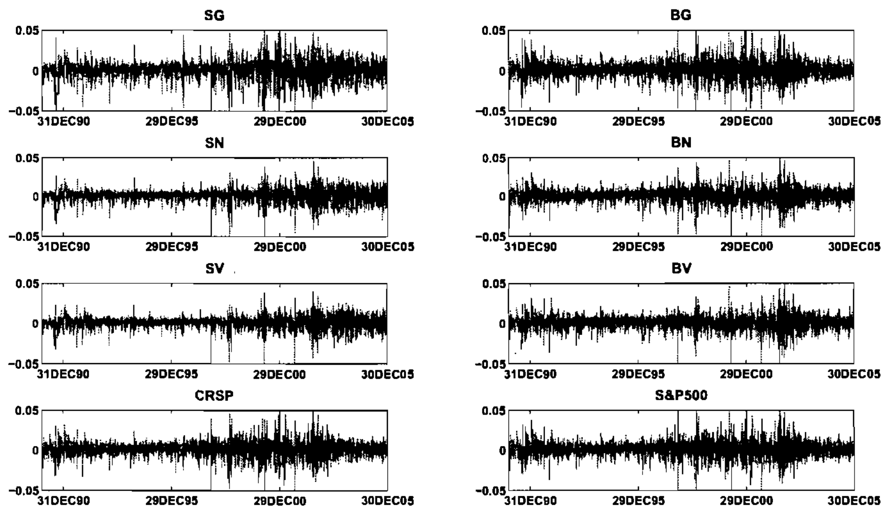


Table 4.9: **Summary Statistics for Strike Price and Maturity Categories.**

(a) Summary statistics by moneyness						
	<0.95	**0.95 to 0.975**	**0.975 to 1**	**1 to 1.025**	**>1.025**	**all**
Number of Contracts	3192	2351	3851	3048	3216	15658
Average Call Price	29.412	32.625	37.292	47.294	81.372	45.986
Average IV	0.195	0.194	0.194	0.202	0.232	0.203

(b) Summary statistics by Maturity						
	1	**2**	**3**	**4-6**	**7-12**	**all**
Number of Contracts	2061	4931	2571	2974	3121	15658
Average Call Price	31.119	38.680	42.018	51.108	65.731	45.986
Average IV	0.207	0.204	0.203	0.202	0.201	0.203

(c) Summary statistics by moneyness and maturities. For each moneyness and strike price category, the first line gives the number of contracts and the second line give the average Implied Volatility

	Moneyness				
Months	**<0.95**	**0.95 to 0.975**	**0.975 to 1**	**1 to 1.025**	**>1.025**
1	7	154	658	699	543
	0.254	0.201	0.187	0.197	0.245
2	277	738	1389	1201	1326
	0.208	0.188	0.189	0.198	0.234
3	439	494	720	454	464
	0.198	0.191	0.197	0.207	0.225
4-6	928	513	574	406	553
	0.193	0.194	0.199	0.207	0.223
7-12	1541	452	510	288	330
	0.193	0.204	0.205	0.213	0.224

Table 4.10: **Estimation of Structural Parameters of Risk-Neutral Processes**
The entries of the table are parameters estimate of different risk-neutral models. These models have been estimate using the two steps non-linear least squared procedure detailed in section 4.6.1.

Param.	SVS1F	SVS1F, $\eta_1 = 0$	SVS2F	Param.	HN	CHJ
β_1^*	-1.808E+01	-1.210E+02	1.348E+02	γ^*	2.410E+02	1.223E-06
η_1^*	-2.135E-01	0.000E+00	-1.340E-01	ω^*	5.450E-13	1.266E-06
v_1^*	1.347E-02	2.115E-01	5.474E-01	β^*	7.554E-01	9.730E-01
α_1^*	1.177E-04	5.270E-06	1.933E-16	α^*	3.592E-06	1.223E-06
ϕ_1^*	9.989E-01	9.824E-01	9.979E-01	η^*	0.000E+00	-6.218E-02
β_2^*			9.838E-01			
η_2^*			-1.856E-01			
v_2^*			1.108E+00			
α_2^*			4.824E-16			
ϕ_2^*			9.207E-01			
RRMSE	0.078	0.103	0.058		0.092	0.085

Table 4.11: **Relative RMSE by Moneyness and Maturity**
The entries of the table are the relative root-mean squared errors, defined as the ratio between the root mean squared error (RMSE) and the average option price.

	(a) Moneyness					
	<0.95	**0.95 to 0.975**	**0.975 to 1**	**1 to 1.025**	**>1.025**	**all**
SVS1f	0.142	0.089	0.080	0.062	0.055	0.078
SVS1f, $\eta = 0$	0.215	0.122	0.092	0.072	0.072	0.103
SVS2f	0.105	0.052	0.057	0.047	0.044	0.058
HN	0.158	0.129	0.106	0.075	0.057	0.092
CHJ	0.147	0.112	0.094	0.072	0.056	0.085
	(b) Maturity					
	1	**2**	**3**	**4-6**	**7-12**	**all**
SVS1f	0.106	0.066	0.048	0.053	0.092	0.078
SVS1f, $\eta = 0$	0.105	0.069	0.072	0.088	0.125	0.103
SVS2f	0.062	0.038	0.038	0.052	0.070	0.058
HN	0.118	0.081	0.051	0.060	0.109	0.092
CHJ	0.116	0.069	0.062	0.057	0.100	0.085

Table 4.12: **Relative Bias by Moneyness and Maturity**

The entries of the table are the relative option bias, defined as the ratio between the bias and the average option price. The bias is the difference between model's price and observed price.

	(a) Moneyness					
	<0.95	**0.95 to 0.975**	**0.975 to 1**	**1 to 1.025**	**>1.025**	**all**
SVS1f	-0.018	-0.008	0.022	0.018	-0.001	0.005
SVS1f, $\eta = 0$	0.029	-0.020	-0.011	-0.009	0.023	0.006
SVS2f	-0.009	-0.008	0.014	0.014	0.006	0.006
HN	-0.015	0.026	0.036	0.015	-0.016	0.005
CHJ	-0.043	0.009	0.025	0.014	-0.003	0.002
	(b) Maturity					
	1	**2**	**3**	**4-6**	**7-12**	**all**
SVS1f	-0.027	0.015	0.013	0.003	0.003	0.005
SVS1f, $\eta = 0$	-0.065	0.011	0.031	0.020	-0.000	0.006
SVS2f	0.012	0.007	-0.003	0.005	0.007	0.006
HN	-0.015	0.020	0.021	-0.003	-0.004	0.005
CHJ	-0.067	0.007	0.036	0.017	-0.010	0.002

Figure 4.2: **Autocorrelation of Squared Returns.**

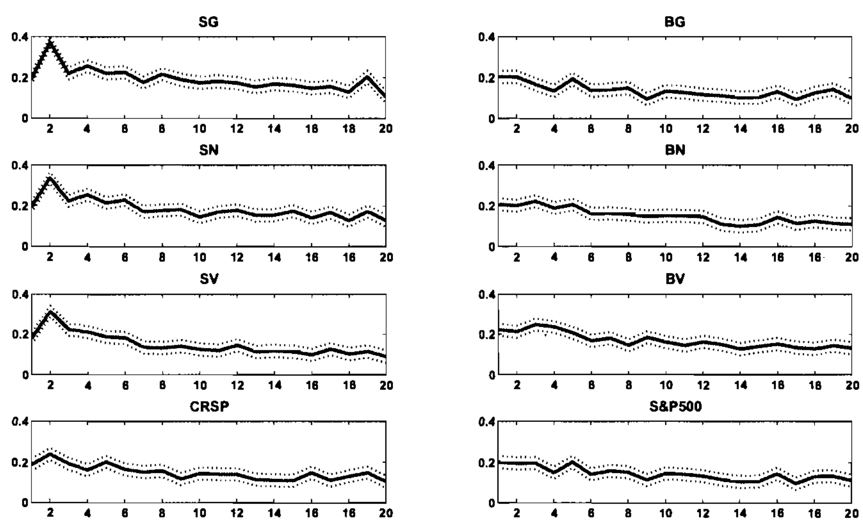
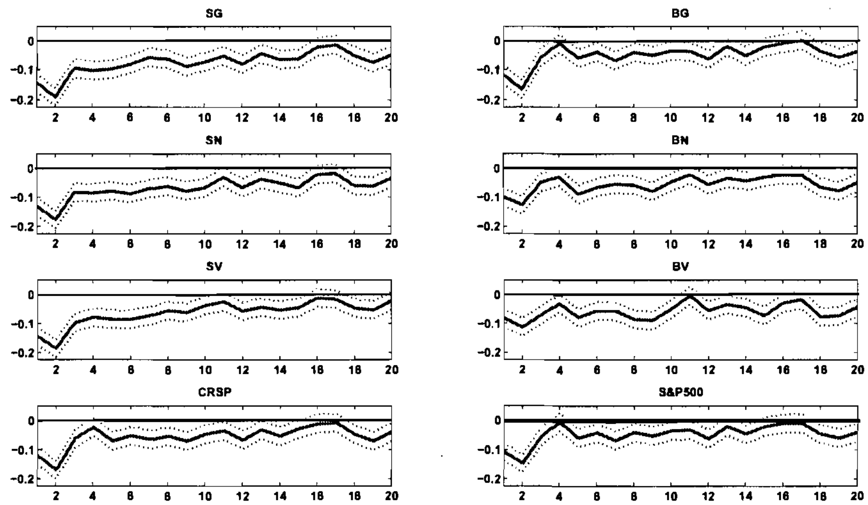


Figure 4.3: **Cross-Correlations Between Returns and Squared Returns.**

Panel A: $Corr\left(r_t, r_{t+j}^2\right)$



Panel B: $Corr\left(r_t^2, r_{t+j}\right)$

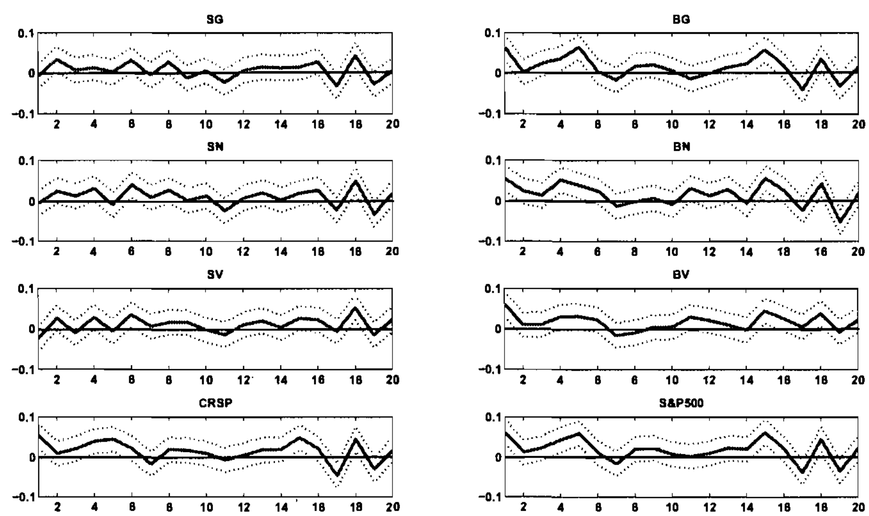
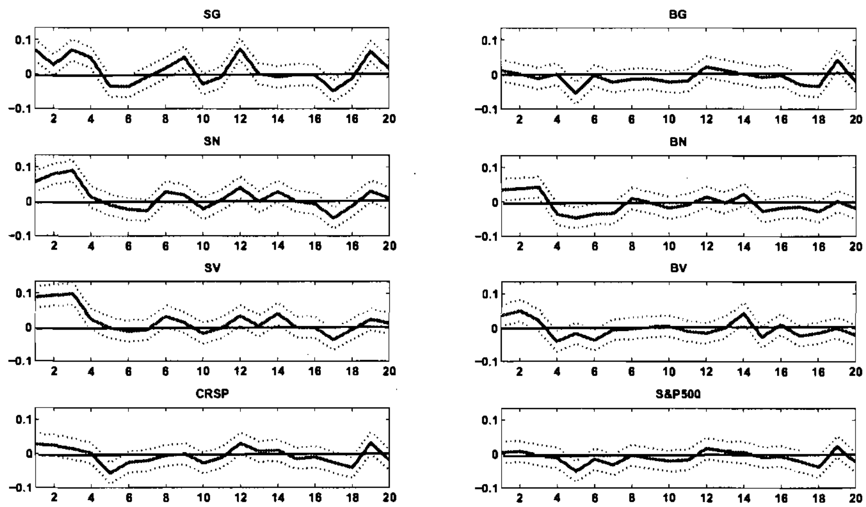


