Non-Negativity, Zero Lower Bound and Affine Interest Rate Models

Guillaume Roussellet

▶ **To cite this version:**

Guillaume Roussellet. Non-Negativity, Zero Lower Bound and Affine Interest Rate Models. General Mathematics [math.GM]. Université Paris Dauphine - Paris IX, 2015. English. NNT : 2015PA090012 . tel-01171952

HAL Id: tel-01171952
https://theses.hal.science/tel-01171952

Submitted on 6 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Dauphine
Centre de Recherche en Mathématiques de la Décision UMR CNRS 7534

Non-Negativity, Zero Lower Bound and Affine Interest Rate Models

Positivité, séjours en zéro et modèles affines de taux d'intérêt

Thèse présentée et soutenue publiquement le

15 juin 2015

en vue de l'obtention du

Doctorat en Sciences ; discipline : Mathématiques Appliquées

par

Guillaume Roussellet

Directeur de Recherche :	**Monsieur Alain Monfort** Professeur CNAM ENSAE-CREST
Rapporteurs :	**Monsieur Eric Renault** Professeur des Universités *Brown University*
	Monsieur Olivier Scaillet Professeur des Universités Université de Genève et *Swiss Finance Institute*
Suffragants :	**Monsieur Christian Gouriéroux** Professeur des Universités Université Paris-Dauphine, CREST, *University of Toronto*
	Monsieur Nour Meddahi Professeur des Universités *Toulouse School of Economics*
	Monsieur Serge Darolles Professeur Associé Université Paris-Dauphine, CREST

L'université Paris-Dauphine n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leur auteur.

À Catherine, à Gisèle.

Remerciements

Je tiens tout d'abord à adresser mes plus chaleureux remerciements à Alain Monfort, mon directeur de thèse. Grâce à ses constants encouragements et à son écoute tout au long de ces trois années, Alain m'a permis d'acquérir les connaissances, la rigueur et la passion nécessaires au travail de recherche. Travailler sous sa direction a été pour moi une opportunité formidable.

Je remercie très sincèrement Olivier Scaillet et Éric Renault d'avoir accepté d'être les rapporteurs de ma thèse, et de m'avoir soutenu ces dernières années. Je suis reconnaissant envers Christian Gouriéroux, Nour Meddahi et Serge Darolles d'avoir accepté de faire partie de mon jury.

Dès mon arrivée à la Banque de France, j'ai bénéficié d'un encadrement scientifique "au jour le jour" par l'investissement permanent de mes collègues et co-auteurs Jean-Paul Renne, Fulvio Pegoraro et Simon Dubecq. Leur dévouement, leur patience et de nombreuses conversations de couloir m'ont permis d'apprendre constamment de nouvelles choses. Merci pour ce soutien sans failles.

La Banque de France a constitué un environnement privilégié pour mon apprentissage. Je remercie particulièrement Benoît Mojon, Jean-Stéphane Mésonnier et Caroline Jardet de m'avoir fourni une telle opportunité. J'ai aussi grandement bénéficié d'échanges avec les chercheurs de Recfin et Pomone, notamment Eric Mengus, Jérome Dugast et Sarah Mouabbi, avec qui j'ai partagé mon bureau.

Le temps de recherche passé au Crest a aussi beaucoup apporté à mon travail. Je remercie Jean-Michel Zakoian, Christian Francq, Serge Darolles et Gaëlle Le Fol pour leurs encouragements. Bien que ne faisant pas partie de l'auto-proclamé "Burex", je remercie Jérémy Dudek, Erwan Koch, et Jean-Cyprien Héam pour avoir contribué à mon intégration immédiate au sein du laboratoire.

Pendant ces trois ans, mes amis sont probablement ceux qui ont le plus observé mon

humeur au jour le jour au cours de cette aventure. Pour leurs questions, leur intérêt, leur soutien et leur capacité à me faire changer d'air, je remercie Aurélie, Claire et Marko, collègues mais avant tout amis ; Rémy, Juju (aka *crassou*), Clouze, Anissa, Clémence, Gab' et Tamtam pour les retrouvailles au Mad ; Romain, Jeanne, Julie, Violaine et Tim pour leur présence permanente ; et tous les membres de l'orchestre *La Sirène* pour le grand bol d'air dominical hebdomadaire. Je remercie aussi tous ceux qui ont contribué à la relecture, notamment Max, Maud, JB, Tim, Anissa, et Jeanne.

Enfin, ma famille a constitué un aide et une écoute précieuse, et m'a toujours encouragé dans ce processus. Je remercie particulièrement mon père Jean-François et mon frère Valentin pour leur intérêt, et mes grands-parents Lucien, Robert et Micheline. Merci aussi à Elisabeth et Olivier pour leur soutien sans faille.

Contents

List of Tables

List of Figures

Résumé

Les modèles de taux d'intérêt à absence d'opportunité arbitrage ont pour but de reproduire le comportement dans le temps de prix d'obligations à différentes maturités. Parmi ceux-ci, les modèles affines de structure par terme (ATSM) forment la classe de modèles la plus utilisée depuis leur introduction par Duffie and Kan [1996]. Ils permettent en effet de représenter les taux d'intérêt des obligations zéro-coupon à toutes les maturités par une fonction affine d'un nombre réduit de facteurs, disponible en formule fermée. Le modèle affine le plus couramment utilisé est incontestablement le modèle gaussien, dans lequel les rendements des obligations zéro-coupon sont des combinaisons affines de processus VAR gaussiens standards. En conséquence, les taux d'intérêt à toutes maturités sont aussi gaussiens et ont pour support la droite des réels. Cette propriété est en contradiction avec de nombreuses caractéristiques théoriques des taux d'intérêt obligataires. En particulier, il est toujours possible de substituer de la monnaie à un investissement obligataire, conservant ainsi un taux d'intérêt nominal nul. En appliquant l'hypothèse de non-arbitrage, on obtient que les taux d'intérêt des obligations zéro-coupon sont théoriquement toujours positifs ou nuls. Toutefois, en pratique, on peut observer des taux d'intérêt négatifs causés par des frictions financières (par exemple si la possession de liquidités a un coût positif). Les taux d'intérêt peuvent donc posséder une borne inférieure négative, que l'hypothèse gaussienne ne peut pas non plus prendre en compte. Les ATSM positifs constituent une réponse ciblée à ce problème, en assurant par construction la positivité de l'ensemble de la courbe des taux. On considère dans cette thèse plusieurs extensions des modèles affines positifs de structure par terme des taux d'intérêt. Après une introduction générale, une première partie, composée des chapitres 2 et 3, se consacre à l'étude du modèle de structure par terme quadratique (QTSM). Par la suite, les chapitres 4 et 5 forment une seconde partie dédiée à la modélisation des taux d'intérêt pour reproduire le comportement de "taux au plancher".

Le chapitre introductif détaille les principales hypothèses utilisées dans la construction d'un modèle affine de taux d'intérêt. On suppose notamment que le vecteur de facteurs endogènes est un processus affine - ou composé auto-régressif (Car) - sous la mesure risque-neutre, c'est-à-dire que sa transformée de Laplace conditionnée a son passé est une

fonction exponentielle-affine de ses valeurs passées. Lorsque l'on spécifie le taux d'intérêt à court terme (et l'intensité de défaut dans le cas des obligations risquées) comme une combinaison affine de ces facteurs, on obtient tous les taux d'intérêt des obligations zéro-coupon en formule fermée. Cette propriété est particulièrement utile, notamment lors de l'estimation du modèle. La spécification du facteur d'escompte stochastique permet de dégager la dynamique du vecteur de facteurs sous la mesure historique, et de formuler le modèle sous une forme espace-état.

Comme démontré par Cheng and Scaillet [2007], les QTSM font partie de la classe des modèles affines de taux d'intérêt. Représentés sous une forme espace-état, ces modèles rassemblent des équations de transition affines (VAR gaussien standard) et des équations de mesure linéaire-quadratiques. Lorsque le vecteur de facteurs est partiellement ou totalement inobservable, le filtre de Kalman standard ne peut plus être utilisé à cause de la non-linéarité des équations de mesure. Afin de résoudre ce problème, plusieurs filtres non-linéaires ont été développés dans un cadre général. Ces filtres peuvent être divisés en deux classes distinctes. D'une part, certains filtres utilisent des simulations pour approximer les distributions non-gaussiennes issues des non-linéarités, et obtenir un filtrage plus précis. Cependant, les simulations employées ont pour conséquence une complexité computationnelle fortement accrue. D'autre part, les filtres dits *déterministes* s'appuient sur des développements de Taylor des équations non-linéaires. Néanmoins, la plupart de ces algorithmes (notamment les filtres de Kalman étendus et le filtre de Kalman inodore[1]) sont exprimés dans le cadre le plus général possible et ne sont pas spécifiquement adaptés au cas linéaire-quadratique. Le chapitre 2 propose de résoudre ces inadéquations grâce à un nouvel algorithme de filtrage appelé *filtre de Kalman quadratique* (QKF). Comme ce nouvel outil appartient à la classe des filtres déterministes, il est à la fois rapide et facile à implémenter. La formalisation de l'algorithme dépend principalement des propriétés des processus affines, notamment du résultat suivant : si le processus du vecteur de facteurs est un VAR gaussien standard, alors le processus empilé composé du même vecteur et de tous les produits possibles de ses composantes est affine. En formulant la dynamique de ce nouveau processus empilé sous une forme VAR à chocs hétéroscédastiques non-gaussiens, on obtient un modèle espace-état augmenté affine dans lequel le filtre de Kalman linéaire peut être utilisé. De la même manière, l'algorithme de lissage peut être obtenu immédiatement dans le modèle espace-état augmenté. Afin d'analyser la performance du QKF en termes de filtrage et d'estimation par quasi maximum de vraisemblance, on réalise des simulations Monte Carlo et montre que notre algorithme surpasse ses concurrents dans les deux exercices.

1. Traduction littérale de *Unscented Kalman filter*.

Dans le chapitre 3, on utilise cette nouvelle technique d'estimation et de filtrage pour un QTSM des écarts de taux interbancaires en zone Euro. Ces derniers correspondent à l'écart entre les taux interbancaires risqués Euribor et les taux *swaps overnight* (OIS), considérés comme sans risques. Deux risques majeurs influencent les mouvements de ces écarts de taux : risque de défaut et risque de liquidité, tous deux contenus dans les taux Euribor. En conséquence, ces taux d'intérêt risqués incorporent une compensation pour le risque positive et les écarts de taux Euribor-OIS sont toujours positifs. Contrairement aux différentes modélisations utilisées dans la littérature, le QTSM permet d'imposer cette restriction théorique sur les écarts de taux. On construit ainsi un modèle à intensités de défaut et de liquidité, où celles-ci sont des fonctions quadratiques de deux facteurs gaussiens. On peut identifier chacun des facteurs respectivement comme risque de défaut et de liquidité grâce à des variables observables que nous savons liées à ces deux types de risques. Nos estimations permettent de décomposer les écarts de taux à chaque maturité en leurs différentes composantes afin d'analyser l'impact des politiques monétaires non conventionnelles menées par la BCE. Nos résultats indiquent que malgré l'explosion du risque de liquidité sur le marché interbancaire européen après la faillite de *Lehman Brothers*, les *very long-term refinancing operations* et les *outright monetary transactions* ont contribué à sa diminution récente. A l'aide de notre modèle, on peut aussi obtenir les probabilités de défaut moyennes des banques de la zone Euro, et l'on montre leur diminution récente à la suite des politiques non conventionnelles citées précédemment.

Le chapitre 4 aborde un problème plus général lié à la construction des modèles multivariés affines de taux d'intérêt. Une méthode simple consiste à considérer un ensemble de processus univariés affines et indépendants entre eux, et de les empiler dans un même vecteur afin de former un processus affine multivarié. L'hypothèse forte d'indépendance peut la plupart du temps être remplacée par une hypothèse d'indépendance des différentes composantes, conditionnellement au passé du processus, sans changer le caractère affine du processus multivarié. En revanche, introduire de la dépendance conditionnelle entre les composantes s'avère en général plus ardu. Dans ce chapitre, on propose une nouvelle technique pour construire des processus affines multivariés sans s'appuyer sur l'hypothèse d'indépendance conditionnelle. Ces processus, qu'on appelle *récursifs affines*, peuvent être définis à l'aide de leur transformée de Laplace conditionnelle. Supposons le vecteur de facteurs divisé en plusieurs blocs. Chaque bloc est un processus récursif affine si sa transformée de Laplace conditionnelle au passé du processus et aux blocs situés avant lui dans le vecteur est une fonction exponentielle-affine des variables conditionnantes. Cette méthode est suffisamment générale pour être appliquée dans de nombreux exemples : processus VAR gaussiens à volatilité stochastique, processus multivariés à valeurs discrètes, ou encore processus vectoriel auto-régressif gamma. On montre ensuite que les processus récursifs affines sont

aussi des processus affines, et leurs deux premiers moments conditionnels au passé peuvent être exprimés par des combinaisons affines disponibles en formule fermée. On peut aussi incorporer les processus récursifs affines aux modèles de valorisation des actifs à l'aide d'un facteur d'escompte stochastique exponentiel-affine. Dans ce cas, la formule de passage de la transformée de Laplace conditionnelle sous les mesures historique et risque-neutre est obtenue à l'aide des transformées d'Esscher, et est disponible en formule fermée. Enfin, on discute des méthodes d'estimation disponibles pour cette classe de processus suivant l'observabilité de celui-ci.

Lorsqu'on modélise des variables à support positif comme les taux d'intérêt, il est naturel de considérer des processus auto-régressifs gamma. Dans le chapitre 5, on introduit une nouvelle classe de processus auto-régressifs gamma appelés gamma-zéro (ARG_0), à support sur $\mathbb{R}_+$, qui peuvent atteindre zéro et y rester pendant plusieurs périodes. Ces processus sont particulièrement adaptés pour reproduire le comportement et la persistance des taux d'intérêt au plancher (ZLB). A l'aide de la méthode développée dans le chapitre 4, on construit un processus multivarié affine dont chacune des composantes peuvent individuellement atteindre zéro et y rester. Grâce à ce nouveau processus multivarié, on formule le premier modèle affine de taux d'intérêt délivrant simultanément un taux d'intérêt à court terme pouvant rester à zéro, et des taux d'intérêt positifs ou nuls à toutes les maturités exprimables comme fonctions affines des facteurs, disponibles en formules fermées. Si on observe des taux d'intérêt négatifs, notre modélisation autorise la spécification d'une borne basse négative pour le taux court. En utilisant des données de taux d'obligations souveraines japonaises de 1995 à 2014, on montre que notre modèle reproduit précisément le niveau et les volatilités des taux d'intérêt, particulièrement lorsque les taux sont au plancher. Les propriétés du processus vectoriel auto-régressif gamma-zéro nous permettent aussi l'obtention des probabilités de *lift-off* (c'est-à-dire les probabilités que le taux court reprenne des valeurs positives) en formules fermées aussi bien sous la mesure historique que risque-neutre. Nos estimations montrent que les différences entre les probabilités obtenues sous les deux mesures peuvent être substantielles, dues à la présence de primes de risque conséquentes.

Abstract

No-arbitrage interest rate models are devoted to reproduce the behavior of bond prices of several maturities through time. Among them, the class of affine term structure models (ATSM) has become increasingly popular since its introduction by Duffie and Kan [1996]. Indeed, this class of model is able to produce zero-coupon bond interest rates at each maturity that are closed-form affine functions of a reduced number of factors. The most widely used ATSM is undoubtedly the Gaussian affine term structure model, where bond yields at all maturities turn out to be affine combinations of a standard multivariate Gaussian VAR process. This implies in particular that all yields at all maturities have a Gaussian distribution, thus possess a statistical support on the entire real axis. This feature is well-known to be inconsistent with several theoretical properties of bond yields, such as the zero lower bound: since holding cash (providing a zero interest rate) is always an alternative to investing in bonds, no-arbitrage arguments imply that bond interest rates cannot theoretically go below zero. In practice however, we can observe negative bond interest rates due to financial frictions such as positive cash storage costs. Yields might have a negative lower bound which the Gaussian assumption still fails to reproduce. In comparison, the positive ATSMs are designed to ensure positivity of the whole yield curve, overcoming this issue. This thesis proposes several extensions to the existing positive affine term structure models. After a general introduction, a first part composed of Chapters 2 and 3 studies the quadratic term structure model (QTSM). A second part composed of Chapters 4 and 5 is centered on reproducing the behavior of interest rates during the zero lower bound.

In the introductory chapter, we detail the main assumptions required to build an affine term structure model (ATSM) of bond interest rates. The vector of endogenous factors is assumed to be an *affine* – or compound autoregressive (Car) – process under the risk-neutral measure, i.e. its conditional Laplace transform given its past is an exponential-affine function of its past values. Together with an short-term interest rate specified as an affine combination of these factors (and an affine risk intensity in the case of defaultable bonds), we obtain closed-from pricing formulas for zero-coupon bond yields, a desirable property, especially for estimation purposes. The specification of a stochastic discount

factor also allows to retrieve the physical dynamics of the vector of factors, and formulate a state-space model.

As shown by previous work, QTSMs are indeed embedded in the class of ATSMs. When considering a quadratic term structure model (QTSM), the state-space model possesses affine transition equations and linear-quadratic measurement equations. When the factors are latent, the standard Kalman filter cannot be employed. Several non-linear filters have been developed for general non-linear state-space models. A first class of non-linear filters use simulations to approximate the non-Gaussian distributions and produce more accurate filtered variables. However, this comes at the cost of a high computational complexity. A second class of non-linear filters can be called deterministic. They mostly rely on Taylor expansions of the state-space model's non-linearities. However, these algorithms (such as the extended Kalman filter, or the unscented Kalman filter) are general and not specifically fitted to the linear-quadratic state-space model. In Chapter 2, we develop a new filtering algorithm that we call the *Quadratic Kalman Filter* (QKF), in order to circumvent these issues. Since the QKF belongs to the class of deterministic filters, it is computationally very fast and simple. The development of the algorithm relies entirely on the affine processes properties. Indeed, when the state vector follows a Gaussian VAR(1), the augmented vector composed of the original state and all possible products of the state vector elements is known to be an affine process. Using the VAR representation of the augmented vector, we express the state-space model in a linear form, bringing back the possibility to use the linear Kalman filter algorithm. As a consequence, the derivation of the smoothing algorithm is straightforward. Using Monte-Carlo simulations, we assess the QKF performance for both filtering and quasi-maximum likelihood estimation. We show that it outperforms its deterministic competitors in both exercises.

This new estimation technique is used in Chapter 3 to estimate a QTSM on the Euro area interbank spreads. These spreads are defined by the difference between the risky Euribor interest rates, and the nearly risk-less overnight indexed swap rates (OIS). The sources of fluctuations of these spreads are assumed to be only related to default and liquidity risks. Contrary to most existing models of interbank spreads, we specifically impose that the spreads are always positive through the quadratic specification. Indeed, the defaultable zero-coupon bonds interest rates incorporate a positive compensation for the credit and liquidity risks whereas the risk-less bonds interest rates do not. We build an intensity-based model to take these two risks into account. Both intensities as quadratic combinations of two different Gaussian factors. These factors are related to credit and liquidity observable proxies in order to identify both components in the spreads. Our estimates allow us to decompose the spreads into their different components at each ma-

turity, and to provide an insight on the effect of the ECB unconventional monetary policy actions on the Euro area interbank market. We show that even though most of the tensions after Lehman failure are liquidity-related, the recent ECB actions (mostly the *long term refinancing operations* and the *outright monetary transactions*) have contributed to greatly reduce these concerns. Our framework also allows to derive the average default probabilities in the Eurozone. We show that those probabilities have decreased as well after the recent episodes of unconventional monetary policies.

In chapter 4, we consider a broader issue related to the construction of affine interest rate models. Given several univariate affine process, we can easily build a multivariate affine process stacking the different components together and assuming that they are independent. The independence assumption can often be replaced by a conditional independence assumption between components given the past without changing the affine nature of the multivariate process. However, introducing conditional dependence between the different components while preserving the affine property of the multivariate process can be complicated. In this chapter, we provide a new simple method to define multivariate affine processes without relying on conditional independence between components. We call these processes *recursive affine*. They can be defined recursively as follows. Let the vector of risk factors be partitioned in several blocks. This vector is recursive affine if the conditional Laplace transform of each block given the past of the whole process and the present of blocks located above it in the vector is an exponential-affine function of the conditioning variables. This technique can be applied to a wide variety of examples, such as stochastic volatility VAR processes, discrete-value processes, or vector autoregressive gamma processes. These multivariate processes are shown to be affine, hence their first two conditional moments can be expressed with simple closed-form affine expressions. Assuming an exponential-affine stochastic discount factor, the transition between the physical and risk-neutral conditional Laplace transforms is available in closed-form using Esscher transforms. We last discuss the estimation of such processes depending on its observability.

Considering autoregressive gamma processes are of particular use when modeling positive variables. In Chapter 5, we develop a new class of autoregressive gamma processes that we call *autoregressive gamma-zero* (ARG_0), that is not only positive, but can reach zero and stay there for several periods. These processes are particularly fitted to reproduce the interest rates stylized facts observed during the zero lower bound (ZLB), specifically the persistence of the ZLB state. Using the method developed in Chapter 4, we build the multivariate affine counterpart of the ARG_0 processes the components of which can individually reach and stay at zero for several periods. This new process allows us to

build an affine term structure model which, contrary to existing models, can provide simultaneously non-negative interest rates, a short-term interest rate that stays at zero for several periods, and closed-form interest rate formulas. in the model, the lower bound on the short-term interest rate can also be negative consistently with observed data. We apply our model on Japanese Government Bonds data from 1995 to 2014. The estimated model shows nice performances in reproducing the level and conditional volatilities of yields at all maturities. The properties of the vectorial autoregressive gamma zero process also allow us to compute in closed-form the so-called *lift-off probabilities*, that is the probabilities that the short-term interest rate goes back to positive values permanently. We provide evidence that there can be a large bias if the assessment of these probabilities is made without purging the risk premia estimates from the observed yields.

Chapter 1

Introduction

Abstract

This first chapter summarizes the notations and the assumptions most commonly used in affine interest rate models. It is accompanied by a review of the associated literature. In Section 1.1, we present the conditional Laplace transform and define the class of affine processes, which are both of utmost importance for asset pricing models. Section 1.2 introduces the no-arbitrage assumption along with the stochastic discount factor, which allows for the identification of the risk premia contained in asset prices. In section 1.3, we develop general pricing formulas for risk-less and defaultable zero-coupon bonds. In the case of affine interest rate models, these general formulas are extremely simple, as shown by Section 1.4. This section also covers the most popular examples of affine interest rate models. Section 1.5 reviews the properties, the estimation techniques, and the recent developments of the class of quadratic term structure models. This model proves particularly useful to reproduce the positivity of interest rates. This positivity property alone is not sufficient to reproduce the fact that interest rates can reach zero for a long period (*zero lower bound*). Section 1.6 summarizes the recent interest rate models employed to that respect. Last, Section 1.7 announces the following chapters.

1.1 The general discrete-time setup

1.1.1 Information of investors and factor dynamics: notations

Consider a discrete-time economy composed of a representative investor and n risk factors denoted by X_t. The set of information available to the investor at time t is denoted by $\underline{X_t} = (X_t, X_{t-1}, \ldots)$. These factors can take values on the entire real line or on a subset. The definition of the risk factors is central in asset pricing models. First, we assume that the past and present values of the process X_t correspond to the full information set of the investor, and is always observable by him. Second, any asset payoff is defined as a function of the risk factors $g(\underline{X_t})$, the form of $g(\bullet)$ depending on the considered asset type. When pricing financial assets, the representative investor incorporates all available information on present and past values of X_t and his expectations of the future values of the process.

The risk factors in asset pricing models can be either partially or totally observed by the econometrician. Observable variables include for instance prices of financial assets or macroeconomic variables such as GDP growth or inflation (see for instance Ang and Piazzesi [2003]). When the factors are unobserved, the econometrician has to resort to filtering techniques to infer their values from observed data.

The probability distribution of the risk factors in the real world is called the historical or physical measure. The dynamics of the risk factors may be known via their conditional probability density function.

Definition 1.1.1. *Let X_t be a $\mathbb{R}^n$-valued process. The dynamics of X_t under the physical measure are defined by the p.d.f. of X_{t+1} given all the present and past values of X_t, denoted by $\underline{X_t} = (X_t, X_{t-1}, \ldots)$. This p.d.f. is denoted by $f^{\mathbb{P}}(x_{t+1}|\underline{X_t})$, $x_{t+1} \in \mathbb{R}^n$.*

Throughout this thesis, we will particularly focus on interest rates on bonds. We will consider mostly continuously-compounded interest rates (or *log*-interest rates) defined as the log-ratio of bond prices taken at two different dates. We concentrate on zero-coupon bonds, that is bonds that yield one unit of currency at maturity. We define the short-term interest rate between t and $t+1$ as the log-return of a one-period zero-coupon bond, denoted by r_t. This interest rate is known when contracting the bond at time t and is thus risk-less. This short-term interest rate is a function of the risk factors hence r_t will be used as a notation abuse for $r(\underline{X_t})$.

1.1.2 The conditional Laplace transform

Though the conditional p.d.f. of the risk factors characterizes their dynamics, a more convenient representation can be based on the conditional Laplace transform of X_{t+1} given $\underline{X_t}$.

Definition 1.1.2. *Let X_t be a $\mathbb{R}^n$-valued process. Let the historical dynamics of X_t be defined by the conditional p.d.f $f^{\mathbb{P}}(x_{t+1}|\underline{X_t})$. The conditional Laplace transform of X_{t+1} given $\underline{X_t}$ is denoted by $\varphi_t^{\mathbb{P}}(u)$ and is given by:*

$$\forall u \in \mathcal{S}\,, \quad \varphi_t^{\mathbb{P}}(u) := \mathbb{E}^{\mathbb{P}}[\exp(u'X_{t+1})|\underline{X_t}] = \int_{\mathbb{R}^n} \exp(u'x_{t+1}) f^{\mathbb{P}}(x_{t+1}|\underline{X_t})\,\mathrm{d}x_{t+1}\,, \tag{1.1}$$

where $\mathbb{E}^{\mathbb{P}}(\bullet)$ is the expectation operator under the physical probability measure, and $\mathcal{S}$ is a subset of $\mathbb{R}^n$ such that the conditional Laplace transform is well-defined.

The conditional log-Laplace transform of X_{t+1} given $\underline{X_t}$ is denoted by $\psi_t^{\mathbb{P}}(u)$ and is given by:

$$\forall u \in \mathcal{S}\,, \quad \psi_t^{\mathbb{P}}(u) := \log\left(\mathbb{E}^{\mathbb{P}}[\exp(u'X_{t+1})|\underline{X_t}]\right) = \log\left(\varphi_t^{\mathbb{P}}(u)\right)\,. \tag{1.2}$$

Knowing the conditional Laplace transform of X_{t+1} given $\underline{X_t}$ is sufficient to characterize the whole dynamics of X_t. One can therefore use either the conditional p.d.f. or the conditional Laplace transform equivalently. However, the latter has several advantages. First, the k^{th} derivative of $\varphi_t^{\mathbb{P}}(u)$, setting $u = 0$, exactly gives the k^{th} conditional moment of X_{t+1} given $\underline{X_t}$. Indeed, the conditional Laplace transform is also called the conditional *moment-generating function*. Similarly, the conditional log-Laplace transform's k^{th} derivative, setting $u = 0$, gives the k^{th} conditional cumulant of X_{t+1} given $\underline{X_t}$, hence $\psi_t^{\mathbb{P}}(u)$ is also called the *cumulant-generating function*.

1.1.3 Affine processes and multihorizon Laplace transforms

So far, we have defined the one-period conditional distribution of the risk factors. However, financial assets often pay off several periods after the purchasing date. This is the case for fixed-income securities which are contracted for a certain time-to-maturity, denoted by h in this thesis. As a consequence, it is useful to consider not only the one-period conditional Laplace transform, but the multi-horizon conditional Laplace transform of $(X_{t+1}, \ldots, X_{t+h})$ given $\underline{X_t}$.

Definition 1.1.3. *Let X_t be a $\mathbb{R}^n$-valued process. The multi-horizon conditional Laplace transform is denoted by $\varphi_{t,t+h}^{\mathbb{P}}(u_1, \ldots, u_h)$ and is defined by:*

$$\forall h \in \mathbb{N}^*\,, \forall (u_1, \ldots, u_h) \in \mathcal{S}^h\,, \quad \varphi_{t,t+h}^{\mathbb{P}}(u_1, \ldots, u_h) := \mathbb{E}^{\mathbb{P}}\left[\exp(u_1'X_{t+1} + \ldots + u_h'X_{t+h})|\underline{X_t}\right]\,,$$

where $\mathcal{S}$ is a subset of $\mathbb{R}^n$ such that the conditional Laplace transform is well-defined.

The multi-horizon conditional Laplace transform generalizes the one-period conditional Laplace transform. Indeed, for $u = (u_1, 0, \ldots, 0)$, the multi-horizon conditional Laplace transform is exactly equal to $\varphi_t^{\mathbb{P}}(u)$. More generally, we can define the h-period conditional Laplace transform putting all u_i to zero except u_h. Provided that (X_t) is strongly stationary, it is also easy to obtain the marginal distribution of the risk factors setting all u_i equal to zero except u_h, and letting h tend to infinity. The h-period conditional Laplace transform converges to the marginal Laplace transform of X_t.

The multi-horizon conditional Laplace transform is extensively used in asset-pricing models (see next section). In that respect, some conditional distributions are easier to manipulate due to the specific form of their Laplace transforms. We define the class of *affine* processes as the set of processes that possess exponential-affine conditional Laplace transforms in X_t, or equivalently affine conditional log-Laplace transforms in X_t (see Darolles, Gouriéroux, and Jasiak [2006]).

Definition 1.1.4. *The process $(X_t)_{t\in\mathbb{Z}}$ is said to be affine (under the $\mathbb{P}$-measure) – or compound autoregressive of order 1 – if and only if the conditional Laplace transform (resp. conditional log-Laplace transform) of X_{t+1} given $\underline{X_t}$ is an exponential-affine (resp. affine) function of X_t.*

$$\forall u \in \mathcal{S}\,, \quad \varphi_t^{\mathbb{P}}(u) = \exp\left[a(u)'X_t + b(u)\right] \tag{1.3}$$

$$\psi_t^{\mathbb{P}}(u) = a(u)'X_t + b(u)\,, \tag{1.4}$$

where $a(u) : \mathcal{S} \longrightarrow \mathbb{R}^n$ and $b(u) : \mathcal{S} \longrightarrow \mathbb{R}$ are deterministic functions of u.

The class of affine processes is wide enough to include e.g. the Gaussian VAR, the autoregressive gamma process of Gouriéroux and Jasiak [2006], or the autoregressive Poisson process. Inside the class of affine processes, the conditional distributions are entirely defined by functions $a(u)$ and $b(u)$. For instance, if X_t follows a Gaussian VAR:

$$X_{t+1} = \mu + \Phi X_t + \Sigma^{1/2}\varepsilon_{t+1}\,,$$

where $\mu \in \mathbb{R}^n$, Φ and Σ matrices in $\mathbb{R}^{n\times n}$, and ε_{t+1} is a multivariate i.i.d. normalized Gaussian shock, the conditional Laplace transform of X_{t+1} given X_t is exponential-affine in X_t and defined by:

$$a(u) = \Phi' u \quad \text{and} \quad b(u) = u'\mu + \frac{1}{2}u'\Sigma u\,.$$

By combining the properties of conditional Laplace transforms with the specific formulation of affine processes, simple properties of conditional moments and cumulants can easily be derived.

Proposition 1.1.1. *Let (X_t) be an affine process under the physical measure. Conditional cumulants of X_{t+1} given $\underline{X_t}$ of any order are affine functions of X_t. In particular, the first two conditional cumulants and the first two conditional moments are the same and are equal to:*

$$\mathbb{E}^{\mathbb{P}}(X_{t+1}|\underline{X_t}) = \left.\frac{\partial a(u)'}{\partial u}\right|_{u=0} X_t + \left.\frac{\partial b(u)}{\partial u}\right|_{u=0} \tag{1.5}$$

$$\mathbb{V}^{\mathbb{P}}(X_{t+1}|\underline{X_t}) = \sum_{i=1}^{n} \left(\frac{\partial^2 a_i(u)}{\partial u \partial u'}\right)_{u=0} X_{i,t} + \left.\frac{\partial^2 b(u)}{\partial u \partial u'}\right|_{u=0}, \tag{1.6}$$

where $a_i(u)$ is the i^{th} element of vector $a(u)$, and $\mathbb{V}^{\mathbb{P}}(\bullet)$ is the variance operator under the physical measure.

This convenient property helps deriving closed-form formulas for the first two conditional moments of an affine process: as long as $a(u)$ and $b(u)$ are known-in closed-form, it is easy to derive conditional cumulants of any order, including the first two which are equal to the first two conditional moments. Affine processes also possess exponential-affine multi-horizon conditional Laplace transforms.

Proposition 1.1.2. *Let (X_t) be an affine process under the $\mathbb{P}$-measure. The multi-horizon h-period ahead conditional Laplace transform given $\underline{X_t}$ is an exponential-affine function of X_t.*

$$\forall h \in \mathbb{N},\ \forall (u_1,\ldots,u_h) \in \mathcal{S}^h,\quad \varphi^{\mathbb{P}}_{t,t+h}(u_1,\ldots,u_h) = \exp\left[\mathcal{A}^{(h)}_h(u_1,\ldots,u_h)'X_t + \mathcal{B}^{(h)}_h(u_1,\ldots,u_h)\right],$$

where $\mathcal{A}^{(h)}_h(u_1,\ldots,u_h) : \mathcal{S}^h \longrightarrow \mathbb{R}^n$ and $\mathcal{B}^{(h)}_h(u_1,\ldots,u_h) : \mathcal{S}^h \longrightarrow \mathbb{R}$ can be computed recursively, initializing at $\mathcal{A}^{(1)}_h = a(u_h)$ and $\mathcal{B}^{(1)}_h = b(u_h)$, as follows:

$$\begin{aligned} \textit{For}\quad k \in \{2,\ldots,h\},\ \mathcal{A}^{(k)}_h(u_1,\ldots,u_h) &= a(u_{h-k+1} + \mathcal{A}^{(k-1)}_h(u_1,\ldots,u_h)) \\ \textit{and}\quad \mathcal{B}^{(k)}_h(u_1,\ldots,u_h) &= \mathcal{B}^{(k-1)}_h(u_1,\ldots,u_h) + b(u_{h-k+1} + \mathcal{A}^{(k-1)}_h(u_1,\ldots,u_h)). \end{aligned}$$

Again, as long as $a(u)$ and $b(u)$ are available in closed-form, the recursive formulas of the multi-horizon conditional Laplace transform are easily obtained. In the next section, we use these convenient properties for the modeling of asset prices.

1.2 General results of asset pricing modeling

In this section, we describe the absence of arbitrage (AOA) assumption on which asset pricing is based, and we introduce the notions of stochastic discount factor and risk-neutral dynamics.

1.2.1 The absence of arbitrage and the stochastic discount factor (SDF)

Consider the general setup developed in the previous section. The new information available at time t is given by X_t. Consider also an asset that yields $g(\underline{X_{t+h}})$ at time $t+h$. An asset-pricing model maps the payoff of this asset, to the price P_t at which the asset trades today. To do so, we must consider several assumptions.

First, we assume that the prices at which the assets trade at each period exist and are unique (*law of one price*). Second, we assume that the payoff function $g(\bullet)$ belongs to the set of square integrable functions at each date. Third, we assume that the price is linear and continuous in the payoff of two different assets. Suppose we have a portfolio of assets 1 and 2 in quantities γ_1 and γ_2 yielding payoffs at $t+h$ equal to $g_1(X_{t+h})$ and $g_2(X_{t+h})$ respectively. The total payoff in $t+h$ will be equal to $\gamma_1 g_1(X_{t+h}) + \gamma_2 g_2(X_{t+h})$. The linearity assumption states that the price of the portfolio should be equal to the linear combination of the prices of each asset loaded by (γ_1, γ_2). The continuity assumption states that an asset with a payoff tending towards zero should have a price converging to zero as well. The last and most important assumption that we consider for the pricing of financial assets is the *absence of arbitrage opportunities* or *no-arbitrage assumption* (AOA). Intuitively, this assumption states that the prices of assets today instantaneously adjust such that no trading strategy can provide a strictly positive excess return over another trading strategy with probability one.