Figure 4.4: **Cross-Correlations Between Returns and Cubed Returns.**

Panel A: $Corr\left(r_t, r_{t+j}^3\right)$



Panel B: $Corr\left(r_t^3, r_{t+j}\right)$

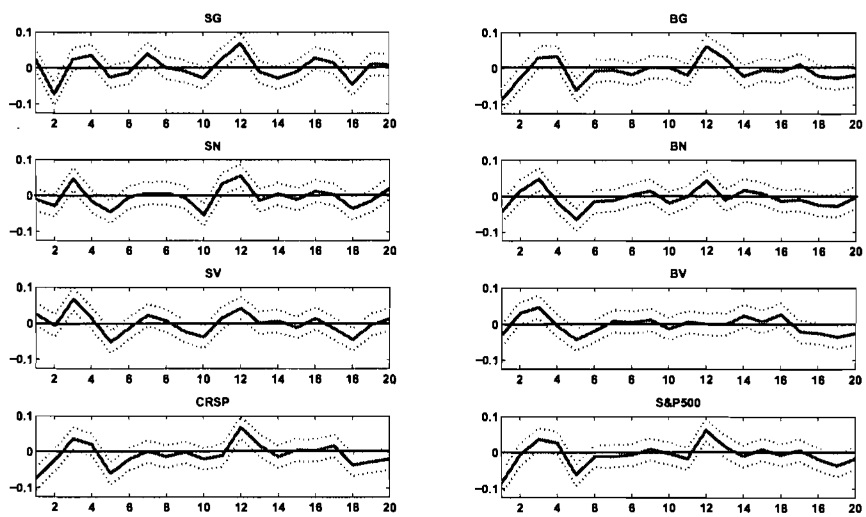


Figure 4.5: **Portfolios Volatility and Skewness: Market Indexes**

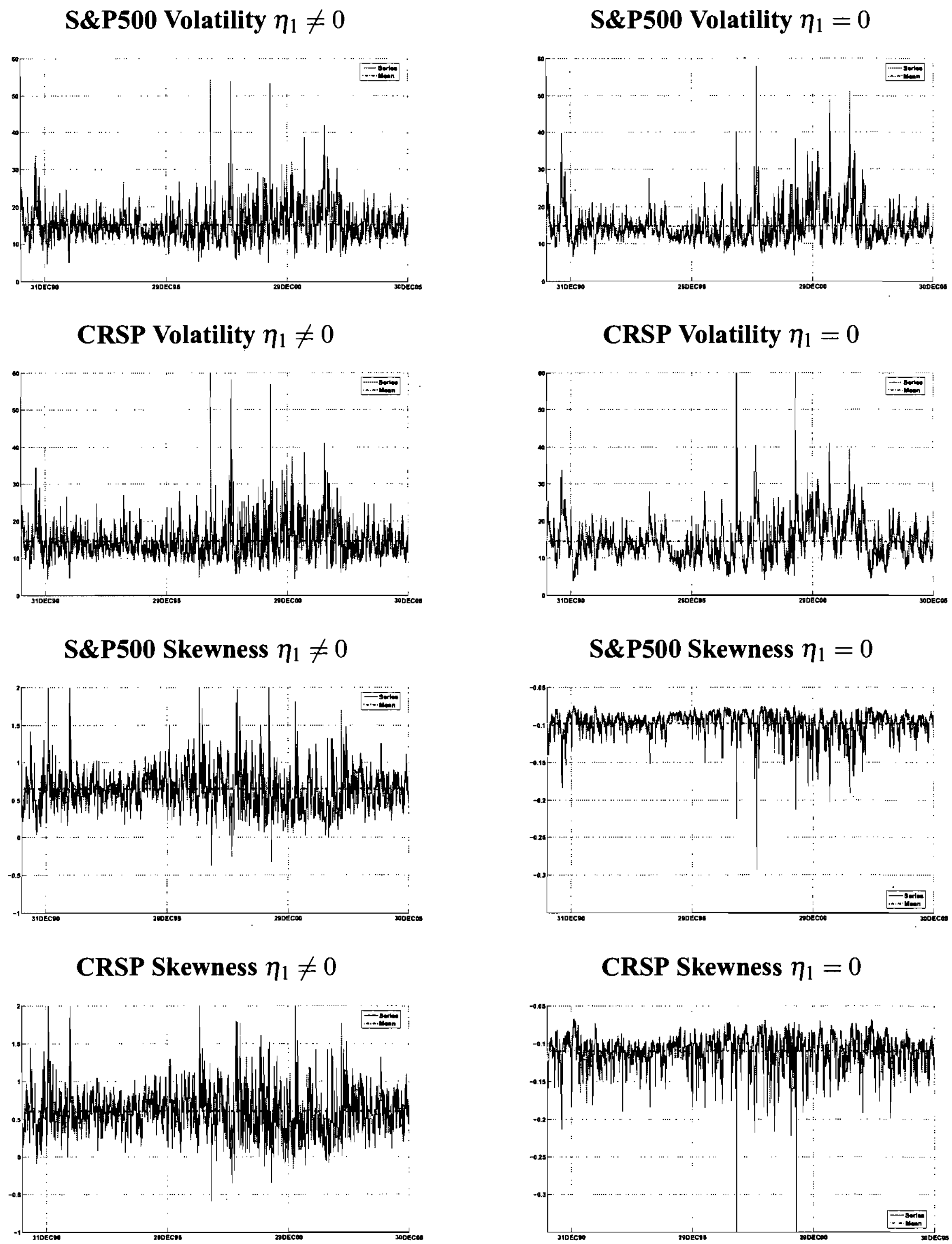


Figure 4.6: **Portfolios Volatility and Skewness: Small and Big Stocks**

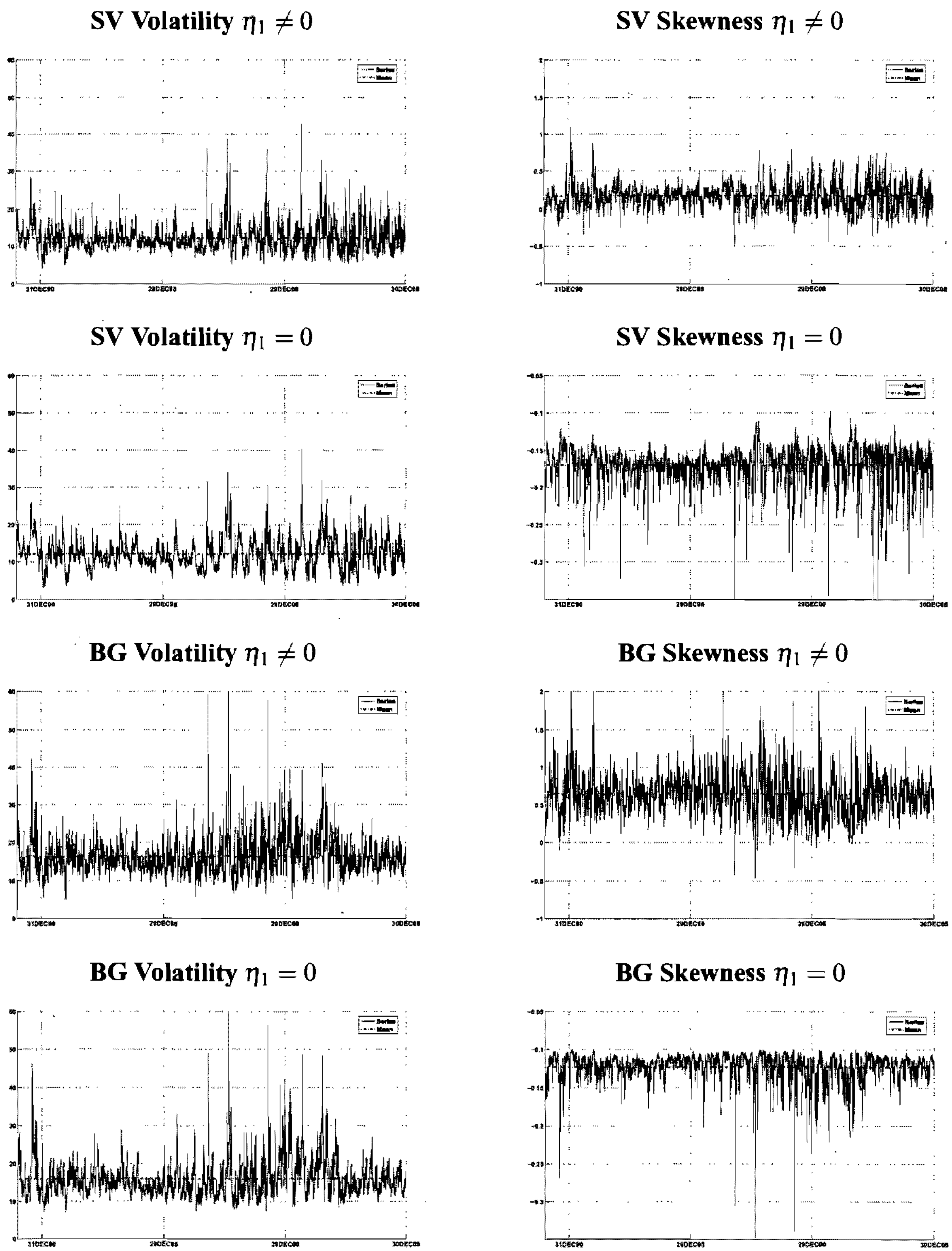
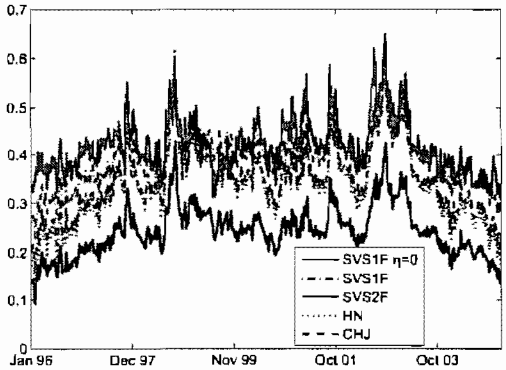
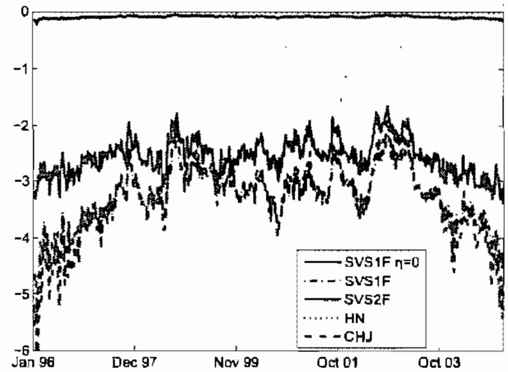


Figure 4.7: **Risk-Neutral Volatility and Conditional Skewness**

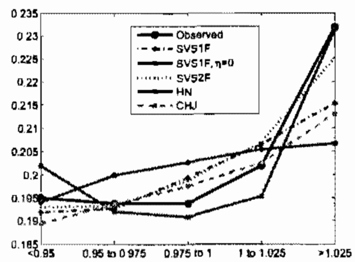


Volatility

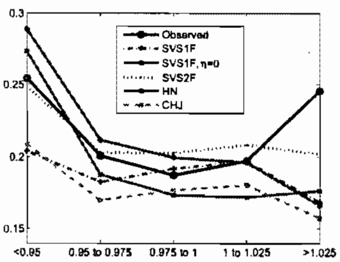


Conditional Skewness

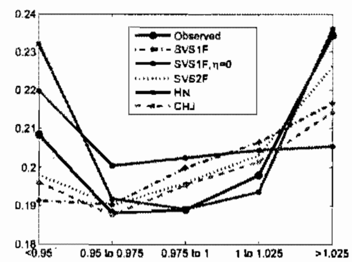
Figure 4.8: **Implied BSM volatility by Moneyness, Maturity and Model**



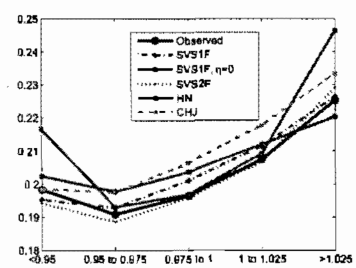
all



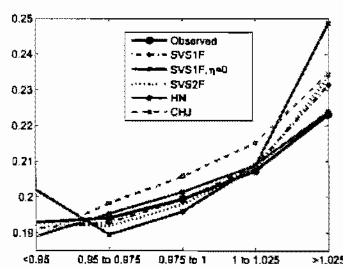
1 month



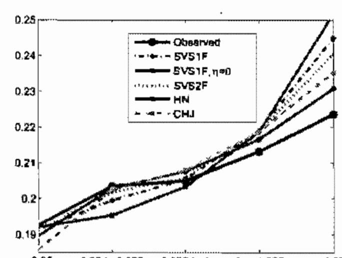
2 months



3 months

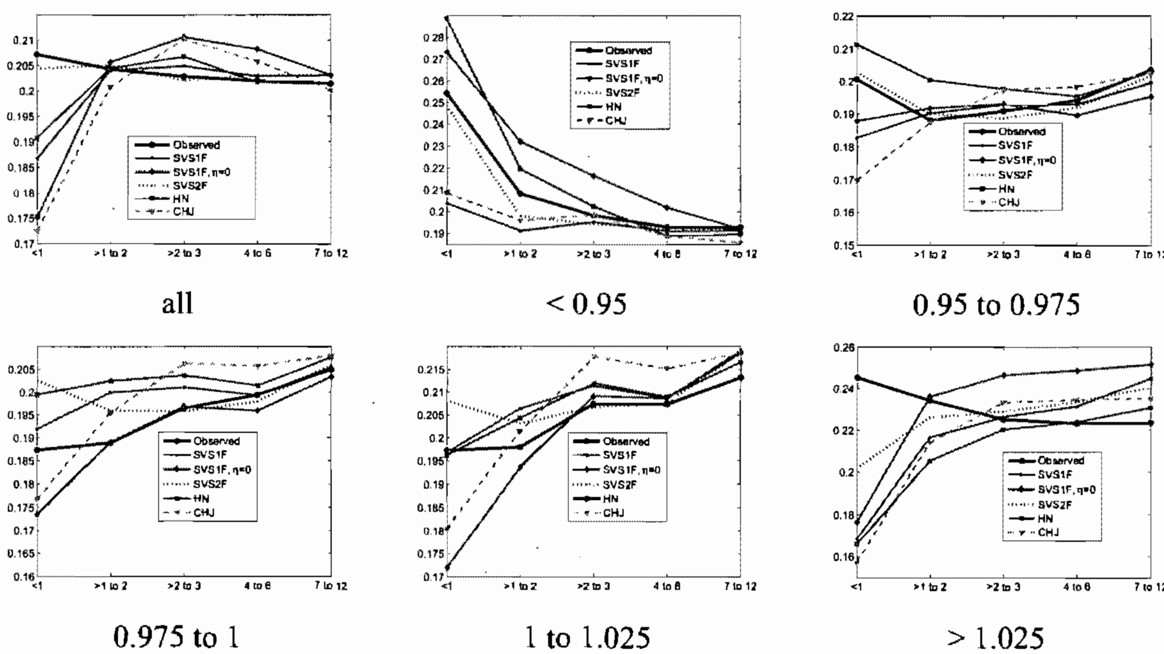


4-6 months



7-12 months

Figure 4.9: **Implied BSM volatility by Maturity, Moneyness and Model**



CONCLUSION GÉNÉRALE

Dans cette thèse, nous avons généralisé les modèles affines en introduisant une composante du type moyenne mobile dans la fonction cumulant conditionnelle. Cette extension est importante théoriquement parce que des modèles existant comme les modèles ARMA ne sont pas affines, par ailleurs nous montrons comment construire des modèles affines d'ordre infini parcimonieux. Cette extension est aussi importante empiriquement comme le montre nos trois exemples empiriques[7]. En particulier notre exemple empirique sur la structure à terme des taux d'intérêt montre qu'un modèle affine généralisé sur les traditionnels trois facteurs prédit mieux la courbe des taux qu'un modèle affine sur les trois facteurs et les variables macroéconomiques. Ce qui nous permet de conclure qu'ajouter une composante MA permet de prendre en compte les facteurs qui ne sont pas éléments de l'espace vectoriel engendré par les trois facteurs. Il existe une approche alternative qui conduit à des processus non-Markoviens affines. Elle utilise la transformée de Laplace conditionnelle du processus x_t défini comme $\mathscr{L}_t(u) = \exp(\psi_t(u))$ en lieu et place de la fonction cumulant conditionnelle. Le modèle affine traditionnel est caractérisé par

$$\mathscr{L}_t(u) = \exp(\omega(u) + \alpha(u)x_t).$$

Dans un travail en cours, nous étudions présentement le processus défini comme suit

$$\mathscr{L}_t(u) = \gamma(u) + \exp(\omega(u) + \alpha(u)x_t) + \beta(u)\mathscr{L}_{t-1}(u).$$

Une application plus approfondie des modèles affines généralisés a été faite en structure à terme des taux d'intérêt, en étudiant le modèle VARMA. D'un point de vue théorique nous montrons comment étendre un modèle VAR d'absence d'opportunité d'arbitrage à un modèle VARMA. Dans le modèle VARMA, pour une maturité donnée, le taux d'intérêt est une fonction affine de la variable d'état et de sa moyenne conditionnelle, par conséquent il dépend de toutes les réalisations passées de la variable d'état. En utilisant une procédure d'estimation à deux étapes, le modèle s'estime aisément à l'aide

7. Le premier exemple a été présenté dans le chapitre 1, et les deux autres sont fournis en appendice I.

du maximum de vraisemblance dans la première étape, pour l'estimation des paramètres de la dynamique historique de la variable d'état. Cette étape est suivie par l'estimation des paramètres du taux d'escompte stochastique par la méthode des moindres carrés non linéaires contrainte. Il existe des similarités entre les structures à terme VAR et VARMA, en particulier elles s'accordent sur le signe de l'inflation et du niveau d'activité espérées sur la courbe des taux. Mais elles divergent sur les composantes imprévisibles de ces deux agrégats. Finalement nos résultats indiquent qu'un modèle d'absence d'opportunité d'arbitrage VARMA sur les taux à maturités 1, 24 et 96 mois est meilleur que le modèle "Nelson–Siegel avec facteurs dynamiques AR(1)".

Le troisième chapitre nous fourni des résultats sur l'évaluation des titres contingents en temps discret et avec un espace-état de dimension infini. Notre résultat d'évaluation s'applique sur une large classe de processus avec innovation conditionnellement normale et non-normale, moyenne et variance conditionnelle variables et potentiellement des variations du prix du risque. Ce cadre généralise les résultats de Duan(1995) dans le sens où nous ne nous restreignons pas à la normalité conditionnelle, et au prix du risque constant. Nos résultats s'appliquent à des processus populaires en temps discret tels que les modèles GARCH. Nous appliquons aussi notre approche aux modèles discrets avec innovations multiples. Pour la classe de processus infiniment divisibles, les dynamiques risque-neutre et historiques sont dans la même famille. Nous donnons quelques intuitions de ces résultats en faisant le lien avec la risque-neutralisation faite en temps continue.

Pour démontrer la pertinence empirique de notre approche, nous fournissons en appendice III une analyse d'un modèle GARCH avec innovation conditionnelle qui suit une distribution variance gamma asymétrique construite à l'aide de la convolution de deux lois gamma. Par conséquent la skewness et la kutosis conditionnelle entrent directement et distinctement comme paramètres de cette nouvelle distribution. Nous l'estimons par la méthode du maximum de vraisemblance et montrons que le modèle est largement préféré au modèle conditionnellement normal. Une analyse des "smirks" sur les options démontre que ce modèle offre plus de flexibilité pour évaluer les options. Notre approche présente une manière de résoudre des questions méthodologiques et empiriques importantes en évaluation des options. Il existe sans aucun doute d'autres approches. Au niveau empirique, une combinaison de la non normalité et de l'hétéroscédasticité permet de corriger les biais reliés au GARCH normaux. Ces biais sont similaires à ceux engendrés par le modèle d'Heston(1993a), en temps continu ces biais se corrigent en introduisant des sauts dans les rendements et la volatilité. Ce chapitre est dès lors relié aux études empiriques sur les modèles avec sauts. Au point de vue méthodologique, il est probable que notre risque-neutralisation se dérive en spécifiant le taux d'escompte stochastique en lieu et place d'une mesure de martingale équivalente (MME), ou en utilisant la transformée d'Esscher, mais nous commençons plutôt par la spécification d'une famille de densité de Radon-Nikodym et nous caractérisons la MME dans cette famille. Certaines questions restent encore inexplorées. Premièrement, bien que nous obtenons une MME unique pour ce choix de famille de densité de Radon-Nikodym, nous ne pouvons pas exclure le fait que pour une spécification donnée de la prime de risque, il existe d'autres MMEs correspondant à différents choix de la forme fonctionnelle de la densité de Radon-Nikodym. Deuxièmement, ce serait intéressant d'explorer plus en détail la relation entre nos résultats et la littérature sur la dominance stochastique qui entraîne les bornes sur les prix d'options. Troisièmement, bien que nous préconisions la distinction entre un principe d'évaluation et le modèle d'équilibre général qui le sous-tend, les fondements économiques de nos résultats sont bien sûr très importants. Cependant, cette problématique n'est pas évidente, et fera l'objet de recherches futures.

Dans le quatrième chapitre de cette thèse, nous avons fourni un nouveau modèle affine multivarié à variables latentes pour les rendements journaliers. Dans ce modèle, la variance et l'asymétrie conditionnelles sont des combinaisons linéaires de facteurs stochastiques. Nous avons caractérisé ces moments conditionnels critiques tels que perçus par l'agent économique, ainsi que leurs contreparties telles que vues par l'économètre. Le modèle permet d'obtenir des formules analytiques aussi bien pour les moments en population des rendements que pour les prix d'actifs financiers. Nous développons ensuite une procédure d'estimation par la méthode des moments généralisée. Nous argumentons que cette procédure présente un énorme avantage par rapport à l'estimation par maximum de vraisemblance. En outre elle permet de reproduire parfaitement des moments critiques des rendements tels que l'asymétrie et l'aplatissement tandis que la plupart des méthodes y échouent. Nous avons appliqué cette nouvelle procédure d'estimation au cas univarié de notre modèle et avons estimé le facteur latent grâce à une variante du filtre de Kalman non linéaire. Les résultats ont montré que l'asymétrie inconditionnelle est déterminante pour l'évaluation d'actifs financiers. Plus frappant encore, une asymétrie positive de la distribution des rendements courants conditionnellement à la volatilité courante est nécessaire et suffisante pour reproduire l'asymétrie et les effets de levier inconditionnels, mais engendre une asymétrie positive de la distribution des rendements courants conditionnellement aux rendements passés, ce qui est contraire aux résultats empiriques connus. Ce résultat étonnant et robuste demande d'examiner plus rigoureusement la question de savoir si un modèle reproduisant parfaitement les asymétries inconditionnelles génèrerait une asymétrie conditionnelle négative. Cette dernière question constitue une recherche en cours, ainsi que l'estimation du modèle bivarié et ses implications pour la valorisation des produits dérivés et la structure à terme des taux d'intérêt.