Definition 1.2.1. *The absence of arbitrage opportunities assumption ensures that at any date t, it is not possible to build a portfolio of future payoffs – possibly modified at subsequent dates – such that:*

- *the price of the portfolio at time t is zero,*
- *payoffs at subsequent dates are non-negative,*
- *there is at least one subsequent date such that the net payoff at this date is strictly positive with a non zero conditional probability at t.*

Using the previous assumption, Riesz representation theorem ensures the following statements:

Proposition 1.2.1. *Under the previous assumptions, there exists a unique discount factor between t and $t+h$, denoted by $M_{t,t+h}(\underline{X_{t+h}})$, such that any asset with a payoff given by $g(\underline{X_{t+h}})$ has a price P_t given by:*

$$P_t = \mathbb{E}^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}})g(\underline{X_{t+h}})\middle|\underline{X_t}\right] . \tag{1.7}$$

For the specific case of infinitely lived assets the payoff function is equal to the price at $t+h$ hence $P_t = \mathbb{E}^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}})P_{t+h}\middle|\underline{X_t}\right]$. In addition, $M_{t,t+h}$ is:

- *positive given the AOA assumption;*
- *equal to $\prod_{i=0}^{h-1} M_{t+i,t+i+1}(\underline{X_{t+i+1}})$ the product of all one-period discount factors;*

This factor $M_{t,t+1}(\underline{X_{t+1}})$ is called the *stochastic discount factor* (SDF) and depends on the risk factors $\underline{X_{t+1}}$. Putting together the previous results on the SDF, the prices of financial assets today are equal to the expected discounted value of their future payoffs, where the discount factor is a random variable which is not equal to the risk-less interest rate most of the times. This risk-less interest rate would be the prevailing discount factor provided the representative agent were risk-neutral. Therefore, whenever the SDF is random and different from the risk-less interest rate, the representative agent shows non-zero risk aversion. A direct consequence of the previous proposition is that the price of a zero-coupon bond maturing at $t+h$ is $\mathbb{E}^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}})\middle|\underline{X_t}\right]$. Indeed, using the pricing formulas, the price of an asset paying $g(\underline{X_{t+h}})$ at time $t+h$ can be decomposed:

$$\begin{aligned} P_t &= \mathbb{E}^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}})g(\underline{X_{t+h}})\middle|\underline{X_t}\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}})\middle|\underline{X_t}\right]\mathbb{E}^{\mathbb{P}}\left[g(\underline{X_{t+h}})\middle|\underline{X_t}\right] + \mathbb{C}ov^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}}), g(\underline{X_{t+h}})\middle|\underline{X_t}\right] \\ &= B(t,h)\ \underbrace{\mathbb{E}^{\mathbb{P}}\left[g(\underline{X_{t+h}})\middle|\underline{X_t}\right]}_{\text{expected component}} + \underbrace{\mathbb{C}ov^{\mathbb{P}}\left[M_{t,t+h}(\underline{X_{t+h}}), g(\underline{X_{t+h}})\middle|\underline{X_t}\right]}_{\text{risk premium}}, \end{aligned}$$

where $B(t,h)$ is the price of a risk-less zero-coupon bond of maturity h at time t. The price of any asset is therefore equal to the weighted sum of the conditional expectation of the asset payoff called *expected component*, and of the conditional covariance between the SDF up to maturity and the future payoff, known as the *risk premium*.

1.2.2 Modeling with the risk-neutral dynamics

Assume that we are in the setup of the previous sections. It is possible to formulate a new probability measure such that the observed prices are equal to the conditional expectation of the future payoffs with respect to this new measure, discounted by the risk-less

interest rate. This probability measure is called *risk-neutral*.

Definition 1.2.2. *Let X_t be a $\mathbb{R}^n$-valued process. Let $M_{t,t+1}(\underline{X_{t+1}})$ be the stochastic discount factor. The risk-neutral probability measure is defined by $f^{\mathbb{Q}}(x_{t+1}|\underline{X_t})$ the risk-neutral conditional densities of the risk-factors which are given by:*

$$f^{\mathbb{Q}}(x_{t+1}|\underline{X_t}) = f^{\mathbb{P}}(x_{t+1}|\underline{X_t})\frac{M_{t,t+1}(\underline{x_{t+1}})}{\mathbb{E}^{\mathbb{P}}\left[M_{t,t+1}(\underline{X_{t+1}})\middle|\underline{X_t}\right]} = f^{\mathbb{P}}(x_{t+1}|\underline{X_t})M_{t,t+1}(\underline{x_{t+1}})\exp(r_t)\,, \tag{1.8}$$

where $r_t = r(\underline{X_t})$ is the risk-less one-period interest rate. Any asset with a single payoff at $t+h$ given by $g(\underline{X_{t+h}})$ has a price P_t given by:

$$P_t = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\sum_{i=0}^{h-1} r_{t+i}\right) g(\underline{X_{t+h}})\middle|\underline{X_t}\right]\,. \tag{1.9}$$

Equation (1.9) is the equivalent of Equation (1.7) in the risk-neutral world. Under $\mathbb{P}$, the discounting of future payoffs is made with the SDF whereas under $\mathbb{Q}$, the probability distribution is shifted such that the discounting is made with the compounded h-period risk-less interest rate, namely $\exp(-r_t - \ldots - r_{t+h-1})$. If one is interested only in the pricing of financial instruments, leaving aside the risk premia decomposition, it is possible to work only under the risk-neutral measure. As long as the conditional risk-neutral p.d.f and the function $r(\underline{X_t})$ are known, asset prices can be computed.

1.2.3 Modeling strategies

Consider now that the function $r(\underline{X_t})$ – or equivalently $\mathbb{E}^{\mathbb{P}}[M_{t,t+1}(\underline{X_{t+1}})|\underline{X_t}]$ – is set. We can obtain prices with the physical measure specifying both $f^{\mathbb{P}}(x_{t+1}|\underline{X_t})$ and $M_{t,t+1}(\underline{X_{t+1}})$, or with the risk-neutral measure specifying only $f^{\mathbb{Q}}(x_{t+1}|\underline{X_t})$. If we are not only interested in pricing financial assets but also in obtaining the risk premia components, the specification of the risk-neutral density alone is not enough. However, the transition formula (1.8) between the physical and the risk-neutral world essentially states that it is possible to obtain such a decomposition specifying only two elements out of the physical conditional density, the risk-neutral conditional density, and the SDF. For instance, if we specify both the physical and the risk-neutral conditional densities, the SDF can be deduced as a residual.

This particular relationship has led Bertholon, Monfort, and Pegoraro [2008] to create a typology of the specification strategies in asset pricing models. These strategies are summarized in Table 1.1.

	$f^{\mathbb{P}}(x_{t+1}\|\underline{X_t})$	$f^{\mathbb{Q}}(x_{t+1}\|\underline{X_t})$	$M_{t,t+1}(x_{t+1}\|\underline{X_t})$
Direct Modeling	specified	deduced	specified
Risk-Neutral Constrained Direct Modeling	specified	specified	deduced
Back Modeling	deduced	specified	specified

Table 1.1: Asset Pricing Modeling Strategies

In the *direct modeling* strategy, the physical conditional density and the SDF are directly specified. In the *risk-neutral constrained direct modeling* strategy, the econometrician specifies both the physical and the risk-neutral conditional densities and obtain the SDF as a residual. Last, in the *back modeling* strategy, both the risk-neutral conditional density and the SDF are specified. The adopted modeling strategy often depends on the focus of the asset-pricing model and whether the econometrician wants to control for the historical dynamics and/or the pricing. If some of the risk factors are observed with, for instance, macro variables, it seems more natural to focus on the historical dynamics specifying directly the physical conditional p.d.f., using either the direct or risk-neutral constrained modeling. If on the contrary the model is entirely focused on the pricing of assets with latent factors, either the risk-neutral constrained or the back modeling are preferred.

1.2.4 Exponential-affine specification of the stochastic discount factor

So far, we have presented the physical and the risk-neutral representations of the formation of asset prices. In the previous section, we have emphasized that the transition between the two worlds is made simple with the SDF. In this section, we present a convenient specification of the stochastic discount factor to obtain simple transition formulas.

A key feature of the SDF is that it should always be positive. A simple representation has been given by e.g. Duffee [2002], where the SDF is an exponential-affine function of the risk factors.

Definition 1.2.3. *Let X_t be a $\mathbb{R}^n$-valued process. Let $M_{t,t+1}(\underline{X_{t+1}})$ be the SDF. The SDF is exponential-affine if it is of the form:*

$$M_{t,t+1}(\underline{X_{t+1}}) = \exp\left[\alpha_t'(\underline{X_t})X_{t+1} + \beta_t(\underline{X_t})\right] , \tag{1.10}$$

where $\alpha_t(\underline{X_t})$ and $\beta_t(\underline{X_t})$ are functions of the past and present values of X_t.

When the SDF is exponential-affine, the risk-less interest rate can be expressed simply

with the use of the physical conditional Laplace transform of X_{t+1} given $\underline{X_t}$. This implies that the transition formula between the physical and risk-neutral conditional densities can also be expressed easily.

Proposition 1.2.2. *Let X_t be a $\mathbb{R}^n$-valued process. Let $M_{t,t+1}(\underline{X_{t+1}})$ be the exponential-affine SDF given by Equation (1.10). The short-term risk-less interest rate is given by:*

$$\begin{aligned} r(\underline{X_t}) &= -\log \mathbb{E}^{\mathbb{P}}\left[M_{t,t+1}(\underline{X_{t+1}})\middle|\underline{X_t}\right] = -\beta_t(\underline{X_t}) - \psi_t^{\mathbb{P}}\left[\alpha_t(\underline{X_t})\right] \\ \textit{hence} \quad M_{t,t+1}(\underline{X_{t+1}}) &= \exp\left[-r(\underline{X_t}) + \alpha_t'(\underline{X_t})X_{t+1} - \psi_t^{\mathbb{P}}(\alpha_t(\underline{X_t}))\right] , \end{aligned}$$

and the transition equation (1.8) can be rewritten:

$$f^{\mathbb{Q}}(x_{t+1}|\underline{X_t}) = f^{\mathbb{P}}(x_{t+1}|\underline{X_t})\frac{\exp\left[\alpha_t'(\underline{X_t})X_{t+1}\right]}{\varphi_t^{\mathbb{P}}\left[\alpha_t(\underline{X_t})\right]} . \tag{1.11}$$

This convenient mapping between the risk-neutral and the physical conditional densities can also be expressed in terms of conditional Laplace transforms of X_{t+1} given $\underline{X_t}$. Again, the transition formula turns out to be simple.

Proposition 1.2.3. *Let X_t be a $\mathbb{R}^n$-valued process. Let $M_{t,t+1}(\underline{X_{t+1}})$ be the exponential-affine SDF given by Equation (1.10). The risk-neutral conditional Laplace transform and the conditional log-Laplace transform of X_{t+1} given $\underline{X_t}$ are respectively denoted by $\varphi_t^{\mathbb{Q}}(u)$ and $\psi_t^{\mathbb{Q}}(u)$, and are given by:*

$$\forall u \in \mathcal{S}, \quad \varphi_t^{\mathbb{Q}}(u) := \mathbb{E}^{\mathbb{Q}}\left[\exp(u'X_{t+1})\middle|\underline{X_t}\right] = \frac{\varphi_t^{\mathbb{P}}\left[u + \alpha_t(\underline{X_t})\right]}{\varphi_t^{\mathbb{P}}\left[\alpha_t(\underline{X_t})\right]} \tag{1.12}$$

$$\psi_t^{\mathbb{Q}}(u) := \log\left(\varphi_t^{\mathbb{Q}}(u)\right) = \psi_t^{\mathbb{P}}\left[u + \alpha_t(\underline{X_t})\right] - \psi_t^{\mathbb{P}}\left[\alpha_t(\underline{X_t})\right] . \tag{1.13}$$

When the risk factor process is affine, either under $\mathbb{P}$ or $\mathbb{Q}$, the transition between the two worlds given by Equation (1.13) is facilitated by the affine form of the conditional log-Laplace transform of X_{t+1} given $\underline{X_t}$.

Proposition 1.2.4. *Let X_t be an affine process under the physical measure with a conditional Laplace transform given in Definition 1.1.4. Let $M_{t,t+1}(\underline{X_{t+1}})$ be the exponential-affine SDF given by Equation (1.10). The risk-neutral conditional log-Laplace transform of X_{t+1} given $\underline{X_t}$ is given by:*

$$\forall u \in \mathcal{S}, \quad \psi_t^{\mathbb{Q}}(u) = \left\{a\left[u + \alpha_t(\underline{X_t})\right] - a\left[\alpha_t(\underline{X_t})\right]\right\}' X_t + b\left[u + \alpha_t(\underline{X_t})\right] - b\left[\alpha_t(\underline{X_t})\right] .$$

We will show in the following section that it is important that the risk factor process is affine under the $\mathbb{Q}$-measure to obtain closed-form pricing formulas. In the context of risk-neutral constrained or back modeling, we just have to specify the risk-neutral distribution

of the risk factors such that the process is affine under $\mathbb{Q}$. In the context of direct modeling however, assuming that X_t is affine in the physical world, we can see from the previous proposition that not all functions $a(\bullet)$, $b(\bullet)$ and $\alpha_t(\bullet)$ yield affine risk-neutral dynamics for X_t.

1.3 Risk-less and risky bond pricing

In this section, we use the framework that has been presented above to price both risk-less and defaultable bonds. After defining interest rates on zero-coupon bonds, we derive general pricing formulas using the exponential-affine SDF specification.

1.3.1 Risk-less bond pricing

A risk-less zero-coupon bond is a financial instrument that provides one unit of currency at maturity. We denote by $B(t,h)$ the date t price of a zero-coupon bond maturing in h periods and $R(t,h)$ the corresponding continuously-compounded rate. As emphasized previously, the short-term risk-less interest rate r_t is equal to the one-period bond interest rate $R(t,1)$. Using the definition of log interest rates, the relationship between $R(t,h)$ and $B(t,h)$ is given by:

$$R(t,h) = -\frac{1}{h}\log\left[B(t,h)\right] .$$

Using the no-arbitrage assumption presented previously, the prices of zero-coupon bonds are given either under the physical of the risk-neutral measure by:

$$\begin{aligned} B(t,h) &= \mathbb{E}^{\mathbb{P}}\left[M_{t,t+1}(\underline{X_{t+1}})\dots M_{t+h-1,t+h}(\underline{X_{t+h}})\middle|\underline{X_t}\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\exp(-r_t-\dots-r_{t+h-1})\middle|\underline{X_t}\right] . \end{aligned} \tag{1.14}$$

Let us consider an econometric setup where the short-term risk-less rate is a parametric function of the risk factors governed by a set of parameters denoted by $\theta_r \in \Theta_r$, that is $r_t = r(\underline{X_t};\theta_r)$. We consider the direct modeling strategy and specify the physical conditional density of X_{t+1} given $\underline{X_t}$ as a parametric function of a set of parameters $\theta^{\mathbb{P}} \in \Theta^{\mathbb{P}}$, and the p.d.f. is denoted by $f^{\mathbb{P}}(x_{t+1}|\underline{X_t};\theta^{\mathbb{P}})$. We consider an exponential-affine SDF as in Equation (1.10) and specify the loadings as a parametric function of a set of parameters $\theta_\alpha \in \Theta_\alpha$, which are denoted by $\alpha_t(\underline{X_t};\theta_\alpha)$.[1] Using the transition equation between the two worlds, the risk-neutral conditional p.d.f. of X_{t+1} given $\underline{X_t}$ is a parametric function of $(\theta_\alpha,\theta^{\mathbb{P}})$ and is denoted by $f^{\mathbb{Q}}(x_{t+1}|\underline{X_t};\theta_\alpha,\theta^{\mathbb{P}})$. The econometrician is interested in esti-

1. The function $\beta_t(\underline{X_t};\theta_\alpha)$ can be deduced from the specification of $\alpha_t(\underline{X_t};\theta_\alpha)$ and $r_t(\underline{X_t};\theta_r)$.

mating $\theta = (\theta_r, \theta^{\mathbb{P}}, \theta_\alpha) \in \Theta = (\Theta_r \times \Theta^{\mathbb{P}} \times \Theta_\alpha)$.

Using the previous pricing formulas, we can link the observed bond prices to the parametric specification.

$$\begin{aligned} B(t,h) &= \mathbb{E}^{\mathbb{P}(\theta^{\mathbb{P}})}\left[\exp\left(\sum_{i=0}^{h-1}\alpha_{t+i}(\underline{X_{t+i}};\theta_\alpha)'X_{t+i+1} - r(\underline{X_{t+i}};\theta_r) - \psi_t^{\mathbb{P}(\theta^{\mathbb{P}})}\left[\alpha_{t+i}(\underline{X_{t+i}};\theta_\alpha)\right]\right)\big|\underline{X_t}\right] \\ &= \mathbb{E}^{\mathbb{Q}(\theta_\alpha,\theta^{\mathbb{P}})}\left[\exp\left(-\sum_{i=0}^{h-1} r(\underline{X_{t+i}};\theta_r)\right)\big|\underline{X_t}\right]. \end{aligned}$$

The entire set θ intervenes in the pricing formulas both under the physical or the risk-neutral expression which, provided identification conditions are met, allows the econometrician to estimate parameters from the bond prices. The denomination *risk-less* for the bonds stems from the fact that the terminal payoff at time $t+h$ is equal to one and is known *ex-ante*. Therefore, if an investor buys a risk-less zero-coupon bond at time t and holds the bond up to maturity, he knows that he will earn an interest rate $R(t,h)$. Still, between the buying and the maturity dates, the successive prices of the bond are not known at time t and are random variables. In other words, if the investor decides *ex-ante* to sell the bond before maturity, he will earn a random interest rate. This uncertainty on the future value of long-term bonds is called *term risk* or *interest rate risk*, and corresponds to the uncertainty on future realizations of the short-term risk-less rate r_t between t and $t+h$.

1.3.2 Defaultable bonds and intensity models

A defaultable zero-coupon bond is a financial instrument that provides one unit of currency at maturity provided the issuing entity has not defaulted. The payoff of such bonds is either 1 if the entity is still alive at time $t+h$ and zero otherwise (zero recovery rate). We denote by $\widetilde{B}(t,h)$ the date t price of a defaultable zero-coupon bond maturing in h periods if the entity is still alive at date t, and $\widetilde{R}(t,h)$ the corresponding continuously-compounded rate. We also denote by d_t the binary variable indicating whether the issuing entity is alive at time t (default: $d_t = 1$). For the sake of clarity, even though the default dummy is a risk factor and belongs to the investor's information set, we separate X_t from d_t in this section. We make several additional assumptions.

Definition 1.3.1. *Let $d_t \in \{0,1\}$ be the discrete process indicating whether default has occurred.*

1. *the default state is absorbing such that once the issuing entity defaults, it stays in default state forever;*

2. *the transition probabilities under the risk-neutral measure are given by* $\mathbb{Q}(d_{t+1} = 0|\underline{X_{t+1}}, d_t = 0) = \exp[-\lambda_{t+1}(\underline{X_{t+1}})]$, *where* $\lambda_{t+1}(\underline{X_{t+1}})$ *is the so-called default intensity; and*
3. d_t *does not Granger-cause the risk factors* X_t *under the risk-neutral measure* $\mathbb{Q}$ *(see Monfort and Renne [2013]).*

In the following, by a notation abuse, we write λ_{t+1} or $\lambda_{t+1}(\underline{X_{t+1}})$ indifferently. This intensity plays the role of a survival rate and depends only on the factors X_{t+1} (see e.g. Jarrow, Lando, and Turnbull [1997], Duffie and Singleton [1999], Duffie and Lando [2001], or Jarrow and Yu [2001]). Using an exponential survival rate proves useful for deriving the pricing formulas. Using no-arbitrage arguments, under the risk-neutral measure, the price $\widetilde{B}(t,h)$ must be equal to its expected present value in the next period.

$$\widetilde{B}(t,h) = \exp(-r_t)\mathbb{E}^{\mathbb{Q}}\left[\widetilde{B}(t+1,h-1) \times (1-d_{t+1})\big|\underline{X_t}, d_t = 0\right] . \tag{1.15}$$

In $t+1$, the risky bond has value $\widetilde{B}(t+1,h-1)$ if the entity has not defaulted ($d_{t+1}=0$), and zero value if the entity has defaulted ($d_{t+1}=1$). To obtain general pricing formulas, we iterate backwards from the maturity date. We know that $\widetilde{B}(t+h,0)=1$ provided no default occurred. At the previous period, the price was:

$$\widetilde{B}(t+h-1,1) = \exp(-r_{t+h-1})\mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\lambda_{t+h}(\underline{X_{t+h}})\right)\big|\underline{X_{t+h-1}}, d_{t+h-1}=0\right] .$$

Using the fact that d_t does not Granger-cause X_t (see previous definition), $\widetilde{B}(t+h-1,1)$ is a function of only X_{t+h-1} and r_{t+h-1}, but not of d_{t+h-1}. At date $t+h-2$, using Equation (1.15), we obtain:

$$\widetilde{B}(t+h-2,2) = \exp(-r_{t+h-2})\mathbb{E}^{\mathbb{Q}}\left[\widetilde{B}(t+h-1,1)(1-d_{t+h-1})\big|\underline{X_{t+h-2}}, d_{t+h-2}=0\right] .$$

Conditioning first with respect to $(\underline{X_{t+h-1}}, r_{t+h-1})$ and using the fact that $\widetilde{B}(t+1,h-1)$ is function of this information only, we get:

$$\begin{aligned}\widetilde{B}(t+h-2,2) &= \exp(-r_{t+h-2})\mathbb{E}^{\mathbb{Q}}\Big[\exp(-r_{t+h-1}-\lambda_{t+h}(\underline{X_{t+h}}) \\ &\quad - \lambda_{t+h-1}(\underline{X_{t+h-1}})\big|\underline{X_{t+h-2}}, d_{t+h-2}=0\Big] .\end{aligned}$$

Iterating on maturities, the pricing formulas for defaultable bonds of any maturity h are given by:

$$\begin{aligned}\widetilde{B}(t,h) &= \mathbb{E}^{\mathbb{Q}}\left[\exp(-r_t-\lambda_{t+1}-\ldots-r_{t+h-1}-\lambda_{t+h})\big|\underline{X_t}, d_t=0\right] \\ \text{and}\quad \widetilde{R}(t,h) &= -\frac{1}{h}\log\left\{\mathbb{E}^{\mathbb{Q}}\left[\exp(-r_t-\lambda_{t+1}-\ldots-r_{t+h-1}-\lambda_{t+h})\big|\underline{X_t}, d_t=0\right]\right\} .\end{aligned} \tag{1.16}$$

This formula resembles closely the pricing formula (1.14) for risk-less zero coupon bonds. The main difference is that the term $\exp(-r_{t+i})$ is replaced with $\exp[-(r_{t+i}+\lambda_{t+i+1})]$ under the $\mathbb{Q}$-expectation. Under the risk-neutral measure, the terminal payoff is discounted by the the short-term risk-less interest rate adjusted by the default intensity. A more general case can be considered, where the recovery rate is non-zero and is random (see e.g. Duffie and Singleton [1999]). In this framework, the recovery rate is defined as a fraction of the defaultable bond price that would have been observed were the issuing entity not defaulting. This framework will prove particularly useful in Chapter 3 when we model the default and liquidity risks embedded in the interbank market interest rates.

Consider now the same econometric setup as in the previous section. We assume that the intensity is a parametric function of the risk factors governed by a set of parameters denoted by $\theta_\lambda \in \Theta_\lambda$, that is: $\lambda_t = \lambda(\underline{X_t}; \theta_\lambda)$. The entire set of parameters becomes $\theta = (\theta_r, \theta_\lambda, \theta^{\mathbb{P}}, \theta_\alpha) \in \Theta = (\Theta_r \times \Theta_\lambda \times \Theta^{\mathbb{P}} \times \Theta_\alpha)$. We can rewrite the previous formulas to incorporate the parametric econometric specification.

$$\widetilde{B}(t,h) = \mathbb{E}^{\mathbb{Q}(\theta_\alpha,\theta^{\mathbb{P}})}\left[\exp\left(-\sum_{i=0}^{h-1}\left[r(\underline{X_{t+i}};\theta_r)+\lambda(\underline{X_{t+i+1}};\theta_\lambda)\right]\right)\big|\underline{X_t}, d_t = 0\right]. \tag{1.17}$$

Note that the short-term risk-less interest rate and the default intensity are functions of factors at different dates. This implies that observing only the prices of defaultable bonds can be sufficient to identify both components. However, it is more common to model jointly the risk-less and defaultable bonds in a single framework.

1.4 Affine term structure models (ATSM)

In this section, we present the class of affine term structure models of interest rates (ATSM). These models rely on the assumptions that X_t is an affine process under the risk-neutral measure, and that the short-term risk-less interest rate (and the risk intensity) is an affine combination of X_t. After presenting the construction of ATSM, we detail two popular distributions often used in practice to build ATSMs.

1.4.1 Building affine models

In the previous section, we have expressed the prices of risk-less and risky zero-coupon bonds as the conditional expectation under the risk-neutral measure of the exponential of minus future values of the short-term interest rate for risk-less bonds, added with the risk intensity for defaultable bonds. When r_t and λ_t are affine functions of the risk factors

X_t, the prices of bonds are equal to a conditional multi-horizon Laplace transform of $(X_{t+1}, \ldots, X_{t+h})$ given $\underline{X_t}$ under the risk-neutral measure.

Definition 1.4.1. *Let us consider the econometric framework presented in the previous section. In affine term structure models, the parametric specifications of the short-term risk-less interest rate r_t and of the risk intensity λ_t are given by:*

$$\begin{aligned} r(\underline{X_t}; \theta_r) &= \delta_0 + \delta' X_t \quad \textit{with} \quad \theta_r = (\delta_0, \delta')' \\ \lambda(\underline{X_t}; \theta_\lambda) &= \lambda_0 + \lambda' X_t \quad \textit{with} \quad \theta_\lambda = (\lambda_0, \lambda')' , \end{aligned}$$

where δ_0 and λ_0 are scalars, and δ and λ are vectors of parameters of size n. In that case, the pricing formulas can be written:

$$\begin{aligned} B(t,h) &= \mathbb{E}^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})} \left[\exp\left(- \sum_{i=0}^{h-1} (\delta_0 + \delta' X_{t+i}) \right) \Big| \underline{X_t} \right] \\ &= \exp(-h\delta_0 - \delta' X_t)\, \varphi_{t,t+h}^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})} (\delta, \ldots, \delta, 0) \\ \widetilde{B}(t,h) &= \mathbb{E}^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})} \left[\exp\left(- \sum_{i=0}^{h-1} (\delta_0 + \delta' X_{t+i} + \lambda_0 + \lambda' X_{t+i+1}) \right) \Big| \underline{X_t}, d_t = 0 \right] \\ &= \exp(-h\delta_0 - h\lambda_0 - \delta' X_t)\, \varphi_{t,t+h}^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})} (\delta - \lambda, \ldots, \delta - \lambda, -\lambda) \ . \end{aligned}$$

Proposition 1.4.1. *Let X_t be an affine process under the risk-neutral measure, that is there exists deterministic functions $a^{\mathbb{Q}}(u)$ and $b^{\mathbb{Q}}(u)$ such that $\psi_t^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u) = a^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u)' X_t + b^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u)$. If the short-term interest rate and the risk intensity are given by Definition 1.4.1, the prices and interest rates of zero-coupon bonds are given by:*

$$B(t,h) = \exp\left(A_h' X_t + B_h\right) \quad \textit{and} \quad R(t,h) = -\frac{A_h}{h}' X_t - \frac{B_h}{h} \tag{1.18}$$

$$\widetilde{B}(t,h) = \exp\left(\widetilde{A}_h' X_t + \widetilde{B}_h\right) \quad \textit{and} \quad \widetilde{R}(t,h) = -\frac{\widetilde{A}_h'}{h} X_t - \frac{\widetilde{B}_h}{h} , \tag{1.19}$$

where the loadings of the previous equations can be computed recursively using $a^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u)$ and $b^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u)$, and are parametric functions of θ_α, $\theta^{\mathbb{P}}$, δ_0, λ_0, δ, and λ.

The ATSMs are called *affine* since they provide yield formulas for all maturities that are affine functions of the risk factors X_t. Given a set of parameters, the interest rates' loadings are easily computable and the pricing formulas are available in closed-form. This is very useful for estimation purposes, since we do not need to simulate the risk factors process to obtain the bond prices, which is computationally burdensome.

1.4.2 The Gaussian ATSM (GATSM)

One of the most popular ATSM is obtained by setting the conditional distribution of the risk factors X_t as a Gaussian VAR(1). We present the model in the context of direct modeling, which is often adopted in the macro-finance literature. Let us assume that the $\mathbb{P}$-dynamics of the risk factors are given by:

$$X_t = \mu + \Phi X_{t-1} + \Sigma^{1/2}\varepsilon_t \quad \text{with} \quad \varepsilon_t \overset{\mathbb{P}}{\sim} \mathcal{IIN}(0, I_n)\,. \tag{1.20}$$

The set of parameters μ, Φ, and Σ composes $\theta^{\mathbb{P}}$ that entirely defines the dynamics under the physical measure. The process $(X_t)_{t\in\mathbb{Z}}$ belongs to the affine class of processes since the conditional Laplace transform of X_{t+1} given $\underline{X_t}$ is given by:

$$\varphi_t^{\mathbb{P}}(u) = \exp\left[u'(\mu + \Phi X_t) + \frac{1}{2}u'\Sigma u\right]\,, \tag{1.21}$$

which is an exponential-affine function of X_t. Both the short-term risk-less interest rate r_t and the risk intensity λ_t are specified as in definition 1.4.1. To obtain the risk-neutral conditional distribution of the factors, we must specify a SDF. We use a slightly modified version of the form given by Equation (1.10), and specify the SDF as an exponential-affine function of the shocks of the VAR(1). More specifically, the SDF is given by:

$$M_{t,t+1} = \exp\left[-r_t - \Gamma_t'\varepsilon_{t+1} - \frac{1}{2}\Gamma_t'\Gamma_t\right]\,, \tag{1.22}$$

where Γ_t are the so-called market prices of risk, and the last term in the exponential is the Jensen convexity adjustment term such that $\mathbb{E}_t^{\mathbb{P}}[M_{t,t+1}] = \exp(-r_t)$. Relating to the notations of Equation (1.10), this implies that $\alpha_t(\underline{X_t}) = \Sigma^{-1/2}\Gamma_t$ and $\beta_t(\underline{X_t}) = -r_t - \Gamma_t'\Sigma^{-1}(\mu + \Phi X_t) + \frac{1}{2}\Gamma_t'\Sigma^{-1}\Gamma_t$. Duffee [2002] dubs *essentially affine* such a model where the market prices of risk Γ_t are specified as:

$$\Gamma_t = \gamma_0 + \gamma X_t\,, \tag{1.23}$$

where γ_0 is a vector of size n and γ is a $(n \times n)$ matrix. The SDF is therefore entirely specified by the loadings γ_0 and γ, which compose the set of parameters θ_α of the general econometric framework. It can be easily shown that the risk-neutral dynamics of the risk factors are still given by a Gaussian VAR(1) with different parameters.

$$\begin{aligned} X_t &= \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma^{1/2}\varepsilon_t^{\mathbb{Q}} \quad \text{where} \quad \varepsilon_t^{\mathbb{Q}} \overset{\mathbb{Q}}{\sim} \mathcal{IIN}(0, I_n) \\ \mu^{\mathbb{Q}} &= \mu - \Sigma^{1/2}\gamma_0 \quad \text{and} \quad \Phi^{\mathbb{Q}} = \Phi - \Sigma^{1/2}\gamma\,. \end{aligned} \tag{1.24}$$

As in the physical world, the risk-neutral dynamics of X_t is affine. Using the same notation as in the previous sections, we obtain the following pricing formulas:

$$\begin{aligned} R(t,h) &= -\frac{A_h'}{h}X_t - \frac{B_h}{h} \\ A_h &= -\delta + \Phi^{\mathbb{Q}'}A_{h-1}\,, \quad A_0 = 0\,, \\ B_h &= -\delta_0 + B_{h-1} + A_{h-1}'\mu^{\mathbb{Q}} + \frac{1}{2}A_{h-1}'\Sigma A_{h-1}\,, \quad B_0 = 0\,, \end{aligned}$$

and,

$$\begin{aligned} \widetilde{R}(t,h) &= -\frac{\widetilde{A}_h'}{h}X_t - \frac{\widetilde{B}_h}{h} \\ \widetilde{A}_h &= -\delta + \Phi^{\mathbb{Q}'}\left(\widetilde{A}_{h-1} - \lambda\right)\,, \quad A_0 = 0\,, \\ \widetilde{B}_h &= -\delta_0 - \lambda_0 + \widetilde{B}_{h-1} + \left(\widetilde{A}_{h-1} - \lambda\right)'\mu^{\mathbb{Q}} \\ &\quad + \frac{1}{2}\left(\widetilde{A}_{h-1} - \lambda\right)'\Sigma\left(\widetilde{A}_{h-1} - \lambda\right)\,, \quad B_0 = 0\,. \end{aligned}$$

Put together, the Gaussian assumption and the availability of closed-form pricing formulas explain the great popularity of this model, most notably through the simplicity of its estimation with for instance the linear Kalman filter.

However, the simplicity of the GATSM comes at the cost of unpleasant properties. First, the physical and risk-neutral dynamics of the risk factors X_t are such that both the conditional and marginal distributions of X_t are Gaussian. Therefore X_t takes values in the entire $\mathbb{R}^n$ which implies that both r_t and λ_t can take values on the entire real axis. However, first, nominal interest rates on bonds are supposed to be bounded below by 0, being constrained by the so-called *zero lower bound*. This is directly linked to the fact that holding cash is always an alternative to buying a short-term nominal bond (see Black [1995]). Under the assumption of no storage costs, the short-term risk-less interest rate r_t cannot theoretically go below 0.[2] Second, when $\lambda_t < 0$, the conditional default probability given by the transition matrix of the default dummy d_t goes below 0 which is theoretically impossible. The following section presents an ATSM with positive factors to circumvent this issue.

1.4.3 Cox, Ingersoll and Ross-type ATSM

Based on the positive factor process developed by Cox, Ingersoll, and Ross [1985] in continuous-time, the Cox, Ingersoll, Ross (CIR henceforth) model has also been widely

2. Though the exact value of the lower bound can be debatable, it is commonly agreed that the nominal interest rates are bounded below and that such a lower bound exists.

used in practice. In their seminal paper, the authors develop an affine continuous-time diffusion univariate process taking only positive values. For X_t a scalar process, the diffusion equation writes:

$$\mathrm{d}X_t = \kappa(\mu - X_t)\,\mathrm{d}t + \sigma\sqrt{X_t}\,\mathrm{d}W_t$$

where W_t is a Brownian motion, κ is the mean reversion speed parameter, and μ is the unconditional mean of X_t. In addition, if $2\kappa\mu \geqslant \sigma^2$, the zero value is precluded (Feller condition). The original interest rate model developed by Cox, Ingersoll and Ross assumes that $r_t = X_t$, so that the whole term structure of interest rates is governed by a single factor only. Multifactor CIR diffusions can also be used to build term structure models ($A_n(n)$ models in the terminology of Dai and Singleton [2000]). The multivariate CIR diffusion can be defined as:

$$\mathrm{d}X_t = \mathcal{K}(\mu - X_t)\,\mathrm{d}t + \Sigma\,\mathrm{diag}\left(\sqrt{S_t}\right)\,\mathrm{d}W_t$$

where $S_t = \omega_0 + \Omega X_t$, ω_0 being a vector of size n, Ω a $(n \times n)$ matrix, $\mathcal{K}$ is a $(n \times n)$ matrix, μ is a vector of size n, Σ is a $(n \times n)$ matrix, and W_t is a multivariate Brownian motion with independent increments. Dai and Singleton [2000] note that due to the positivity of X_t components, this multivariate process can only possess non-negative conditional and unconditional correlations between components. In terms of estimation, Ait-Sahalia and Kimmel [2010] show that it is possible to estimate yield curve affine models containing these multivariate processes with closed-form approximations of the likelihood function. In this thesis, we will consider the discrete-time equivalent of the multivariate CIR process that we present hereafter.

Gouriéroux and Jasiak [2006] derive the univariate discrete-time equivalent of the continuous-time CIR process, that they call *autoregressive gamma process* (ARG). This process possesses a positive statistical support, time-varying volatility, and belongs to the class of affine processes. We present hereafter a multivariate generalization of this ARG process.

Definition 1.4.2. *Let X_t be a $\mathbb{R}_+^n$-valued process, and Z_t be a $\mathbb{N}^n$-valued mixing random variable. X_t is said to be a vectorial autoregressive gamma process with conditionally independent elements if:*

$$\begin{aligned} \forall j \in \{1,\ldots,n\}, \quad Z_{j,t+1}\big|\underline{X_t} &\sim \mathcal{P}\left(\phi_j' X_t\right) \\ \text{and} \quad X_{j,t+1}\big|\underline{Z_{j,t+1}}, \underline{X_t} &\sim \gamma_{\nu_j + Z_{j,t+1}}(\zeta_j), \end{aligned}$$

where $\phi = [\phi_j]_{\{j=1,\ldots,n\}}$ is a $(n \times n)$ matrix and $\nu = [\nu_j]_{\{j=1,\ldots,n\}}$ and $\zeta = [\zeta_j]_{\{j=1,\ldots,n\}}$ are

size-n vectors. ν is called the shape or degree-of-freedom parameter, and ζ is the scale parameter. The conditional Laplace transform of X_{t+1} given $\underline{X_t}$ is given by:

$$\forall u < \frac{1}{\zeta}, \quad \mathbb{E}\left[\exp\left(u' X_{t+1}\right)\right] = \exp\left[\left(\frac{u \odot \zeta}{1 - u \odot \zeta}\right)' \phi' X_t - \nu' \log(1 - u \odot \zeta)\right], \qquad (1.25)$$

where $\odot$ is the Hadamard product (component-wise), while, with abuse of notations, the division and log operators work component-wise.