Appendix I

Additional Empirical results : Chapter 1

The Term Structure of Realized Risk

This example is studied in Feunou and Meddahi (2007a). There are two goals. The first one is to model the joint dynamics of the returns and the realized variance. The second goal is to compute the term structure of the value-at-risk, i.e. to characterize the quantile function of the aggregated returns, $\sum_{i=1}^{h} r_{t+i}$, when h varies.

We consider the daily realized variance computed as the sum of squared intra-daily returns, five-minutes and thirty-minutes returns in our empirical application. The recent literature on volatility shows the importance of such measures. The basic theory on realized volatility assumes that the underlying process is in continuous time and shows that the realized variance converges to the integrated variance when the length of intra-day returns goes to zero. In our empirical analysis, we specify the model in discrete time and we do not make the formal connection between the realized variance and the daily returns. We will specify discrete models, affine or generalized affine, and allow the data to select the best model. We will, however, use some insights from continuous time when we specify the discrete model. In what follows the conditioning information is $I_t = \sigma(r_\tau, RV_\tau, \tau \leq t)$ where r_t is the daily returns.

We start our analysis by modeling the realized variance as either an affine process or a generalized one. Consider the affine model given by

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t. \tag{I.1}$$

Given the non-negativity of the realized variance process, we will consider two examples. The first one corresponds to the Inverse Gaussian case while the second is the Gamma case, which corresponds to the exact discretization of the square-root process, studied in Gourieroux and Jasiak (2006)):

$$\text{Inverse Gaussian}: \quad \omega(u) = \nu(1-\sqrt{1-2u\mu}), \; \alpha(u) = \frac{\rho}{\mu}(\exp(1-\sqrt{1-2u\mu})-1) \tag{I.2}$$

$$\text{Gamma}: \quad \omega(u) = -\nu\log(1-u\mu), \; \alpha(u) = \frac{\rho u}{1-u\mu}. \tag{I.3}$$

When we extend our analysis to the generalized affine case, i.e.,

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t + \beta\psi_{t-1}(u), \tag{I.4}$$

we still consider the same two examples of Inverse-Gaussian and Gamma. We prove in Section 2 that this leads to a proper cumulant function.

We use the maximum likelihood method to estimate the four models (two models on realized variance only, and two on joint realized variance and returns). Joint estimation or estimation on realized variance only yield quite the same estimator for the realized variance dynamic. Also, inverse gaussian and gamma model provide very similar results. For this reason in this paper we report only the estimation of the dynamic of returns conditional on realized variance and the dynamic of realized variance given by the maximization of the joint likelihood of returns and realized variance in the inverse gaussian case. These empirical results are provided in Table I.1. The main empirical result is that the coefficient β is non-zero whatever the model or the realized volatility measure (based on five-minutes or thirty-minutes). In particular, the increase of the log-likelihood is substantial when one allows β to be non-zero. Another interesting result is that the inverse Gaussian model describes better the data for the two frequencies.

We now want to specify a joint model for the returns and the realized variance. When one considers a continuous time stochastic volatility model

$$d\log p_u = (a+b\sigma_u^2)du + \sigma_u dW_u,$$

and assumes that there is no leverage effect, one gets that the daily return $r_{t+1} = \log(p_{t+1}) -$

$\log(p_t)$ has the following distribution:

$$r_t|\sigma(p_\tau,\sigma_s,\tau\leq t,s\leq t+1) \sim \mathcal{N}(a+bIV_{t+1},IV_{t+1}),$$

which suggests the following discrete time model that we study:

$$r_{t+1} \mid \sigma(r_\tau,RV_\tau,RV_{t+1},\tau\leq t) \sim \mathcal{N}(a+bRV_{t+1},c+dRV_{t+1}). \tag{I.5}$$

We assume that RV_{t+1} follows (I.4) where $\alpha(u)$ follows either (I.2) or (I.3). By denoting the joint cumulant function of (r_{t+1},RV_{t+1}) as $\psi_{r,RV;t}(v,u)$ defined by

$$\psi_{r,RV;t}(v,u) \equiv \log E_t[\exp(vr_{t+1}+uRV_{t+1})],$$

one gets

$$\psi_{r,RV;t}(v,u) = (va+v^2c/2)+\psi_t(vb+v^2d/c+u).$$

Hence, the joint process r_t,RV_t is indeed a generalized affine process because one has

$$\psi_{r,RV;t}(v,u) = \tilde{\omega}(v,u)+\tilde{\alpha}(v,u)RV_t+\beta\psi_{r,RV;t-1}(v,u), \tag{I.6}$$

where

$$\tilde{\omega}(v,u) = (va+v^2c/2)(1-\beta)+\omega(vb+v^2d/2+u) \tag{I.7}$$

$$\tilde{\alpha}(v,u) = \alpha(vb+v^2d/2+u). \tag{I.8}$$

We compute the term structure of the Value-at-Risk, i.e., we compute the 5%-quantile of

$$\bar{r}_{t+1:t+h} \equiv \frac{1}{\sqrt{h}}\sum_{i=1}^{h} r_{t+i}.$$

For this purpose, we derive the conditional characteristic function of $\bar{r}_{t+1:t+h}$ and then we invert it to get the cumulative distribution function. This approach has been used in the affine case and continuous time by Duffie and Pan (2001).

In practice, the value at risk of $\bar{r}_{t+1:t+h}$ will depend on RV_t its lagged values. In order to graphically present the results, one needs to choose RV_t. We proceed by taking from the data three values for RV_t: a small value (low case), a median one (median case) and a large one (high case). Then, we use the lagged values of each of them to plot the term structure of the value-at-risk (VaR).

Figures I.2 to I.5 present and compare Affine and Generalized affine term structure of the value-at-risk. Figure I.3 shows that in a low variance day, the VaR increases with the maturity and that the affine model overestimates the VaR. In contrast, in a high or median volatility day, affine model overestimates the VaR for lowest maturity and underestimates it for longer maturities. Underestimation of the VaR could lead to important risk management problems; see Feunou and Meddahi (2007a) for more discussions. Likewise, we show in Feunou and Meddahi (2007a) that it is useful to consider realized variances, i.e., we did the same approach with the Heston and Nandi (2000) daily model and show that the model with realized volatility is the best one. We also provide in Feunou and Meddahi (2007a) the term structure of another risk measure called the expected shortfall.

Realized Option Pricing model

This subsection hinges on Feunou, Christoffersen, Jacobs and Meddahi (2007). We used the model developed in the first empirical example and used the option pricing formulas derived in Section 1.3.2 where h_t equals the realized variance RV_t. We model jointly the dynamics of the return r_t and realized variance RV_t in the same way as in the previous example, with a slight modification of the distribution of the stock log-returns r_t conditional on realized variance RV_t. Following Christoffersen et al (2006), Feunou (2006) and Feunou and Tedongap (2007), we used a skewed inverse Gaussian distribution, which nests the normal distribution. This extension is empirically important.

The model is given by

$$r_{t+1} \mid \sigma(r_\tau, RV_\tau, RV_{t+1}, \tau \leq t) \sim a + bRV_{t+1} - \eta(c + dRV_{t+1}) + \frac{1}{\eta} y_{t+1}, \qquad \text{(I.9)}$$

with $y_{t+1} \sim \mathscr{IG}(\eta^2(c+dRV_{t+1}))$. $\mathscr{IG}$ means the standard inverse gaussian distribution. The conditional cumulant function of the return r_{t+1} conditional on I_t and RV_{t+1} is given by

$$E[\exp(ur_{t+1})|RV_{t+1}, I_t] = \exp(\omega_0(u) + \alpha_0(u)RV_{t+1}),$$

with

$$\omega_0(u) = u(a-c\eta)+c\eta^2\left(1-\sqrt{1-\frac{2u}{\eta}}\right), \text{ and } \alpha_0(u) = u(b-d\eta)+d\eta^2\left(1-\sqrt{1-\frac{2u}{\eta}}\right).$$

In the affine case, the conditional cumulant function of RV_{t+1} given I_t is given by (I.1) where $\omega(u)$ and $\alpha(u)$ are defined either by (I.2) for the inverse gaussian case or by (I.3) for the gamma case. We extend this affine case to the generalized affine of order (1,2) as follows

$$\psi_t(u) = \log E_t[\exp(uRV_{t+1})] = \omega(u) + \alpha(u)RV_t + \beta_1\psi_{t-1}(u) + \beta_2\psi_{t-2}(u). \quad \text{(I.10)}$$

Consequently, the joint cumulant function of (r_{t+1}, RV_{t+1}) given I_t is

$$\psi_{r,RV;t}(v,u) = \omega_0(v) + \psi_t(u + \alpha_0(v)).$$

Eq. (I.10) implies that the joint process (r_t, RV_t) is a generalized affine process

$$\psi_{r,RV;t}(v,u) = \tilde{\omega}(v,u) + \tilde{\alpha}(v,u)RV_t + \beta_1\psi_{r,RV;t-1}(v,u) + \beta_2\psi_{r,RV;t-2}(v,u), \quad \text{(I.11)}$$

with $\tilde{\omega}(v,u) = \omega_0(v)(1-\beta_1-\beta_2) + \omega(u+\alpha_0(v))$ and $\tilde{\alpha}(v,u) = \alpha(u+\alpha_0(v))$.

We assume that the generalized affine model is defined under the risk-neutral probability measure. The estimation is done by minimizing the MSE of the implied Black-Scholes volatility from the option (IVMSE) defined as

$$IVMSE = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i - \sigma_i(\theta))^2,$$

where the implied volatilities are obtained as

$$\sigma_i = BS^{-1}\left(C_i, T_i, X_i, S, r\right) \quad \text{and} \quad \sigma_i(\theta) = BS^{-1}\left(C_i(\theta), T_i, X_i, S, r\right),$$

with BS^{-1} being the inverse of the Black-Scholes formula, T_i the time to maturity, X_i the strike price, S the price of the underlying stocks and r the riskless interest rate. Figures I.6, I.7 and I.8 represent the daily implied volatility bias, option price bias and implied volatility RMSE. The generalized affine model clearly outperforms the affine model in terms of pricing errors. This result holds whatever the maturity of the moneyness; see Tables I.2 and I.3.