This multivariate process is such that the components of X_t are always positive due to the positive support of the gamma distribution, while staying in the affine class since the conditional Laplace transform of X_{t+1} given $\underline{X_t}$ is an exponential-affine function of X_t. Though the elements of X_{t+1} are conditionally independent given $\underline{X_t}$, Granger-causality between the different components of X_t can be easily taken into account if ϕ is not diagonal. The conditional independence assumption is not necessary to keep the affine property of such multivariate processes. Indeed, in Chapter 4, we introduce a new recursive method to build instantaneously correlated multivariate affine processes.

Let us now assume that the $\mathbb{P}$-dynamics of X_t are given by an autoregressive gamma process with parameters $\phi^{\mathbb{P}}$, $\zeta^{\mathbb{P}}$ and ν, that is $\theta^{\mathbb{P}} = (Vec(\phi^{\mathbb{P}})', \zeta^{\mathbb{P}'}, \nu')'$. Both the short-term risk-less interest rate and the risk-intensity are affine functions of X_t given by Definition 1.4.1. We also assume that the SDF is an exponential-affine function of X_{t+1} with constant prices of risk, that is $\alpha_t(\underline{X_t}) = \gamma_0$.

$$M_{t,t+1}(\underline{X_{t+1}}) = \exp\left[-r_t + \gamma_0' X_{t+1} - \psi_t^{\mathbb{P}}(\gamma_0)\right].$$

We therefore have $\theta_\alpha = \gamma_0$. With these assumptions, it can be shown that the risk-neutral dynamics of X_t are given by a vectorial autoregressive gamma process with parameters $\phi^{\mathbb{Q}}$, $\zeta^{\mathbb{Q}}$ and ν (see Chapter 5), such that:

$$\phi^{\mathbb{Q}} = \operatorname{diag}\left(\frac{\zeta^{\mathbb{P}}}{1 - \gamma_0 \odot \zeta^{\mathbb{P}}}\right) \times \phi^{\mathbb{P}} \quad \text{and} \quad \zeta^{\mathbb{Q}} = \operatorname{diag}\left(\frac{\zeta^{\mathbb{P}}}{1 - \gamma_0 \odot \zeta^{\mathbb{P}}}\right) \times \zeta^{\mathbb{P}},$$

where the vector of degree of freedom ν is the same under both measures. Again, assuming that $r_t = \delta_0 + \delta' X_t$ and $\lambda_t = \lambda_0 + \lambda' X_t$, bond yields at all maturities are affine functions of X_t and, using the same notations as in the previous sections, the pricing formulas are

given by:

$$\begin{aligned} R(t,h) &= -\frac{A_h}{h}' X_t - \frac{B_h}{h} \\ A_h &= -\delta + \phi^{\mathbb{Q}}\left(\frac{A_{h-1} \odot \zeta^{\mathbb{Q}}}{1 - A_{h-1} \odot \zeta^{\mathbb{Q}}}\right) \\ B_h &= -\delta_0 + B_{h-1} - \nu' \log\left(1 - A_{h-1} \odot \zeta^{\mathbb{Q}}\right) \end{aligned}$$

and,

$$\begin{aligned} \widetilde{R}(t,h) &= -\frac{\widetilde{A}_h}{h}' X_t - \frac{\widetilde{B}_h}{h} \\ \widetilde{A}_h &= -\delta + \phi^{\mathbb{Q}}\left(\frac{\left(\widetilde{A}_{h-1} - \lambda\right) \odot \zeta^{\mathbb{Q}}}{1 - \left(\widetilde{A}_{h-1} - \lambda\right) \odot \zeta^{\mathbb{Q}}}\right) \\ \widetilde{B}_h &= -\delta_0 - \lambda_0 + \widetilde{B}_{h-1} - \nu' \log\left[1 - \left(\widetilde{A}_{h-1} - \lambda\right) \odot \zeta^{\mathbb{Q}}\right] . \end{aligned}$$

Within this framework, the whole term structure of interest rates is positive at all times as long as δ and λ have positive entries, and that δ_0 and λ_0 are positive.

1.4.4 Empirical assessment of multifactor CIR models

Though we focused on the discrete-time equivalent of the CIR-process and the associated discrete-time term structure model, empirical assessments have been mostly performed on the continuous-time model. The CIR term structure models have been largely employed in the yield curve literature, with different fields of application. Brown and Dybvig [1986] investigate the implications of the CIR one-factor term structure model of interest rates. Longstaff and Schwartz [1992] develop the first two-factor general equilibrium model with CIR processes to price bonds as well as options. Generalizing the class of affine models, Dai and Singleton [2000] include CIR models with any number of factors in the affine specifications. Le, Singleton, and Dai [2010] use a back modeling approach to obtain richer historical dynamics with Dai and Singleton's affine class of models while keeping tractable pricing formulas. More recently, Backus, Chernov, and Zin [2014] apply the ARG framework to model a consumption process in an asset pricing model.

While the one-factor model has been acknowledged to provide a limited fit and flexibility, the multivariate version of the CIR process has been preferred, often with the assumption that the components of the risk factors are independent (see e.g. Duffie and Singleton [1997]). This independence assumption is not mandatory but greatly simplifies the maximum likelihood estimation procedure at the cost of over-identification (see Dai

and Singleton [2000]). Whereas a term structure model with several positive risk factors provides a better fit to the data, it is known to have difficulties providing reliable risk premia estimates (see for instance Backus, Foresi, and Telmer [2001]). Dai and Singleton [2002] mostly attribute this failure to the close relationship between the conditional mean and conditional variance of univariate CIR processes, whereas Le, Singleton, and Dai [2010] consider that using an exponential-affine SDF imposes too many constraints on the physical dynamics when using the back modeling strategy.

In Chapter 4, we provide a more general technique to create multivariate affine processes that can be instantaneously cross-correlated as well as including Granger causality. We define the class of recursive affine processes as the multivariate processes in which each factors' components $X_{j,t}$ has a conditional Laplace transform given $\underline{X_{t-1}}$ and $(X_{1,t}, \ldots, X_{j-1,t})$ which is an exponential-affine function of the conditioning variables. Using the recursive specification, it is easy to build for instance multivariate ARG processes, with a richer and more general correlation structure. We also derive the closed-form transition formulas between the physical and the risk-neutral world for this class of processes using an exponential-affine SDF. While keeping tractable dynamics under both measures, this generalization of affine processes in discrete-time helps to address some of the drawbacks of the CIR process presented above.

1.4.5 Applying the intensity-based ATSM to modeling default and liquidity risks

A first part of the risky bond literature is mostly interested in the sole modeling of credit risk in defaultable bonds or CDS. Using the ATSM framework, several authors have emphasized its flexibility and simplicity for the joint pricing of risk-less and risky bonds (see e.g. Duffee [1999], Collin-Dufresne and Solnik [2001], Chen, Cheng, and Wu [2011], Gouriéroux, Monfort, and Polimenis [2006] or Jarrow, Li, Liu, and Wu [2010]). The ATSM for the pricing of defaultable bonds is also particularly fitted to explain the so-called *credit-spread puzzle*, that is the fact that risky rates contain significant and time-varying default risk premia (see e.g. Collin-Dufresne, Goldstein, and Martin [2001], Driessen [2005] or Chen, Collin-Dufresne, and Goldstein [2009]). These default risk premia reflect the investor's aversion for bearing an asset containing default risk.

Extending the credit risk applications, some authors also considered the joint modeling and the decomposition of both credit and liquidity risks in sovereign and corporate bonds, or in CDS. Liquidity risk can represent an important pricing factor for risky bonds especially during distress periods where capital is known to be particularly scarce and asset

prices drop sharply (resp. funding liquidity and market liquidity issues, see e.g. Brunnermeier and Pedersen [2009]). Liu, Longstaff, and Mandell [2006] develop a five-factor ATSM to decompose credit and liquidity risks from swaps, and extract the associated risk premia. Longstaff, Mithal, and Neis [2005] measure default and non-default components on the CDS market and show that the default component represents the vast majority of the spread. Feldhutter and Lando [2008] jointly price Treasuries, corporate bond and swaps to decompose the swap rates into a credit component, a convenience yield (liquidity-related), and a swap idiosyncratic factor. Monfort and Renne [2014] develop a framework where the Eurozone government bond spreads with respect to German bonds contain both credit and liquidity risks. The two risks are introduced via credit and liquidity intensities. The identification is performed using credit and liquidity risk proxies, such as the KfW-bund spread. In a related approach, Filipovic and Trolle [2013] use CDS and overnight indexed swaps on LIBOR to decompose the term structure of interbank risk into a default and a non-default component.

1.5 Quadratic term structure models (QTSM)

A natural idea to deal with the positivity of interest rates is to specify the short-term interest rate r_t as a quadratic combination of Gaussian risk factors X_t. The same specification can be used for the risk intensities λ_t to provide default probabilities constrained between 0 and 1. This section presents the framework of the quadratic term structure model and its general properties in terms of pricing.

1.5.1 The quadratic framework

The Quadratic Term Structure model (QTSM) was first introduced in the 1990s to combine the nice characteristics of the affine framework of Duffie and Kan [1996] with the positivity of interest rates and the time-varying conditional volatility as in the CIR model of Cox, Ingersoll, and Ross [1985]. Ahn, Dittmar, and Gallant [2002] develop the general QTSM framework and derive its theoretical properties. Following their assumptions, we present hereafter its main characteristics.

Let us assume that the $\mathbb{P}$-dynamics and the SDF specification are respectively given by Equations (1.20) and (1.22). The risk-neutral dynamics of the risk factors are thus given by the Gaussian VAR(1) of Equation (1.24). The main difference with the GATSM is that the short-term interest rate and the risk-intensity are specified as quadratic combinations of X_t.

Definition 1.5.1. *In a quadratic term structure model, the short-term interest rate* $r(\underline{X}_t; \theta_r)$

and the risk intensity $\lambda(\underline{X_t};\theta_\lambda)$ *are specified as:*

$$r_t = \delta_0 + \delta_1' X_t + X_t' \delta_2 X_t \tag{1.26}$$
$$\lambda_t = \lambda_0 + \lambda_1' X_t + X_t' \lambda_2 X_t \,, \tag{1.27}$$

where δ_0 *and* λ_0 *are scalar values,* δ_1 *and* λ_1 *are size* n *vectors,* δ_2 *and* λ_2 *are symmetric* $(n \times n)$ *matrices, and* $\theta_r = (\delta_0, \delta_1', Vec(\delta_2)')'$ *and* $\theta_\lambda = (\lambda_0, \lambda_1', Vec(\lambda_2)')'$.

For identification purposes, δ_2 and λ_2 are imposed symmetric. It is also easy to impose that r_t and λ_t are positive valued processes: for example for r_t, a necessary and sufficient condition is that δ_2 is positive definite and that δ_0 is bounded below by $\frac{1}{4}\delta_1'\delta_2^{-1}\delta_1$. The same conditions are applicable for the intensity specification.[3]

1.5.2 Quadratic models are affine

With the short-term interest rate specification given by Equation (1.26), QTSMs do not belong to the class of affine models listed by Dai and Singleton [2000]. However, Ahn, Dittmar, and Gallant [2002] or Leippold and Wu [2007] show that closed-form pricing formulas can be preserved. This result is mainly based on the following property shown by Cheng and Scaillet [2007]:

Proposition 1.5.1. *Let* X_t *be a Gaussian VAR(1) process which* $\mathbb{P}$*-dynamics are given by Equation* (1.20)*. Let us define by* $Z_t = (X_t', Vec(X_t X_t')')'$ *the augmented vector of factors of size* $n + n^2$*. The process* $(Z_t)_{t\in\mathbb{Z}}$ *is an affine process and the conditional Laplace transform of* Z_{t+1} *given* $\underline{Z_t}$ *is given by:*

$$\begin{aligned} \forall (u,V) \in \mathcal{S}_u \times \mathcal{S}_V \,, \quad \varphi_t^{\mathbb{P}}(u,V) &= \exp\left[a_1(u,V)' X_t + X_t' a_2(u,V) X_t + b(u,v)\right] \\ &= \exp\left[a(u,V)' Z_t + b(u,V)\right] \,, \end{aligned}$$

where $\mathcal{S}_u$ *is the set of admissible vectors* u *of size* n *and* $\mathcal{S}_V$ *is the set of admissible*

3. The Wishart term structure model of Gouriéroux and Sufana [2011] based on the Wishart autoregressive process of Gouriéroux, Jasiak, and Sufana [2009] is closely related to the QTSM. In these models, the Wishart process is defined as the sum of K Gaussian VAR(1) outer-products. The short rate is specified as a linear combination of all the Wishart matrix elements. Using notations of Equation (1.26), when $\delta_1 = 0$ and the drift μ of the dynamics of X_t is equal to 0, the QTSM is equivalent to a Wishart term structure model with one degree of freedom ($K = 1$). Applications of Wishart models of interest rates include Filipovic and Teichmann [2002], Gouriéroux and Sufana [2010], Buraschi, Cieslak, and Trojani [2008], or Jin and Maheu [2013].

symmetric matrices V of size $(n \times n)$, and:

$$\begin{aligned} a_1(u,V) &= \Phi'(I_n - 2V\Sigma)^{-1}(u + 2V\mu) \\ a_2(u,V) &= \Phi' V (I_n - 2\Sigma V)^{-1}\Phi \\ b(u,V) &= u'(I_n - 2\Sigma V)^{-1}\left(\mu + \frac{1}{2}\Sigma u\right) + \mu' V (I_n - 2\Sigma V)^{-1}\mu - \frac{1}{2}\log|I_n - 2\Sigma V| \,. \end{aligned}$$

First, since the $\mathbb{Q}$-dynamics of X_t are also given by a Gaussian VAR when the exponential-affine SDF is given by Equations (1.22) and (1.23), Z_t is also an affine process under the risk-neutral measure. Second, note that both r_t and λ_t are affine functions of Z_t. Put together, the QTSM is an affine model where the total set of risk factors is given by Z_t. Using the properties of affine models presented in the previous sections, we obtain closed-form pricing formulas for both risk-less and defaultable bonds, and the yields at each maturity are affine functions of Z_t, or quadratic combinations of the Gaussian risk factors X_t.

$$\begin{aligned} R(t,h) &= -\frac{1}{h}\left(A_h + B_h' X_t + X_t' C_h X_t\right) \\ \widetilde{R}(t,h) &= -\frac{1}{h}\left(\widetilde{A}_h + \widetilde{B}_h' X_t + X_t' \widetilde{C}_h X_t\right) . \end{aligned}$$

The factor loadings can be recursively computed as:

$$\begin{aligned} A_h &= A_{h-1} - \delta_0 + B_{h-1}'(I_n - 2\Sigma C_{h-1})^{-1}\left(\mu^{\mathbb{Q}} + \frac{1}{2}\Sigma B_{h-1}\right) \\ &\qquad + \mu^{\mathbb{Q}\prime} C_{h-1}(I_n - 2\Sigma C_{h-1})^{-1}\mu^{\mathbb{Q}} - \frac{1}{2}\log|I_n - 2\Sigma C_{h-1}| \,, \quad A_0 = 0 \\ B_h &= -\delta_1 + \Phi^{\mathbb{Q}\prime}\left[(I_n - 2C_{h-1}\Sigma)^{-1}(B_{h-1} + 2C_{h-1}\mu^{\mathbb{Q}})\right] \,, \quad B_0 = 0 \\ C_h &= -\delta_2 + \Phi^{\mathbb{Q}\prime} C_{h-1}(I_n - 2\Sigma C_{h-1})^{-1}\Phi^{\mathbb{Q}} \,, \quad C_0 = 0 \\ \widetilde{A}_h &= \widetilde{A}_{h-1} - \lambda_0 - \delta_0 + (\widetilde{B}_{h-1} - \lambda_1)'(I_n - 2\Sigma(\widetilde{C}_{h-1} - \lambda_2))^{-1}\left(\mu^{\mathbb{Q}} + \frac{1}{2}\Sigma(\widetilde{B}_{h-1} - \lambda_1)\right) \\ &\qquad + \mu^{\mathbb{Q}\prime}(\widetilde{C}_{h-1} - \lambda_2)(I_n - 2\Sigma(\widetilde{C}_{h-1} - \lambda_2))^{-1}\mu^{\mathbb{Q}} - \frac{1}{2}\log\left|I_n - 2\Sigma(\widetilde{C}_{h-1} - \lambda_2)\right| \,, \quad \widetilde{A}_0 = 0 \\ \widetilde{B}_h &= -\delta_1 + \Phi^{\mathbb{Q}\prime}\left[(I_n - 2(\widetilde{C}_{h-1} - \lambda_2)\Sigma)^{-1}(\widetilde{B}_{h-1} - \lambda_1 + 2(\widetilde{C}_{h-1} - \lambda_2)\mu^{\mathbb{Q}})\right] \,, \quad \widetilde{B}_0 = 0 \\ \widetilde{C}_h &= -\delta_2 + \Phi^{\mathbb{Q}\prime}(\widetilde{C}_{h-1} - \lambda_2)(I_n - 2\Sigma(\widetilde{C}_{h-1} - \lambda_2))^{-1}\Phi^{\mathbb{Q}} \,, \quad \widetilde{C}_0 = 0 \,. \end{aligned}$$

Filipovic [2002] shows that a term structure model embedding a short rate specified as a polynomial combination of order k of Gaussian variables belongs to the class of ATSM if and only if $k \leqslant 2$. In other words, the QTSM is the maximal polynomial term structure model with closed-form pricing formulas. Reciprocally, Chen, Filipovic, and Poor [2004] show that a QTSM can only admit Orstein-Uhlenbeck diffusion for the risk factors X_t in

a continuous-time setting (corresponding to Gaussian VAR discrete-time dynamics).

1.5.3 Estimation methods

Up to this point, the composition of the risk factors X_t have been left undetailed. For estimation purposes, we must distinguish two different cases. First, if X_t is composed of components that are observed at all times, the parameter estimation can often be performed with standard techniques, such as exact or pseudo maximum likelihood, or GMM. When some components of X_t are unobserved however, the econometrician has to *(i)* estimate the parameters, and *(ii)* retrieve the values of the unobserved factors with filtering or smoothing techniques. Points *(i)* and *(ii)* usually go hand-in-hand and most parameter estimation techniques are combined with filtering procedures as well. We focus on the case when X_t is partially or entirely latent.

A natural procedure which has been largely employed in the Gaussian ATSM case with latent factors is to resort to the linear Kalman filter (see Kalman [1960] for the original algorithm, and Duan and Simonato [1999] and De Jong [2000] for term structure applications). In the QTSM however, the state-space model gathering the risk factors $\mathbb{P}$-dynamics and the observable variables measurement equations is linear-quadratic, thus non-linear algorithms must be employed.

The first non-linear filtering algorithms are based on Taylor expansions of the non-linear equations of the state-space model (namely the measurement equations which are quadratic in the QTSM case), and are called *extended Kalman filters*. The most popular is the first-order extended Kalman filter (a first-order Taylor expansion is used, see e.g. Jazwinski [1970]) whereas the second-order Kalman filter is rarely used in practice (see Athans, Wishner, and Bertolini [1968]). Among the papers estimating QTSMs, Brandt and Chapman [2003], Inci and Lu [2004], or Kim and Singleton [2012] use the first-order extended Kalman filter, whereas Lund [1997] use an iterated extended Kalman filter which is supposed to correct some undesirable features of the first-order extended Kalman filter. Baadsgaard, Nielsen, and Madsen [2000] estimate their term structure model with a version of the second-order extended Kalman filter called the *truncated* second-order filter (see e.g. Maybeck [1982]). These methods can be used in any non-linear state-space model, but rely on crude approximations when the model is highly non-linear.

A second strand of the literature prefers to use the *unscented Kalman filter* to estimate non-linear term structure models. This algorithm was developed for physics applications by Van Der Merwe and Wan [2001], Julier [2002], or Julier and Uhlmann [2004]. The

authors advocate a better performance than the extended Kalman filters both in terms of filtering, accuracy, and computational time. Christoffersen, Dorion, Jacobs, and Karoui [2014] are the first to analyze the performance of the unscented Kalman filter with respect to term structure modeling applications. Leippold and Wu [2007] are the first to use this estimation technique in a QTSM framework, followed by Chen, Cheng, Fabozzi, and Liu [2008] and Doshi, Ericsson, Jacobs, and Turnbull [2013].

Other estimation methods have been less used in the literature: Ahn, Dittmar, and Gallant [2002] use the EMM with a reprojection technique to retrieve the factors, or Andreasen and Meldrum [2013] use the sequential regression approach developed in Andreasen and Meldrum [2015]. The set of available estimation methods are usually not specifically fitted to the QTSM case.

In Chapter 2, we develop a specific method that we call the *Quadratic Kalman Filter* to increase the estimation and filtering performance in linear-quadratic state-space models in general, and in QTSMs in particular. This new method builds mainly on the affine property of the augmented process Z_t. In the following section, we detail the different applications associated with the QTSM specification.

1.5.4 Assessment and applications

The first introduction of quadratic term structure models are based on extending the CIR term structure models. Longstaff [1989] first introduces the double square-root process in continuous-time to model the short-term interest rate, and its diffusion is given by:

$$\mathrm{d}r_t = \kappa(\mu - \sqrt{r_t})\mathrm{d}t + \sigma\sqrt{r_t}\,\mathrm{d}W_t$$

This diffusion stems from a Vasicek risk factor X_t, and the short-term interest rate being specified as a quadratic function of X_t. This specification is used to build a yield curve model which is able to generate richer patterns than the original CIR model of interest rates.[4] Estimating a one-factor model on U.S. Treasury bills from 1964 to 1986, Longstaff shows evidence that the risk-free rate specified as a quadratic function of risk factors is supported by the data. Beaglehole and Tenney [1992] reassess the model of Longstaff showing that his pricing formulas only work when the underlying risk process X_t

4. In his paper, Longstaff also derives the different properties associated with this short-term interest rate process and the associated term structure. He shows that for $\mu > 0$, negative interest rates are precluded, and zero is a reflecting barrier. The speed of mean-reversion is asymmetric due to the term in $\mu - \sqrt{r_t}$. When the underlying factor X_t has a stationary distribution, then the short-term interest rate r_t has a Weibull stationary distribution. The whole term structure of interest rates can be expressed as an affine combination of r_t and $\sqrt{r_t}$, thus as a quadratic combination of X_t.

is Orstein-Uhlenbeck. Also noting the limitations of the CIR term structure model, Constantinides [1992] develops a *squared-autoregressive independent variable nominal term structure model* (SAINTS). In this model, the risk-factors X_t are independent Orstein-Uhlenbeck diffusion processes and the stochastic discount factor is specified such that the short-term interest rate is a quadratic function of X_t. This authorizes closed-form formulas for both nominal bond prices and European option prices. A one-factor model is calibrated and Constantinides shows the variety of yield curve patterns that the model can generate. Originally developed in continuous time, the model was adapted in discrete time by Realdon [2006].

From a theoretical point of view, these models all belong to the class of QTSM introduced by Ahn, Dittmar, and Gallant [2002]. After proposing a canonical form of QTSM and comparing them to the canonical ATSM of Dai and Singleton [2000], they estimate the model on U.S. zero-coupon monthly nominal interest rates from 1946 to 1991. They compare different specifications contained in the class of QTSMs and ATSMs to conclude that the QTSM provides a better fit to the interest rate data. Concerned with the assessment of the QTSM abilities, Brandt and Chapman [2003] compare the possibilities of Dai and Singleton's affine class of models and of the QTSMs to reproduce a set of moment conditions, expected excess returns (that is to forecast the behavior of the nominal term premia), and the conditional yield's volatilities. Estimating the models on U.S. nominal monthly zero-coupon yields from 1953 to 1998, they conclude in the superiority of the QTSM with respect to these criteria.

With established performance among existing term structure models and correcting undesirable features of CIR term structure models, the QTSM has been widely used for the modeling of the risk-free yield curve. A first set of applications regards the joint modeling of international risk-free rates. Inci and Lu [2004] derive a quadratic term structure model in continuous-time on U.S./German yield curves on the one hand and U.S./U.K. yield curves on the other hand. Using the covered interest rate parity, the exchange rate is a quadratic function of the risk factors and both countries' yield curves are quadratic. They show that the QTSM empirically performs better than the CIR international term structure model of Backus, Foresi, and Telmer [2001] with respect to the fitting properties (term structures and exchange rate), and the out-of-sample forecasting ability. Leippold and Wu [2007] generalize this framework for international term structure models building a class of $(m+n)$ *Multi-Currency Quadratic Models* (MCQM). They distinguish two orthogonal sets of risk factors of respective size m and n and intervening respectively in the term structures of yields and in currency risk premia. Using both Japanese and U.S. data, they analyze the empirical results of a preferred $(6+1)$ MCQM and conclude in a

high flexibility for the modeling of international term structures of interest rates when an independent currency factor is considered.

A second set of applications regards the modeling of nominal risk-free yield curve with respect to monetary policy and macro-finance issues. Since the QTSM is able to provide positivity of the entire yield curve, it is natural to assess its performance in a context of very low interest rates such as the zero lower bound period. Kim and Singleton [2012] compare the performance of a broad set of models, including both CIR models and QTSMs (and their *shadow-rate* versions, see Section 1.6.2) on the Japanese government bond yield data between 1995 and 2008. Their performance is assessed with respect to the ability of fitting data, reproducing the time-varying volatility behavior of yields, and predicting excess returns. Andreasen and Meldrum [2013] perform the same type of comparison on U.S. data from 1961 to 2013. QTSM can also include observable variables to reproduce a Taylor-type monetary policy rule, as introduced by Ang and Piazzesi [2003] in the Gaussian ATSM case. Ang, Boivin, Dong, and Loo-Kung [2011] introduce a model where the short-term nominal interest rate is specified as a linear combination of output gap and inflation rate, where the two associated loadings are time-varying. The entire set of 4 risk factors is hence composed of both the interest rate loadings and the macroeconomic variables and follow Gaussian VAR(1) dynamics. The short-term interest rate is therefore a quadratic combination of both observed and unobserved factors. Their model authorizes both the comparison of the size of the interest rate loadings, the yields forecasts, the precision of expected excess returns, and the impulse-response of an adverse inflation or GDP shock on the monetary policy reaction function. Campbell, Sunderam, and Viceira [2013] introduce a QTSM of both the nominal and real term structure of interest rates to allow for time-varying conditional covariances in both term structures.

A last application domain regards the modeling of the term structure of risky interest rates. Chen, Filipovic, and Poor [2004] first introduce the QTSM for defaultable interest rates in a continuous-time framework. A direct application of this framework concerns the modeling of CDS spreads as in e.g. Chen, Cheng, Fabozzi, and Liu [2008]. Based on the approach of Duffie, Pan, and Singleton [2000], they specify both the short-term risk-less interest rate and the default intensity as a quadratic function of Gaussian risk factors in order to obtain positive spread values. Using CDS data for more than 1,000 entities, they estimate the model on three years of daily data and are able to separate the default risk premium from the observed CDS spread. In a similar application, Doshi, Ericsson, Jacobs, and Turnbull [2013] model CDS spreads with default intensities specified as quadratic functions of macroeconomic variables and firm observable characteristics. They are able to show that the introduction of these observable explanatory variables contribute

to a better empirical performance of their no-arbitrage QTSM, both for pricing and explaining the evolution of CDS spreads. Last, Hordahl and Tristani [2013] construct a model for Euro-area sovereign spreads, taking into account the default probability of each sovereign. They impose positivity of the intensities through the quadratic specification, and include national debt-to-GDP ratios and GDP growth to explain the evolution of the spreads through time.

In Chapter 3, we investigate further this branch of applications. We consider the spreads between interbank risky and risk-less rates. These spreads contain both credit **and** liquidity risks representing *(i)* the default risk of the borrowing bank and *(ii)* the liquidity needs of the lending bank. We develop an intensity-based term structure model of interbank spreads as in Monfort and Renne [2014], but with quadratic credit and liquidity intensities to impose for the positivity of the spreads. We estimate the model on EURIBOR-OIS spreads (European data) and derive the default and liquidity interbank risk premia in the Eurozone, as well as the default probabilities under both the risk-neutral and the physical measures. We show that the decrease of spreads after the unconventional monetary policy phases in the Eurozone are mostly liquidity-related.

1.6 Modeling the zero lower bound (ZLB)

In the previous sections, we have presented several models that can allow risk-less and risky interest rates to be positive at all times. While easy to manipulate, these models miss the possibility to reproduce closely the behavior of yields at the zero lower bound (ZLB). In this situation, the short-term interest rate r_t reaches zero and stays at zero for an extensive period of time. In both the CIR model and the QTSM, the short-term interest rate bounces back to positive values immediately after reaching the zero value. In this section, we present extensions that have been operated to model the yield curve at the ZLB.

1.6.1 The ZLB modeling problem

In a seminal article, Black [1995] emphasizes that provided there are no frictions in financial markets, nominal bond interest rates should always be bounded below by zero.[5] Indeed, for an investor, holding cash-money is always an option which pays off a zero nominal interest rate. Therefore, using no-arbitrage arguments, nominal bonds should always provide at least a zero interest rate since they entail more risk than holding paper

5. If there are frictions, interest rates can reach negative values. However, there should still exist a lower bound, even though it is negative.

money. This imposes a first theoretical constraint that the nominal yield curve should always be non-negative at all times.

Second, it is usually the case that when the short-term nominal interest rates are at zero, they stay at zero for a long period. To provide more economic intuition with this stylized fact, consider that after a bad economic shock, the central bank's short-term interest rate often decreases sharply to keep inflation expectations well-anchored and/or restore GDP growth. When reaching the ZLB, the central bank is not able anymore to control inflation expectations and/or GDP growth through the nominal interest rate. In addition, putting interest rates to zero may not be sufficient for the economy to recover from the bad shock, leading to a long period of ZLB.

Third, it is of crucial importance for policy-makers and investors to know when the ZLB period will end, and to calculate the so-called *lift-off* probabilities, that is the probabilities that the future short-term interest rate goes back to positive values. To be consistent with the zero lower bound, a term structure model should be able to gather the three previous stylized facts.

Let the economy be represented by the set of risk factors X_t. The modeling problem for the ZLB is the following:

Can we find a short-term interest rate specification $r(\underline{X_t}, \theta_r)$ *and some dynamics under the physical and risk-neutral measures,* $\varphi_t^{\mathbb{P}(\theta^{\mathbb{P}})}(u)$ *and* $\varphi_t^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}(u)$ *such that:*

1. *the short-term interest rate is always non-negative (or bounded below);*
2. *the short-term interest rate can stay at the lower bound for several periods; and*
3. *the bond pricing formulas are easily computable in closed-form.*

Under AOA, the risk factors X_t possess the same statistical support under the physical and the risk-neutral measures. Imposing constraints on the risk-neutral or the physical conditional distribution of X_{t+1} given $\underline{X_t}$ is hence the same with respect to the first two points. The third point raises the question of finding an affine model able to cope with the first two points. The most commonly-used ZLB-consistent model, called the *shadow rate* model, does not consider the third point. We present this model in the following section.

1.6.2 The shadow-rate models

Shadow rate models are built to provide non-negative interest rates at all maturities and a short-term risk-less interest rate that can stay extensively at zero. Consider

an economy where the risk factors evolve under the physical measure accordingly with Equation (1.20), and the SDF is given by Equations (1.22) and (1.23). The risk-neutral dynamics of X_t are given by the Gaussian VAR(1) of Equation (1.24). Consider that there exists in the economy a so-called *shadow rate* denoted by $s_t = s(\underline{X_t}, \theta_s)$ that can take positive as well as negative values. This shadow rate can be economically interpreted as the short-term interest rate that would have prevailed were it not bounded below by zero. In most models, this shadow rate is specified as an affine function of the risk factors (Kim and Singleton [2012] however consider a two-factor quadratic shadow-rate model).

Definition 1.6.1. *Let X_t be a $\mathbb{R}^n$-valued process. The affine shadow-rate is defined by:*

$$s(\underline{X_t}; \theta_s) = \delta_0 + \delta' X_t \,, \tag{1.28}$$

with $\theta_s = (\delta_0, \delta')'$. The effective short-term interest rate $r_t = r(\underline{X_t}, \theta_r)$ is defined by:

$$r(\underline{X_t}; \theta_r) = \max\left[s(\underline{X_t}, \theta_s), 0\right] \,. \tag{1.29}$$

Here, $\theta_r = \theta_s$ and the set of parameters governing the shadow short rate is the same as the set of parameters governing the effective short-term interest rate. Note that instead of putting zero in the max function, we can estimate a lower bound parameter $\underline{r}$. Here, the short-term interest rate is bounded below by 0, and can be exactly equal to zero as long as the shadow rate evolves on the negative axis. Economically, this approach is appealing for a Taylor-rule-type interpretation: s_t would be the rate implied by a standard Taylor rule whereas r_t is the effective rate that the central bank decides to implement. Also, when s_t is high and positive, the effective short-term rate r_t essentially behaves like in a GATSM.

This concept of shadow-rate was first introduced by Black [1995] and was further developed by Krippner [2013] in a term structure framework. However, the max(•) function of Equation (1.29) implies that the short-term rate process is not affine. Therefore, closed-form formulas for bond prices at all maturities are **not** available. To compute the bond prices, we must return to the no-arbitrage general formulas given by Equation (1.14):

$$B(t,h) = \mathbb{E}^{\mathbb{Q}(\theta_\alpha, \theta^{\mathbb{P}})}\left[\exp(-r(\underline{X_t}; \theta_r) - \ldots - r(\underline{X_{t+h-1}}; \theta_r))\middle|\underline{X_t}\right] \,. \tag{1.30}$$

To calculate each bond price, one must simulate the process X_t under the risk-neutral measure, calculate the shadow rate and the effective short-term interest rate, and calculate the conditional expectation using Monte-Carlo. This is often problematic for most shadow-rate term structure models in terms of both estimation and flexibility, and including more than two latent factors can be a computationally challenging task.

Recent papers have proposed estimation methods to overcome this issue. Krippner [2012] proposes a method based on the valuation of options indexed on the shadow rate. Christensen and Rudebusch [2013] apply this estimation method on a Nelson-Siegel arbitrage-free shadow-rate term structure model on Japanese data, therefore estimating a three factor model. Kim and Singleton [2012] and Bauer and Rudebusch [2013] estimate the shadow-rate models resorting to the extended Kalman filter with Monte-Carlo simulations to calculate the measurement equations loadings. Recently, Priebsch [2013] proposes a method based on a truncation of the conditional Laplace transform of the effective short-term interest rate to compute approximated closed-form formulas in the shadow-rate framework.

In the current context of near zero short-term interest rates in the biggest economic zones, the shadow-rate models have been mostly compared to other models providing positive yield curves, such as the multivariate CIR model. Kim and Singleton [2012] benchmark the shadow-rate models comparing them to the QTSMs as well as CIR-type models. On Japanese government bond data, they find that the quadratic-shadow-rate model with two factors performs better than the others in terms of fit, conditional volatility, and prediction of bond excess returns. On an extended sample of Japanese data, Christensen and Rudebusch [2013] estimate different shadow-rate specifications and show that the filtered values of the shadow-rate vary strongly depending on the specification. They hence bring caution to the interpretation of the shadow-rate as a measure of the stance of monetary policy, since its values are not robust to the specification. The same message is held by Bauer and Rudebusch [2013] on U.S. data, where the authors compare the results of shadow-rate models with different numbers of factors, including different numbers of observable macroeconomic variables. They also use different specifications to compare the lift-off date distribution, i.e. the most probable future date at which the economy is expected to leave the zero lower bound. However, they do not purge the risk premium contained in bond prices when computing the lift-off probabilities. The knowledge of lift-off probabilities has become of crucial importance for central bankers, reflecting market expectations on the future recovery of the economy or the credibility of monetary policy. Building on a framework with only macroeconomic variables, Wu and Xia [2013] estimate a shadow-rate model to compute the effects of the unconventional monetary policies put into action by the Fed. Filipovic, Larsson, and Trolle [2013] use a linear pricing model to derive swaption pricing formulas with CIR processes in a zero lower bound context.

As emphasized previously, the existing models are not able to match simultaneously the positivity of the yield curve, the prolonged periods where the short-term interest rate is at

zero, and the closed-form bond pricing formulas. While multivariate CIR models abandon the ability to stay at zero, shadow rate models prefer to leave aside the closed-form pricing formulas. In Chapter 5, we construct a new affine process based on the autoregressive gamma processes, which is able to provide positive yields at all maturities as well as staying at zero for several periods. The conditional distribution of the risk factor process entails a zero point mass which is particularly adapted to the modeling of the zero lower bound. We present a multifactor term structure model on Japanese Government bond data.

1.7 Following chapters

To sum up, the remainder of the thesis is organized as follows.

In Chapter 2, we present a new filtering and approximate maximum likelihood techniques to estimate quadratic term structure models. This methods that we call *quadratic Kalman filter* (QKF) relies on the fact that, if X_t is a Gaussian VAR(1), the conditional Laplace transform of $Z_t = (X_t', Vec(X_t X_t')')'$ given $\underline{Z_{t-1}}$ is exponential-affine. With Monte Carlo simulations, we show that this new method outperforms its related competitors, namely the extended and unscented Kalman filters.