Table I.1: **MLE Estimation ARIG**

The data is the Deutsche mark (DM) / US dollar (USD) exchange rate returns and realized variance. Sample period is 1986:12:01 to 1996:12:01 with a total of 2449 observations

	30 min				5 min			
	Affine		G-Affine		Affine		G-Affine	
par	Est	STD	Est	STD	Est	STD	Est	STD
β			0.6111	0.0396			0.5449	0.0419
ρ	0.3255	0.0203	0.1754	0.0179	0.3444	0.0193	0.2150	0.0192
μ	0.2341	0.0114	0.1834	0.0087	0.1642	0.0071	0.1328	0.0058
v	1.2565	0.0398	0.5045	0.0545	2.0818	0.0647	0.9380	0.0961
a	0.0063	0.0139	0.0063	0.0140	0.0064	0.0180	0.0064	0.0180
c	-0.0214	0.0448	-0.0214	0.0449	-0.0180	0.0433	-0.0180	0.0434
b	1.74E-08	5.709E-06	1.44E-08	6.024E-06	4.98E-08	1.093E-05	1.54E-08	5.777E-06
d	0.9282	0.0290	0.9282	0.0290	0.7551	0.0236	0.7551	0.0236
LIK	-1600.0932		-1547.4719		-1838.6743		-1790.0531	
BIC	0.8069		0.7850		0.9234		0.9034	

Table I.2: **Implied volatilities, Option prices RMSEs and bias by Moneyness**

We estimate the models on a total of 16, 506 contracts with an average call price of 46.05 and average implied volatility of 20.26. The estimation have been done by minimizing the Black-Scholes IVRMSE

Moneyness					
	S/X<0.975	0.975<S/X<1	1<S/X<1.025	1.025<S/X	All
Model	IVRMSE (%)				
Affine	3.8809	4.2988	4.4313	5.0642	4.3768
G-Affine	2.9471	2.9476	3.2201	3.7181	3.1915
Model	IV bias (%)				
Affine	0.2134	-0.0661	0.0346	-0.3556	-0.0166
G-Affine	0.0211	0.1873	0.3723	-0.2216	0.0694
Model	Option price bias				
Affine	0.4357	-0.3601	-0.4654	-1.2721	-0.3124
G-Affine	0.1809	0.0281	0.0990	-0.9638	-0.1342

Table I.3: **Implied volatilities, Option prices RMSEs and bias by Maturity**
We estimate the models on a total of 16, 506 contracts with an average call price of 46.05 and average implied volatility of 20.26. The estimation have been done by minimizing the Black-Scholes IVRMSE

Maturity					
	DTM<30	30<DTM<90	90<DTM<180	180<DTM	All
Model	IVRMSE (%)				
Affine	5.4750	4.3179	3.9963	3.8295	4.3768
G-Affine	3.9103	3.1432	2.8824	2.9301	3.1915
Model	IV bias (%)				
Affine	0.5038	-0.2423	-0.0693	0.1737	-0.0166
G-Affine	0.8447	0.1338	-0.2344	-0.4328	0.0694
Model	Option price bias				
Affine	0.3352	-0.5526	-0.5843	0.0199	-0.3124
G-Affine	0.8570	0.3845	-0.5387	-1.8517	-0.1342

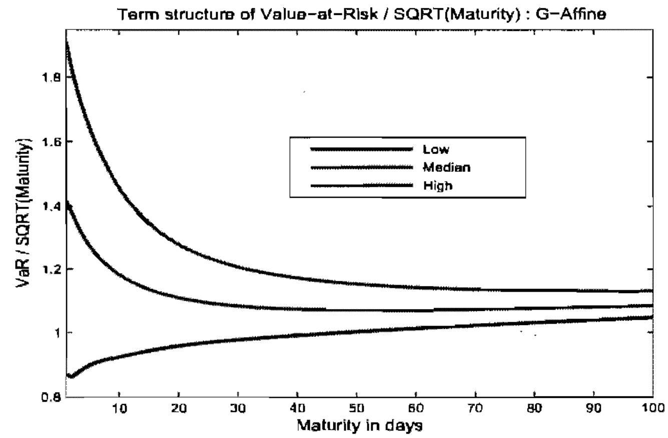


Figure I.1: **GARNIG term structure of VaR**

We use parameters estimated from the MLE to compute the term structure of value at risk. Several cases have been considered depending on the day where the term structure is evaluated. The cases are High volatility day (day with higher realized variance), Median volatility day and Low volatility day

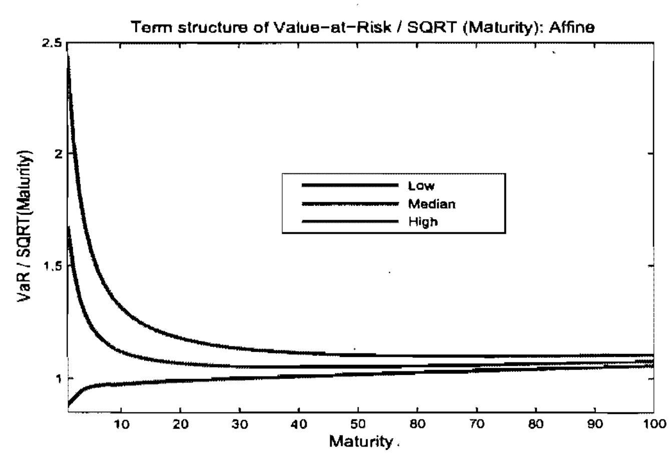


Figure I.2: **ARNIG term structure of VaR**

We use parameters estimated from the MLE to compute the term structure of value at risk. Several cases have been considered depending on the day where the term structure is evaluated. The cases are High volatility day (day with higher realized variance), Median volatility day and Low volatility day

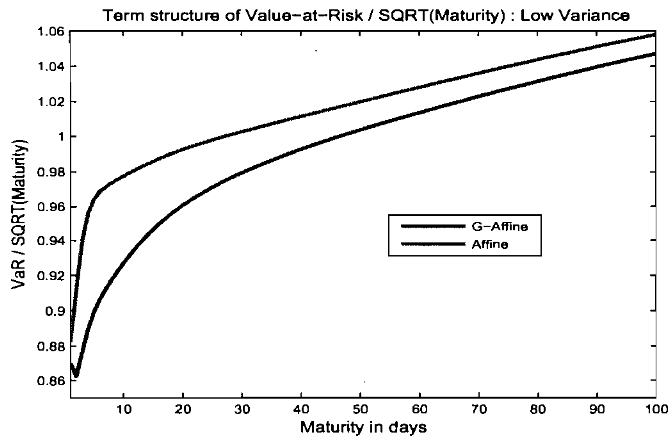


Figure I.3: **Term structure of VaR: low variance day**

We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure

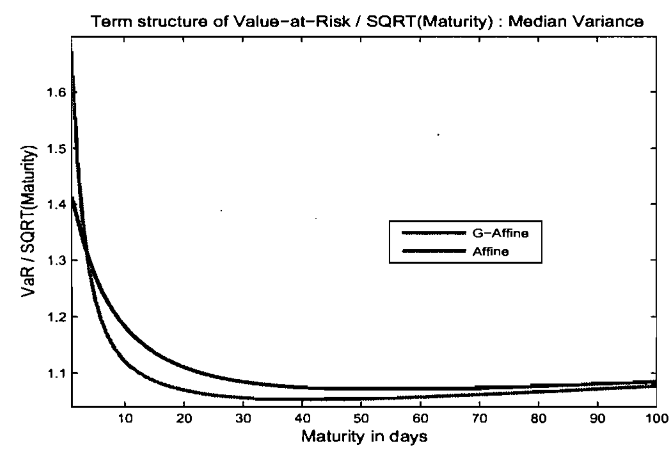


Figure I.4: **Term structure of VaR: median variance day**

We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure

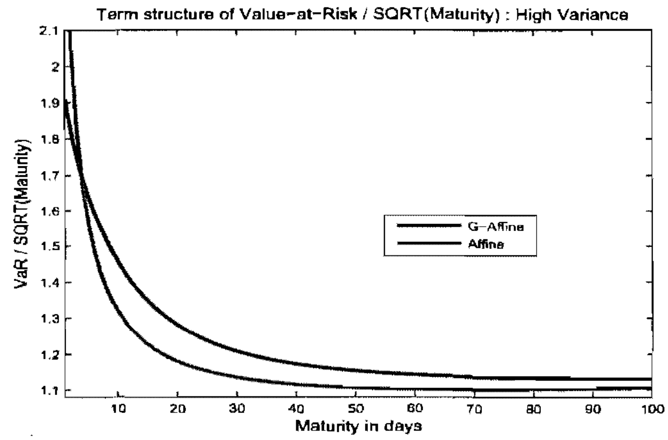


Figure I.5: **Term structure of VaR: high variance day**

We use parameters estimated from the MLE to compute the term structure of value at risk. We compared Affine and Generalized affine term structure.

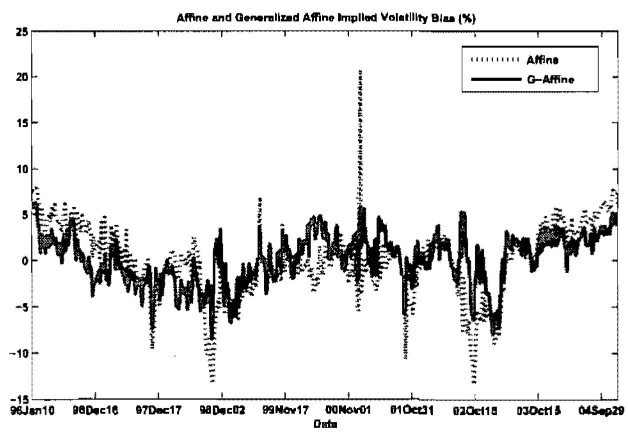


Figure I.6: **Implied Volatility Bias**

The figure displays implied volatility bias as a function of the day at which option is priced. Implied volatility bias is the difference between model and observed black scholes implied volatility. For each day we compute average available Implied volatility bias

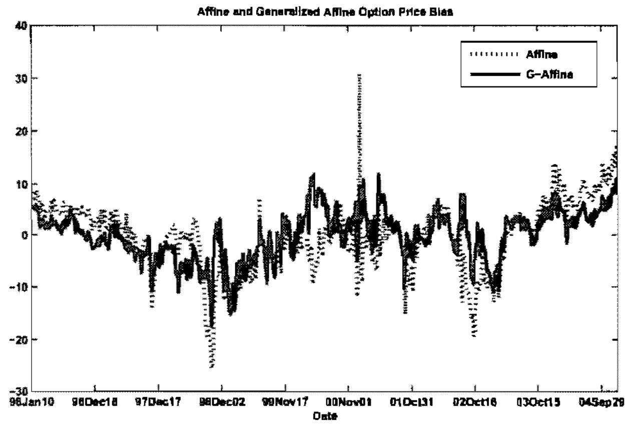


Figure I.7: **Option price Bias**

The figure displays Option price bias as a function of theb day at which option is priced. Option price bias is the difference between model and observed Option price. For each day we compute average available option bias

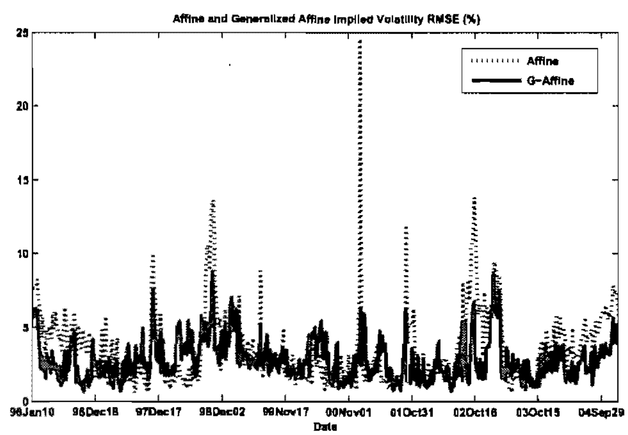


Figure I.8: **Implied Volatility Root mean squared error**

The figure displays IVRMSE as a function of day at which option is priced. IVRMSE is the square-root of the average squared difference between model and observed black scholes implied volatility.

Appendix II

Technical Appendix of Chapter 3: Option Bounds

We derive bounds on option prices from discrete time SV and GARCH models using the approach of Cochrane and Saa-Requejo (2000), generalized by Bondarenko and Longarela (2004). [1] These bounds are derived using a distance between a given stochastic discount factor (SDF) M_t and a benchmark SDF $\tilde{M}_t$, defined as

$$d_t\left(M_t, \tilde{M}_t\right) = \left(E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t} - 1\right)^2\right]\right)^{1/2} = \left(E_{t-1}\left(\frac{M_t}{\tilde{M}_t}\right)^2 - 1\right)^{1/2}.$$

Note that all expectations are taken under the benchmark risk-neutral measure, unless otherwise indicated.

When using a linear EMM and a GARCH stock price dynamic, the resulting price is unique. Therefore, to meaningfully analyze bounds we have to either generalize the assumption on the EMM, or work with a more complex stock price dynamic. We first investigate bounds on option prices using the discrete-time stochastic volatility model in Section 7.1, while maintaining a linear EMM. Subsequently, we analyze bounds for a GARCH model, under the assumption of a quadratic EMM. For convenience, we rewrite the implications of our choice of EMM in terms of the SDF M_t, using the fact that $M_t = \frac{\frac{dQ_t}{dP_t}}{\frac{dQ_{t-1}}{dP_{t-1}}}$.