In Chapter 3, we apply this estimation technique to model the Eurozone interbank spreads of Euribor v.s. overnight-indexed swaps rates. Using a quadratic term structure model, we specify credit and liquidity intensities to decompose the spreads at all maturities between their different risk components. Using a no-arbitrage framework also allows us to extract the expected and risk premia components. We use this framework to provide an insight on the efficiency of the ECB's unconventional monetary policy actions targeting the interbank market.

In Chapter 4, we provide a general recursive way of specifying affine processes which components can be instantaneously correlated. These recursive affine autoregressive processes are built assuming that the conditional Laplace transform of a component $X_{j,t}$ given $\underline{X_{t-1}}$ and $(X_{1,t}, \ldots, X_{j-1,t})$ is an exponential-affine function of the conditioning variables. Still, it remains that the whole process $(X_t)_{t\in\mathbb{Z}}$ is affine and transition formulas between the physical and the risk-neutral measures with an exponential-affine SDF are derived in closed-form. We detail a wide range of possible processes that enter this class, building in particular vectorial autoregressive gamma processes which components can be instantaneously cross-correlated.

In Chapter 5, we generalize the class of autoregressive gamma processes of Gouriéroux and Jasiak [2006] to allow them to have a zero point mass. We call these processes *autoregressive gamma zero*. Using the results of the previous chapter, we use a multivariate version of these generalized ARG processes to build a term structure model consistent with the zero lower bound. In particular, our framework authorizes to derive closed-form formulas for calculating lift-off probabilities. An application is performed on Japanese government bond yields, and the model performs well in fitting both yield levels, volatilities and excess returns.

Chapter 2

Estimation and Filtering in Quadratic Factor Models

This chapter is based on the article "A Quadratic Kalman Filter" of Monfort, Renne, and Roussellet (forthcoming in the Journal of Econometrics).

Abstract

We propose a new filtering and smoothing technique for non-linear state-space models. Observed variables are quadratic functions of latent factors following a Gaussian VAR. Stacking the vector of factors with its vectorized outer-product, we form an augmented state vector whose first two conditional moments are known in closed-form. We also provide analytical formulae for the unconditional moments of this augmented vector. Our new quadratic Kalman filter (QKF) exploits these properties to formulate fast and simple filtering and smoothing algorithms. A first simulation study emphasizes that the QKF outperforms the extended and unscented approaches in the filtering exercise showing up to 70% RMSEs improvement of filtered values. Second, we provide evidence that QKF-based maximum-likelihood estimates of model parameters always possess lower bias or lower RMSEs that the alternative estimators.[1]

1. *Functions for the Quadratic Kalman Filter are implemented with the* `R`*-software and are available on the* `runmycode`*-website at* `http://www.runmycode.org/companion/view/313`.

Résumé

Pour reproduire l'évolution d'un ensemble de taux d'intérêt d'actifs financiers, l'utilisation de modèles à facteurs inobservables est fréquente. En effet, ceux-ci sont parcimonieux et expliquent une large dimension longitudinale de données par un nombre réduit de facteurs. Ils peuvent être représentés sous une forme *espace-état*, comprenant deux ensembles d'équations. Le premier contient les équations de *transition*, donnant la loi de probabilité d'évolution des facteurs inobservables. Le second contient les équations de *mesure*, qui relient les données observées à chaque date aux facteurs inobservables et à un bruit de mesure. Aussi bien les équations de transition que les équations de mesure sont paramétrées. Les deux enjeux de l'économètre sont *(1)* d'estimer les paramètres des deux ensembles d'équations, et *(2)* d'obtenir une évaluation des facteurs *filtrés* et *lissés*, i.e. de déduire la valeur optimale des facteurs sachant les observations disponibles.

Le filtre de Kalman [1960] standard est la méthode d'estimation et de filtrage optimale si les équations de transition sont décrites par un modèle VAR linéaire gaussien, et les variables observables sont des fonctions affines des facteurs auxquelles sont additionnés des bruits de mesure gaussiens. Dans ce cas, aussi bien la loi conditionnelle des observables étant donné leur passé que la loi des facteurs inobservables conditionnellement à leur passé sont gaussiennes. Les paramètres du modèle peuvent être estimés aisément par maximum de vraisemblance, et les facteurs filtrés sont obtenus par régression linéaire. Si l'un des deux ensembles d'équations devient non-linéaire, il n'est plus possible d'utiliser le filtre de Kalman standard comme méthode optimale, car les distributions conditionnelles précédentes sont difficilement exprimables en formules fermées. Une première possibilité est d'utiliser des méthodes simulées (Monte Carlo, méthodes bayesiennes), permettant d'approximer précisément les distributions conditionnelles précédentes, à un coût computationnel élevé. Une seconde possibilité est d'utiliser des filtres dits *déterministes*, comme les filtres de Kalman étendus ou inodore.[2] Ces filtres approximent les non-linéarités des équations de transition et de mesure à l'aide de développements de Taylor, permettant une estimation simple et rapide, au prix de la disparition des propriétés asymptotiques des estimateurs.

Dans ce chapitre, nous développons une nouvelle méthode d'estimation et de filtrage pour les modèles espace-état dans lesquels les équations de transition sont données par un VAR linéaire gaussien, et les variables observables sont des combinaisons linéaires-quadratiques des facteurs inobservables, auxquelles sont additionnés des bruits de mesure gaussiens. Notre nouvel algorithme, le *Filtre de Kalman Quadratique* (QKF), est dérivé

2. Traduction littérale de l'anglais *Unscented Kalman Filter*.

d'une propriété fondamentale montrée par Cheng and Scaillet [2007] : lorsque la dynamique d'un vecteur de facteurs est donnée par un VAR linéaire gaussien, le processus multivarié constitué de l'empilement de ce vecteur et de tous les produits possibles de ses composantes est un processus affine (au sens de composé auto-régressif, voir Darolles, Gouriéroux, and Jasiak [2006]). Ce nouveau processus empilé appelé *vecteur de facteurs augmenté* possède ainsi une espérance et une matrice de variance-covariance conditionnelles à son passé exprimables en formules fermées sous la forme de fonctions affines de son passé. Une première partie de ce chapitre est dédiée à la dérivation des deux premiers moments conditionnels et marginaux du vecteur de facteurs augmenté.

Dans un second temps, nous transformons le modèle espace-état pour lui redonner une forme linéaire. En effet, les équations de mesure, qui étaient des combinaisons linéaires-quadratiques des facteurs initiaux, deviennent des combinaisons affines des éléments du vecteur de facteurs augmenté. Les équations de transition peuvent être exprimées sous une forme VAR, dans laquelle les chocs sont des différences de martingale de moyenne nulle et de variance unitaire. En approximant la loi conditionnelle de ces chocs étant donné le passé des facteurs par une loi gaussienne, on obtient un *modèle espace-état augmenté*, constitué d'équations de transitions du vecteur de facteurs augmenté dont la dynamique est donnée par un VAR gaussien hétéroscédastique, et dont les équations de mesure sont données par des fonctions affines du vecteur de facteurs augmenté. On peut ainsi appliquer directement l'algorithme du filtre de Kalman à ce nouveau modèle espace-état, constituant notre algorithme du QKF.

Une dernière section est dédiée à l'analyse des performances empiriques de ce nouvel algorithme en comparaison des filtres déterministes existants mentionnés ci-avant. L'analyse empirique s'appuie sur des simulations du modèle espace-état linéaire-quadratique. On considère un modèle de référence où le processus latent et les variables observables sont des scalaires. L'équation de transition est donnée par un AR(1) gaussien de moyenne nulle et de variance conditionnelle normalisée à 1. On fait varier le paramètre auto-régressif pour considérer différents cas. L'équation de transition est une fonction quadratique du facteur latent plus un bruit de mesure. On paramétrise l'équation de mesure telle que la variance marginale de l'observable soit unitaire. Deux paramètres contrôlent la part de la variance de l'observable attribuée à la variance du facteur par rapport à celle du bruit de mesure, et la part de la variance attribuée au terme linéaire du facteur par rapport au terme quadratique. On totalise ainsi trois paramètres (un pour l'équation de transition et deux pour l'équation de mesure) permettant de considérer un ensemble vaste de modèles espace-état linéaire-quadratiques.

Le premier exercice consiste à simuler le modèle de référence sur des trajectoires longues (un million d'observations) pour différentes valeurs des paramètres. On calcule les facteurs filtrés par les différents algorithmes en leur indiquant la vraie valeur des paramètres utilisés pour la simulation. On compare ensuite les erreurs moyennes des facteurs filtrés par chacun des algorithmes par rapport aux facteurs simulés. Dans tous les cas considérés, le QKF produit des erreurs de filtrage inférieures ou égales à ses concurrents. Les différences sont particulièrement frappantes lorsque les paramètres du modèle attribuent une forte part de la variance de l'observable au terme quadratique : le QKF produit des erreurs de filtrage moyennes jusqu'à 70% plus petites que celles de ses concurrents.

Le second exercice consiste à simuler le modèle de référence un grand nombre de fois pour un jeu de paramètres donné, sur des trajectoires courtes (200 observations). On utilise les différents algorithmes de filtrage pour estimer les paramètres du modèle par quasi-maximum de vraisemblance. L'utilisation de plusieurs trajectoires courtes simulées par jeu de paramètres permet d'obtenir la moyenne et variance relatives à la distribution à distance finie des estimateurs de chaque algorithme de filtrage. Nos résultats montrent la meilleure performance du QKF face à ses compétiteurs, en produisant des estimations moins biaisées et des erreurs quadratiques moyennes plus petites dans la plupart des cas considérés.

2.1 Introduction

This paper proposes a new discrete-time Kalman filter for state-space models where the transition equations are linear and the measurement equations are quadratic. We call this method the *Quadratic Kalman Filter* (QKF). While this state-space model have become increasingly popular in the applied econometrics literature, existing filters are either highly computationally intensive, or not specifically fitted to the linear-quadratic case. We begin by building the augmented vector of factors stacking together the latent vector and its vectorized outer-product. To the best of our knowledge, this paper is the first to derive analytically and provide closed-form formulae of both the conditional and the unconditional first-two moments of this augmented vector.[3] Using these moments, the transition equations of the augmented vector are expressed in an affine form. Similarly, the measurement equations are rewritten as affine functions of the augmented vector of factors. We thus obtain an *augmented state-space model* that is fully linear.

We perform the derivation of the QKF filtering and smoothing algorithms by applying the linear Kalman algorithms to the augmented state-space model. To do so, we approximate the conditional distribution of the augmented vector of factors given its own past by a multivariate Gaussian distribution. Since no adaptation of the linear algorithm is needed, the QKF combines simplicity of implementation and fast computational speed. We apply the same method for the derivation of the Quadratic Kalman Smoothing algorithm (QKS). Indeed, since the QKF and QKS requires no simulations, it represents a convenient alternative to particle filtering.

To compare our filter with the popular existing *traditional filters* (see Tanizaki [1996]), namely the first- and second-order extended and the unscented Kalman filters, we implement a Monte-Carlo experiment. In order to explore a broad range of cases, we build a benchmark state-space model with different values for *(i)* the persistence of the latent process, *(ii)* the importance of noise variance in the observable, and *(iii)* the importance of quadratic terms in the observables. RMSE measures are computed in each case. We compare the filters with respect to two different criteria: filtering, i.e. retrieving latent factors precisely from a fixed set of parameters, and parameter estimation, i.e. the capacity to estimate the state-space model parameters.

First, these computations provide evidence of the superiority of the QKF filtering over its competitors in all cases. When the measurement equations are fully quadratic, the QKF is

3. Buraschi, Cieslak, and Trojani [2008] provide formulae of conditional first-two moments for the specific case of centred Wishart processes.

the only filter able to capture the non-linearities and to produce time-varying evaluations of the latent factors. This results in up to 70% lower RMSEs for the QKF compared to the other filters, all cases considered. For measurement equations with both linear and quadratic terms, the QKF still results – to a smaller extent – in lower filtering RMSEs. These results are robust to the persistence degree of the latent process and the size of the measurement noise. Also, we emphasize that the first-order extended Kalman filter performs particularly poorly in some cases and should therefore be discarded for filtering in the linear-quadratic model.

Second, the QKF-based maximum-likelihood estimates of model parameters always possess lower bias or lower RMSEs that the alternative estimators. We provide evidence that this superiority is robust to the degree of persistence of the latent process, to the degree of linearity of the measurement equations, and to the size of the measurement errors. We conclude that the QKF results in the best bias/variance trade-off for the quasi maximum likelihood estimation.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the non-linear filtering literature and its applications. Section 3 presents the state-space model and builds the QKF. Section 4 performs a comparison of the QKF with popular competitors using Monte-Carlo experiments. Section 5 concludes. Proofs are gathered in the Appendices.

2.2 Literature review

The existing *traditional* non-linear filters use linearization techniques to transform the state-space model. First and second-order extended Kalman filters build respectively on first and second-order Taylor expansions of transition and measurement equations. The first-order extended Kalman filter is extensively covered in Anderson and Moore [1979]. To reduce the errors linked to the first-order approximations, Athans, Wishner, and Bertolini [1968] develop a second-order extended Kalman filter.[4] In the general non-linear case, both methods require numerical approximations of gradients and Hessian matrices, potentially increasing the computational burden.[5] The unscented Kalman filter belongs more to the class of *deterministic density estimators*, and was originally implemented as an alternative to the previous techniques for applications in physics. It is a derivative-

4. This method is treated in continuous and continuous-discrete time in Maybeck [1982]. Bar-Shalom, Kirubarajan, and Li [2002] propose a complete description of this second-order filter.

5. Gustafsson and Hendeby [2012] build a derivative-free version of the second-order extended Kalman filter which avoids issues due to numerical approximations, but shows a similar computational complexity.

free method which is shown to be computationally close to the second-order extended Kalman filter in terms of complexity (see Julier, Uhlmann, and Durrant-Whyte [2000] or Julier and Uhlmann [2004]).[6] Whereas many other filters exist, both the extended and unscented filters have been the most widely used in recent econometric applications.

We consider here a specification in which the transition equations are affine and the measurement equations are quadratic. This quadratic framework is particularly suited to numerous dynamic economic models. While first-order linearization is standard and largely employed in the dynamic stochastic general equilibrium (DSGE) literature, the algorithm we develop is fitted to exploit second-order approximations.[7]

As for finance, an important field of applications of our filter is the modelling of term structures of interest rates.[8] The standard and popular Gaussian affine term-structure model (GATSM) provides yields which are affine combinations of dynamic linear autoregressive factor processes. As these models include latent factors, the linear Kalman filter[9] has gained overwhelming popularity compared to other estimation techniques (see e.g. Duan and Simonato [1999] or Joslin, Singleton, and Zhu [2011]). A natural extension of the GATSM is to assume that yields are quadratic functions of factor processes. The bulk of the papers using QTSMs considers the dynamics of government-bond yield curves (e.g. Leippold and Wu [2007] and Kim and Singleton [2012]). QTSMs have also been shown to be relevant to model the dynamics of positive risk intensities and their implied term structures: while default intensities are considered in the credit-risk literature (see e.g. Doshi, Ericsson, Jacobs, and Turnbull [2013] and Dubecq, Monfort, Renne, and Roussellet [2014]), mortality intensities have also been modelled in this framework (Gouriéroux and Monfort [2008]). In order to estimate QTSMs involving latent variables, a wide range of techniques are considered in the existing literature: Inci and Lu [2004] and Li and Zhao [2006] use the extended Kalman filter, Leippold and Wu [2007], Doshi, Ericsson, Jacobs, and Turnbull [2013] or Chen, Cheng, Fabozzi, and Liu [2008] employ the unscented Kalman filter and Andreasen and Meldrum [2011] opt for the particle filter.[10] Finally, Dubecq, Monfort, Renne, and Roussellet [2014] use the QKF filter that is

6. A general version of the algorithm is provided in Appendix 2.G.

7. Our approach could for instance be exploited to estimate the standard asset-pricing model of Burnside [1998] considered e.g. by Collard and Juillard [2001].

8. See Dai and Singleton [2003] for a survey of interest-rate term-structure modelling literature.

9. See Kalman [1960] for the original linear filter derivation. Properties are developed in e.g. Harvey [1991] or Durbin and Koopman [2012].

10. Ahn, Dittmar, and Gallant [2002] resort to the efficient method of moments (EMM). However, Duffee and Stanton [2008] show that, compared to maximum likelihood approaches, EMM has poor finite sample properties when data are persistent, a typical characteristic of bond yields. Moreover, while EMM is used to estimate model parameters, it does not directly provide estimates of the latent factors. Gallant and Tauchen [1998] however propose a reprojection method to recover latent variables after

developed hereafter.

The quadratic state-space framework that we consider in the present paper is also well-suited to work with Wishart processes. These processes have been used in various empirical-finance studies. In most cases, they are employed in multivariate stochastic volatility models (see e.g. Jin and Maheu [2013] or Rinnergschwentner, Tappeiner, and Walde [2011]). Wishart processes have also been exploited in several QTSMs (Filipovic and Teichmann [2002], Gouriéroux, Monfort, and Sufana [2010], Gouriéroux and Sufana [2011]).

2.3 The Quadratic Kalman Filter (QKF) and Smoother (QKS)

2.3.1 Model and notations

We are interested in a state-space model with affine transition equations and quadratic measurement equations. We consider the following model involving a latent (or state) variable X_t of size n and an observable variable Y_t of size m. X_t might be only partially latent, that is, some components of X_t might be observed.

Definition 2.3.1. *The linear-quadratic state-space model is defined by:*

$$X_t = \mu + \Phi X_{t-1} + \Omega\varepsilon_t \tag{2.1a}$$

$$Y_t = A + BX_t + \sum_{k=1}^{m} e_k X_t' C^{(k)} X_t + D\eta_t. \tag{2.1b}$$

where ε_t and η_t are independent Gaussian white noises with unit variance-covariance matrices, $\Omega\Omega' = \Sigma$ and $DD' = V$. e_k is the column selection vector of size m whose components are 0 except the k^{th} one, which is equal to 1. μ and Φ are respectively a n-dimensional vector and a square matrix of size n. A and B are respectively a vector of size m and a $(n \times m)$ matrix. All $C^{(k)}$'s are without loss of generality square symmetric matrices of size $m \times m$.

This formulation of the state-space model is for instance the typical quadratic term structure framework explored by Ahn, Dittmar, and Gallant [2002]. A component-by-component version of the measurement equations (2.1b) is:

$$Y_{t,k} = A_k + B_k X_t + X_t' C^{(k)} X_t + D_k \eta_t, \quad \forall k \in \{1, \dots, m\}, \tag{2.2}$$

having estimated the model parametrization by means of EMM.

where $Y_{t,k}$, A_k, B_k, D_k are respectively the kth row of Y_t, A, B, and D. Note that μ, Φ, Σ, A, B, $C^{(k)}$, and D might be functions of $(Y_{t-1}, Y_{t-2}, \ldots)$, that are the past values of the observable variables. The quadratic measurement equations notably imply that the observable variables feature conditional heteroskedasticity.

Our objective is twofold: *(i)* filtering and smoothing of X_t, which consist in retrieving the values of X_t conditionally on, respectively, past and present values of Y_t, and all the observed values of $(Y_t)_{t=1,\ldots,T}$; and *(ii)* estimation of the parameters appearing in μ, Φ, Ω, A, B, $C^{(k)}$, D. Note that Ω and D are defined up to the right multiplication by an orthogonal matrix. These matrices can be fixed by imposing $\Omega = \Sigma^{1/2}$ and $D = V^{1/2}$.[11]

Throughout the paper, we use the following notations. At date t, past observations of the observed vector are denoted by $\underline{Y_t} = \{Y_t, Y_{t-1}, Y_{t-2}, \ldots, Y_1\}$, and for any process W_t:

$$\begin{aligned} W_{t|t} &\equiv \mathbb{E}\left[W_t|\underline{Y_t}\right], & P^W_{t|t} &\equiv \mathbb{V}\left[W_t|\underline{Y_t}\right], \\ W_{t|t-1} &\equiv \mathbb{E}\left[W_t|\underline{Y_{t-1}}\right], & P^W_{t|t-1} &\equiv \mathbb{V}\left[W_t|\underline{Y_{t-1}}\right], \\ \mathbb{E}_{t-1}(W_t) &\equiv \mathbb{E}\left[W_t|\underline{W_{t-1}}\right]. & \mathbb{V}_{t-1}(W_t) &\equiv \mathbb{V}\left[W_t|\underline{W_{t-1}}\right]. \end{aligned}$$

We also introduce the notation $M_{t|t-1} \equiv \mathbb{V}\left[Y_t|\underline{Y_{t-1}}\right]$ and:

$$Z_t = \left(X_t', Vec(X_t X_t')'\right)'.$$

Z_t is the vector stacking the components of X_t and its vectorized outer-product. This vector Z_t, called the **augmented state vector** (see Cheng and Scaillet [2007]), will play a key role in our algorithms. We first study the conditional moments of this vector given past information.

2.3.2 Conditional moments of Z_t

It can be shown (see Bertholon, Monfort, and Pegoraro [2008]) that when μ, Φ and Σ do not depend on $\underline{Y_{t-1}}$, the process (Z_t) is Compound Autoregressive of order 1 –or Car(1)–, that is to say, the conditional log-Laplace transform, or cumulant generating function defined by:

$$\log \varphi_t(u) = \log \mathbb{E}\left[\exp(u' Z_t)|\underline{Z_{t-1}}\right]$$

is affine in Z_{t-1}. This result crucially depends on the assumption stating that the matrix Ω

11. Ω and D can be rectangular when Σ or V are not of full-rank.

(hence Σ) does not depend on past values of X_{t-1}.[12] This affine property implies, in particular, that the conditional expectation $\mathbb{E}_{t-1}(Z_t)$ and the conditional variance-covariance matrix $\mathbb{V}_{t-1}(Z_t)$ of Z_t given $\underline{Z_{t-1}}$ are affine functions of Z_{t-1}. Moreover, $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ have closed-form expressions given in the following proposition.

Proposition 2.3.1. $\mathbb{E}_{t-1}(Z_t) = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}$ *and* $\mathbb{V}_{t-1}(Z_t) = \widetilde{\Sigma}_{t-1}$, *where:*

$$\widetilde{\mu} = \begin{pmatrix} \mu \\ Vec(\mu\mu' + \Sigma) \end{pmatrix}, \quad \widetilde{\Phi} = \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array}\right)$$

$$\widetilde{\Sigma}_{t-1} \equiv \widetilde{\Sigma}(Z_{t-1}) = \left(\begin{array}{c|c} \Sigma & \Sigma\Gamma'_{t-1} \\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma'_{t-1} + \left(I_{n^2} + \Lambda_n\right)(\Sigma \otimes \Sigma) \end{array}\right)$$

$$\Gamma_{t-1} = I_n \otimes (\mu + \Phi X_{t-1}) + (\mu + \Phi X_{t-1}) \otimes I_n$$

Λ_n *being the* $n^2 \times n^2$ *matrix, partitioned in* $(n \times n)$ *blocks, such that the* (i, j) *block is* $e_j e_i'$ *(see Appendix 2.B for* Λ_n *properties).*

Proof. See Appendix 2.C. ■

Note that $\widetilde{\Sigma}_{t-1}$ is a $n(n+1)\times n(n+1)$ matrix whereas $\widetilde{\Sigma}(\bullet)$ is a $\mathbb{R}^{n(n+1)} \longmapsto \mathcal{M}_{n(n+1)\times n(n+1)}$ function, $\mathcal{M}_{n(n+1)\times n(n+1)}$ being the space of symmetric positive definite matrices of size $n(n+1)$. If μ, Φ, and Σ are functions of $\underline{Y_{t-1}}$, Proposition 2.3.1 still holds replacing $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ by $\mathbb{E}(Z_t|\underline{Z_{t-1}}, \underline{Y_{t-1}})$ and $\mathbb{V}(Z_t|\underline{Z_{t-1}}, \underline{Y_{t-1}})$, respectively.

$\widetilde{\Sigma}_{t-1}(\bullet)$ is clearly a quadratic function of X_{t-1} and an affine function of Z_{t-1}, denoted by $\widetilde{\Sigma}(Z_{t-1})$ (Proposition 2.3.1). In the filtering algorithm, we have to compute $\mathbb{E}[\widetilde{\Sigma}(Z_{t-1})|\underline{Y_{t-1}}]$. This quantity is easily computable as $\widetilde{\Sigma}(Z_{t-1|t-1})$ only once the affine form of the function $\widetilde{\Sigma}(Z)$ is explicitly available. Proposition 2.3.2 details this affine form.

Proposition 2.3.2. *We denote* $\widetilde{\Sigma}_{t-1}^{(i,j)}$ *for* i *and* j *being* $\{1, 2\}$ *the* (i, j) *block of* $\widetilde{\Sigma}_{t-1}$. *Each*

12. This property is of particular use for term-structure modelling, allowing for closed-form formulas to price long-term bonds in QTSM of the form of Ahn, Dittmar, and Gallant [2002].

block of $\widetilde{\Sigma}$ *is affine in* Z_{t-1} *and we have:*

$$\begin{aligned}
Vec\left(\widetilde{\Sigma}_{t-1}^{(1,1)}\right) &= Vec(\Sigma) \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(1,2)}\right) &= [\Sigma \otimes (I_{n^2} + \Lambda_n)]\,[Vec(I_n) \otimes I_n]\left\{\mu + \widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(2,1)}\right) &= [(I_{n^2} + \Lambda_n) \otimes \Sigma]\,(I_n \otimes \Lambda_n)\,[Vec(I_n) \otimes I_n]\left\{\mu + \widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(2,2)}\right) &= [(I_{n^2} + \Lambda_n) \otimes (I_{n^2} + \Lambda_n)]\,[(I_n \otimes \Lambda_n \otimes I_n)(Vec(\Sigma) \otimes I_{n^2})]\left\{\mu \otimes \mu + \widetilde{\Phi}_2 Z_{t-1}\right\} \\
&\quad + [I_{n^2} \otimes (I_{n^2} + \Lambda_n)]\,Vec(\Sigma \otimes \Sigma)
\end{aligned} \tag{2.3}$$

Where $\widetilde{\Phi}_1$ *and* $\widetilde{\Phi}_2$ *are respectively the upper and lower blocks of* $\widetilde{\Phi}$ *and* Λ_n *is defined as in Proposition 2.3.1. This particularly implies:*

$$Vec\left[\mathbb{V}_{t-1}(Z_t)\right] = Vec\left[\widetilde{\Sigma}(Z_{t-1})\right] = \nu + \Psi Z_{t-1},$$

where ν *and* Ψ *are permutations of the multiplicative matrices in Equation 2.3, and are detailed in Appendix 2.D.*

Proof. See Appendix 2.D. ■

These results extend the computations of Buraschi, Cieslak, and Trojani [2008]. While these authors express the conditional first-two moments of a central Wishart autoregressive process (see Appendix C. of Buraschi, Cieslak, and Trojani [2008]), we derive the first two-conditional moments of our augmented vector Z_t in a more general case (where $\mu \neq 0$).

2.3.3 Unconditional moments of Z_t and stationarity conditions

The analytic derivation of the first two unconditional moments of Z_t can, in particular, be exploited to initialize the filter. In the following subsection, we consider the standard case where μ, Φ and Σ are not depending on $\underline{Y_{t-1}}$. If the eigenvalues of Φ have a modulus strictly smaller than 1, the process (X_t) is strictly and, a fortiori, weakly stationary. Since Z_t is a function of X_t the same is true for the process (Z_t). The unconditional or stationary distribution of X_t is the normal distribution $\mathcal{N}(\mu^u, \Sigma^u)$ where:

$$\mu^u = (I - \Phi)^{-1}\mu \quad \text{and} \quad \Sigma^u = \Phi\Sigma^u\Phi' + \Sigma \tag{2.4}$$

Equivalently, we can write $Vec(\Sigma^u) = (I - \Phi \otimes \Phi)^{-1}\,Vec(\Sigma)$. The stationary distribution of Z_t is the image of $\mathcal{N}(\mu^u, \Sigma^u)$ by the function f defined by $f(x) = (x', Vec(xx')')'$. In order

to initialize our filter, we need the first two moments of this stationary distribution, that is to say the unconditional expectation $\mathbb{E}(Z_t)$ and the unconditional variance-covariance matrix $\mathbb{V}(Z_t)$ of Z_t.

Proposition 2.3.1 gives the expressions of the conditional moments of Z_t given Z_{t-1}, namely $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$. In general, the sole knowledge of these conditional moments does not allow to compute the unconditional moments $\mathbb{E}(Z_t)$ and $\mathbb{V}(Z_t)$. However, it is important to note that, here, the affine forms of $\mathbb{E}_{t-1}(Z_t)$ and $\mathbb{V}_{t-1}(Z_t)$ make these computations feasible analytically. More precisely, starting from any value Z_0 of Z_t at $t = 0$, the sequence $[\mathbb{E}(Z_t)', Vec(\mathbb{V}(Z_t))']'$, for $t = 1, 2, \ldots$ satisfies a first-order linear difference equation defined in the following proposition.

Proposition 2.3.3. *We have:*

$$\begin{bmatrix} \mathbb{E}(Z_t) \\ Vec[\mathbb{V}(Z_t)] \end{bmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \nu \end{pmatrix} + \Xi \begin{bmatrix} \mathbb{E}(Z_{t-1}) \\ Vec[\mathbb{V}(Z_{t-1})] \end{bmatrix} \quad \text{where} \quad \Xi = \begin{pmatrix} \widetilde{\Phi} & 0 \\ \Psi & \widetilde{\Phi} \otimes \widetilde{\Phi} \end{pmatrix}. \tag{2.5}$$

where $\widetilde{\mu}$ and $\widetilde{\Phi}$ are defined in Proposition 2.3.1, and ν, Ψ are defined according to Proposition 2.3.2.

Proof. See Appendix 2.E. ■

This linear difference equation is convergent since all the eigenvalues of Ξ have a modulus strictly smaller than 1. This is easily verified: Ξ is block triangular thus its eigenvalues are the eigenvalues of $\widetilde{\Phi}$ and $\widetilde{\Phi} \otimes \widetilde{\Phi}$. Using the same argument, $\widetilde{\Phi}$ has the same eigenvalues as Φ and $\Phi \otimes \Phi$ (see Proposition 2.3.1). Moreover the eigenvalues of the Kronecker product of two square matrices are given by all the possible products of the first and second matrices eigenvalues. Therefore, since Φ has eigenvalues inside the unit circle, so have $\widetilde{\Phi}$, $\widetilde{\Phi} \otimes \widetilde{\Phi}$, and Ξ.

We deduce that the unconditional expectation $\widetilde{\mu}^u$ and variance-covariance $\widetilde{\Sigma}^u$ (or rather $Vec(\widetilde{\Sigma}^u)$) of Z_t are the unique solutions of:

$$\begin{pmatrix} \widetilde{\mu}^u \\ Vec\left(\widetilde{\Sigma}^u\right) \end{pmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \nu \end{pmatrix} + \begin{pmatrix} \widetilde{\Phi} & 0 \\ \Psi & \widetilde{\Phi} \otimes \widetilde{\Phi} \end{pmatrix} \begin{pmatrix} \widetilde{\mu}^u \\ Vec\left(\widetilde{\Sigma}^u\right) \end{pmatrix}. \tag{2.6}$$

We get the following corollary:

Corollary 2.3.3.1. *The unconditional expectation $\widetilde{\mu}^u$ and variance-covariance $\widetilde{\Sigma}^u$ of Z_t are given by:*

$$\begin{aligned} \widetilde{\mu}^u &= \left(I_{n(n+1)} - \widetilde{\Phi}\right)^{-1} \widetilde{\mu} \\ Vec\left(\widetilde{\Sigma}^u\right) &= \left(I_{n^2(n+1)^2} - \widetilde{\Phi} \otimes \widetilde{\Phi}\right)^{-1} (\nu + \Psi \widetilde{\mu}^u) \\ &= \left(I_{n^2(n+1)^2} - \widetilde{\Phi} \otimes \widetilde{\Phi}\right)^{-1} Vec\left[\widetilde{\Sigma}\left(\widetilde{\mu}^u\right)\right], \end{aligned}$$

where $\widetilde{\mu}$ and $\widetilde{\Phi}$ are defined in Proposition 2.3.1.

These closed-form expressions of $\widetilde{\mu}^u$ and $\widetilde{\Sigma}^u$ will make easy the initialization of our algorithms. Note that the computation of $Vec[\widetilde{\Sigma}(\widetilde{\mu}^u)]$ requires the explicit affine expression of Appendix 2.D given by $Vec[\widetilde{\Sigma}(\widetilde{\mu}^u)] = \nu + \Psi\widetilde{\mu}^u$.

2.3.4 Conditionally Gaussian approximation of (Z_t)

Proposition 2.3.1 shows that Z_t satisfies:

$$Z_t = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1} + \widetilde{\Omega}(Z_{t-1})\xi_t, \tag{2.7}$$

where $\widetilde{\Omega}(Z_{t-1})$ is such that $\widetilde{\Omega}(Z_{t-1})\widetilde{\Omega}(Z_{t-1})' = \widetilde{\Sigma}(Z_{t-1})$ and (ξ_t) is a martingale difference process, with a unit conditional variance-covariance matrix (i.e. $\mathbb{E}_{t-1}(\xi_t) = 0$ and $\mathbb{V}_{t-1}(\xi_t) = I_{n(n+1)}$). In the sequel, we approximate the process (ξ_t) by a Gaussian white noise. In the standard case where μ, Φ and Σ are time-invariant, the process Z_t^*, $t = 0, 1, \ldots$, defined by $Z_0^* \sim \mathcal{N}(\widetilde{\mu}^u, \widetilde{\Sigma}^u)$ and

$$Z_t^* = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}^* + \widetilde{\Omega}(Z_{t-1}^*)\xi_t^*,$$

where (ξ_t^*) is a standard Gaussian white noise, has exactly the same second-order properties as process (Z_t). This statement is detailed in Proposition 2.3.4.

Proposition 2.3.4. *If μ, Φ and Σ are time-invariant, the processes Z_t and Z_t^* have the same second-order properties, i.e. the same means, variances, instantaneous covariances, serial correlations, and serial cross-correlations.*

Proof. It is easy to check that, for both processes, the mean, variance-covariance matrix, and lag-h covariance matrix are respectively $\widetilde{\mu}^u$, $\widetilde{\Sigma}^u$ and $\widetilde{\Phi}^h\widetilde{\Sigma}^u$. ■

2.3.5 The filtering algorithm

Using the augmented state vector Z_t we can rewrite the state-space model of Definition 2.3.1 as an augmented state-space model.

Definition 2.3.2. *The augmented state-space model associated with the linear-quadratic state-space model is defined by:*

$$\begin{cases} Z_t &= \widetilde{\mu} + \widetilde{\Phi} Z_{t-1} + \widetilde{\Omega}_{t-1}\xi_t, \\ Y_t &= A + \widetilde{B} Z_t + D\eta_t, \end{cases} \tag{2.8}$$

where η_t, A, and D are defined as in Definition 2.3.1, $\widetilde{\Omega}_{t-1}$ is such that $\widetilde{\Omega}_{t-1}\widetilde{\Omega}'_{t-1} = \widetilde{\Sigma}_{t-1}$, and $\widetilde{\mu}$, $\widetilde{\Phi}$, are defined as in Proposition 2.3.1. Matrix $\widetilde{B} \in \mathbb{R}^{m\times n(n+1)}$ is:

$$\widetilde{B} = \left[\begin{array}{c|c} & Vec\left[C^{(1)}\right]' \\ B & \vdots \\ & Vec\left[C^{(m)}\right]' \end{array}\right]$$

Approximating the process (ξ_t) by a standard Gaussian white-noise and noting that the transition and measurement equations in Formula (2.8) are respectively linear in Z_{t-1} and Z_t, the resulting state-space model is linear Gaussian. Whereas numerous existing filters rely on an approximation of the conditional distribution of X_t given $\underline{Y_{t-1}}$ (see e.g. the EKF and UKF in the next section), the QKF builds on an approximation of the conditional distribution of Z_t given $\underline{Z_{t-1}}$ or, equivalently, of Z_t given $\underline{X_{t-1}}$. Proposition 2.3.4 shows that this approximation is exact up to the second order.[13] The conditional variance-covariance matrix of the transition noise, i.e. $\widetilde{\Omega}_{t-1}\widetilde{\Omega}'_{t-1} = \widetilde{\Sigma}_{t-1}$, is a linear function of Z_{t-1} (see Proposition 2.3.2), which will be replaced in the standard linear Kalman filter by $\widetilde{\Sigma}(Z_{t-1|t-1})$. At each iteration, we emphasize that this computation should always be made using the formulae of Proposition 2.3.2 where the affine forms in Z_{t-1} are made completely explicit (see the discussion below Proposition 2.3.1). Finally, we get the Quadratic Kalman Filter algorithm displayed in Table 2.1.