Bounds on Option Prices in Discrete-Time Stochastic Volatility Models with a Linear EMM

The EMM used in the case of the discrete-time SV model is

$$\frac{dQ}{dP}|F_t = \exp\left(-\sum_{i=1}^{t}\left(\nu_{1,i}z_{1,i} + \nu_{2,i}z_{2,i} + \Psi_i(\nu_{1,i}, \nu_{2,i})\right)\right).$$

1. See Bernardo and Ledoit (2000) for a related approach.

Under the assumption of normal innovations with correlation ρ we have

$$\Psi_t(u_1,u_2)=\frac{1}{2}\left[(u_1+\rho u_2)^2+\left(1-\rho^2\right)u_2^2\right].$$

From this we can write the ratio of the SDFs as

$$\frac{M_t}{\tilde{M}_t}=\exp\left(\begin{array}{c}-(v_{1,t}-\tilde{v}_{1,t})z_{1,t}-(v_{2,t}-\tilde{v}_{2,t})z_{2,t}\\ -\frac{1}{2}\left(v_{1,t}^2-\tilde{v}_{1,t}^2\right)-\frac{1}{2}\left(v_{2,t}^2-\tilde{v}_{2,t}^2\right)-(v_{1,t}v_{2,t}-\tilde{v}_{1,t}\tilde{v}_{2,t})\rho\end{array}\right).$$

For pricing we need the distribution of the innovations under the benchmark risk-neutral measure

$$\left(\begin{array}{c}z_{1,t}\\ z_{2,t}\end{array}\right)\sim N\left(\left((-(\tilde{v}_{1,t}+\rho\tilde{v}_{2,t}),-(\rho\tilde{v}_{1,t}+\tilde{v}_{2,t}))',\left(\begin{array}{cc}1&\rho\\ \rho&1\end{array}\right)\right)\right).$$

This gives

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}\right)^2\right]=\exp\left(\begin{array}{c}2(v_{1,t}-\tilde{v}_{1,t})(\tilde{v}_{1,t}+\rho\tilde{v}_{2,t})+2(v_{2,t}-\tilde{v}_{2,t})(\rho\tilde{v}_{1,t}+\tilde{v}_{2,t})\\ +2(v_{1,t}-\tilde{v}_{1,t})^2+2(v_{2,t}-\tilde{v}_{2,t})^2+4\rho(v_{1,t}-\tilde{v}_{1,t})(v_{2,t}-\tilde{v}_{2,t})\\ -\left(v_{1,t}^2-\tilde{v}_{1,t}^2\right)-\left(v_{2,t}^2-\tilde{v}_{2,t}^2\right)-2(v_{1,t}v_{2,t}-\tilde{v}_{1,t}\tilde{v}_{2,t})\rho\end{array}\right).$$

$$\begin{aligned}E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right]&=\exp\left((v_{2,t}-\tilde{v}_{2,t})^2-(v_{1,t}-\tilde{v}_{1,t})^2\right)-1\\ &=\exp\left(\left(1-\rho^2\right)(v_{2,t}-\tilde{v}_{2,t})^2\right)-1.\end{aligned}$$

We now investigate the following bound

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right]\leq A_t^2,$$

where A_t is known at time $t-1$ but pertains to the SDF for time t. Bondarenko and Longarela (2004) show that this type of bound is equivalent to a bound on a generalized version of the Sharpe ratio.

We first establish the following proposition.

Proposition II.0.1.

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right] \leq A_t^2 \Longleftrightarrow \left|\nu_{2,t}-\tilde{\nu}_{2,t}\right| \leq \sqrt{\frac{\ln\left(1+A_t^2\right)}{1-\rho^2}}. \tag{II.1}$$

Proof.

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right] \leq A_t^2 \Longleftrightarrow \exp\left(\left(1-\rho^2\right)\left(\nu_{2,t}-\tilde{\nu}_{2,t}\right)^2\right)-1 \leq A_t^2.$$

$$\begin{aligned}
&\Longleftrightarrow \quad \exp\left(\left(1-\rho^2\right)\left(\nu_{2,t}-\tilde{\nu}_{2,t}\right)^2\right) \leq 1+A_t^2 \\
&\Longleftrightarrow \quad \left(1-\rho^2\right)\left(\nu_{2,t}-\tilde{\nu}_{2,t}\right)^2 \leq \ln\left(1+A_t^2\right) \\
&\Longleftrightarrow \quad \left(\nu_{2,t}-\tilde{\nu}_{2,t}\right)^2 \leq \frac{\ln\left(1+A_t^2\right)}{1-\rho^2} \\
&\Longleftrightarrow \quad \left|\nu_{2,t}-\tilde{\nu}_{2,t}\right| \leq \sqrt{\frac{\ln\left(1+A_t^2\right)}{1-\rho^2}}.
\end{aligned}$$

□

This proposition states that a bound relative to a benchmark SDF $\tilde{M}_t$ can be re-written in terms of a bound relative to the corresponding benchmark price of volatility risk $\tilde{\nu}_{2,t}$. We now use this result to find the maximum option price ($\overline{C}(\overline{\nu_{2,t}})$) and the minimum option price ($\underline{C}(\underline{\nu_{2,t}})$ corresponding to a given radius around the benchmark price of risk $\tilde{\nu}_{2,t}$.

Corollary 2. *The maximum and the minimum call price* $\overline{C}(\overline{\nu_{2,t}})$ *and* $\underline{C}(\underline{\nu_{2,t}})$ *in the following radius of the benchmark price of risk* $\tilde{\nu}_2$

$$\left|\nu_{2,t}-\tilde{\nu}_{2,t}\right| \leq \sqrt{\frac{\ln\left(1+A_t^2\right)}{1-\rho^2}}, \tag{II.2}$$

are given by

$$\overline{\nu_{2,t}} = \tilde{\nu}_{2,t} - \sqrt{\frac{\ln\left(1+A_t^2\right)}{1-\rho^2}} \qquad \underline{\nu_{2,t}} = \tilde{\nu}_{2,t} + \sqrt{\frac{\ln\left(1+A_t^2\right)}{1-\rho^2}}.$$

We now consider the special case of an affine stochastic volatility model, which implies $\nu_{2,t} = \nu_2 \sigma_t$. This gives

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t} - 1\right)^2\right] \leq A_t^2 \Longleftrightarrow |\nu_2 - \tilde{\nu}_2| \leq \sqrt{\frac{\ln\left(1+A_t^2\right)}{(1-\rho^2)\,\sigma_t^2}}.$$

If we also require the radius around $\tilde{\nu}_2$ to be constant, we get

$$A_t^2 = \exp\left(A^2\sigma_t^2\right) - 1,$$

and thus

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t} - 1\right)^2\right] \leq \exp\left(A^2\sigma_t^2\right) - 1 \Longleftrightarrow |\nu_2 - \tilde{\nu}_2| \leq \frac{A}{\sqrt{1-\rho^2}}.$$

Under these conditions we therefore have the following corollary

Corollary 3. *The maximum option price ($\overline{C}(\overline{\nu_2})$) and the minimum option price ($\underline{C}(\underline{\nu_2})$) in the radius of a benchmark SDF given by* $E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t} - 1\right)^2\right] \leq \exp\left(A^2\sigma_t^2\right) - 1$ *are obtained with*

$$\overline{\nu_2} = \tilde{\nu}_2 - \frac{A}{\sqrt{1-\rho^2}} \qquad \underline{\nu_2} = \tilde{\nu}_2 + \frac{A}{\sqrt{1-\rho^2}}.$$

Given an economically relevant SDF $\tilde{M}_t$, it is therefore straightforward to find the option bounds, because we can simply plug the upper bound and the lower bound on the price of risk into the pricing formula.

Bounds on GARCH Option Prices with a Quadratic EMM

In the case of the stochastic volatility models, we obtain some very elegant results for the bounds. The bound on the pricing kernel can be expressed as a bound on the price of volatility risk, which facilitates the computation of the bounds, and which provides more intuition for the results. We now investigate bounds for another incomplete-market setup, with the stock price process given by a GARCH model, and a quadratic EMM rather than a linear EMM. This case is of substantial interest, because quadratic EMMs have thus far not been considered in the literature. However, the quadratic EMM greatly complicates the derivation of the bounds.

We assume normally distributed returns

$$R_t|F_{t-1} \sim N\left(\mu_t - \tfrac{1}{2}\sigma_t^2, \sigma_t^2\right).$$

For a given sequence $\left\{\nu_{1,t}, \nu_{2,t}\right\}$, consider the following EMM from Section 3.3

$$\frac{dQ}{dP}\bigg|F_t = \exp\left(-\sum_{i=1}^{t}\left(\nu_{1,i}\varepsilon_i + \nu_{2,i}\varepsilon_i^2 + g_i\right)\right), \tag{II.3}$$

with

$$\begin{aligned} g_i &= \frac{1}{2}\left(\nu_{1,i}^2\sigma_i^{*2} - \ln\left(\sigma_i^2/\sigma_i^{*2}\right)\right), \text{ where} && \text{(II.4)} \\ \sigma_i^{*2} &= Var_{i-1}^{Q}(\varepsilon_i) = \frac{\sigma_i^2}{1+2\nu_{2,i}\sigma_i^2}, \text{ and} && \text{(II.5)} \\ \nu_{1,i} &= \left[\frac{\mu_i}{\sigma_i^2} - \frac{r_i}{\sigma_i^{*2}}\right] + 2\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\nu_{2,i}. && \text{(II.6)} \end{aligned}$$

The corresponding SDF is

$$M_t = \exp\left(-\nu_{1,t}\varepsilon_t - \nu_{2,t}\varepsilon_t^2 - g_t\right).$$

We now consider the following benchmark SDF

$$\tilde{M}_t = \exp\left(-\tilde{\nu}_{1,t}\varepsilon_t - \tilde{\nu}_{2,t}\varepsilon_t^2 - \tilde{g}_t\right),$$

where similarly

$$\tilde{g}_t = \frac{1}{2}\left(\tilde{\nu}_{1,t}^2\tilde{\sigma}_t^{*2} - \ln\left(\sigma_t^2/\tilde{\sigma}_t^{*2}\right)\right), \text{ and} \tag{II.7}$$

$$\tilde{\sigma}_t^{*2} = \frac{\sigma_i^2}{1+2\tilde{\nu}_{2,t}\sigma_t^2}, \text{ and } \tilde{\nu}_{1,t} = \left[\frac{\mu_t}{\sigma_t^2} - \frac{r_t}{\tilde{\sigma}_t^{*2}}\right] + 2\left(\mu_t - \frac{1}{2}\sigma_t^2\right)\tilde{\nu}_{2,t}. \tag{II.8}$$

This gives

$$\begin{aligned}
\frac{M_t}{\tilde{M}_t} &= \exp\left(\left(\tilde{\nu}_{1,t} - \nu_{1,t}\right)\varepsilon_t + \left(\tilde{\nu}_{2,t} - \nu_{2,t}\right)\varepsilon_t^2 + \left(\tilde{g}_t - g_t\right)\right) \\
\left(\frac{M_t}{\tilde{M}_t}\right)^2 &= \exp\left(2\left(\tilde{\nu}_{1,t} - \nu_{1,t}\right)\varepsilon_t + 2\left(\tilde{\nu}_{2,t} - \nu_{2,t}\right)\varepsilon_t^2 + 2\left(\tilde{g}_t - g_t\right)\right).
\end{aligned}$$

We now need to do some tedious computations to compute conditional expectations, where we use the following standard result. Given $\varepsilon \sim N\left(\mu, \sigma^2\right)$, we have

$$E\exp\left(a\varepsilon + b\varepsilon^2 + c\right) = \exp\left(c - \frac{1}{2}\ln\left(1-2b\sigma^2\right) - \frac{a^2}{4b} + \frac{b\left(\mu + \frac{a}{2b}\right)^2}{1-2b\sigma^2}\right). \tag{II.9}$$