Starting the algorithm at $t = 1$, we need the initial values $Z_{0|0}$ and $P^Z_{0|0}$. As emphasized previously, one can take the unconditional moments $Z_{0|0} = \widetilde{\mu}^u$ and $P_{0|0} = \widetilde{\Sigma}^u$. Note that using Equations (2.6), we have $\widetilde{\mu}^u = \widetilde{\mu} + \widetilde{\Phi}\mu^u$ and $\widetilde{\Sigma}^u = \widetilde{\Phi}\widetilde{\Sigma}^u\widetilde{\Sigma}' + \widetilde{\Sigma}(\mu^u)$ and, therefore, $Z_{1|0} = Z_{0|0}$, $P^Z_{1|0} = P^Z_{0|0}$. In other words we can also start the algorithm by the prediction of Y_t, for $t = 1$, using the initial values $Z_{1|0} = \widetilde{\mu}^u$ and $P^Z_{1|0} = \widetilde{\Sigma}^u$.

For large values of n, the QKF algorithm might suffer from the curse of dimensionality since

13. Proposition 2.3.4 implies in particular that the linear regressions of Z_t and Z_t^* on past observables are the same functions. However, it does not imply that $\mathbb{E}(Z_t|\underline{Y_t})$ and $\mathbb{E}(Z_t^*|\underline{Y_t})$ are equal since ξ_t is not Gaussian and the conditional covariance matrix $\widetilde{\Sigma}_{t-1}$ depends on past values of the latent variable.

Table 2.1: Quadratic Kalman Filter (QKF) algorithm

Initialization:		$Z_{0\|0} = \widetilde{\mu}^u$ and $P^Z_{0\|0} = \widetilde{\Sigma}^u$.
State prediction:	$Z_{t\|t-1}$	$\widetilde{\mu} + \widetilde{\Phi} Z_{t-1\|t-1}$
	$P^Z_{t\|t-1}$	$\widetilde{\Phi} P^Z_{t-1\|t-1} \widetilde{\Phi}' + \widetilde{\Sigma}(Z_{t-1\|t-1})$
Measurement prediction:	$Y_{t\|t-1}$	$A + \widetilde{B} Z_{t\|t-1}$
	$M_{t\|t-1}$	$\widetilde{B} P^Z_{t\|t-1} \widetilde{B}' + V$
Gain:	K_t	$P^Z_{t\|t-1} \widetilde{B}' M^{-1}_{t\|t-1}$
State updating:	$Z_{t\|t}$	$Z_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^Z_{t\|t}$	$P^Z_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: $\widetilde{\mu}^u$ and $\widetilde{\Sigma}^u$ are respectively the unconditional mean and variance of process Z_t (that are given in Corollary 2.3.3.1). Note that the implied value of $[(XX')_{t|t} - X_{t|t}X'_{t|t}]$, that is a covariance matrix, should be a non-negative matrix. When it is not the case, we replace its negative eigenvalues by 0 and recompute $(XX')_{t|t}$ accordingly. Such a correction is not needed in the state-prediction step: indeed, using the expression of matrices $\widetilde{\mu}$ and $\widetilde{\Phi}$, we get that $[(XX')_{t+1|t} - X_{t+1|t}X'_{t+1|t}] = \Phi((XX')_{t|t} - X_{t|t}X'_{t|t})\Phi + \Sigma$, which is then positive.

the total filtered vector is of size $n(n+1)$. In that case, it is likely that other traditional filters will result in faster computations. However, note that the $n(n+1)$-dimensional vector Z_t could be replaced by the smaller vector $[X_t', Vech(X_tX_t)']'$ of size $n(n+3)/2$. This transformation barely changes the augmented state space model, premultiplying the lower block of Z_t by a selection matrix H_n such that $Vech(X_tX_t') = H_n Vec(X_tX_t')$. The formal definition of the selection matrix is given in Appendix 2.F. The computation of conditional moments using $Vech$ is thus straightforward.

2.3.6 The smoothing algorithm

Contrary to most existing non-linear filters, that are presented in the next section, our QKF approach has a straightforward smoothing extension. Indeed, since our basic state-space model is linear, we just have to use the standard backward fixed-interval algorithm. Note however that the variance-covariance matrices $P^Z_{t+1|t}$ computed with the filtering algorithm using $Vec(\bullet)$ are not of full-rank since at least one component of Z_t is redundant when $n \geqslant 2$. Consequently, the smoothing algorithm must be expressed with

the $Vech(\bullet)$ operator. Let us introduce the following matrices:

$$\widetilde{H}_n = \begin{pmatrix} I_n & 0 \\ 0 & H_n \end{pmatrix} \quad \text{and} \quad \widetilde{G}_n = \begin{pmatrix} I_n & 0 \\ 0 & G_n \end{pmatrix},$$

that are respectively the $\frac{n(n+3)}{2} \times n(n+1)$ and $n(n+1) \times \frac{n(n+3)}{2}$ matrices using the selection and duplication matrices H_n and G_n defined in Appendix 2.F. We have:

$$Vec(X_t X_t') = G_n Vech(X_t X_t') \quad \text{and} \quad Vech(X_t X_t') = H_n Vec(X_t X_t')$$

$\widetilde{H}_n$ is defined such that $\widetilde{H}_n Z_t = [X_t', Vech(X_t X_t')']'$. The sandwich multiplication $\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'$ drops the redundant rows and columns. We get the following smoothing algorithm:

$$\begin{aligned} F_t &= \left(\widetilde{H}_n P^Z_{t|t} \widetilde{H}_n'\right) \left(\widetilde{H}_n \widetilde{\Phi} \widetilde{G}_n\right)' \left(\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'\right)^{-1} \\ \widetilde{H}_n Z_{t|T} &= \widetilde{H}_n Z_{t|t} + F_t \left(\widetilde{H}_n Z_{t+1|T} - \widetilde{H}_n Z_{t+1|t}\right) \\ \left(\widetilde{H}_n P^Z_{t|T} \widetilde{H}_n'\right) &= \left(\widetilde{H}_n P_{t|t} \widetilde{H}_n'\right) + F_t \left[\left(\widetilde{H}_n P^Z_{t+1|T} \widetilde{H}_n'\right) - \left(\widetilde{H}_n P^Z_{t+1|t} \widetilde{H}_n'\right)\right] F_t' \end{aligned}$$

The initial values $Z_{T|T}$ and $P^Z_{T|T}$ are obtained from the filtering algorithm.

2.4 Usual non-linear filters

Among the popular non-linear filters, two main classes of algorithms are widely used: the extended Kalman filter (Ekf) and the unscented Kalman filter (Ukf). Both approximate the non-linear measurement or transition equations using linearization techniques but their spirit differ radically. This section presents these algorithms applied to the linear-quadratic state-space model of Definition 2.3.1. They will further be used as competitors compared to the Qkf in the performance assessment.

Two versions of the Ekf have been used, namely the first and second order – Gaussian – filters. Their derivations are respectively based on first- and second-order Taylor expansions of the measurement equations around $X_{t|t-1}$ at each iteration. For simplicity, we use the following notations:

$$\begin{aligned} h(X_t) &\equiv A + BX_t + \sum_{k=1}^{m} e_k X_t' C^{(k)} X_t \\ \mathcal{G}_{t|t-1} &\equiv \frac{\partial h}{\partial X_t'}(X_{t|t-1}) = B + 2\sum_{k=1}^{m} e_k X_{t|t-1}' C^{(k)} \end{aligned}$$

Table 2.2 details both EKF algorithms in the quadratic measurement case.[14] A general non-linear version is provided in Appendix 2.G (see also Jazwinski [1970] and Anderson and Moore [1979] for the EKF1, and Athans, Wishner, and Bertolini [1968] or Maybeck [1982] for the EKF2).

Table 2.2: EKF algorithms in the quadratic case

		EKF1	EKF2
Initialization:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$	
State prediction:	$X_{t\|t-1}$	$\mu + \Phi X_{t-1\|t-1}$	
	$P^X_{t\|t-1}$	$\Phi P^X_{t-1\|t-1}\Phi' + \Sigma$	
Measurement prediction:	$Y_{t\|t-1}$	$h(X_{t\|t-1})$	$h(X_{t\|t-1}) + \sum_{k=1}^{m} e_k \mathrm{Tr}\left(P^X_{t\|t-1} C^{(k)}\right)$
	$M_{t\|t-1}$	$\mathcal{G}_{t\|t-1} P^X_{t\|t-1} \mathcal{G}'_{t\|t-1} + V$	$\mathcal{G}_{t\|t-1} P^X_{t\|t-1} \mathcal{G}'_{t\|t-1} + V + 2\sum_{k,j=1}^{m} e_k e_j' \mathrm{Tr}\left(C^{(k)} P^X_{t\|t-1} C^{(j)} P^X_{t\|t-1}\right)$
Gain:	K_t	$P^X_{t\|t-1}\mathcal{G}'_{t\|t-1} M^{-1}_{t\|t-1}$	
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$	
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$	

Note: See above for the definition of $\mathcal{G}_{t|t-1}$ and $h(x)$.

In the EKF1 algorithm, both $Y_{t|t-1}$ and $M_{t|t-1}$ are grossly approximated, whereas the EKF2 incorporates the so-called *bias correction terms* which are expected to reduce the error on these moments evaluation (see fourth and fifth rows of Table 2.2). Even if the Taylor expansion of the measurement equation is exact in the EKF2, it implicitly approximates the conditional distribution of (Y_t, X_t) given $\underline{Y_{t-1}}$ by a Gaussian distribution, which also induces errors in the recursions.

In comparison, the UKF belongs to the class of *density-based* filters and uses a set of

14. Another version of the second order filter called the *truncated* second-order filter is presented in Maybeck [1982]. However, it makes the assumption that the third and higher-order conditional moments of X_t given $\underline{Y_{t-1}}$ are sufficiently small to be negligible and set to 0. As a consequence, the calculation of $M_{t|t-1}$ in this algorithm can yield non positive-definite matrices showing far less computational stability than the Gaussian second-order extended filter. We thus left it aside in our comparison exercise. Also, higher-order extended filters can be derived with statistical linearization techniques, but are rarely used in practice (see Gelb [1974]).

vectors called *sigma points.*[15]

Definition 2.4.1. *Let $X \in \mathbb{R}^n$ a random vector and define $m = \mathbb{E}(X)$ and $P = \mathbb{V}(X)$. Let $(\sqrt{P})_i$ denote the i^{th} column of the lower-triangular Cholesky decomposition of P. The sigma set associated with X is composed of $2n+1$ sigma points $(\mathcal{X}_i(m,P))_{i=\{0,\dots,2n\}}$ and 2 sets of $2n+1$ weights $(\mathcal{W}_i)_{i=\{0,\dots,2n\}}$ and $(\mathcal{W}_i^{(c)})_{i=\{0,\dots,2n\}}$ defined by:*

$$\mathcal{X}_i = \begin{cases} m & \text{for} \quad i = 0 \\ m + \left(\sqrt{(n+\lambda)P}\right)_i & \text{for} \quad i \in [\![1, n]\!] \\ m - \left(\sqrt{(n+\lambda)P}\right)_{i-n} & \text{for} \quad i \in [\![n+1, 2n]\!] \end{cases} \qquad \mathcal{W}_i = \begin{cases} \lambda/(\lambda+n) & \text{for} \quad i = 0 \\ 1/[2(\lambda+n)] & \text{for} \quad i \neq 0 \end{cases}$$

$$\mathcal{W}_i^{(c)} = \begin{cases} \mathcal{W}_i + 1 - \alpha^2 + \beta & \text{for} \quad i = 0 \\ \mathcal{W}_i & \text{for} \quad i \neq 0, \end{cases}$$

where (α, κ, β) is a vector of tuning parameters and $\lambda = \alpha^2(n+\kappa) - n$. It is easy to see that for any (α, κ, β) we have:

$$\sum_{i=0}^{2n} \mathcal{W}_i \mathcal{X}_i = m \quad \text{and} \quad \sum_{i=0}^{2n} \mathcal{W}_i \left(\mathcal{X}_i - m\right)\left(\mathcal{X}_i - m\right)' = \sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left(\mathcal{X}_i - m\right)\left(\mathcal{X}_i - m\right)' = P\,.$$

The sigma set of Definition 2.4.1 is then used to approximate the moments of the non-linear transformation $h(X)$. The algorithm in the quadratic measurement equation case is given in Table 2.3. A general non-linear version is also provided in Appendix 2.G.[16]

The tuning parameters (α, κ, β) are set by the user and depend on the applied filtering problem specificities (dimension size n, number of periods T, and prior knowledge on distributions). Usual values when the distribution of X_t given $\underline{Y_{t-1}}$ is assumed Gaussian are $\beta = 2$, $\kappa = 3 - n$ or 0, and $\alpha = 1$ for low dimensional problems.

2.5 Performance comparisons using Monte Carlo experiments

We simulate a linear-quadratic state-space model and compare the performance of the QKF filter against other popular non-linear filters. We distinguish two exercises, namely filtering and parameter estimation.

15. The name *density-based filter* belongs to the terminology of Tanizaki [1996].

16. For an extensive description of the unscented Kalman filter, see Julier, Uhlmann, and Durrant-Whyte [2000], Julier [2002], or Julier and Uhlmann [2004], and applications in Kandepu, Foss, and Imsland [2008], or Christoffersen, Dorion, Jacobs, and Karoui [2014]. For the square-root version, see Van Der Merwe and Wan [2001] or Holmes, Klein, and Murray [2008] for a square-root filtering application.

Table 2.3: UKF algorithm in the quadratic case

Initialization:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$ and choose (α, κ, β).
State prediction:	$X_{t\|t-1}$	$\mu + \Phi X_{t-1\|t-1}$
	$P^X_{t\|t-1}$	$\Phi P^X_{t-1\|t-1}\Phi' + \Sigma$
Sigma points:		$\left\{\mathcal{X}_{i,t\|t-1}(X_{t\|t-1}, P^X_{t\|t-1})\right\}_{i=\{1,\dots,2n\}}$ according to Definition 2.4.1.
Measurement prediction:	$Y_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i\, h(\mathcal{X}_{i,t\|t-1})$
	$M_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]\left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' + V$
Gain:	K_t	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\mathcal{X}_{i,t\|t-1} - X_{t\|t-1}\right]\left[h(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: Weights $\mathcal{W}_i$ and $\mathcal{W}_i^{(c)}$ are given in Definition 2.4.1.

2.5.1 A simple example

To emphasize the specificity of the QKF compared to both EKFs and UKF, let us consider a very simple state-space model where analytical computations are feasible. Assume that $X_t = \varepsilon_t \sim \mathcal{IIN}(0, \sigma_\varepsilon^2)$. The measured univariate Y_t is given by $Y_t = X_t^2$ and is perfectly measured without noise or, equivalently, the noise is infinitely small. The natural method to retrieve X_t from Y_t is straightforward inverting the previous formula. The only uncertainty remaining is the sign of $\pm\sqrt{Y_t}$ which is impossible to infer. In that model, the distribution of Y_t is a $\gamma(1/2, 2\sigma_\varepsilon^2)$ distribution, with mean and variance respectively given by σ_ε^2 and $2\sigma_\varepsilon^4$. [17]

We compute the filtering formulae of the four aforementioned filters and compare them. The results are presented in Table 2.4. Despite the simplicity of the model, the EKF1 is unable to reproduce the moments of Y_t (second column of Table 2.4). Both the QKF and the EKF2 give the exact formulation of Y_t moments, whereas the computation of $M_{t|t-1}$ for the UKF depends on the tuning parameters (α, κ, β) (see 3^{rd} and 4^{th} rows). More importantly, looking at the last-two rows of Table 2.4, we see that the QKF is the

17. Recall that the density of a $\gamma(k, \rho)$ is given by $f(x) = \frac{1}{\Gamma(k)\rho^k} x^{k-1} \exp(-x/\rho)$.

only filter to update the state variables correctly in the squared components, since the second component of $Z_{t|t}$ is exactly the observed Y_t. However, all filters including the QKF produce $X_{t|t} = 0$ for all periods. Therefore the QKF is the only considered filter to jointly *(i)* correctly reproduce Y_t first-two moments, and *(ii)* produce time-varying estimates of the latent factors. We systematize this comparison to different state-space models using simulations in the next section.

TABLE 2.4 – Example : computation of filters' formulae

	QKF	EKF 1	EKF 2	UKF
$X_{t\|t-1}$ (or $Z_{t\|t-1}$ for the QKF)	$\begin{pmatrix} 0 \\ \sigma_\varepsilon^2 \end{pmatrix}$	0	0	0
$P^X_{t\|t-1}$ (or $P^Z_{t\|t-1}$ for the QKF)	$\begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 2\sigma_\varepsilon^4 \end{pmatrix}$	σ_ε^2	σ_ε^2	σ_ε^2
$Y_{t\|t-1}$	σ_ε^2	0	σ_ε^2	σ_ε^2
$M_{t\|t-1}$	$2\sigma_\varepsilon^4$	0	$2\sigma_\varepsilon^4$	$(\alpha^2\kappa + \beta)\sigma_\varepsilon^4$
$X_{t\|t}$ (or $Z_{t\|t}$ for the QKF)	$\begin{pmatrix} 0 \\ Y_t \end{pmatrix}$	0	0	0
$P^X_{t\|t}$ (or $P^Z_{t\|t}$ for the QKF)	$\begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 0 \end{pmatrix}$	σ_ε^2	σ_ε^2	σ_ε^2

Notes : The state-space model is defined by $X_t \sim \mathcal{IIN}(0, \sigma_\varepsilon^2)$ and $Y_t = X_t^2$. 'QKF' is the Quadratic Kalman filter, 'EKF 1' and 'EKF 2' are respectively the first- and second-order extended Kalman filters, 'UKF' is the unscented Kalman filter.

2.5.2 Comparison of filtering performance

We compare the filtering performance of the QKF against the EKF 1 and EKF 2, and the UKF in a linear-quadratic state-space model. We parameterize state-space model as follows :

$$\begin{aligned} X_t &= \Phi X_{t-1} + \varepsilon_t \\ Y_t &= \sqrt{\theta_2(1-\theta_1)}\sqrt{1-\Phi^2}\, X_t + \sqrt{(1-\theta_2)(1-\theta_1)}\frac{1-\Phi^2}{\sqrt{2}}\, X_t^2 + \sqrt{\theta_1}\eta_t \end{aligned} \tag{2.9}$$

where both ε_t and η_t are zero-mean normalized Gaussian white-noises, and both X_t and Y_t are scalar variables ($n = m = 1$). Comparing with Equations (2.1a) and (2.1b), we have set $\mu = 0$ and $A = 0$ for simplicity. It is straightforward to see that the unconditional variance of Y_t is equal to 1. Therefore, the weights $(\theta_1, \theta_2) \in [0, 1]^2$, should be interpreted in the following way : θ_1 is the proportion of Y_t variance explained by the measurement noise, the rest (i.e. $1 - \theta_1$) being explained by the state variables in the measurement equation. θ_2 is the proportion of the variance of Y_t explained by the linear term, within the part explained by the state variables.

The performance of the different filters are assessed with respect to values of Φ, θ_1 and θ_2. We successively set $\Phi = \{0.3, 0.6, 0.9, 0.95\}$ controlling from low to very high persistence of X_t process, $\theta_1 = \{0.2, 0.25, 0.3, \ldots, 0.8\}$ and $\theta_2 = \{0, 0.25, 0.5, 0.75\}$ (for a total of 208 cases). For instance, a combination of $(\theta_1, \theta_2) = (0.2, 0.25)$ should be interpreted as *20% of Y_t variance can be attributed to the measurement noise and 80% to the latent factors, of which 25% is attributed to the linear term and 75% to the quadratic term.*[18] Degenerated cases where either $\theta_1 = 0$, or $\theta_1 = 1$ are not considered (they correspond respectively to situations with no measurement noise or no explanatory variables in the measurement equation). Also, the case where $\theta_2 = 1$ is left aside as the measurement equation becomes linear, and all the considered filters boil down to the linear Kalman filter.[19] For each value of Φ, we simulate paths of the latent process X_t of $T = 1,000,000$ periods with a starting value of $X_0 = 0$. We then simulate the measurement noises η_t and compute implied observable variables Y_t for each combination of (θ_1, θ_2). The filtering exercise is performed for each filter, initial values being known.[20] For the UKF, we set $\alpha = 1$, and $\beta = 2$ as in Christoffersen, Dorion, Jacobs, and Karoui [2014]. For those values of (α, β) and scalar processes, it can be shown that $\kappa = 0$ implies the exact same recursions as the EKF2.[21] We therefore set $\kappa = 3 - n = 2$.

We denote by $\widehat{X_{t|t}}$, $\widehat{X^2_{t|t}}$ and $\widehat{P_{t|t}}$ the filtered values resulting from any filtering algorithm. The different filters are compared with respect to three measures of performance. First, we compute the RMSEs of filtered values $\widehat{X_{t|t}}$ compared to X_t. Second, we calculate RMSEs of the quadratic process $\widehat{X^2_{t|t}}$. Whereas the QKF evaluates this quantity directly in the algorithm, we recompute its underlying value for the other filters with the formula $\widehat{X^2_{t|t}} = \widehat{X_{t|t}}^2 + \widehat{P_{t|t}}$. The RMSE measures for any of our estimated values are normalized

18. Note that in the general quadratic models that we consider here, we have $Cov(X_t, Vec(X_t X_t')) = 0$.
19. This is in fact not obvious for the UKF, and the proof is provided in Appendix 2.H.
20. Thus we set, $X_{0|0} = 0$ and $P^X_{0|0} = 0$ for the EKFs and UKF, and $Z_{0|0} = 0_{\mathbb{R}^2}$ and $P^Z_{0|0} = 0_{\mathbb{R}^{2\times 2}}$ for the QKF.
21. See Appendix 2.I for a proof.

by the standard deviation of the simulated process :

$$\overline{RMSE}_W = \frac{RMSE_W}{\sigma_W} = \left[\frac{T^{-1} \sum_{t=1}^{T} (W_t - \widehat{W_{t|t}})^2}{\mathbb{V}(W_t)} \right]^{1/2}$$

where $W_t = X_t$ or X_t^2 and $\widehat{W_{t|t}} = \widehat{X_{t|t}}$ or $\widehat{X_{t|t}^2}$. This measure converges to 1 if the filtered values are equal to the unconditional mean of the latent process for all periods. Consequently, if any filter yields a normalized RMSE greater than 1, a better filtering result would be obtained by setting $\widehat{W_{t|t}} = \mathbb{E}(W_t)$, for all t. Lastly, we compare the filters capacities to discriminate between the explanatory process and the measurement noise by computing non-normalized RMSEs of implied $\widehat{\eta}_t$. The results are respectively presented on Figures 2.1, 2.2, and 2.3.

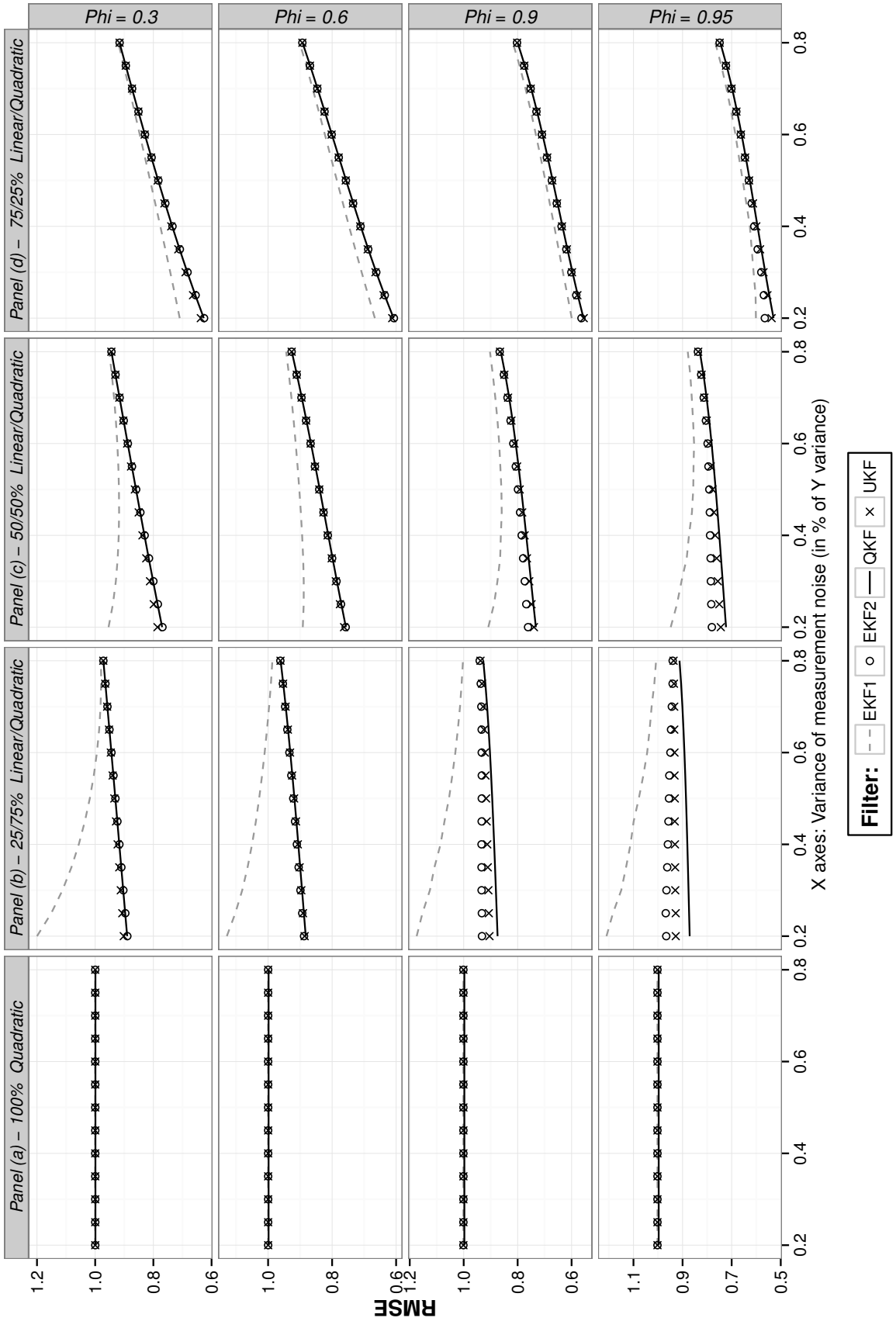


FIGURE 2.1 – RMSE of $\widehat{X_{t|t}}$

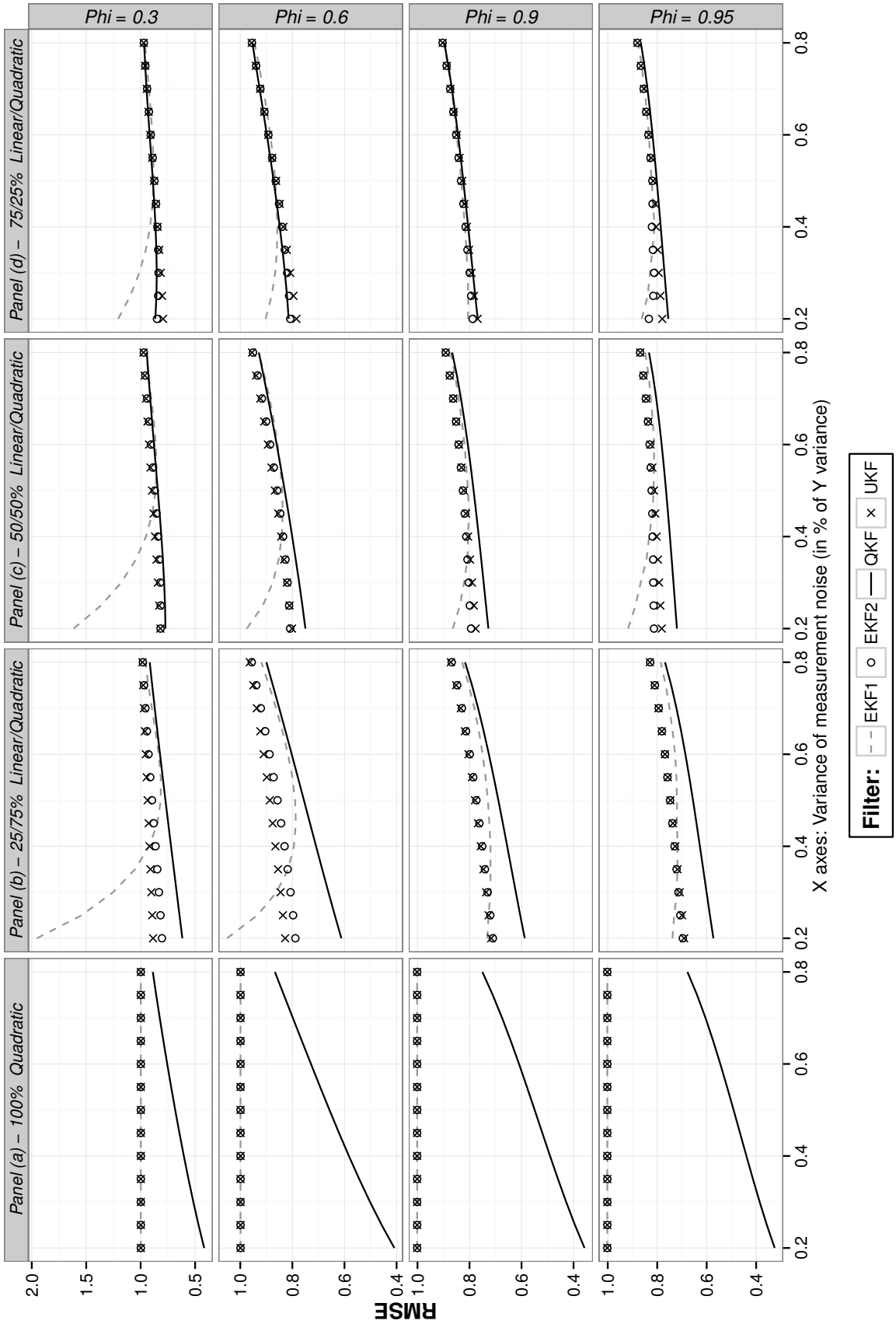
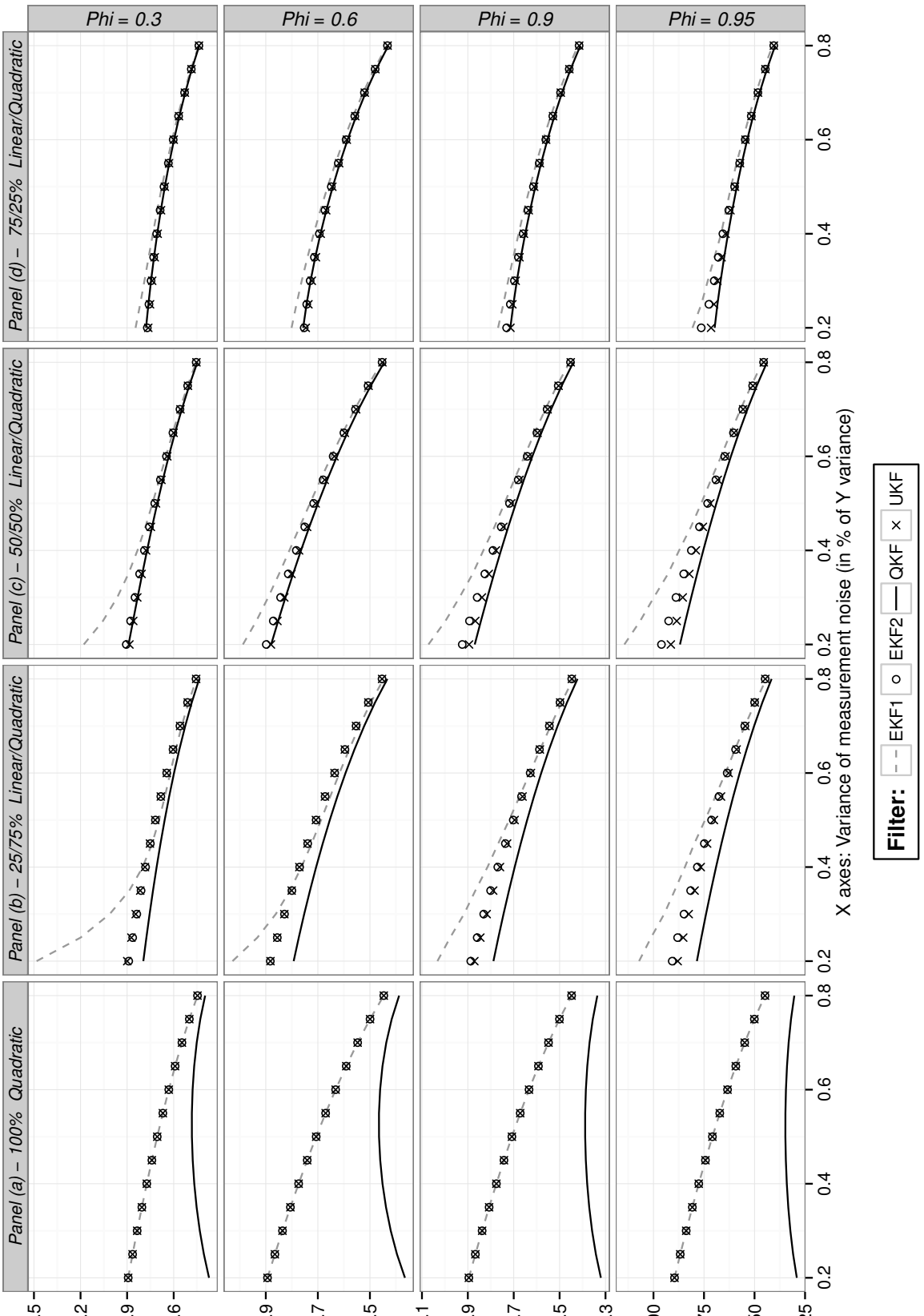


FIGURE 2.2 – RMSE of $\widehat{X^2_{t|t}}$

FIGURE 2.3 – RMSE of $\widehat{\eta}_t$



Result 1. *When the measurement equation is fully quadratic ($\theta_2 = 0$), The* QKF *is the only considered filter capable of both :*

(i) Filtering out a substantial part of the measurement noise,

(ii) Yielding accurate evaluations of $X^2_{t|t}$.

We first analyse the case where the measurement equation is only quadratic ($\theta_2 = 0$, left column of all figures). As already noted for a specific case in the previous section, all filters are "blind" on the evaluation of $X_{t|t}$ producing a flat $\widehat{X_{t|t}} = 0$, and normalized RMSEs are equal to 1 whatever the values of Φ and θ_1 (see Figure 2.1, left column). However, looking at Figure 2.2, we see that for any relative size of the measurement errors and any persistence, the QKF yields more accurate evaluations of $X^2_{t|t}$ than the other filters, showing 5% to 60% smaller RMSEs depending on the case. Two patterns can be observed here. First, the smaller the measurement errors, the stronger the outperformance of the QKF filter compared to the others. Second, the outperformance of the QKF increases with the persistence of the latent process. [22] This better performance is confirmed by looking at the evaluation of the measurement noise, where the QKF also provides the smallest RMSEs for all values of (Φ, θ_1) (see Figure 2.3, first column). The reduction in the measurement noise RMSEs for the QKF compared to the others can reach 70%. This result emphasizes the substantial improvement of the fitting properties of the QKF compared to those of the other filters.

Result 2. *For measurement equations where the linearity degree goes from 25% to 50%, the* QKF *beats the other filters, especially for the evaluation of $X^2_{t|t}$. Eventually, for levels of about 75% of linearity in the measurement equation, the RMSEs of all filters converge to the same values.*

We turn now to the cases where the measurement equation has from 25% to 50% of linearity degree ($\theta_2 = \{0.25, 0.5\}$, second and third columns of all figures). We first leave aside the EKF 1 (see result 3). For $\widehat{X_{t|t}}$, normalized RMSEs are more or less the same for the EKF 2 and the UKF in all cases. In comparison, the QKF is either equivalent, either showing smaller RMSEs for high-persistent cases ($\Phi = 0.9$ or $\Phi = 0.95$, third and fourth rows of Figure 2.1). This better performance is confirmed when looking at Figure 2.2. In all cases, the QKF possesses lowest RMSEs for $\widehat{X^2_{t|t}}$. For example, for $\Phi = 0.9$, $\theta_1 = 0.2$ and $\theta_2 = 0.25$, the QKF shows RMSEs slightly below 60% of X^2_t standard deviation whereas the others are all above 70% (see Figure 2.2, third row of panel (b)).
Unsurprisingly, this evidence places the QKF ahead of its competitors for the de-noising exercise : for panels (b) and (c) of Figure 2.3, RMSEs of $\widehat{\eta_t}$ are always below the others

22. We see this as a pleasant feature for term-structure modelling applications where yields are typically highly persistent and measured with low errors.

for the QKF. Looking at panel (d) where the measurement equation is 75% linear (fourth column of all figures), we see that all RMSEs eventually converge to each other for all filters. This is consistent with the fact that all filters reduce to the standard Kalman filter when the measurement equation is fully linear.

Result 3. *The* EKF 1 *should be discarded for filtering, especially when the variance of the measurement errors is low (cases where* θ_1 *is low).*

Looking at Figures 2.1 and 2.2, we notice a very unpleasant behaviour of the EKF 1. For low measurement errors, RMSEs of both $\widehat{X_{t|t}}$ and $\widehat{X^2_{t|t}}$ can reach values greater than 1, especially in panels (b) and (c) where the measurement equation shows medium linearity degree (see second and fourth columns of Figures 2.1 and 2.2). This catastrophic performance can be particularly observed for low persistence, low linearity degree, and low measurement errors : when $\Phi = 0.3$, $\theta_1 = 0.2$ and $\theta_2 = 0.25$, $\widehat{X_{t|t}}$ and $\widehat{X^2_{t|t}}$ show respectively 120% and 200% normalized RMSE values. That is to say filtered values yielded by the EKF 1 prove to be very poor in some cases.

This Monte-Carlo experiment provides evidence that in terms of filtering, the QKF largely dominates both EKFs and the UKF for evaluating X_t and X_t^2, as well as for de-noising the observable y_t. This is particularly the case when the degree of linearity in the measurement equation is low. Increasing the degree of linearity produces closer RMSEs for the QKF, the EKF2, and the UKF ; the EKF1 shows a very unstable behaviour. In the next section, we explore the characteristics of the different techniques in terms of parameter estimation.

2.5.3 Quasi maximum likelihood parameter estimation

To compare the filters with respect to parameter estimation, we simulate the same benchmark model given in Equation (2.9). We estimate the vector of parameters $\beta = (A, B, C, D, \Phi)$ (see the notations of Equations (2.1a - 2.1b)) for some specific values of (Φ, θ_1, θ_2). To explore the finite sample properties of the different estimators, we set $T = 200$ and simulate 1000 dynamics for a given set of (Φ, θ_1, θ_2). This provides us with the empirical marginal distributions of the estimators. As usual in non-linear filter estimation, the technique is only quasi maximum likelihood as the distribution of Y_t given $\underline{Y_{t-1}}$ is approximated as a Gaussian. [23]

To avoid local maxima, a two-step estimation is performed. First, a stochastic maximization algorithm is launched to select a potential zone for the global maximum. Second, a

23. It should be noted here that none of the filters produce exact first-two conditional moments of Y_t given $\underline{Y_{t-1}}$. The asymptotic properties of the quasi maximum likelihood are therefore not relevant.

simplex algorithm is used to refine the estimates in the selected zone.[24] This procedure makes the results reliable at the cost of extended computational burden. This particular reason leads us to select three paradigmatic cases for the simulated processes. The first considered case is fully quadratic with high persistence and low measurement error variance ($\Phi = 0.9$, $\theta_1 = 0.05$, and $\theta_2 = 0$). In the second case, we decrease the persistence of the latent process and increase the size of measurement errors setting $\Phi = 0.6$, $\theta_1 = 0.2$, and keeping $\theta_2 = 0$. In the last case we introduce a linear component in the measurement equation, with the parametrization : $\Phi = 0.6$, $\theta_1 = 0.2$, and $\theta_2 = 0.25$. More linear cases ($\theta_2 > 0.25$) were not considered as we emphasized in Section 2.5.2 that the four filters yield closer results in those cases. For identification purpose, we also impose $\hat{B} > 0$. Results in terms of bias, standard errors, and RMSEs are presented in Table (2.5). Comparisons of filters on average across panels are provided in Table 2.6.

24. The stochastic algorithm used is the articificial bee colony of Karaboga and Basturk [2007]. In order to limit computational time, we consider 256 people in the population, and only 10 iterations. Then, the best member of the population is selected as initial conditions for the Nelder-Mead algorithm.

TABLE 2.5 – Estimation results

		Panel (a)					*Panel (b)*					*Panel (c)*				
	MODEL	A	B	C	Φ	$\sqrt{\theta_1}$	A	B	C	Φ	$\sqrt{\theta_1}$	A	B	C	Φ	$\sqrt{\theta_1}$
		0	0	0.131	0.9	0.224	0	0	0.405	0.6	0.447	0	0.358	0.351	0.6	0.447
Bias	QKF	0.177*	0.188*	−0.003*	−0.063*	−0.037*	0.147*	0.269*	−0.076*	−0.074*	−0.151	0.065*	−0.089*	−0.048*	−0.029*	−0.132
	EKF 1	0.350	0.256	0.016	−0.067	0.076	0.566†	0.547	1.318†	−0.140†	0.052*	0.460†	0.158	0.309†	−0.094†	0.055*
	EKF 2	0.363	0.412	−0.022	−0.122	−0.107	0.260	0.580†	−0.143	−0.098	−0.276	0.165	0.194†	−0.107	−0.058	−0.244
	UKF	0.414†	0.415†	−0.038†	−0.124†	−0.170†	0.336	0.561	−0.200	−0.084	−0.321†	0.246	0.194†	−0.157	−0.052	−0.304†
Std.	QKF	0.319	0.202†	0.038	0.087	0.091	0.210†	0.316	0.164	0.176*	0.290	0.202†	0.299	0.144	0.163*	0.275
	EKF 1	0.211*	0.192	0.104†	0.132†	0.082*	0.115*	0.321†	18.218†	0.251†	0.309†	0.114*	0.304†	10.725†	0.236†	0.277†
	EKF 2	0.389†	0.114	0.030	0.095	0.105†	0.204	0.224	0.137	0.213	0.252	0.189	0.213	0.122	0.201	0.259
	UKF	0.277	0.108*	0.019*	0.077*	0.086	0.167	0.215*	0.100*	0.208	0.229*	0.173	0.194*	0.092*	0.186	0.233*
RMSE	QKF	0.365*	0.276*	0.038	0.108*	0.098*	0.257*	0.415*	0.181*	0.191*	0.327	0.212*	0.312	0.152*	0.166*	0.305
	EKF 1	0.409	0.320	0.105†	0.148	0.112	0.577†	0.635†	3.656†	0.287†	0.314*	0.474†	0.343†	2.042†	0.254†	0.282*
	EKF 2	0.532†	0.427	0.037*	0.155†	0.150	0.331	0.621	0.198	0.235	0.374	0.250	0.288	0.162	0.209	0.356
	UKF	0.498	0.429†	0.042	0.146	0.190†	0.375	0.601	0.224	0.225	0.395†	0.301	0.275*	0.182	0.193	0.383†

Notes : For each set of parameters, estimations were performed simulating 1,000 different dynamics. All computed values in the table are averages across simulations.EKF 1 and EKF 2 stand respectively for the first and second order extended Kalman filters. Panel (a) to (c) vary with respect to the parameters' value provided on the first row. For each simulation, the bias is calculated as $\hat{\beta} - \beta$, hence a positive value indicates an average overestimation. '*' indicates the best value among filters, '†' indicates the worst value among filters.

Result 4. *The* QKF *quasi maximum likelihood estimates are either the less biased, either possess the lowest RMSEs for all parameters. In addition, on average across panels, the* QKF *is the less biased filter and possesses lowest RMSEs. This superiority is robust to the degree of persistence of the latent process, to the degree of linearity of the measurement equation, and to the size of the measurement errors.*

Over the three panels, the results of Tables 2.5 and 2.6 are in favour of our QKF maximum likelihood estimates. We first concentrate on panel (a) results. For the five estimated parameters, the QKF shows smaller bias than the other filters : for $\widehat{A}$, $\widehat{B}$, and $\widehat{\Phi}$, the bias of the QKF estimates corresponds to half the bias of the EKF 2 and the UKF. In addition, for four out of the five parameters, the QKF estimates yield smaller RMSEs even though it often entails higher standard deviation than its competitors (see Table 2.5, panel (a)). The same general pattern can be observed for panel (b), where persistence degree is smaller. Consistently with the intuition, the QKF always outperforms its competitors for estimating parameters B and C. This shows a better capacity to discriminate the influence of linear and quadratic terms in the observable. While panel (c) introduces some linearity in the measurement equation ($B \neq 0$), the QKF still beats the other filters for four (resp. three) out of five parameters in terms of bias (resp. RMSEs). In the end, looking at Table 2.6, we observe the superiority of the QKF across all cases : on 13 (resp. 11) out of 15 parameters, the QKF estimates possess the lowest bias (resp. RMSEs) compared to the others.

Result 5. *On average across cases,*

- *The* QKF *never yields the worst bias or RMSEs of all filters.*
- *The* EKF 1 *estimates possess the largest RMSEs and standard deviations.*
- *The* UKF *estimates possess the lowest standard deviations, but are the most biased.*
- *The* EKF 2 *is rarely the best in terms of both bias, standard deviations and RMSEs, but is also rarely the worst.*

We turn now to comparing the average results of the different filters. Table 2.6 presents the number of times each filter is best and worst in terms of bias, standard deviations and RMSEs. We have already emphasized that the QKF estimates surpass the others on average in terms of bias and RMSEs. A striking feature presented in Table 2.6 is also that QKF estimates are never the most biased, neither possess the biggest RMSEs (see first column). Overall, these results underline a better bias/variance trade-off for the QKF compared to the other filters.

The results of Tables 2.5 and 2.6 also confirm the concerns about the EKF 1 performance : out of 15 estimates, 6 are the most biased, 10 possess the biggest standard deviations,

TABLE 2.6 – Maximum likelihood performance over the three panels

	QKF	EKF 1	EKF 2	UKF
Number of times being less biased	13	2	0	0
Number of times being most biased	0	6	2	7
Number of times having smallest std.	2	4	0	9
Number of times having biggest std.	3	10	2	0
Number of times having smallest RMSEs	11	2	1	1
Number of times having biggest RMSEs	0	9	2	4

Notes : Cases are taken from Table 2.5 estimates. Total number of estimated parameters are 15. Note that the sum of the second row however yields a result of 16 due to equality of the EKF 2 and the UKF possessing the worst bias.

and 9 possess the highest RMSEs. This poor performance is particularly observable for the estimation of C in panel (b) and (c) of Table 2.5 : the standard deviations of the estimates are respectively 18 and 10, and their RMSEs are more than 10 times bigger than those of the other filters. This can be explained by the fact that the curvature of the EKF 1 log-likelihood along the C-axis is very close to zero. Hence the estimate $\widehat{C}$ can move a lot along the line with very little change in the log-likelihood. This corroborates the incapacity of the EKF 1 to deal with high non-linearities in the measurement equation, as already noted in the filtering performance comparison (see previous section).

Interestingly, the UKF also shows some concerning features for parameter estimation. It is the most biased for 8 parameters out of 15, which is the worst bias performance among all filters. However, it is also the filter that produces on average the smallest standard deviations for 9 parameters (see last column of Table 2.6). Looking at Table 2.5), we observe that those cases where the standard deviation is low tend to correspond to cases where the bias is highest. This bias/variance trade-off hands up being very poor regarding the RMSEs : the UKF is the best only once, and four times the worst out of the 15 parameters. Consequently, we argue that the use of the UKF should be made with caution in the linear-quadratic state-space model since it tends to result in parameter estimates that are "tightly" distributed around biased values.

Finally, the EKF 2 seems to yield better average results than both the EKF 1 (unsurprisingly) and the UKF : although it is never the less biased and possesses the lowest RMSE for only one estimate, it is also rarely the most biased or rarely shows the biggest RMSE (see Table 2.6). Still, those results are far less encouraging than those of the QKF and the latter should be preferred in linear-quadratic state-space model estimations.

On the whole, for most estimates, the QKF is less biased and possesses the lowest RMSEs. Despite a slightly poorer performance on the standard deviations, the QKF maximum likelihood estimates show a better bias/variance trade-off than its competitors. Also, the consideration of 3 different panels provide evidence that these results are neither altered by the degree of curvature in the measurement equation, nor by the persistence of the latent process or by the size of the measurement errors. These finite-sample estimation properties emphasize the superiority of the QKF for practical applications.

2.6 Conclusion

In this paper, we develop the quadratic Kalman filter (QKF), a fast and efficient technique for filtering and smoothing state-space models where the transition equations are linear and the measurement equations are quadratic. Building the *augmented vector of factors* stacking together the latent vector with its vectorized outerproduct, we provide analytical formulae of its first-two conditional and unconditional moments. With this new expression of the latent factors, we show that the state-space model can be expressed in a fully linear form with non-Gaussian residuals. Using this new formulation of the linear-quadratic state-space model, we adapt the linear Kalman filter to obtain the Quadratic Kalman Filter and Smoother algorithms (resp. QKF and QKS). Since no simulation is required in the computations, both QKF and QKS algorithms are computationally fast and stable. We compare performance of the QKF against the extended and unscented versions of the Kalman filter in terms of filtering and parameter estimation. Our results suggest that for both filtering and quasi maximum likelihood estimation, the QKF outperforms its competitors. For filtering, the higher the curvature of the measurement equation, the more effective the QKF compared to the other filters. For parameter estimation, the QKF shows either smaller bias or smaller RMSEs than its competitors.

Appendices to Chapter 2

2.A Useful algebra

We detail hereby some properties of both the Kronecker product and the $Vec(\bullet)$ operator. Their proofs are available in Magnus and Neudecker [1988]. These properties will be used extensively in the proofs presented in Appendices 2.B, 2.C and 2.D.

Proposition 2.A.1. *Let m_1 and m_2 be two size-n vectors, M_1 and M_2 be two square matrices of size n. Let also P, Q, R, and S be four matrices with respective size $(p \times q)$, $(q \times r)$, $(r \times s)$, and $(s \times t)$. We have:*

(i) $Vec(m_1 m_2') = m_2 \otimes m_1$.

(ii) $Vec(M_1 \otimes M_2) = (I_n \otimes \Lambda_n \otimes I_n)\,[Vec(M_1) \otimes Vec(M_2)]$ Λ_n *given in Lemma 2.B.1.*

in particular: $Vec(M_1 \otimes m_1) = Vec(M_1) \otimes m_1$,

and $Vec(M_1 \otimes m_1') = (I_n \otimes \Lambda_n)\,[Vec(M_1) \otimes m_1]$

(iii) $Vec(PQR) = (R' \otimes P)Vec(Q)$

(iv) $Vec(PQ) = (I_r \otimes P)Vec(Q) = (Q' \otimes I_p)Vec(P)$

(v) $(PQ) \otimes (RS) = (P \otimes R)(Q \otimes S)$.

2.B Properties of the commutation matrix

Lemma 2.B.1. *Let Λ_n be the $(n^2 \times n^2)$ commutation matrix partitioned in $(n \times n)$ blocks, whose (i,j) block is $e_j e_i'$. Let M_1 and M_2 be two square matrices of size n, and m be a vector of size $(n \times 1)$. We have:*

(i) $\Lambda_n = \sum_{i,j=1}^{n} (e_i e_j') \otimes (e_j e_i')$

(ii) Λ_n *is orthogonal and symmetric:* $\Lambda_n^{-1} = \Lambda_n' = \Lambda_n$

(iii) $\Lambda_n Vec(M_1) = Vec(M_1')$

(iv) $\Lambda_n (M_1 \otimes M_2) \Lambda_n = M_2 \otimes M_1$

(v) $\Lambda_n (M_1 \otimes m) = m \otimes M_1$.

Proof. (i) Straightforward by definition.

(ii) Λ_n is symmetric:

$$\Lambda_n' = \sum_{i,j=1}^{n} (e_j e_i') \otimes (e_i e_j') = \Lambda_n.$$

Λ_n is orthogonal:

$$\Lambda_n\Lambda_n' = \sum_{i,j=1}^{n}[(e_ie_j')\otimes(e_je_i')][(e_je_i')\otimes(e_ie_j')] = \sum_{i,j=1}^{n}(e_ie_i')\otimes(e_je_j') = I_{n^2}.$$

(iii)

$$\begin{aligned}\Lambda_n Vec(M_1) &= \sum_{i,j=1}^{n}\left[(e_ie_j')\otimes(e_je_i')\right]Vec(M_1)\\ &= \sum_{i,j=1}^{n}Vec\left[(e_je_i')M_1(e_ie_j')'\right]\\ &= \sum_{i,j=1}^{n}Vec\left[e_jM_1^{(i,j)}e_i'\right] = Vec(M_1').\end{aligned}$$

(iv) By definition,

$$M_1\otimes M_2 = \sum_{i,j=1}^{n}(M_1^{(,i)}\otimes M_2^{(,j)})(e_i'\otimes e_j'),$$

where $M_1^{(,i)}$ and $M_2^{(,j)}$ are respectively the i^{th} and j^{th} columns of matrices M_1 and M_2. Therefore we have:

$$\begin{aligned}\Lambda_n(M_1\otimes M_2)\Lambda_n &= \sum_{i,j=1}^{n}\Lambda_n(M_1^{(,i)}\otimes M_2^{(,j)})(e_i\otimes e_j)'\Lambda_n\\ &= \sum_{i,j=1}^{n}\left[\Lambda_n(M_1^{(,i)}\otimes M_2^{(,j)})\right][\Lambda_n(e_i\otimes e_j)]'\\ &= \sum_{i,j=1}^{n}\left[\Lambda_n Vec(M_2^{(,j)}M_1^{(,i)'})\right][\Lambda_n Vec(e_je_i')]'\\ &= \sum_{i,j=1}^{n}(M_2^{(,j)}\otimes M_1^{(,i)})(e_j\otimes e_i)'\\ &= M_2\otimes M_1.\end{aligned}$$

(v) With the same notations,

$$\begin{aligned}
\Lambda_n(M_1 \otimes m) &= \Lambda_n \sum_{i=1}^{n} (M_1^{(,i)} e_i') \otimes m \\
&= \Lambda_n \sum_{i=1}^{n} (M_1^{(,i)} \otimes m) e_i' \\
&= \Lambda_n \sum_{i=1}^{n} Vec(m M_1^{(,i)'}) e_i' \\
&= \sum_{i=1}^{n} Vec(M_1^{(,i)} m') e_i' \\
&= \sum_{i=1}^{n} (m \otimes M_1^{(,i)}) e_i' \\
&= m \otimes M_1.
\end{aligned}$$

■

2.C Z_t conditional moments calculation

Lemma 2.C.1. *If $\varepsilon \sim \mathcal{N}(0, I_n)$, we have*

$$\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right] = I_{n^2} + \Lambda_n,$$

where Λ_n is given in Lemma 2.B.1.

Proof.

$$Vec(\varepsilon\varepsilon') = \left[(\varepsilon\varepsilon_1)', (\varepsilon\varepsilon_2)', \ldots, (\varepsilon\varepsilon_n)'\right]'.$$

$\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right]$ is a $(n^2 \times n^2)$ matrix, partitioned in $(n \times n)$ blocks, whose (i,j) block is $V_{i,j} = cov(\varepsilon\varepsilon_i, \varepsilon_j\varepsilon)$. The (k,ℓ) entry of $V_{i,j}$ is $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$. Two cases can be distinguished:

- *Case 1:* if $i \neq j$, then the only non-zero terms among the $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$ are obtained for $k = j$ and $i = \ell$. By the properties of standardized Gaussian distribution, we have $\text{cov}(\varepsilon_i\varepsilon_j, \varepsilon_i\varepsilon_j) = \mathbb{V}(\varepsilon_i\varepsilon_j) = 1$. Finally, $V_{i,j} = e_j e_i'$.
- *Case 2:* if $i = j$, then the non-zero terms among the $cov(\varepsilon_k\varepsilon_i, \varepsilon_j\varepsilon_\ell)$ are obtained for $k = \ell = i$ and its value is 2, or for $k = \ell \neq i$, and its value is 1. Finally, $V_{i,i} = I_n + e_i e_i'$.

Putting case 1 and 2 together, we get $\mathbb{V}\left[Vec(\varepsilon\varepsilon')\right] = I_{n^2} + \Lambda_n$. ■

Proposition 2.3.1 $\mathbb{E}_{t-1}(Z_t) = \widetilde{\mu} + \widetilde{\Phi} Z_{t-1}$ and $\mathbb{V}_{t-1}(Z_t) = \widetilde{\Sigma}_{t-1}$, where:

$$\widetilde{\mu} = \begin{pmatrix} \mu \\ Vec(\mu\mu' + \Sigma) \end{pmatrix}, \quad \widetilde{\Phi} = \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array}\right)$$

$$\widetilde{\Sigma}_{t-1} \equiv \widetilde{\Sigma}(Z_{t-1}) = \left(\begin{array}{c|c} \Sigma & \Sigma\Gamma_{t-1}' \\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma_{t-1}' + \left[I_{n^2} + \Lambda_n\right](\Sigma \otimes \Sigma) \end{array}\right)$$

$$\Gamma_{t-1} = I_n \otimes (\mu + \Phi X_{t-1}) + (\mu + \Phi X_{t-1}) \otimes I_n,$$

Λ_n being the $n^2 \times n^2$ matrix, defined in Lemma 2.B.1.

Proof.

$$\begin{aligned} \mathbb{E}_{t-1}(X_t) &= \mu + \Phi X_{t-1} \\ \mathbb{E}_{t-1}[X_t X_t'] &= \mathbb{E}_{t-1}\left(\mu + \Phi X_{t-1} + \Omega\varepsilon_t\right)\left(\mu + \Phi X_{t-1} + \Omega\varepsilon_t\right)' \\ &= \mu\mu' + \mu X_{t-1}'\Phi' + \Phi X_{t-1}\mu' + \Phi X_{t-1}X_{t-1}'\Phi' + \Sigma \end{aligned}$$

Using the $Vec(\bullet)$ operator properties of Proposition 2.A.1, (iii), we obtain:

$$\mathbb{E}_{t-1}\left[Vec(X_t X_t')\right] = Vec(\mu\mu' + \Sigma) + (\mu \otimes \Phi)X_{t-1} + (\Phi \otimes \mu)X_{t-1} + (\Phi \otimes \Phi)Vec(X_{t-1}X_{t-1}')$$

Finally,

$$\boxed{\mathbb{E}_{t-1}(Z_t) = \begin{pmatrix} \mu \\ Vec(\mu\mu' + \Sigma) \end{pmatrix} + \left(\begin{array}{c|c} \Phi & 0 \\ \hline \mu \otimes \Phi + \Phi \otimes \mu & \Phi \otimes \Phi \end{array}\right) Z_{t-1}}$$

For the conditional variance-covariance matrix, we have $\mathbb{V}_{t-1}(X_t) = \Sigma$ and

$$\begin{aligned} \mathbb{V}_{t-1}\left[Vec(X_t X_t')\right] = \mathbb{V}_{t-1}\big[Vec(&\overbrace{\mu\mu' + \mu X_{t-1}'\Phi' + \Phi X_{t-1}\mu' + \Phi X_{t-1}X_{t-1}'\Phi'}^{\text{deterministic given } \underline{Z_{t-1}}} \\ &+ (\mu + \Phi X_{t-1})\varepsilon_t'\Omega' + \Omega\varepsilon_t(\mu' + X_{t-1}'\Phi') + \Omega\varepsilon_t\varepsilon_t'\Omega')\big] \end{aligned}$$

Using properties Proposition 2.A.1, $(iii-iv)$,

$$\begin{aligned}
\mathbb{V}_{t-1}\left[Vec(X_tX_t')\right] &= \mathbb{V}_{t-1}\left[(I_n\otimes\mu+\mu\otimes I_n+I_n\otimes\Phi X_{t-1}+\Phi X_{t-1}\otimes I_n)\Omega\varepsilon_t+Vec(\Omega\varepsilon_t\varepsilon_t'\Omega)\right]\\
&\equiv \mathbb{V}_{t-1}\left[\Gamma_{t-1}\Omega\varepsilon_t+Vec(\Omega\varepsilon_t\varepsilon_t'\Omega)\right]\\
&= \Gamma_{t-1}\Sigma\Gamma_{t-1}'+\mathbb{V}_{t-1}\left[(\Omega\otimes\Omega)Vec(\varepsilon_t\varepsilon_t')\right] \quad \text{as} \quad \mathrm{Cov}_{t-1}[\varepsilon_t,Vec(\varepsilon_t\varepsilon_t')]=0\\
&= \Gamma_{t-1}\Sigma\Gamma_{t-1}'+(\Omega\otimes\Omega)(I_{n^2}+\Lambda_n)(\Omega'\otimes\Omega') \quad \text{(using Lemma 2.C.1.)}
\end{aligned}$$

Proposition 2.A.1, (v) implies that $(\Omega\otimes\Omega)(\Omega\otimes\Omega)'=\Sigma\otimes\Sigma$. Therefore, we have:

$$\begin{aligned}
&(\Omega\otimes\Omega)(\Omega\otimes\Omega)'=(\Omega\otimes\Omega)\Lambda_n(\Omega\otimes\Omega)'\Lambda_n \quad \text{(using Lemma 2.B.1, } (iv))\\
\iff\ &(\Omega\otimes\Omega)\Lambda_n(\Omega\otimes\Omega)'=\Lambda_n(\Sigma\otimes\Sigma) \quad \text{since} \quad \Lambda_n=\Lambda_n^{-1}\\
\iff\ &(\Omega\otimes\Omega)(I_{n^2}+\Lambda_n)(\Omega\otimes\Omega)'=(I_{n^2}+\Lambda_n)(\Sigma\otimes\Sigma).
\end{aligned}$$

Hence:

$$\mathbb{V}_{t-1}\left[Vec(X_tX_t')\right]=\Gamma_{t-1}\Sigma\Gamma_{t-1}'+(I_{n^2}+\Lambda_n)(\Sigma\otimes\Sigma).$$

Using again the fact that ε_t and $Vec(\Omega\varepsilon_t\varepsilon_t'\Omega')$ are non-correlated, we have:

$$\begin{aligned}
\mathrm{cov}_{t-1}\left[Vec(X_tX_t'),X_t\right] &= \mathrm{cov}_{t-1}\left[\Gamma_{t-1}\Omega\varepsilon_t,\Omega\varepsilon_t\right]\\
&= \Gamma_{t-1}\Sigma
\end{aligned}$$

Finally, the conditional variance-covariance matrix of Z_t given X_{t-1} is

$$\boxed{\widetilde{\Sigma}_{t-1}=\left(\begin{array}{c|c}\Sigma & \Sigma\Gamma_{t-1}'\\ \hline \Gamma_{t-1}\Sigma & \Gamma_{t-1}\Sigma\Gamma_{t-1}'+(I_{n^2}+\Lambda_n)(\Sigma\otimes\Sigma)\end{array}\right).} \tag{2.10}$$

■

2.D Proof of Proposition 2.3.2

We want to explicitly disclose the affine form of $\widetilde{\Sigma}(Z_{t-1})$. In order to achieve this, we consider the four blocks of the matrix in Equation (2.10) and express the vectorized form of each block. First, let us show that $Vec(\Gamma_{t-1}\Sigma)$ is affine in Z_{t-1}. We have:

$$\begin{aligned}
\Gamma_{t-1} &= I_n\otimes(\mu+\Phi X_{t-1})+(\mu+\Phi X_{t-1})\otimes I_n\\
&= I_n\otimes(\mu+\Phi X_{t-1})+\Lambda_n\left[I_n\otimes(\mu+\Phi X_{t-1})\right] \quad \text{(using Lemma 2.B.1, } (v))\\
&= (I_{n^2}+\Lambda_n)\left[I_n\otimes(\mu+\Phi X_{t-1})\right].
\end{aligned}$$

Therefore we have:

$$
\begin{aligned}
Vec(\Gamma_{t-1}\Sigma) &= Vec\left\{\left(I_{n^2}+\Lambda_n\right)\left[I_n\otimes\left(\mu+\Phi X_{t-1}\right)\right]\Sigma\right\} \\
&= \left[\Sigma\otimes\left(I_{n^2}+\Lambda_n\right)\right]Vec\left\{I_n\otimes\left(\mu+\Phi X_{t-1}\right)\right\} && \text{(Prop. 2.A.1, } (iii)\text{)} \\
&= \left[\Sigma\otimes\left(I_{n^2}+\Lambda_n\right)\right]\left[Vec(I_n)\otimes\left(\mu+\Phi X_{t-1}\right)\right] && \text{(Prop. 2.A.1, } (ii)\text{)} \\
&= \left[\Sigma\otimes\left(I_{n^2}+\Lambda_n\right)\right]Vec\left[\left(\mu+\Phi X_{t-1}\right)Vec(I_n)'\right] && \text{(Prop. 2.A.1, } (i)\text{)} \\
&= \left[\Sigma\otimes\left(I_{n^2}+\Lambda_n\right)\right]\left[Vec(I_n)\otimes I_n\right]\left(\mu+\Phi X_{t-1}\right) && \text{(Prop. 2.A.1, } (iv)\text{)} \\
Vec(\Sigma\Gamma_{t-1}') &= Vec\left\{\Sigma\left[\left(I_{n^2}+\Lambda_n\right)\left(I_n\otimes\left(\mu+\Phi X_{t-1}\right)\right)\right]'\right\} \\
&= Vec\left\{\Sigma\left[I_n\otimes\left(\mu+\Phi X_{t-1}\right)'\right]\left(I_{n^2}+\Lambda_n\right)\right\} \\
&= \left[\left(I_{n^2}+\Lambda_n\right)\otimes\Sigma\right]Vec\left[I_n\otimes\left(\mu+\Phi X_{t-1}\right)'\right] && \text{(Prop. 2.A.1, } (iii)\text{)} \\
&= \left[\left(I_{n^2}+\Lambda_n\right)\otimes\Sigma\right]\left(I_n\otimes\Lambda_n\right)\left[Vec(I_n)\otimes\left(\mu+\Phi X_{t-1}\right)\right] && \text{(Prop. 2.A.1, } (ii)\text{)} \\
&= \left[\left(I_{n^2}+\Lambda_n\right)\otimes\Sigma\right]\left(I_n\otimes\Lambda_n\right)\left[Vec(I_n)\otimes I_n\right]\left(\mu+\Phi X_{t-1}\right) && \text{(Prop. 2.A.1, } (i-iv)\text{).}
\end{aligned}
$$

We turn now to the lower-right block of the conditional variance-covariance matrix of Z_t.

We have:

$$
\begin{aligned}
& Vec(\Gamma_{t-1}\Sigma\Gamma'_{t-1}) \\
= \; & Vec\left\{(I_{n^2}+\Lambda_n)\left[I_n\otimes(\mu+\Phi X_{t-1})\right]\Sigma\left[I_n\otimes(\mu+\Phi X_{t-1})'\right](I_{n^2}+\Lambda_n)\right\} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\left[I_n\otimes(\mu+\Phi X_{t-1})\right]\Sigma\left[I_n\otimes(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. 2.A.1, } (iii)\text{)} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\left[\Sigma\otimes(\mu+\Phi X_{t-1})\right]\left[I_n\otimes(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. 2.A.1, } (v)\text{)} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad Vec\left\{\Sigma\otimes\left[(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right]\right\} && \text{(Prop. 2.A.1, } (v)\text{)} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\}\right] && \text{(Prop. 2.A.1, } (ii)\text{)} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]Vec\left[Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\}\times Vec(\Sigma)'\right] && \text{(Prop. 2.A.1, } (i)\text{)} \\
= \; & \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\times \\
& \quad \left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes I_{n^2}\right]Vec\left\{(\mu+\Phi X_{t-1})(\mu+\Phi X_{t-1})'\right\} && \text{(Prop. 2.A.1, } (iv)\text{)}
\end{aligned}
$$

Finally we obtain the affine formulae for the four blocks of the conditional variance-covariance matrix $\widetilde{\Sigma}_{t-1}^{(i,j)}$ for $i,j=\{1,2\}$:

$$
\begin{aligned}
Vec\left(\widetilde{\Sigma}_{t-1}^{(1,1)}\right) &= Vec(\Sigma) \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(1,2)}\right) &= \left[\Sigma\otimes(I_{n^2}+\Lambda_n)\right]\left[Vec(I_n)\otimes I_n\right]\left\{\mu+\widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(2,1)}\right) &= \left[(I_{n^2}+\Lambda_n)\otimes\Sigma\right](I_n\otimes\Lambda_n)\left[Vec(I_n)\otimes I_n\right]\left\{\mu+\widetilde{\Phi}_1 Z_{t-1}\right\} \\
Vec\left(\widetilde{\Sigma}_{t-1}^{(2,2)}\right) &= \left[(I_{n^2}+\Lambda_n)\otimes(I_{n^2}+\Lambda_n)\right]\left[I_n\otimes\Lambda_n\otimes I_n\right]\left[Vec(\Sigma)\otimes I_{n^2}\right]\left\{\mu\otimes\mu+\widetilde{\Phi}_2 Z_{t-1}\right\} \\
&\quad + \left[I_{n^2}\otimes(I_{n^2}+\Lambda_n)\right]Vec(\Sigma\otimes\Sigma),
\end{aligned}
$$

where $\widetilde{\Phi}_1$ and $\widetilde{\Phi}_2$ are respectively the upper and lower blocks of $\widetilde{\Phi}$, thus $\widetilde{\Phi}_1=\begin{pmatrix}\Phi & 0\end{pmatrix}$ and $\widetilde{\Phi}_2=\begin{pmatrix}\mu\otimes\Phi+\Phi\otimes\mu & \Phi\otimes\Phi\end{pmatrix}$.

It should be noted that the computation of $Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]$ – i.e. the analytical expressions of ν and Ψ – involves, in theory, a permutation of the previous vectorized-blocks formulae ; however, we describe hereafter a simple and pragmatic method to reconstruct $Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]$ in the QKF algorithm:

1. Use the formulae of Proposition 2.3.2 to construct the four vectorized blocks of $\widetilde{\Sigma}(Z_{t-1})$ as explicit affine functions of Z_{t-1} (or $\widetilde{\Sigma}(Z_{t-1|t-1})$ as affine functions of $Z_{t-1|t-1}$ in the QKF algorithm).
2. Reconstruct the square matrix $\widetilde{\Sigma}(Z_{t-1})$ from the previous vectorized blocks.
3. Vectorize the reconstructed matrix.

Using the aforementioned method does not require an analytical expression of ν and Ψ and is a fast technique to calculate both the conditional and unconditional variances in the algorithm.

2.E Unconditional moments of Z_t

Proposition 2.3.3 We have:

$$\begin{bmatrix} \mathbb{E}(Z_t) \\ Vec[\mathbb{V}(Z_t)] \end{bmatrix} = \begin{pmatrix} \widetilde{\mu} \\ \nu \end{pmatrix} + \begin{pmatrix} \widetilde{\Phi} & 0 \\ \Psi & \widetilde{\Phi} \otimes \widetilde{\Phi} \end{pmatrix} \begin{bmatrix} \mathbb{E}(Z_{t-1}) \\ Vec[\mathbb{V}(Z_{t-1})] \end{bmatrix}$$

Proof. The first set of equation is immediately obtained from the state-space representation. For the second set, the variance decomposition writes:

$$\begin{aligned} \mathbb{V}(Z_t) &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \mathbb{V}\left[\mathbb{E}(Z_t|\underline{Z_{t-1}})\right] \\ &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \mathbb{V}(\widetilde{\mu} + \widetilde{\Phi} Z_{t-1}) \\ &= \mathbb{E}\left[\mathbb{V}(Z_t|\underline{Z_{t-1}})\right] + \widetilde{\Phi}\mathbb{V}(Z_{t-1})\widetilde{\Phi}' \\ &= \mathbb{E}\left[\widetilde{\Sigma}(Z_{t-1})\right] + \widetilde{\Phi}\mathbb{V}(Z_{t-1})\widetilde{\Phi}' \\ \Longrightarrow \quad & Vec\left[\mathbb{V}(Z_t)\right] = \mathbb{E}\left\{Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]\right\} + (\widetilde{\Phi} \otimes \widetilde{\Phi}) Vec\left[\mathbb{V}(Z_{t-1})\right] \end{aligned}$$

Denoting $Vec[\widetilde{\Sigma}(Z_{t-1})]$ by $\nu + \Psi Z_{t-1}$ we get:

$$\mathbb{E}\left\{Vec\left[\widetilde{\Sigma}(Z_{t-1})\right]\right\} = \nu + \Psi\mathbb{E}(Z_{t-1})$$

and the result follows.

■

2.F Selection and duplication matrices

Definition 2.F.1. *Let P be a $(n \times n)$ symmetric matrix. Let us define a partition of $I_n = [u_n, U_n]$ where u_n is the first column of I_n and U_n is the $(n \times (n-1))$ other sub-matrix. Let Q_n be a $(n^2 \times n^2)$ matrix defined as $Q_n = (Q_{1,n}, Q_{2,n})$ such that:*

$$Q_{1,n} = I_n \otimes u_n \qquad \text{and} \qquad Q_{2,n} = I_n \otimes U_n$$

A duplication matrix G_n and a selection matrix H_n are such that:

$$\begin{aligned} Vec(P) &= G_n Vech(P) \\ Vech(P) &= H_n Vec(P) \end{aligned}$$

and can be expressed recursively by:

$$G_{n+1} = Q_{n+1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_n & 0 \\ 0 & I_n & 0 \\ 0 & 0 & G_n \end{bmatrix} \qquad \text{and} \qquad H_{n+1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & H_n \end{bmatrix} Q'_{n+1}$$

with $G_1 = H_1 = Q_1 = 1$. These definitions can be found in Magnus and Neudecker [1980] or Harville [1997].