For the expectation $E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}\right)^2\right]$ that we have to solve, we get

$$\begin{aligned}
a &= 2\left(\tilde{\nu}_{1,t} - \nu_{1,t}\right) = 4\left(\mu_t - r_t - \frac{\sigma_t^2}{2}\right)\left(\tilde{\nu}_{2,t} - \nu_{2,t}\right) \\
b &= 2\left(\tilde{\nu}_{2,t} - \nu_{2,t}\right) \\
c &= 2\left(\tilde{g}_t - g_t\right) = \left(\tilde{\nu}_{1,t}^2\tilde{\sigma}_t^{*2} - \nu_{1,t}^2\sigma_t^{*2}\right) + \ln\left(\frac{\tilde{\sigma}_t^{*2}}{\sigma_t^{*2}}\right).
\end{aligned}$$

Under the risk-neutral benchmark measure we have $\varepsilon_t|F_{t-1} \sim N\left(r_t - \mu_t - \frac{\tilde{\sigma}_t^{*2}-\sigma_t^2}{2}, \tilde{\sigma}_t^{*2}\right)$. We can now use $(II.9)$ and the relationship between $\nu_{1,t}$ and $\nu_{2,t}$ to derive

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}\right)^2\right] = \frac{\tilde{\sigma}_t^{*2}}{\sqrt{\sigma_t^{*2}\left(2\tilde{\sigma}_t^{*2} - \sigma_t^{*2}\right)}}\exp\left(\frac{\left(\tilde{\sigma}_t^{*2} - \sigma_t^{*2}\right)^2}{4\left(2\tilde{\sigma}_t^{*2} - \sigma_t^{*2}\right)}\right).$$

This result can then be used to derive the bounds

$$\begin{aligned} E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right] &\leq A_t^2 \Longleftrightarrow E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}\right)^2\right] \leq 1+A_t^2 \\ E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right] &\leq A_t^2 \Longleftrightarrow E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}\right)^2\right]^2 \leq \left(1+A_t^2\right)^2 \\ &\Longleftrightarrow \frac{\tilde{\sigma}_t^{*4}}{\sigma_t^{*2}\left(2\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)}\exp\left(\frac{\left(\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)^2}{2\left(2\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)}\right) \leq \left(1+A_t^2\right)^2. \end{aligned}$$

We have therefore established the following proposition

Proposition II.0.2.

$$E_{t-1}\left[\left(\frac{M_t}{\tilde{M}_t}-1\right)^2\right] \leq A_t^2 \Longleftrightarrow \frac{\tilde{\sigma}_t^{*4}}{\sigma_t^{*2}\left(2\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)}\exp\left(\frac{\left(\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)^2}{2\left(2\tilde{\sigma}_t^{*2}-\sigma_t^{*2}\right)}\right) \leq \left(1+A_t^2\right)^2. \tag{II.10}$$

This result is similar to the one obtained for the stochastic volatility model with a linear pricing kernel in (II.1). We can establish pricing bounds by using a lower bound $\underline{\sigma}_t^{*2}$ and an upper bound $\bar{\sigma}_t^{*2}$in the pricing formula, as in the case of (II.2) in the stochastic volatility model. The difference is that it is not possible to further simplify the expression in (II.10), as was done for the stochastic volatility model with linear pricing kernel in (II.1). The lower bound $\underline{\sigma}_t^{*2}$ and an upper bound $\bar{\sigma}_t^{*2}$ must be obtained numerically from (II.10).

Appendix III

Empirical illustration using GARCH-SVG model: Chapter 3

In this section we demonstrate how the greater flexibility and generality allowed for by our approach can lead to more realistic option valuation models. To do so, we develop a GARCH-SVG model, which allows for conditional skewness and kurtosis, and which has not yet been analyzed in the literature. We compute option prices using parameters estimated from return data only, subsequently construct option implied volatility smiles, and compare them with option data.

Conditionally skewed variance gamma returns

We now introduce a new model where the conditional skewness, s, and excess kurtosis, k, are given directly by two parameters in the model. [1] Consider the return of the underlying asset specified as follows

$$R_t = \mu_t - \gamma_t + \varepsilon_t = \mu_t - \gamma_t + \sigma_t z_t, \qquad z_t \overset{i.i.d}{\sim} SVG(0,1,s,k).$$

The distribution of the shocks, $SVG(0,1,s,k)$, is a standardized skewed variance gamma distribution which is constructed as a mixture of two gamma variables. [2] The conditional variance, σ_t^2, can take on any GARCH specification.

Let $z_{1,t}$ and $z_{2,t}$ be independent draws from two gamma distributions

$$z_{i,t} \overset{i.i.d}{\sim} \Gamma\left(4/\tau_i^2\right), \qquad i = 1,2,$$

1. In Christoffersen, Heston and Jacobs (2006), conditional skewness and kurtosis are driven by functions of the same parameter.
2. See Madan and Seneta (1990) for an early application of the symmetric and i.i.d. variance gamma distribution in finance.

parameterized as

$$\tau_1 = \sqrt{2}\left(s - \sqrt{\frac{2}{3}k - s^2}\right) \text{ and } \tau_2 = \sqrt{2}\left(s + \sqrt{\frac{2}{3}k - s^2}\right).$$

If we construct the SVG random variable from the two gamma variables as

$$z_t = \frac{1}{2\sqrt{2}}\left(\tau_1 z_{1,t} + \tau_2 z_{2,t}\right) - \sqrt{2}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right),$$

then z_t will have a mean of zero, a variance of one, a skewness of s, and an excess kurtosis of k, thus allowing for conditional skewness and kurtosis in the GARCH model, as intended. [3]

The log moment generating function of ε_t can be derived from the gamma distribution MGF as

$$\Psi_t(u) = \sqrt{2}\left(\tau_1^{-1} + \tau_2^{-1}\right)u\sigma_t - 4\tau_1^{-2}\ln\left(1 + \frac{1}{2\sqrt{2}}\tau_1 u\sigma_t\right) - 4\tau_2^{-2}\ln\left(1 + \frac{1}{2\sqrt{2}}\tau_2 u\sigma_t\right),$$

so that the mean correction variable for the return can be found as $\gamma_t = \Psi_t(-1)$.

In general, there is no analytical solution to the EMM condition for this model. With regard to the sufficient conditions for existence and uniqueness of a solution to the EMM condition in Proposition 2, strict convexity is assured because the distribution is part of the infinitely divisible class (Feller (1968)), but it is not straightforward to verify the condition $u_{1,t} + 1 \leq u_{2,t}$. The boundaries for the SVG model are $u_{1,t} = -\frac{2\sqrt{2}}{\tau_2\sigma_t}$ and $u_{2,t} = -\frac{2\sqrt{2}}{\tau_1\sigma_t}$, and therefore we need to verify $\frac{2\sqrt{2}}{\tau_2} \geq \sigma_t + \frac{2\sqrt{2}}{\tau_1}$. Since by construction $\tau_1 \leq 0$, $\tau_2 \geq 0$, the right hand side may be negative for reasonable values of σ_t, but this is impossible to determine in general.

Using the formula for the risk neutral conditional log MGF

$$\Psi_t^{Q*}(u) = -u\Psi_t'(v_t) + \Psi_t(v_t + u) - \Psi_t(v_t),$$

3. The special cases where τ_1 or τ_2 are zero can be handled easily by drawing from the normal distribution for the relevant mixing variable $z_{1,t}$ or $z_{2,t}$. When both τ_1 and τ_2 are zero then the normal distribution obtains for z_t.

we can show that the risk neutral model is

$$R_t = r_t - \gamma_t^* + \varepsilon_t^*, \tag{III.1}$$

where

$$\Psi_t^{Q*}(u) = \sqrt{2}\left(\tau_1^{-1}\sigma_{1,t}^* + \tau_2^{-1}\sigma_{2,t}^*\right)u - 4\tau_1^{-2}\ln\left(1 + \frac{1}{2\sqrt{2}}\tau_1\sigma_{1,t}^* u\right) - 4\tau_2^{-2}\ln\left(1 + \frac{1}{2\sqrt{2}}\tau_2\sigma_{2,t}^* u\right),$$

with

$$\sigma_{i,t}^* = \frac{\sigma_t}{\sqrt{2} + \frac{1}{2}\tau_i\sigma_t v_t}, \qquad \text{for } i = 1,2. \tag{III.2}$$

We see that $\Psi_t^{Q*}(u)$ is exactly of the same form as $\Psi_t(u)$, and therefore that $\gamma_t^* = \Psi_t^{Q*}(-1)$.

Parameter estimates from index returns

As a benchmark, we use the conditional normal NGARCH model of Engle and Ng (1993)

$$R_t = \mu_t - \gamma_t + \sigma_t z_t, \qquad z_t \overset{i.i.d}{\sim} N(0,1). \tag{III.3}$$

where $\mu_t = r_t + \lambda\sigma_t$, $\gamma_t = \frac{1}{2}\sigma_t^2$, and

$$\sigma_t^2 = \beta_0 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-1}^2\left(z_{t-1} - \beta_3\right)^2.$$

Notice that the β_3 parameter in the GARCH variance specification allows for an asymmetric variance response to positive versus negative shocks, z_{t-1}. This captures the so-called leverage effect, which is another important empirical regularity in daily equity index returns.

Table 1 reports the maximum likelihood estimates of the GARCH parameters. We also report parameter estimates for a version of the model where the GARCH dynamics have been shut down, that is, where $\beta_1 = \beta_2 = \beta_3 = 0$. Notice the large increase in the log-likelihood function from including the GARCH dynamics.

For the implementation of the GARCH-SVG model, μ_t and σ_t^2 are the same as in the conditional normal model in (III.3). We can calibrate the s and k parameters in the GARCH-SVG model by simply equating them to the sample moments from the z_t sequence from the QMLE estimation of the GARCH model. These sample moments are reported in Table 1.

Implied Black-Scholes volatilities from model and data

Armed with the parameter estimates from daily returns, we can transform the physical GARCH-SVG process to the risk-neutral measure and then assess its option pricing implications. Figure 1 contains an illustration using option contracts sampled on three different days. Implied Black-Scholes volatilities from S&P500 call options are shown in circles, and the model-based implied volatilities are in solid lines. Moneyness, defined as strike over spot price is on the horizontal axis in all panels. The three columns of panels correspond to 1-month, 2-month, and 3-month options respectively. Each row of panels corresponds to different levels of spot volatility. The top row contains results for a low volatility day, September 13, 2000, which has an annualized GARCH spot volatility approximately 20% below the average volatility over the 1999-2001 period from which we sample the options. The middle row contains results for a medium volatility day, December 19, 2001, which has a spot volatility close to the average volatility over this period. The bottom row contains results for March 24, 1999, which has an annualized spot volatility about 20% higher than the average.

Figure 1 shows that the GARCH-SVG model fits the observed implied volatilities quite well. The model slightly underprices in-the-money calls at the one-month maturity when volatility is around average, and it slightly underprices options at the 3-month maturity when volatility is high. For the other contracts the model-based IVs are very close to the data.

We conclude that it is possible to build relatively simple models capturing the conditional volatility and non-normality found in index returns data, and that such models provide the flexibility needed to price options.

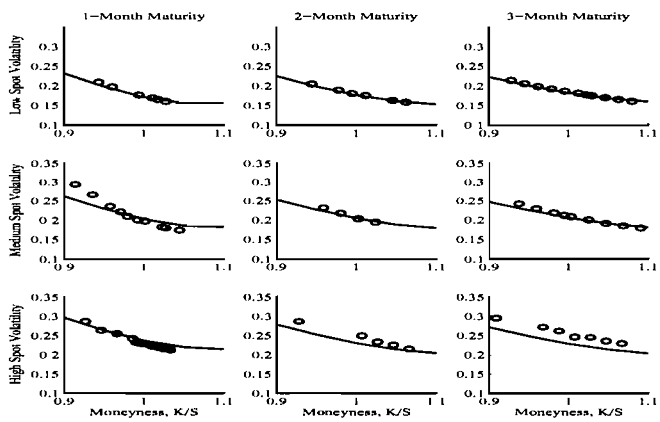
Table III.1: **Parameter Estimates and Model Properties**

Parameters	Independent Returns Estimate	Independent Returns Standard Error	GARCH Returns Estimate	GARCH Returns Standard Error
r	1.370E-04		1.370E-04	
λ	0.0313	0.0129	0.0312	0.0121
β_0	1.111E-04	9.9781E-06	1.516E-06	6.050E-07
β_1			0.8916	0.0274
β_2			0.0617	0.0154
β_3			0.7422	0.0808

Properties	**Independent Returns**	**GARCH Returns**
Log-Likelihood	20,615.00	21,586.28
Volatility Persistence	0	0.9873
Annual Volatility	0.1673	0.1734
Conditional Skewness	-1.2105	-0.4127
Conditional Kurtosis	27.3304	3.4935

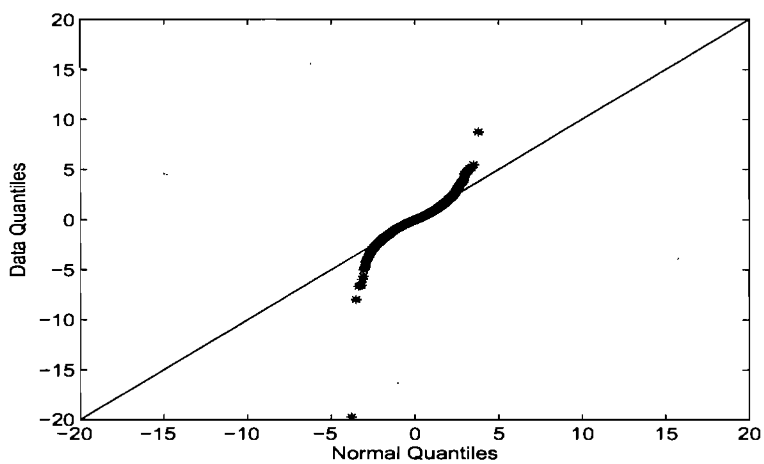
We use quasi maximum likelihood to estimate an independent return and a GARCH return model on daily S&P500 returns from January 2, 1980 to December 30, 2005 for a total of 6,564 observations. We report various properties of the two models including conditional skewness and excess kurtosis which are later used as parameter estimates in the SVG models.

Figure III.1: **Implied Volatility Smirks. Model and S&P500 Index Option Data.**



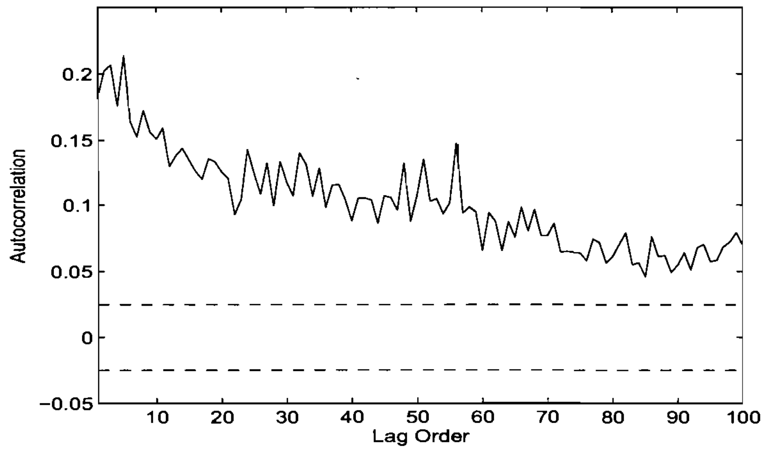
We choose three days from the 1999-2001 period to illustrate the performance of the SVG-GARCH option pricing model. Moneyness, defined as strike over spot price, is on the horizontal axis in all panels. The three columns of panels corresponds to 1-month, 2-month, and 3-month options respectively. The three rows of panels correspond days with different levels of spot volatility. The circles indicate implied Black-Scholes volatilities from S&P500 call options, and the solid lines indicate model-based implied volatilities.

Figure III.2: **Quantile-Quantile Plot of S&P500 Returns Against the Normal Distribution**



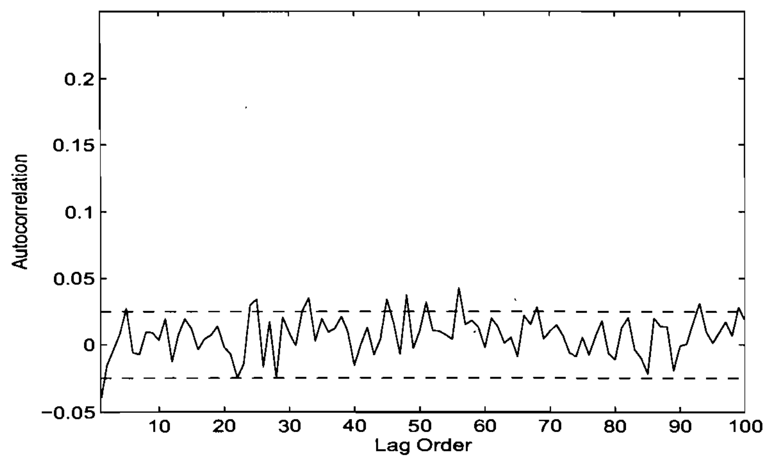
We take daily returns on the S&P500 from January 2, 1980 to December 30, 2005 and standardize them by the sample mean and sample standard deviation. The quantiles of the standardized returns are plotted against the quantiles from the standard normal distribution.

Figure III.3: **Autocorrelation Function of Absolute S&P500 Returns**



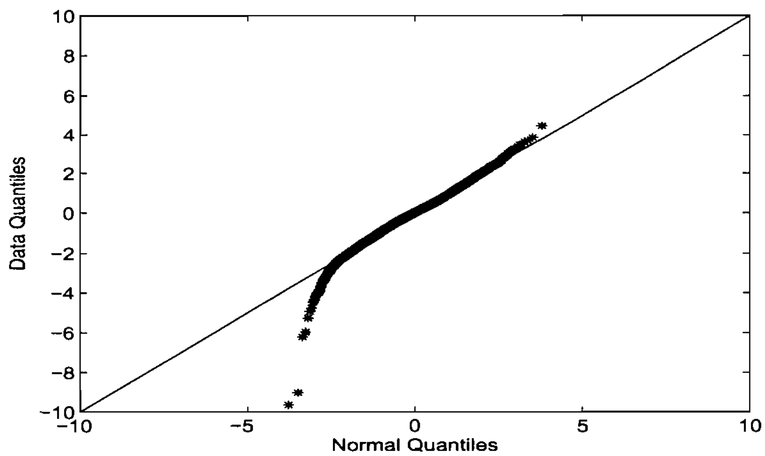
From daily absolute returns on the S&P500 from January 2, 1980 to December 30, 2005 we compute and plot the sample autocorrelations for lags one through 100 days. The horizontal dashed lines denote 95% Bartlett confidence intervals around zero.

Figure III.4: **Autocorrelation Function of Absolute GARCH Innovations**



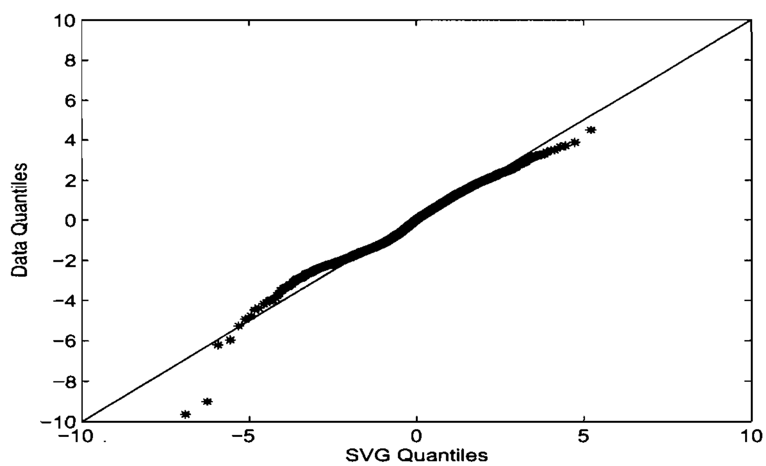
From the estimated GARCH model in Table 1 we construct the absolute standardized sequence of shocks and plot the sample autocorrelations for lags one through 100 days. The horizontal dashed lines denote 95% Bartlett confidence intervals around zero.

Figure III.5: **Quantile-Quantile Plots of GARCH Innovations Against the Normal Distribution**



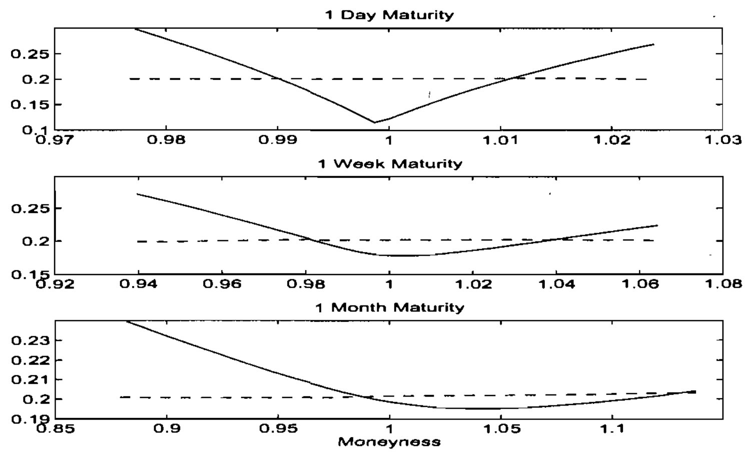
From the estimated GARCH models in Table 1 we compute the time series of dynamically standardized S&P500 returns. The quantiles of these GARCH innovations are plotted against the quantiles from the standard normal distribution.

Figure III.6: **Quantile-Quantile Plots of GARCH Innovations Against the SVG Distribution**



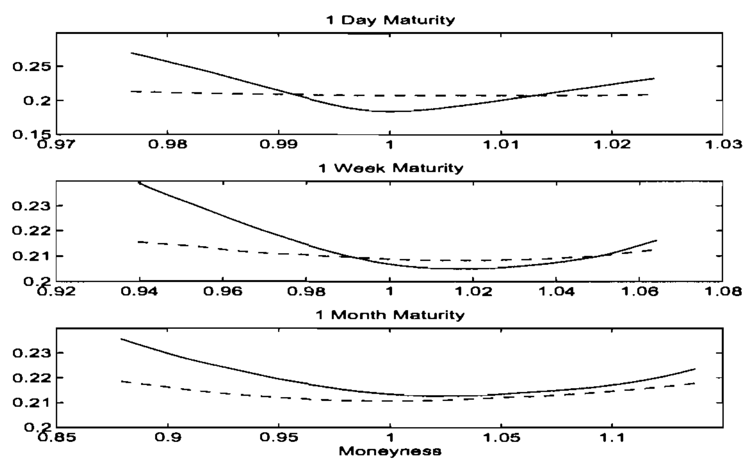
From the estimated GARCH models in Table 1 we compute the time series of dynamically standardized S&P500 returns. The quantiles of these GARCH innovations are plotted against the quantiles from the skewed variance gamma (SVG) distribution.

Figure III.7: **Implied Volatility Functions for Normal and SVG Independent Return Models**



From the estimated independent return model in Table 1 we compute call option prices for various moneyness and maturities and we then compute implied Black-Scholes volatilities from the model option prices. Implied volatility is plotted against moneyness on the horizontal axis. The three panels correspond to maturities of 1 day, 1 week, and 1 month respectively. The solid lines show the i.i.d SVG model and the dashed lines the i.i.d. Normal models.

Figure III.8: **Implied Volatility Functions for Normal and SVG GARCH Models**



From the estimated GARCH model in Table 1 we compute call option prices for various moneyness and maturities and then we compute implied Black-Scholes volatilities from the model option prices. The implied volatilities are plotted with moneyness on the horizontal axis. The three panels correspond to maturities of 1 day, 1 week, and 1 month respectively. The solid lines show the SVG GARCH model and the dashed lines the Normal GARCH model.