2.G Ekf and Ukf general algorithms

Let us consider a state-space model with non-linear transition and measurement equations.

$$X_t = f_t(X_{t-1}) + g_t(X_{t-1})\varepsilon_t \tag{2.11}$$
$$Y_t = h_t(X_t) + d_t(X_t)\eta_t \tag{2.12}$$

where f_t, g_t, h_t, d_t are functions of $\underline{Y_{t-1}}$ and possibly a vector of exogenous variables. Also, $(\varepsilon_t', \eta_t')' \sim \mathcal{II}\mathcal{N}(0, I)$. We use the following notations:

$$\begin{aligned} F_t &= \frac{\partial f_t}{\partial x_{t-1}'}(\widehat{X}_{t-1|t-1}), \quad H_t = \frac{\partial h_t}{\partial x_t'}(\widehat{X}_{t|t-1}) \\ F_{i,t}^{(2)} &= \frac{\partial^2 f_{i,t}}{\partial x_{t-1}\partial x_{t-1}'}(\widehat{X}_{t-1|t-1}), \quad H_{i,t}^{(2)} = \frac{\partial^2 h_t}{\partial x_t \partial x_t'}(\widehat{X}_{t|t-1}) \\ G_t &= g_t(\widehat{X}_{t-1|t-1}), \quad \text{and} \quad D_t = d_t(\widehat{X}_{t|t-1}) \end{aligned}$$

Let us also denote by $e_i^{(k)}$ the vector of size k whose components are equal to 0 except the i^{th} one which is equal to 1. The EKF1 and EKF2 algorithms are respectively given in Tables 2.7 and 2.8. Keeping the same notations, Table 2.9 presents the recursions of the UKF.

Table 2.7: EKF1 algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P_{0\|0}^X = \mathbb{V}(X_0)$.
State prediction:	$X_{t\|t-1}$	$f_t(X_{t-1\|t-1})$
	$P_{t\|t-1}^X$	$F_t P_{t-1\|t-1}^X F_t' + G_t G_t'$
Measurement prediction:	$Y_{t\|t-1}$	$h_t(X_{t\|t-1})$
	$M_{t\|t-1}$	$H_t P_{t\|t-1}^X H_t' + D_t D_t'$
Gain:	K_t	$P_{t\|t-1}^X H_t' M_{t\|t-1}^{-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P_{t\|t}^X$	$P_{t\|t-1}^X - K_t M_{t\|t-1} K_t'$

Note: See Jazwinski [1970], Anderson and Moore [1979], or Gelb [1974] for a proof of the recursions.

Table 2.8: EKF2 algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$.
State prediction:	$X_{t\|t-1}$	$f_t(X_{t-1\|t-1}) + \frac{1}{2}\sum_{i=1}^{n} e_i^{(n)} \operatorname{Tr}\left(F^{(2)}_{i,t} P^X_{t-1\|t-1}\right)$
	$P^X_{t\|t-1}$	$F_t P^X_{t-1\|t-1} F_t' + \frac{1}{2}\sum_{i,j=1}^{n} e_i^{(n)} \operatorname{Tr}\left(F^{(2)}_{i,t} P^X_{t-1\|t-1} F^{(2)}_{j,t} P^X_{t-1\|t-1}\right) e_j^{(n)\prime} + G_t G_t'$
Measurement prediction:	$Y_{t\|t-1}$	$h_t(X_t\|t-1) + \frac{1}{2}\sum_{k=1}^{m} e_k^{(m)} \operatorname{Tr}\left(H^{(2)}_{k,t} P^X_{t\|t-1}\right)$
	$M_{t\|t-1}$	$H_t P^X_{t\|t-1} H_t' + \frac{1}{2}\sum_{k,l=1}^{n} e_k^{(n)} \operatorname{Tr}\left(H^{(2)}_{k,t} P^X_{t\|t-1} H^{(2)}_{l,t} P^X_{t\|t-1}\right) e_l^{(n)\prime} + D_t D_t'$
Gain:	K_t	$P^X_{t\|t-1} H_t' M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: See Athans, Wishner, and Bertolini [1968] or Maybeck [1982] for a proof of the recursions.

Table 2.9: UKF algorithm in the general non-linear case

Initialize:		$X_{0\|0} = \mathbb{E}(X_0)$ and $P^X_{0\|0} = \mathbb{V}(X_0)$ and choose (α, κ, β).
Sigma points:		$\left\{\mathcal{X}_{i,t-1\|t-1}(X_{t-1\|t-1}, P^X_{t-1\|t-1})\right\}_{i=\{1,\dots,2n\}}$ according to Definition 2.4.1.
State prediction:	$X_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i \, f_t(\mathcal{X}_{i,t-1\|t-1})$
	$P^X_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[f_t(\mathcal{X}_{i,t-1\|t-1}) - X_{t\|t-1}\right] \left[f_t(\mathcal{X}_{i,t-1\|t-1}) - X_{t\|t-1}\right]' + G_t G_t'$
Sigma points:		$\left\{\mathcal{X}_{i,t\|t-1}(X_{t\|t-1}, P^X_{t\|t-1})\right\}_{i=\{1,\dots,2n\}}$ according to Definition 2.4.1.
Measurement prediction:	$Y_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i \, h_t(\mathcal{X}_{i,t\|t-1})$
	$M_{t\|t-1}$	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right] \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' + D_t D_t'$
Gain:	K_t	$\sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\mathcal{X}_{i,t\|t-1} - X_{t\|t-1}\right] \left[h_t(\mathcal{X}_{i,t\|t-1}) - Y_{t\|t-1}\right]' M^{-1}_{t\|t-1}$
State updating:	$X_{t\|t}$	$X_{t\|t-1} + K_t(Y_t - Y_{t\|t-1})$
	$P^X_{t\|t}$	$P^X_{t\|t-1} - K_t M_{t\|t-1} K_t'$

Note: See Julier, Uhlmann, and Durrant-Whyte [2000], Julier [2002], Julier [2003], or Julier and Uhlmann [2004] for proofs of the recursions.

2.H The UKF in a linear state-space model

Let us consider a linear state-space model which is given by Equations (2.1a) and (2.1b) putting all the $C^{(k)}$s to 0. Taking the notations of Table 2.9, we have $f_t(X) = \mu + \Phi X$ and $h_t(X) = A + BX$. As for $i = \{1, \dots, 2n\}$, all the weights are equal, the sigma points are symmetrical and $\sum_{i=0}^{2n} \mathcal{W}_i = 1$, we have:

$$
\begin{aligned}
X_{t|t-1} &= \sum_{i=0}^{2n} \mathcal{W}_i \left(\mu + \Phi \mathcal{X}_{i,t-1|t-1}\right) \\
&= \left(\sum_{i=0}^{2n} \mathcal{W}_i\right) \mu + \Phi \left(\sum_{i=0}^{2n} \mathcal{W}_i \mathcal{X}_{i,t-1|t-1}\right) \\
&= \mu + \Phi X_{t-1|t-1} \\
P_{t|t-1} &= \sum_{i=0}^{2n} \mathcal{W}_i^{(c)} \left[\Phi(\mathcal{X}_{i,t-1|t-1} - X_{t-1|t-1})\right] \left[\Phi(\mathcal{X}_{i,t-1|t-1} - X_{t-1|t-1})\right]' + G_t G_t' \\
&= \sum_{i=1}^{n} \mathcal{W}_i^{(c)} \Phi \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_i \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_i' \Phi' \\
&\quad + \sum_{i=n+1}^{2n} \mathcal{W}_i^{(c)} \Phi \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_{i-n} \left(\sqrt{(n+\lambda) P^X_{t-1|t-1}}\right)_{i-n}' \Phi' + G_t G_t' \\
&= 2 \sum_{i=1}^{n} \frac{(n+\lambda)}{2(\lambda+n)} \Phi \left(\sqrt{P^X_{t-1|t-1}}\right)_i \left(\sqrt{P^X_{t-1|t-1}}\right)_i' \Phi' + G_t G_t' \\
&= \Phi P_{t-1|t-1} \Phi' + G_t G_t'
\end{aligned}
$$

which proves the exact matching of the UKF and the Kalman filter for the state prediction phase. The same argument holds by linearity for the measurement prediction phase. In the linear case, the UKF shows exactly the same recursions that the linear Kalman filter, whatever the values of (α, κ, β).

2.I The UKF in a quadratic state-space model: scalar case

Let us consider the quadratic state-space model given by Equations (2.1a) and (2.1b). Let us set the vector of tuning parameters $(\alpha, \kappa, \beta) = (1, 0, 2)$ and $n = m = 1$. From Appendix 2.H, we know that the state prediction phase is exactly the same as in the linear Kalman filter, and is a fortiori the same as in the EKF2. Let us prove that the measurement prediction phase is the same for both filters for those values of (α, κ, β).

First, those tuning parameters imply $\lambda = 0$, thus:

$$\mathcal{X}_{i,t|t-1} = \begin{cases} X_{t|t-1} & \text{for} \quad i=0 \\ X_{t|t-1} + \sqrt{P^X_{t|t-1}} & \text{for} \quad i=1 \\ X_{t|t-1} - \sqrt{P^X_{t|t-1}} & \text{for} \quad i=2, \end{cases} \qquad \mathcal{W}_i = \begin{cases} 0 & \text{for} \quad i=0 \\ 1/2 & \text{for} \quad i \neq 0 \end{cases}$$

$$\mathcal{W}_i^{(c)} = \begin{cases} 2 & \text{for} \quad i=0 \\ 1/2 & \text{for} \quad i \neq 0 \end{cases}$$

Then, using the recursion of the UKF algorithm, we obtain:

$$\begin{aligned}
Y_{t|t-1} &= \frac{1}{2}\left[h(X_{t|t-1} + \sqrt{P^X_{t|t-1}}) + h(X_{t|t-1} - \sqrt{P^X_{t|t-1}})\right] \\
&= \frac{1}{2}\left\{2A + B\left(2X_{t|t-1}\right) + C\left[\left(X_{t|t-1} + \sqrt{P^X_{t|t-1}}\right)^2 + \left(X_{t|t-1} - \sqrt{P^X_{t|t-1}}\right)^2\right]\right\} \\
&= A + BX_{t|t-1} + CX^2_{t|t-1} + CP^X_{t|t-1} \\
&= h(X_{t|t-1}) + CP^X_{t|t-1} \\
M_{t|t-1} &= 2\left[h(X_{t|t-1}) - Y_{t|t-1}\right]^2 \\
&\quad + \frac{1}{2}\left\{\left[h(X_{t|t-1} + \sqrt{P^X_{t|t-1}}) - Y_{t|t-1}\right]^2 + \left[h(X_{t|t-1} - \sqrt{P^X_{t|t-1}}) - Y_{t|t-1}\right]^2\right\} + V \\
&= 2C^2(P^X_{t|t-1})^2 + \frac{1}{2}\Bigg\{\left[B\sqrt{P^X_{t|t-1}} + C\left(P^X_{t|t-1} + 2X_{t|t-1}\sqrt{P^X_{t|t-1}} - P^X_{t|t-1}\right)\right]^2 \\
&\quad + \left[-B\sqrt{P^X_{t|t-1}} + C\left(P^X_{t|t-1} - 2X_{t|t-1}\sqrt{P^X_{t|t-1}} - P^X_{t|t-1}\right)\right]^2\Bigg\} + V \\
&= 2C^2(P^X_{t|t-1})^2 + V + \frac{1}{2}\Bigg\{ \\
&\quad B^2P^X_{t|t-1} + C^2\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)^2 + 2CB\sqrt{P^X_{t|t-1}}\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right) \\
&\quad + B^2P^X_{t|t-1} + C^2\left(-2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)^2 + 2CB\sqrt{P^X_{t|t-1}}\left(2X_{t|t-1}\sqrt{P^X_{t|t-1}}\right)\Bigg\} \\
&= 2C^2(P^X_{t|t-1})^2 + V + B^2P^X_{t|t-1} + 4C^2X^2_{t|t-1}P^X_{t|t-1} + 2BCX_{t|t-1}P^X_{t|t-1} \\
&= \mathcal{G}^2_{t|t-1}P^X_{t|t-1} + 2C^2(P^X_{t|t-1})^2 + V
\end{aligned}$$

Both $Y_{t|t-1}$ and $M_{t|t-1}$ yield the same result as in the EKF2 recursions. Let us now turn to the Kalman gain computation.

$$\begin{aligned}
K_t &= \frac{1}{2}\left\{\left[\sqrt{P^X_{t|t-1}}\sqrt{P^X_{t|t-1}}\left(B + 2CX_{t|t-1}\right) - \sqrt{P^X_{t|t-1}}\sqrt{P^X_{t|t-1}}\left(-B - 2CX_{t|t-1}\right)\right]\right\} M^{-1}_{t|t-1} \\
&= P^X_{t|t-1}\left(B + 2CX_{t|t-1}\right)M^{-1}_{t|t-1} \\
&= P^X_{t|t-1}\mathcal{G}_{t|t-1}M^{-1}_{t|t-1}
\end{aligned}$$

which is also the same gain as in the EKF2. Therefore, for $(\alpha, \kappa, \beta) = (1, 0, 2)$ and scalar transition and measurement equations, The UKF and the EKF2 possess exactly the same

recursions.

Chapter 3

A QTSM for Disentangling Credit and Liquidity Risks in Interbank Spreads

This chapter is based on the article "Credit and Liquidity in Interbank Rates: A Quadratic Approach" of Dubecq, Monfort, Renne, and Roussellet [2014].

Abstract

A bank that lends on the unsecured market requires compensations for facing (a) the risk of default of the borrowing bank (credit risk) and (b) the risk of its own possible future funding needs (liquidity risk). In this paper, we propose a quadratic term-structure model of spreads between unsecured interbank rates and risk-free ones. Our approach allows us to decompose the whole term structure of spreads into credit and liquidity components. Our no-arbitrage econometric framework makes it possible to identify risk premia associated with each of these two risks. Our results shed a new light on the effects of unconventional monetary policy carried out in the Eurosystem. In particular, our findings suggest that most of the recent easing in the euro interbank market comes from a decrease in liquidity-related risk premia.

Résumé

Entre 2007 et 2008, la perception du risque interbancaire, mesurée par l'écart entre les taux des emprunts interbancaires non sécurisés et sécurisés, est passée de proche de zéro à un niveau très élevé. Ce saut a été particulièrement marqué au moment de la faillite de la banque Lehman Brothers, en septembre 2008. Dans ce contexte, les banques centrales ont eu recours à des mesures dites non conventionnelles (MNC) afin de restaurer les mécanismes de transmission de la politique monétaire. Pré-2014, trois différentes phases de mesures non conventionnelles ont été mises en œuvre : le *Securities Market Program* (SMP) en mai 2010 et août 2011 consistant en l'achat stérilisé d'obligations d'état, les *Very Long Term Refinancing Operations* (VLTRO) en décembre 2011 et février 2012 visant à fournir des liquidités aux banques à taux réduit, et les *Outright Monetary Transactions* (OMT) en juillet 2012, assurant l'intervention illimitée de la BCE sur le marché secondaire des obligations d'État si besoin est et conditionnellement à des contraintes budgétaires.

Deux types de risque sont visés par ces mesures : le risque de crédit et le risque de liquidité. Du point de vue du prêteur, la banque débitrice peut faire défaut et ne pas rembourser son emprunt. Par ailleurs, lorsqu'elle prête sur le marché interbancaire, une banque se prive de liquidités dont elle pourrait avoir besoin pendant la durée du prêt pour faire face à de possibles besoins de financement. Les risques de crédit et de liquidité sont tous les deux pris en compte par l'entité prêteuse lors de la négociation des conditions de prêt : plus les risques sont forts, et plus le taux demandé à l'emprunteur est élevé. Ces deux risques se traduisent par des compensations spécifiques présentes dans les taux auxquels les banques se prêtent de façon non sécurisée (sans collatéral). Ces derniers taux sont supposés proches des taux Euribor (*Euro Interbank Offered Rate*), qui sont calculés pour différentes maturités (de quelques jours à quelques semaines).

Il existe des taux Euribor pour chaque maturité allant de une semaine à douze mois, qui sont influencés par ces risques de crédit et de liquidité. Ces deux sources de risque ne sont toutefois pas les seules à déterminer le niveau de ces taux. En effet, ce niveau dépend également des anticipations de taux d'intérêt à court terme : en l'absence de risques de crédit et de liquidité, le taux de l'Euribor serait égal au taux sans risque, ce dernier pouvant être vu comme la moyenne des taux courts qui courront d'ici la maturité considérée. Dans le cadre de notre analyse, le taux sans risque est donné par le taux des OIS (*Overnight Indexed Swap*).

Ce chapitre propose un modèle quadratique de structure par terme pour rendre compte de l'évolution des écarts entre les taux Euribor et OIS, les décomposer en risque de crédit et

liquidité et analyser les effets des MNC menées par la BCE sur chacune des composantes de risque. Notre modélisation prend en compte deux facteurs de risque inobservables pour l'économètre : l'un est relatif au risque de défaut, l'autre au risque de liquidité. Seuls ces facteurs sont endogènes et causent les fluctuations des écarts de taux Euribor-OIS pour chaque maturité. Ils sont construits à partir de variables que nous savons liées aux risques de défaut et de liquidité. Le facteur de crédit est fondé sur les CDS (*Credit Default Swaps*) de banques de la zone euro. Le facteur de liquidité est une combinaison de variables censées capturer différents aspects du risque de liquidité : la liquidité de marché d'une part (*KfW-Bund spread*) et la difficulté à se financer d'autre part (*Bank Lending Survey data*). Les écarts de taux sont modélisés comme des combinaisons quadratiques des facteurs de crédit et liquidité, que nous imposons positives en cohérence avec l'intuition économique.

Exprimé sous la forme espace-état linéaire-quadratique, le modèle peut être estimé à l'aide du *Filtre de Kalman Quadratique* (voir Chapitre 2). Ce modèle nous permet de décomposer les effets de chacun des facteurs de risque sur les écarts de taux Euribor-OIS pour chaque maturité. Deux types de décompositions sont réalisées : les écarts de taux Euribor-OIS observés sont divisés *(1)* en composante de crédit et liquidité, et *(2)* en composante anticipée et prime de risque. Les primes de risque sont des composantes qui seraient absentes de l'écart de taux si (a) les investisseurs étaient neutres aux risques ou si (b) la totalité des risques présents dans l'économie était diversifiable. Nos estimations montrent que la majorité des variations de l'écart de taux est attribuable à des variations du risque de liquidité, notamment au moment de la faillite de Lehman Brothers, fin 2008, et lors de la crise de la dette européenne, fin 2011. La composante de risque de crédit, quant à elle, croît de façon quasi monotone pendant la période. Les primes de risque correspondent à une part substantielle de l'écart de taux, ce qui suggère que les prêts interbancaires contiennent beaucoup de risque non diversifiable.

Si le SMP ne provoque quasiment aucun effet sur les écarts de taux, ni sur leur composition, l'annonce et la mise en place des VLTRO provoque une décroissance substantielle des écarts de taux. Cette baisse s'explique essentiellement par une quasi-disparition du risque de liquidité dans les taux interbancaires après l'annonce des VLTRO puis celle des OMT par Mario Draghi. De plus, les primes de risque ont pratiquement disparu des taux interbancaires à la suite de ces deux événements, soulignant l'importance des actions non conventionnelles de politique monétaire comme outil de couverture face à la nature systématique des risques interbancaires.

3.1 Introduction

Since the beginning of the financial crisis, the interbank market has been carefully scrutinized by commentators and policy-makers, both in Europe and in the US. This paper focuses on the spreads between the *Euro Interbank Offered Rates* (EURIBORs) and their risk-free counterparts, proxied by the *Overnight Indexed Swap* rates (OIS). This spread is considered as a crucial indicator at the very core of the financial crisis: it reveals not only banks' concerns regarding the credit risk of their counterparts, but also their own liquidity needs.

Disentangling those credit and liquidity effects has essential policy implications. If a rise in spreads reflects poor liquidity, policy measures should aim at improving funding facilities. On the other hand credit concerns should be treated by enhancing debtors' solvency (see Codogno, Favero, and Missale [2003]). This question is very important in the euro area, where most of the unconventional monetary operations conducted by the European Central Bank focused on the curbing of interbank risk (see Gonzales-Paramo [2011]). Many attempts have been made to provide a credit/liquidity decomposition of the interbank risk (see next Section), but whereas most studies reckon that liquidity risk has been an important driver of interbank yields during the last 5 years, there is no consensus on the precise size of these effects: Schwarz [2009] estimates that one third of the EURIBOR-OIS 1-month spread is linked to liquidity in January 2008, whereas Filipovic and Trolle [2013] find that nearly all the spread is liquidity-related at that date.

In this paper, we present a new technique to investigate credit and liquidity risks in interbank markets. Our method is based on a reduced-form no-arbitrage term-structure model of the EURIBOR-OIS spreads. Considering the whole term structure of spreads to perform such an exercise is important for at least two reasons. By including several maturities in our sample, we first improve the quality and precision of our model estimation. Second, the term structure dimension of our analysis can be exploited to identify the part of the spreads that corresponds to risk premia, thereby extending the existing literature on interbank risks. Risk premia are the components of yields or spreads that would not exist if (a) economic agents or investors – in our case, banks – were risk-neutral or (b) the risks involved in the considered asset were not systematic, i.e. if they could be diversified away (see e.g. Longstaff, Pan, Pedersen, and Singleton [2011]). Since (a) and (b) are not likely to hold in the case of euro-area interbank risks, risk premia are expected to be present in the EURIBOR-OIS spreads. At the same time, these risk premia cannot account for the whole spreads. Indeed, even if agents were risk-neutral, the EURIBOR-OIS spread would not be zero. In that case, the spreads would just equal expected losses

stemming from the total amount of risk – credit- and liquidity-related – that a bank faces when lending to another bank.

Our methodology aims at decomposing the spreads along two dimensions: credit vs. liquidity and expected vs. risk-premia parts. This is achieved with the use of a no-arbitrage framework involving credit and liquidity intensities. In order to clarify the interpretation of the latter, we propose a stylized interbank-market model where these intensities appear naturally. In this model, when a bank lends to another bank, it is exposed to two kinds of risks: the first corresponds to the default of the borrowing bank and the second pertains to the difficulty to meet potential liquidity needs the lending bank may face over the loan period. In order to respect the non-negativity of the intensities, we take them equal to non-negative quadratic functions of two Gaussian latent factors. Hence, our model belongs to the class of quadratic term-structure models (QTSM). Our identification scheme and the interpretation of the factors rely on credit and liquidity proxies. The estimation is performed using a recently-introduced quadratic Kalman filter (see Monfort, Renne, and Roussellet).

The model is estimated over a 6-year period, between August 2007 and September 2013. Both credit and liquidity components account for the fluctuations of the spreads over that period, with a higher average contribution of liquidity risks. Our results suggest further that both kinds of risk command substantial risk premia, pointing towards the systematic nature of credit and liquidity interbank risks. We illustrate how the existence of credit-risk premia translates into substantial differences between model-implied physical and risk-neutral probabilities of default.

The spreads' decomposition allows us to explore the consequences of unconventional monetary policies conducted by the ECB during this period. Our findings support the claim that the recent 3-year ECB loans to euro commercial banks (i.e. the *Very Long-Term Refinancing Operations*, or VLTROs) and the announcement of the still-unused ECB sovereign-bond purchase program (i.e. the *Outright Monetary Transactions*, or OMTs) have helped to reduce the perception of liquidity risk and its related risk premium. However, we find little evidence that the ECB large-scale asset-purchase programs of 2010 and 2011 (i.e. the *Securities Market Programs*, or SMP 1 & 2) have had any significant impact on the interbank risk. Eventually, we find that unconventional monetary policies have had very modest, if any, effects on the credit part of the spreads.

The remainder of the paper is organized as follows. Section 2 discusses the related literature. Section 3 presents the interest rate data and the proxies. Section 4 develops the

quadratic term-structure model. Section 5 describes the identification strategy and shows the estimation results. Section 6 performs the decomposition of EURIBOR-OIS spreads and discusses the impact of the ECB unconventional monetary policies; it also derives risk-premia-corrected default probabilities of banks. The last section concludes. Proofs are gathered in the Appendices.

3.2 Literature Review

In most term structure models, the authors assume that the default intensity and/or the short-term rate are affine functions of the underlying factors. A quadratic specification however possesses several advantages over the standard affine case. Constantinides [1992] shows that a standard QTSM with a specific quadratic short-term interest rate can generate positive yields for all maturities and more flexibility in the term structure to fit bond data. Leippold and Wu [2002] generalize the quadratic term structure models showing that this specification provides closed-form or semi closed-form formulas for bond pricing of most fixed-income derivatives. Ahn, Dittmar, and Gallant [2002] provide further empirical evidence that QTSM often outperforms the standard affine term structure specification (ATSM). Leippold and Wu [2007] study the joint behavior of exchanges rates and bond yields using QTSM models for Japan and the US. More recently, Andreasen and Meldrum [2011] and Kim and Singleton [2012] exploit the QTSM framework to model the term structure of interest rates in a context of extremely low monetary-policy rates. Turning to the credit literature, Hordahl and Tristani [2012] use a quadratic specification to model euro-area sovereign spreads, and Doshi, Ericsson, Jacobs, and Turnbull [2013] consider a quadratic intensity to price corporate credit default swaps. Our paper also adopts a quadratic approach in order to impose positivity of the risk intensities and spreads, and it takes advantage of a new well-suited technique to estimate the model, namely the quadratic Kalman filter (see Monfort, Renne, and Roussellet).

Our identification scheme follows several studies that rely on reduced-form no-arbitrage models to identify credit and liquidity components in the term structures of yields or spreads (e.g. Liu, Longstaff, and Mandell [2006], Feldhutter and Lando [2008], Longstaff, Mithal, and Neis [2005]). At the heart of these studies are credit/liquidity intensities whose fluctuations affect the whole term structure of spreads. As in Monfort and Renne [2014], the present paper allows for some dynamic interactions between credit and liquidity risks, consistently with the theoretical predictions of, among others, Goldstein and Pauzner [2005] or He and Xiong [2012].

Our paper also relates to the interbank spreads literature. A wide range of studies deals

with the determinants of interbank spreads: Taylor and Williams [2009] claim that counterparty risk was the main driver of the LIBOR-OIS spread, Michaud and Upper [2008] and Gyntelberg and Wooldridge [2008] find that credit and liquidity factors both played a role, while the results by Schwarz [2009] and Filipovic and Trolle [2013] suggest that liquidity risk has accounted for most of the LIBOR-OIS and EURIBOR-OIS spread variations over the period 2007-2009. In comparison, Smith [2010] emphasizes that most of the variation in the risk premia of interbank spreads is explained by credit risk. Finally, Angelini, Nobili, and Picillo [2011] highlight the main role of macro-factors to account for the dynamics of unsecured/secured money-market spreads. The measured impact of unconventional monetary policies is ambiguous: Taylor and Williams [2009] find no effects of the Fed's intervention in 2008, contrary to Christensen, Lopez, and Rudebusch [2014]. According to the latter, Fed's liquidity injections (TAF, for *Term Auction Facility*) reduced significantly the 3-month maturity interbank spread by about 70 basis points. Carpenter, Demiralp, and Eisenschmidt [2014] find that non-standard monetary-policy measures contributed to sustained lending activity by lowering funding volatility. Angelini, Nobili, and Picillo [2011] measure a modest impact of ECB exceptional 3-month refinancing operations, in contradiction with Abbassi and Linzert [2012]. Cecioni, Ferrero, and Secchi [2011] provide a comprehensive review of the quantitative assessment regarding the relative importance of the interbank spreads' drivers, as well as of the effects of unconventional monetary policies in the U.S. and euro-area interbank markets.

3.3 Interbank market rates and risks

3.3.1 The unsecured interbank rates

The interbank money market is at the heart of bank funding issues. It is an over-the-counter market (OTC) where interbank loans are negotiated with maturities ranging from one day to to twelve months. As banks do not possess the same characteristics and underlying risks, there is no uniqueness of interbank rates. Only the disaggregated rates are really representative of the funding issues of each institution. However, such data are not publicly available and in order to conduct an analysis on interbank risks, a more aggregated measure must be considered.

The Euro Interbank Offered Rate (EURIBOR) provides a daily measure of the interest rates at which banks can raise unsecured funds from other financial institutions in the euro wholesale money market, for maturities ranging from one week to twelve months. A daily survey is sent to a panel of 30 to 50 creditworthy banks in the Euro area; the question of the survey is: *what are the rates at which euro interbank term deposits are*

being offered within the Eurozone by one prime bank to another? The EURIBORs are then trimmed means of the contributed rates, the 15% of each tail being erased.

While there are no reliable data on volumes in term money markets, anecdotal evidence suggests that the financial crisis has resulted in a sharp decline in unsecured term money market volumes (see Eisenschmidt and Tapking [2009]). In spite of this, there is evidence that EURIBOR rates remain reliable proxies for bank funding costs. Typically, data collected from the ECB Short Term European Papers (STEP) database suggest that EURIBORs are very close to quotations of certificates of deposits issued by banks.[1] Moreover, using U.S. data, Kuo, Skeie, and Vickery [2012] find that public interbank yield data beyond Libor are moderately informative about bank funding costs.

The loans that underlie the EURIBOR expose lending banks to both credit and liquidity risks. First, these loans are unsecured, that is the lending bank does not receive collateral as protection against default by the borrowing one. Therefore, these rates carry some compensation for solvency issue that we refer to as credit risk. Second, through an interbank loan, a lending bank exposes its funds during the time-to-maturity of the loan although those funds might be needed to cover the bank's own shortfalls. Since an unsecured interbank loan is highly specific to the identity of both counterparties, its unwinding is a costly task. This is taken into consideration by the lending bank at the inception of the loan, which gives rise to an extra (liquidity-related) compensation in the loan rate.

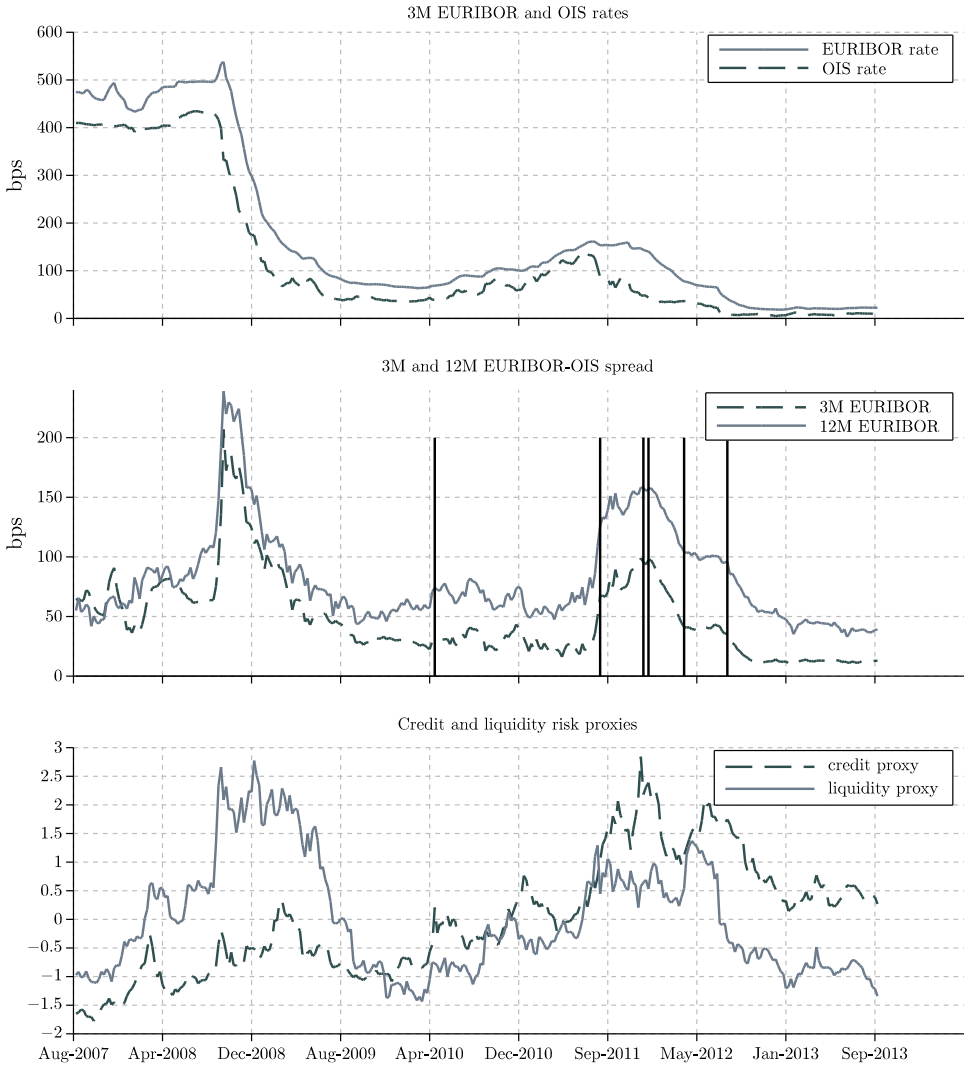
Figure 3.1 presents the evolution of the 3-month EURIBOR from August 2007 to September 2013. During the first year, the rate is stable around 500 basis points. The Lehman bankruptcy of September 2008 is followed by a sharp decline in EURIBOR of about 400 basis points, to 80 basis points. From mid-2010 onwards, the EURIBOR rises slowly to 150 basis points in September 2011 and decays to nearly 20 basis points during the recent period. Table 3.1 presents the descriptive statistics for 3, 6, 9, and 12-month EURIBORs.

3.3.2 The interbank *risk-free* rate

In this paper, the risk-free rates are proxied by the Overnight Indexed Swap (OIS) rates. An OIS is a fixed-for-floating interest rate swap with a floating rate leg indexed on overnight interbank rates, the EONIA (Euro OverNight Index Average) in the euro-area case. OIS have become especially popular hedging and positioning vehicles in euro finan-

1. Indeed, it appears that the average of the spreads between (a) the issuance yields for certificates of deposits with an initial maturity comprised between 101 and 200 days and (b) the 5-month EURIBOR rate was lower than 3 basis points over the 2008-2012 period.

Figure 3.1: Level of 3M rates and spreads



Notes: Top panel: plot of the 3M EURIBOR (light grey) and 3M OIS (dashed dark grey). Middle panel: plot of the 3M (dashed dark grey) and 12M (lighter grey) EURIBOR-OIS spreads. Units are in basis points. Bottom panel: credit (dashed dark grey) and liquidity (light grey) proxies; these proxies are demeaned and standardized. Time ranges from August 31, 2007 to September 13, 2013.

cial markets and grew significantly in importance during the financial turmoil of the last few years. The OIS curve is more and more seen by market participants as a proxy of the risk-free interbank yield curve.[2] As no principal is exchanged, the OIS requires nearly no immobilization of capital. Further, due to netting and credit-enhancement mechanisms (including call margins), the counterparty risk is limited in the case of a swap contract (Bomfim [2003]).

The upper panel of Figure 3.1 displays the 3-month OIS rate from August 2007 to Septem-

2. See e.g. Joyce, Lasaosa, Stevens, and Tong [2011] or Bank of International Settlements [2005].

ber 2013. While this chart shows that EURIBOR and OIS rates present strong common fluctuations, the middle panel also highlights that the spread between the two rates has undergone substantial variations over the last five years. In the next subsection, we discuss the term structure of the EURIBOR-OIS spreads.

3.3.3 Preliminary analysis of the EURIBOR-OIS spreads

Being mostly stable before August 2008, the spread abruptly increased during Lehman crisis until December 2008, the 3-month spread peaking at 200 basis points, where a slow decay begins (see Figure 3.1, middle panel). For sake of comparison, before summer 2007, the EURIBOR-OIS spread was around ten basis points; part of this deviation was accounted for by the fact that the EURIBOR is an offer rate while the OIS is a mid rate (average of bid and ask yields). Then, following a long stabilization period between August 2009 and 2010, a sharp rise stroke again in mid-2011. Since the beginning of 2012, the EURIBOR-OIS spreads have decreased, alternating between a linear decreasing trend and stable phases.

Table 3.1: Descriptive statistics of EURIBOR and OIS rates

	min	max	amplitude	mean	std	skewness	excess kurtosis
			bps				
EURIBOR 3M	18.4	538.1	519.7	172.0	165.1	1.12	−0.35
EURIBOR 6M	29.4	543.1	513.7	190.8	158.9	1.10	−0.32
EURIBOR 9M	38.8	546.3	507.5	202.3	155.1	1.07	−0.31
EURIBOR 12M	47.4	549.3	501.9	213.0	152.1	1.06	−0.29
OIS 3M	4.5	434.6	430.1	123.6	145.7	1.29	−0.03
OIS 6M	2.35	442.85	440.5	125.1	144.9	1.30	0.03
OIS 9M	−0.5	453.5	454	127.9	143.7	1.29	0.05
OIS 12M	−1.1	465.3	466.4	131.2	142.2	1.27	0.07
Spread 3M	9.9	206.9	197	48.4	34.9	1.61	3.37
Spread 6M	19.6	222.5	202.9	65.7	36.5	1.62	3.44
Spread 9M	26.8	227.9	201.1	74.4	38.1	1.63	3.05
Spread 12M	32.9	239	206.1	81.8	40.0	1.54	2.38

Notes: Those figures are computed with weekly data ranging from August, 31 2007 to September, 13 2013.

Standard descriptive statistics of spreads are provided in Table 3.1. The means of spreads increase with respect to maturity, from 48 basis points (3-month maturity) to 82 basis points (12-month maturity). This indicates a positive slope in the term structure of

spreads, which is graphically illustrated by the middle panel in Figure 3.1: except at the very beginning of the sample, the 12-month spread is always larger than the 3-month spread, up to around 50 basis points in late 2011.

Whereas the standard deviations are respectively stable and decreasing with maturity for OIS and EURIBOR rates, the standard deviations of spreads slightly increase with maturity. Regarding higher-order moments, Table 3.1 indicates that spreads are more positively skewed than the rates in level; also, contrary to the latter, spreads are heavy-tailed (positive excess kurtosis). The heavy-tail behavior is typically illustrated during the Lehman crisis on Figure 3.1, where both 3-month and 12-month spreads peak to 207 and 239 basis points, respectively. These levels are about 4 standard-deviation far from their respective sample means.

A principal component analysis performed on the four EURIBOR-OIS spreads proves that the first two principal components captures most of spread fluctuations, explaining 99.7% of the whole variance of the spreads (96.4% and 3.3% for the first and second principal components respectively). This suggest that two factors are sufficient to capture the bulk of spread fluctuations.

3.3.4 Credit and liquidity proxies

In this subsection, we introduce credit and liquidity proxies on which we will base our identification of credit and liquidity parts of EURIBOR-OIS spreads. In the next sections, we relate these proxies to the factors driving our term structure model.

The liquidity proxy we will use in our term-structure model is the first principal component of a set of three liquidity-related variables. These variables are chosen in order to capture different aspects of liquidity pricing, namely market and funding liquidity.[3] Specifically, our first two proxies are mostly related to market liquidity whereas the last one is mostly related to funding liquidity. Nearly 60% of the total variance is explained by the first principal component.

- A first liquidity-pricing factor is the *KfW-Bund spread* (5-year maturity). KfW is a public German agency. KfW bonds are guaranteed by the Federal Republic of Germany. Hence, they embed the same credit quality as their sovereign counterparts, the so-called Bunds. KfW bonds being less liquid than their sovereign counterpart, the KfW-Bund spread essentially reflect liquidity-pricing effects.[4]

3. While market liquidity is reflected by the difference between market and fundamental value of an asset, funding liquidity relates to the scarcity of capital (see Brunnermeier and Pedersen [2009]).

4. See Schwarz [2009], Monfort and Renne [2013] or Schuster and Uhrig-Homburg [2012]). In the

- A second liquidity factor is the Tbill-repo spread, computed as the yield differential between the 3-month German T-bill and the 3-month general-collateral repurchase agreement rate (repo). From an investor point of view, the credit qualities of the two instruments are comparable (as argued by Liu, Longstaff, and Mandell [2006]). The differential between the two rates corresponds to the convenience yield, that can be seen as a premium that one is willing to pay when holding highly-liquid Treasury securities (see Feldhutter and Lando [2008]).
- A third factor is based on the Bank Lending Survey conducted by the ECB on a quarterly basis. This survey is addressed to senior loan officers of a representative sample of around 90 euro-area banks; it addresses issues such as credit standards for approving loans as well as credit terms and conditions applied to enterprises and households. Our indicator is based on the following specific question of the survey: *Over the past three months, how has your bank's liquidity position affected the credit standards as applied to the approval of loans or credit lines to enterprises?*.[5]

The credit proxy is the first principal component of a set of 36 Euro-zone bank CDS. We use 5-year CDS denominated in USD since these are the most traded – and therefore the most liquid – ones. Eight are German, six Italian, five Spanish, four French, four Dutch, three Irish, three Portuguese, two Austrian, and one Belgian. Nearly 72% of the total variance is explained by the first principal component.

3.4 The model

In this section, we propose a reduced-form asset pricing model of the term structure of EURIBOR-OIS spreads. Credit and liquidity risks are introduced by means of the specification of two respective risk intensities. The model implies that spreads are quadratic functions of factors; the estimation of the latter be detailed in the next section.

3.4.1 Notations

We consider the pool of the N banks of the EURIBOR panel. At date t, market participants get the new information $w_t = \{r_t, X_t', d_t', \ell_t'\}'$, where r_t is the short-term risk-free rate between dates t and $t+1$, $X_t = (x_{c,t}, x_{l,t})'$ is a (2×1) vector whose components are respectively credit- and liquidity-related factors, and where d_t and ℓ_t are two N-dimensional vectors of binary variables $d_t^{(i)}$ and $\ell_t^{(i)}$, with $i \in \{1, \ldots, N\}$. While $d_t^{(i)}$ defines

same spirit, Longstaff [2004] computes liquidity premia based on the spread between U.S. Treasuries and government-guaranteed bonds issued by Refcorp.

5. The respondents can answer ++, +, 0, − or −− to that question. We compute the proportion of − and −− as a ratio of total answers. To obtain weekly series, we assign the same value to all weeks in a quarter (step function).

the credit state of bank i at date t, $\ell_t^{(i)}$ defines its liquidity status. In the following, we will make more precise the implications of defaults ($d_t^{(i)} = 1$) or liquidity shocks ($\ell_t^{(i)} = 1$) for interbank-loan payments. For any random vector z_t, we will use the notation $\underline{z_t} = (z_t, z_{t-1}, \dots)$.

3.4.2 Historical and risk-neutral dynamics

Let us first define the historical dynamics of w_t. Following Berndt, Douglas, Duffie, Ferguson, and Schranz [2005], Pan and Singleton [2008] or Longstaff, Pan, Pedersen, and Singleton [2011], we assume that the processes r_t and $(X_t', d_t', \ell_t')'$ are independent. As will be seen in Subsection 3.4.4, we will not have to specify the dynamics of r_t. We assume further that $(d_t', \ell_t')'$ does not cause X_t and that X_t has a VAR(1) dynamics defined by:

$$X_t = \mu + \Phi X_{t-1} + \epsilon_t, \tag{3.1}$$

where $\epsilon_t = (\epsilon_{c,t}, \epsilon_{\ell,t})'$ is a Gaussian white noise such that $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{V}(\epsilon_t) = I$.

Finally, we assume that, conditionally on $(\underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}})$, the vectors d_t and ℓ_t are independent, and such that:

$$\mathbb{P}(d_t^{(j)} = 1 | \underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}}) = \begin{cases} \exp[-\lambda_c(x_{c,t})] & \text{if} \quad d_{t-1}^{(j)} = 0 \\ 0 \quad \text{if} \quad d_{t-1}^{(j)} = 1 \end{cases}$$

$$\mathbb{P}(\ell_t^{(j)} = 1 | \underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}}) = \begin{cases} \exp[-\lambda_\ell(x_{\ell,t})] & \text{if} \quad \ell_{t-1}^{(j)} = 0 \\ 0 \quad \text{if} \quad \ell_{t-1}^{(j)} = 1. \end{cases}$$

The intensity functions $\lambda_c(x_{c,t})$ and $\lambda_\ell(x_{\ell,t})$ will be specified in Subsection 3.4.5.

In order to derive the risk-neutral (R.N.) dynamics of w_t, we introduce a stochastic discount factor between $t-1$ and t of the form:

$$M_{t-1,t} = \exp\left[\Gamma_{t-1}'\epsilon_t - \frac{1}{2}\Gamma_{t-1}'\Gamma_{t-1} - r_{t-1} + g(\underline{r_t})\right],$$

where $g(\underline{r_t})$ is any function such that $\mathbb{E}_{t-1}[g(\underline{r_t})] = 1$, and

$$\Gamma_{t-1} = \Gamma_0 + \Gamma X_{t-1}$$

where Γ_0 is a (2×1) vector and Γ is a (2×2) matrix. it is easily seen (see Appendix 3.A) that with such a SDF, we have the following properties:

- r_t and (X'_t, d'_t, ℓ'_t) are independent in the R.N. world.
- The R.N. dynamics of X_t is defined by:

$$X_t = \mu^* + \Phi^* X_{t-1} + \epsilon^*_t, \tag{3.2}$$

  where $\mu^* = \mu + \Gamma_0$, $\Phi^* = \Phi + \Gamma$ and $\epsilon^*_t \sim IIN^{\mathbb{Q}}(0, I)$. In particular, (d'_t, ℓ'_t) does not cause X_t in the R.N. world.
- The R.N. conditional distribution of (d'_t, ℓ'_t) given $(\underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}})$ is the same as in the historical world; in particular, the R.N. intensities $\lambda^{\mathbb{Q}}_c(x_{c,t})$ and $\lambda^{\mathbb{Q}}_\ell(x_{\ell,t})$ are the same functions as in the historical world; therefore we will denote them by $\lambda_c(x_{c,t})$ and $\lambda_\ell(x_{\ell,t})$.

3.4.3 Intensities and EURIBOR rates

We have assumed that the panel of banks is homogeneous, in the sense that, conditional on $(\underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}})$, the default probabilities and the probabilities of being affected by a liquidity shock are the same for all the banks of the EURIBOR panel. This assumption notably implies that, at each date t, there is a single rate prevailing for interbank unsecured loans between t and a future date $t + h$. This interest rate is denoted by $R^{EUR}_{t,h}$. By definition of this rate, an interbank loan between dates t and $t+h$ of unit face value provides the borrower with the amount $B(t, h) = \exp(-hR^{EUR}_{t,h})$ at date t. Note that the pricing formulas derived in this paper feature continuously-compounded interest rates: denoting by z a market-quoted interest rate and applying the money-market day-count convention (ACT/360), the corresponding continuously-compounded rate is given by $\ln(1+d\times z/360)\times 365/d$ where d is the residual maturity of the considered instrument, expressed in days.

Suppose that, at date t, bank i lends $B(t, h)$ to bank j for a period of length h. The maturity date is $t+h$ and, assuming no premature termination of the loan, the repayment is 1. Now, consider an intermediary date t^* (i.e. $t < t^* \leq t + h$). At date t^*, if bank j defaults or if bank i is hit by a liquidity shock, this terminates the interbank loan and the resulting payoffs are as follows:

- If bank j defaults at date t^* ($d^{(j)}_{t^*} = 1$), then bank i will not obtain full repayment at $t + h$. Instead, at date t^*, it recovers a fraction $\theta_c < 1$ of the "market value" of the loan that would have prevailed at date t^* in the absence of default. This market value corresponds to the face value of the loan discounted by the EURIBOR $R^{EUR}_{t^*,t+h}$. This set up builds on the "recovery at market value" assumption of Duffie and Singleton [1999].

– When bank i is hit by a liquidity shock at date t^* (i.e. $\ell_{t^*}^{(i)} = 1$), bank i has to find some cash in a limited period of time to meet an unexpected liquidity need. It may do so by negotiating a premature termination of the loan with bank j. The latter agrees, but at a discount: the repayment at date t^* is expressed as a fraction $\theta_l < 1$ of the aforementioned "market value" of the loan. Such a mechanism of costly liquidation is in the spirit of Ericsson and Renault [2006] or He and Xiong [2012].

In that context, the value of the loan at date $t+1$ writes:

$$B(t+1,h-1)\left\{(1-d_{t+1}^{(j)}+\theta_c d_{t+1}^{(j)})(1-\ell_{t+1}^{(i)}+\theta_l \ell_{t+1}^{(i)})\right\},$$

where $B(t+1,h-1)$ is the price prevailing in the absence of credit or liquidity event at date $t+1$. Since its price at t is $B(t,h)$, we get:

$$\begin{aligned} &B(t,h) = \exp(-r_t) \times \\ &\mathbb{E}^{\mathbb{Q}}\left[B(t+1,h-1)\left\{(1-d_{t+1}^{(j)}+\theta_c d_{t+1}^{(j)})(1-\ell_{t+1}^{(i)}+\theta_l \ell_{t+1}^{(i)})\right\}|\underline{w_t}\right], \end{aligned} \tag{3.3}$$

where $\mathbb{E}^{\mathbb{Q}}$ denotes the expectation under the risk-neutral (pricing) measure.

Given the definitions of $\lambda_c(x_{c,t})$ and $\lambda_\ell(x_{\ell,t})$, we have:

$$\left\{\begin{array}{rcl} \mathbb{E}^{\mathbb{Q}}(d_{t+1}^{(j)}|\underline{w_t}, X_{t+1}, r_{t+1}, d_t^{(j)}=0, \ell_t^{(i)}=0) &=& 1-\exp[-\lambda_c(x_{c,t+1})] \\ \mathbb{E}^{\mathbb{Q}}(\ell_{t+1}^{(i)}|\underline{w_t}, X_{t+1}, r_{t+1}, d_t^{(j)}=0, \ell_t^{(i)}=0) &=& 1-\exp[-\lambda_\ell(x_{\ell,t+1})]. \end{array}\right. \tag{3.4}$$

$\exp[-\lambda_c(x_{c,t+1})]$ and $\exp[-\lambda_\ell(x_{\ell,t+1})]$ are probabilities and, therefore, $\lambda_c(x_{c,t+1})$ and $\lambda_\ell(x_{\ell,t+1})$ must be positive at all times. When these intensities are small, they are close to the default probabilities and to the probabilities of being hit by the liquidity shock, respectively. Besides, a first order approximation yields:

$$\left\{\begin{array}{rcl} \mathbb{E}^{\mathbb{Q}}((1-\theta_c)d_{t+1}^{(j)})|\underline{w_t}, X_{t+1}) &=& 1-\exp[-(1-\theta_c)\lambda_c(x_{c,t+1})] \\ \mathbb{E}^{\mathbb{Q}}((1-\theta_\ell)\ell_{t+1}^{(i)})|\underline{w_t}, X_{t+1}) &=& 1-\exp[-(1-\theta_\ell)\lambda_\ell(x_{\ell,t+1})], \end{array}\right. \tag{3.5}$$

Introducing the total intensity $\lambda_t = (1-\theta_c)\lambda_c(x_{c,t}) + (1-\theta_\ell)\lambda_\ell(x_{\ell,t})$ in Equation (3.3) implies (see Appendix 3.B):

$$B(t,h) \;=\; \mathbb{E}_t^{\mathbb{Q}}\left[\exp(-r_t - \lambda_{t+1} - \cdots - r_{t+h-1} - \lambda_{t+h})\right]. \tag{3.6}$$

Since $B(t,h) = \exp(-hR_{t,h}^{EUR})$, we have:

$$R_{t,h}^{EUR} = -\frac{1}{h}\ln\left\{\mathbb{E}_t^{\mathbb{Q}}\left[\exp(-r_t - \lambda_{t+1} - \cdots - r_{t+h-1} - \lambda_{t+h})\right]\right\}. \tag{3.7}$$

3.4.4 OIS swap rates and the EURIBOR-OIS spreads

An OIS is an interest-rate derivative that allows for exchanges between a fixed-interest-rate cash flow and a variable-rate cash flow. More precisely, the floating leg of an OIS is indexed on the EONIA. At maturity, the payoff received by the fixed-rate payer is the difference between (a) the notional (W, say) inflated with the date-t OIS (fixed) rate (i.e. $W\exp\left\{hR_{t,h}^{OIS}\right\}$) and (b) the same notional capitalized with the realized short-term rates (i.e. $W\exp\left\{r_t + \ldots + r_{t+h-1}\right\}$). Note that the latter expression implicitly reckons that the OIS reference rate –that is the EONIA rate– corresponds to the risk-free rate r_t, thereby assuming that lending on the overnight interbank market preserves the lending bank from (i) liquidity and (ii) credit risk. The rationale behind (i) and (ii) are the following:

(i) By rolling its cash on the overnight market (at the EONIA rate), a bank is not exposed to the risk of having to liquidate longer-term investments upon the realization of the liquidity shock.

(ii) While the EONIA is an unsecured-transaction rate, the extremely-short maturity of these transactions substantially reduces the credit-risk exposure of the lending bank. This point is corroborated by a comparison of EURIBOR-OIS spreads with spreads between Repo rates – where credit-risk effects are kept at a minimum through collateralization schemes – and OIS rates: over 2007-2013, the mean absolute value of the 3-month Repo-OIS spread is about 10 times smaller than the one of the EURIBOR-OIS spread of the same maturity (the former being of a few basis points).

At the inception date of the swap, there is no cash-flow exchange between the two counterparties, that is, the discounted values of the two legs are initially the same:

$$W\mathbb{E}_t^{\mathbb{Q}}\left[\exp(hR_{t,h}^{OIS})\exp\left\{-r_t - \ldots - r_{t+h-1}\right\}\right] = W,$$

or:

$$R_{t,h}^{OIS} = -\frac{1}{h}\log\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-r_t - \ldots - r_{t+h-1}\right\}\right]. \tag{3.8}$$

Since, as in, e.g., Berndt, Douglas, Duffie, Ferguson, and Schranz [2005], Pan and Singleton [2008] or Longstaff, Pan, Pedersen, and Singleton [2011], we have assumed that the short-term risk-free interest rate and the intensity processes are independent under $\mathbb{Q}$,

denoting by $S(t,h)$ the EURIBOR-OIS spread of maturity h, we have:

$$S(t,h) = R^{EUR}_{t,h} - R^{OIS}_{t,h} = -\frac{1}{h}\log\left(\mathbb{E}^{\mathbb{Q}}_t\left[\exp\left(\sum_{i=1}^{h} -\lambda_{t+i}\right)\right]\right). \tag{3.9}$$

Equation (3.9) shows that, under these assumptions, the study of EURIBOR-OIS spreads does not require the modeling of the short-term risk-free interest rate r_t. In the following, we impose a factor structure and a specification for the modeling of both the credit and liquidity intensities to obtain pricing formulas for the interbank spreads.

3.4.5 Intensity specification

Now, it remains to specify the intensity functions λ_c and λ_ℓ and the factors $(x_{c,t}, x_{\ell,t})'$. In a preliminary analysis, whose results are not reported here for sake of brevity, we postulated a linear relationship between the intensities and the factors, within a standard Gaussian affine term-structure model. However, the results were not satisfying, the model clearly violating the non-negativity of spreads. The model-implied frequencies of generating negative spreads (i.e. considering their marginal densities) was huge and close to 50% for all maturities. This comes from the facts that (a), in such a model, the distribution of model-implied spreads is Gaussian and that (b) consistently with the high persistence of observed spreads, the resulting model-implied variance of the spreads is large. This failure illustrates the inappropriateness of Gaussian ATSM to model such spreads. Therefore, following Doshi, Ericsson, Jacobs, and Turnbull [2013] or Gouriéroux and Monfort [2008], we set a quadratic relationship between the intensities and the associated factors:

$$\lambda_c(x_{c,t}) = \Lambda_c x^2_{c,t} \quad \text{and} \quad \lambda_\ell(x_{\ell,t}) = \Lambda_\ell x^2_{\ell,t}\,. \tag{3.10}$$

This ensures that the underlying probabilities of liquidity and default events are constrained between 0 and 1, both $\lambda_c(x_{c,t})$ and $\lambda_\ell(x_{\ell,t})$ being positive (see Equation 3.4). In turn, this implies that the spreads at any maturity are positive, which can be seen from Equation (3.9). Besides, an additional advantage of this modeling is that it allows to accommodate heteroskedasticity in the spreads (see Ahn, Dittmar, and Gallant [2002]).

3.4.6 Recursive pricing formulas

Putting together the risk-neutral dynamics of X_t given by Equation (3.2) and the intensity specifications of Equation (3.10), it can be shown that our model belongs to the class of Quadratic Term Structure Models (QTSM). We show in Appendix 3.C that the spreads $S(t,h)$ of Equation (3.9) can be expressed as a quadratic combina-

tions $x_{c,t}$ and $x_{\ell,t}$. This results from the fact that the conditional Laplace transform of the vector $(X'_{t+1}, Vec(X_{t+1}X'_{t+1})')'$ given $\underline{X_t}$ is exponential affine in $(x'_t, Vec(x_t x'_t))'$ (see Gouriéroux and Sufana [2011] or Cheng and Scaillet [2007]) and, therefore, the process $(X'_{t+1}, Vec(X_{t+1}X'_{t+1})')'$ is affine. We have:

$$\begin{aligned} S(t,h) &= -\frac{1}{h}\left(\Theta_{0,h} + \Theta'_{1,h}X_t + X'_t\Theta_{2,h}X_t\right) \\ &\triangleq \theta_{0,h} + \theta'_{1,h}X_t + X'_t\theta_{2,h}X_t. \end{aligned} \tag{3.11}$$

The factor loadings $\theta_{0,h}$, $\theta_{1,h}$ and $\theta_{2,h}$ are maturity-dependent and are functions of risk-neutral dynamics parameters and of Λ, which is the (2×2)-dimensional diagonal matrix containing $(1-\theta_c)\Lambda_c$ and $(1-\theta_\ell)\Lambda_\ell$ on its diagonal. The loadings $\Theta_{0,h}$, $\Theta_{1,h}$ and $\Theta_{2,h}$ can be computed recursively as (see Appendix 3.C):

$$\left\{\begin{aligned} \Theta_{0,h} &= \Theta_{0,h-1} + \Theta'_{1,h-1}\left[I_n - 2\left(\Theta_{2,h-1}-\Lambda\right)\right]^{-1}\left(\mu^* + \frac{1}{2}\Theta_{1,h-1}\right) \\ &\quad + \mu^{*'}\left(\Theta_{2,h-1}-\Lambda\right)\left[I_n - 2\left(\Theta_{2,h-1}-\Lambda\right)\right]^{-1}\mu^* - \frac{1}{2}\log\left|I_n - 2\left(\Theta_{2,h-1}-\Lambda\right)\right| \\ \Theta_{1,h} &= \Phi^{*'}\left\{\left[I_n - 2\left(\Theta_{2,h-1}-\Lambda\right)\right]^{-1}\left[\Theta_{1,h-1} + 2\left(\Theta_{2,h-1}-\Lambda\right)\mu^*\right]\right\} \\ \Theta_{2,h} &= \Phi^{*'}\left(\Theta_{2,h-1}-\Lambda\right)\left[I_n - 2\left(\Theta_{2,h-1}-\Lambda\right)\right]^{-1}\Phi^*, \end{aligned}\right. \tag{3.12}$$

where initial conditions are given by $\Theta_{0,0} = 0$, $\Theta_{1,0} = (0,0)'$, and $\Theta_{2,0} = [0]_{i,j\in\{1,2\}}$. One of our main objectives is to decompose spreads into a credit and a liquidity component. A necessary condition to obtain such twofold decomposition is that $\Theta_{2,h}$ is diagonal for all maturities h. This condition constrains Φ^* to be diagonal.

3.5 Estimation procedure

3.5.1 Identification strategy: linking proxies and latent factors

In the following, we relate latent factors $x_{c,t}$ and $x_{\ell,t}$ to our credit and liquidity proxies, that we respectively denote by $P_{c,t}$ and $P_{\ell,t}$. Recall that these proxies are first principal components of sets of credit- and liquidity-related variables.

We assume that – up to a measurement error term – the proxies are quadratic functions of the corresponding latent factors. This relationship, of the same kind of the one relating the latent factors to modeled spreads, is consistent with the fact that several variables

used in the computation of proxies are also homogeneous to interest rates. Formally:

$$\begin{cases} P_{c,t} &= \pi_{c,0} + \pi_{c,1} x_{c,t} + \pi_{c,2} x_{c,t}^2 + \sigma_{\nu_c} \nu_{c,t} \\ P_{\ell,t} &= \pi_{\ell,0} + \pi_{\ell,1} x_{\ell,t} + \pi_{\ell,2} x_{\ell,t}^2 + \sigma_{\nu_l} \nu_{l,t}, \end{cases} \tag{3.13}$$

where $\nu_{c,t}$ and $\nu_{l,t}$ are Gaussian standardized and uncorrelated noises. These measurement errors authorize the proxies to imperfectly represent the underlying corresponding risk, addressing potential concerns regarding the fact that our proxies are not pure measures of credit and liquidity risks. For instance, CDS contracts may be affected by liquidity issues. It is also worth stressing that, even though risk factors $x_{c,t}$ and $x_{\ell,t}$ are contemporaneously uncorrelated, their VAR(1) dynamics authorizes the presence of lagged Granger causality between them. Equations (3.13) therefore imply that the credit (resp. liquidity) proxy is a combination of past (resp. past and current) liquidity shocks, of past and current (resp. past) credit shocks and of an error $\nu_{c,t}$ (resp. $\nu_{l,t}$).

3.5.2 State-space model and estimation strategy

The state-space representation of the model is obtained by gathering: (a) the $\mathbb{P}$-dynamics of the factors $x_{c,t}$ and $x_{\ell,t}$ (Equation (3.1)), (b) the spread formulas (Equation (3.11)) and (c) the proxies measurement equations (Equation (3.13)). More specifically, the measurement equations are:

$$\begin{aligned} S(t,h) &= \theta_{0,h} + \theta_{1,h}' X_t + X_t' \theta_{2,h} X_t + \sigma_\eta \eta_{t,h} \qquad \forall h \in \{13, 26, 39, 52 \text{ weeks}\} \\ P_{i,t} &= \pi_{i,0} + \pi_{i,1} x_{i,t} + \pi_{i,2} x_{i,t}^2 + \sigma_{\nu,i} \nu_{i,t} \qquad \forall i = \{c, \ell\}, \end{aligned} \tag{3.14}$$

where the components of the vector of pricing errors η_t and $\nu_{i,t}$ are independent Gaussian white noises with unit variance. Parameters $\pi_{i,0}$, $\pi_{i,1}$, and $\pi_{i,2}$ are not constrained by model-implied restrictions, contrary to the loadings $\theta_{0,h}$, $\theta_{1,h}$, and $\theta_{2,h}$ that derive from Equations (3.11) and (3.12). Appendix 3.E presents additional restrictions that we impose on model parameters in order to ensure a positive correlation between the proxies and the corresponding intensities.

The estimation data cover the period from August 31, 2007 to September 13, 2013 at the weekly frequency (end of week data). Interest rates and CDS data are extracted from Bloomberg. The EURIBOR-OIS spreads of the following maturities enter the measurement equations: 3, 6, 9, and 12 months.

The model parameters are estimated by maximizing the likelihood function, which is ap-

proximated by means of a Kalman-type algorithm. Whereas recent articles use extensively the so-called Unscented Kalman Filter (UKF, see for instance Filipovic and Trolle [2013] or Christoffersen, Dorion, Jacobs, and Karoui [2014]), we rely on the Quadratic Kalman filter (QKF) of Monfort, Renne, and Roussellet, which is specifically fitted to quadratic measurement equations and which shows nice performances in this context. The filtering algorithm is detailed in Appendix 3.D. Once the model parameters are estimated, a final call of the algorithm provides us with filtered values of the latent factors.

3.5.3 Estimation results

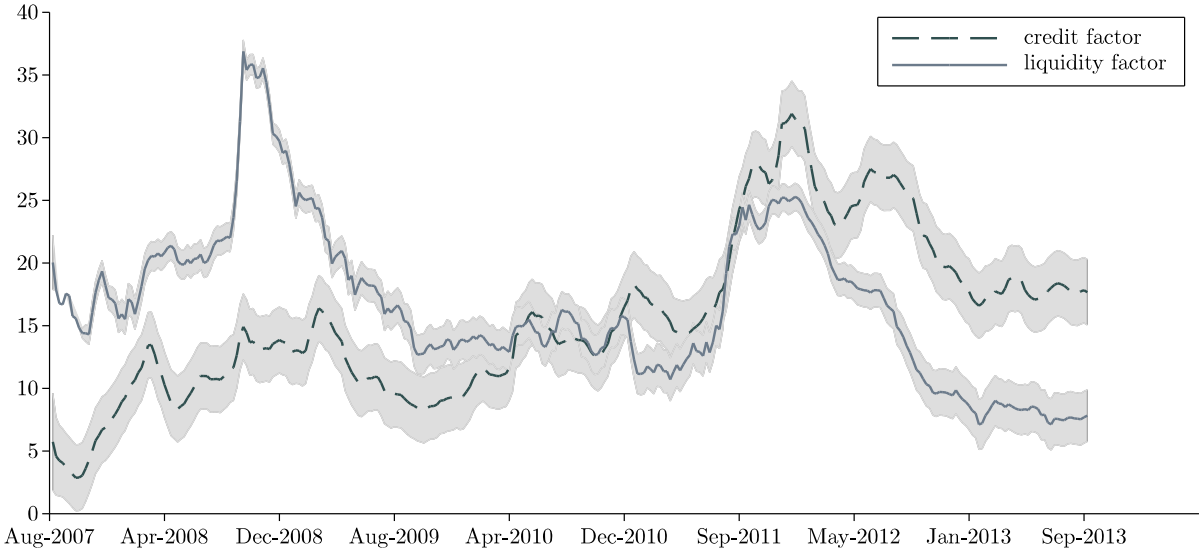
Table 5.3 reports the estimates of the physical and risk-neutral dynamics parameters of $x_{c,t}$ and $x_{\ell,t}$. Both processes are highly persistent, especially under the risk-neutral measure (with eigenvalues of 1 and 0.998). The fact that risk factors are more persistent under the pricing measure than under the physical measure is common in the literature (see e.g. Pan and Singleton [2008]). Intuitively, this feature implies that bad times tend to last longer under $\mathbb{Q}$ than under $\mathbb{P}$, which translates into risk premia. In a preliminary estimation, we found that the Granger causality from credit to liquidity was insignificantly different from zero. Hence it has been imposed exactly to zero in a second pass of maximization. On the other hand, the liquidity factor significantly Granger causes the credit factor, which implies some liquidity feedback in the credit risk. Table 5.3 also reports the market prices of risk parameters, which can be directly backed out from physical and risk-neutral parameters (see Equation (3.2)).

Figure 3.2 presents the filtered time-series of the factors. Whereas they possess roughly the same patterns as the credit and liquidity proxies, the quadratic specification and the measurement errors allows for a greater flexibility in the factor's behavior. In particular, the liquidity factor peaks are shorter in duration than those of the corresponding proxies.

The remaining parameter estimates are gathered in Table 3.3. Both intensities loadings are significantly different from zero, and we observe that $(1-\theta_\ell)\Lambda_\ell > (1-\theta_c)\Lambda_c$ (last row of Table 3.3). This means that liquidity shocks are the main drivers of the short-term fluctuations in the total intensity since the innovations $(\varepsilon_{c,t}, \varepsilon_{\ell,t})$ of factors $(x_{c,t}, x_{\ell,t})$ are of unit variance and given that the total intensity λ_t is given by $(1-\theta_c)\Lambda_c x_{c,t}^2 + (1-\theta_\ell)\Lambda_\ell x_{\ell,t}^2$. However, the credit factor is more persistent than the liquidity one under $\mathbb{Q}$ (see the diagonal elements of Φ^* in Table 5.3); the relative importance of credit in the spread therefore increases with maturity. This will be illustrated below.

The variance estimate $\widehat{\sigma}_\eta^2$ associated with the error terms in the spread equation is 0.007,

Figure 3.2: Filtered credit and liquidity factors



Notes: Time ranges from August 31, 2007 to September 13, 2013. The grey shaded areas are the 95% confidence intervals of the latent factors (this uncertainty is the one associated with the filtering technique).

which translates into an average pricing error of 8 basis points for all maturities. This implies that the model captures 95% of the variation of the spreads.

Besides, the estimated model proves to be able to capture part of the heteroskedasticity in spreads. Indeed, unreported results suggest that the model-implied conditional volatility of spreads exhibits a 60% correlation with realized volatility (measured using daily data on a 2-month rolling window). Note that this is due to our quadratic framework, a standard Gaussian model being unable to generate time-variation in conditional yields' variance.

3.6 Decomposing EURIBOR-OIS spreads

In this section, we present the model-implied decomposition of EURIBOR-OIS spreads for all maturities. We can perform our spread decomposition along two dimensions: credit vs. liquidity on the one hand (as in e.g. Filipovic and Trolle [2013]) and risk premia vs. expected components on the other hand (as in e.g. Pan and Singleton [2008]).

3.6.1 The decomposition method

First, we decompose observed spreads into credit and liquidity components. Remember from Equation (3.9) that the spread of maturity h involves the conditional $\mathbb{Q}$-expectations of both credit and liquidity intensities up to maturity. To obtain the effects on credit only

Table 3.2: Factor parameter estimates

	$\mathbb{P}-$dynamics			$\mathbb{Q}-$dynamics			Market prices of risk		
	μ	$x_{c,t-1}$	$x_{\ell,t-1}$	μ^*	$x_{c,t-1}$	$x_{\ell,t-1}$	Γ_0	$x_{c,t-1}$	$x_{\ell,t-1}$
$x_{c,t}$	0.107	0.960	0.023	1.097	1	0	0.990	0.040	−0.023
	(0.020)	(0.009)	(0.005)	(0.311)	–	–	(0.318)	(0.010)	(0.005)
$x_{\ell,t}$	0.210	0	0.962	0.168	0	0.998	−0.042	0	0.036
	(0.011)	–	(0.004)	(0.040)	–	(0.001)	(0.043)	–	(0.004)

Notes: Standard errors are in parentheses. The '—' sign indicates either that the constraint is binding or that the value is calibrated, thus the parameter is not estimated and its estimator has therefore no standard deviation.

Table 3.3: Parameter estimates of measurement equations

Equation	Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
$P_{c,t}$	$\pi_{c,0}$	−1.977	$\pi_{c,1}$	0.132	$\pi_{c,2}$	0.00001
		(0.217)		(0.024)		(0.0007)
$P_{\ell,t}$	$\pi_{\ell,0}$	−1.370	$\pi_{\ell,1}$	0.039	$\pi_{\ell,2}$	0.002
		(0.092)		(0.008)		(0.0002)
noise	$\sigma^2_{\nu_c}$	0.1	$\sigma^2_{\nu_\ell}$	0.1	σ^2_η	0.007
		–		–		(0.0004)
λ_t	$(1-\theta_c)\Lambda_c$	0.00009	$(1-\theta_\ell)\Lambda_\ell$	0.00134		
		(0.00002)		(0.00009)		

Notes: Standard errors are in parentheses. The '—' sign indicates that the value is calibrated.

(say), we simply put $\Lambda_\ell = 0$ and recompute the counterfactual spread implied by this restriction. More formally, if we denote by $S_c(t,h)$ and $S_\ell(t,h)$ the respective credit and liquidity components of the observed spread, we have:

$$\begin{aligned} S_c(t,h) &= -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{\sum_{i=1}^{h} -(1-\theta_c)\lambda_c(x_{c,t+i})\right\}\right]\right) \\ &\triangleq \theta_{0,h}^{(c)} + \theta_{1,h}^{(c)} x_{c,t} + \theta_{2,h}^{(c)} x_{c,t}^2 \qquad (3.15) \\ S_\ell(t,h) &= -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{\sum_{i=1}^{h} -(1-\theta_\ell)\lambda_\ell(x_{\ell,t+i})\right\}\right]\right) \\ &\triangleq \theta_{0,h}^{(\ell)} + \theta_{1,h}^{(\ell)} x_{\ell,t} + \theta_{2,h}^{(\ell)} x_{\ell,t}^2, \qquad (3.16) \end{aligned}$$

where $\theta_{0,h}^{(c)}$, $\theta_{1,h}^{(c)}$, $\theta_{2,h}^{(c)}$; and $\theta_{0,h}^{(\ell)}$, $\theta_{1,h}^{(\ell)}$, and $\theta_{2,h}^{(\ell)}$ are the entries of $\theta_{0,h}$, $\theta_{1,h}$ and $\theta_{2,h}$ that correspond to credit and liquidity risks, respectively. We then have an exact decomposition of the modeled spread and, for the observed spread we get:

$$S(t,h) = S_c(t,h) + S_\ell(t,h) + \sigma_\eta \eta_{t,h}\,, \qquad (3.17)$$

where $\sigma_\eta \eta_{t,h}$ exactly matches the measurement errors included in the measurement equations (Equation (3.14)). Given their relative small size and following the usual approach, we neglect those measurement errors in the analysis and consider only the decomposition of the modeled spread $S(t,h) - \sigma_\eta \eta_{t,h}$.

Spreads can be split in an other dimension. Indeed, our estimation strategy provides us with both the physical and the risk-neutral dynamics of the factors. This knowledge enables us to extract risk premia from observed spreads. Risk premia are defined as the differentials between observed (or model-implied) spreads and the ones that would prevail if investors were risk-neutral. In the latter case, which corresponds to the expectation hypothesis, spreads would be those obtained by using the physical dynamics to compute the expectation term in Equation (3.9). Using the estimated $\mathbb{P}$-dynamics parameters and the fact that the total intensity λ_t is the same function of X_t under both measures (see Subsection 3.4.2), we calculate a new set of factor loadings under the expectation hypothesis. To perform the credit/liquidity decomposition of this expected component, we use the same formulas as in System (3.15-3.16), replacing the $\mathbb{Q}$ dynamics by the $\mathbb{P}$ dynamics.[6] We denote these components by $S_c^{\mathbb{P}}(t,h)$ and $S_\ell^{\mathbb{P}}(t,h)$.

6. Specifically, we impose respectively $\Lambda_\ell = 0$ and $\Lambda_c = 0$ and then compute associated sets of factor loadings $\theta_{i,h}$ $(i \in \{0,1,2\})$ by using the recursive fomulas of System (3.12).

3.6.2 Decomposition results

The decomposition of the 6- and 12-month maturity spreads are represented in Figure 3.3. On average, the liquidity component accounts for most of the spread averages over the sample period, representing more than 75% of spreads' levels for all maturities (see Table 3.4). The average share of spreads that is associated to credit risks, which is comprised on average between 10% and 25%, increases with respect to maturity. The first row of charts in Figure 3.3 illustrates that the liquidity factor accounts for much of the high-frequency variations in the spreads, in particular during the distress period of late 2008 (after the Lehman collapse) and in end 2011 (in a period of particular strain in the European sovereign markets) no matter the maturity.

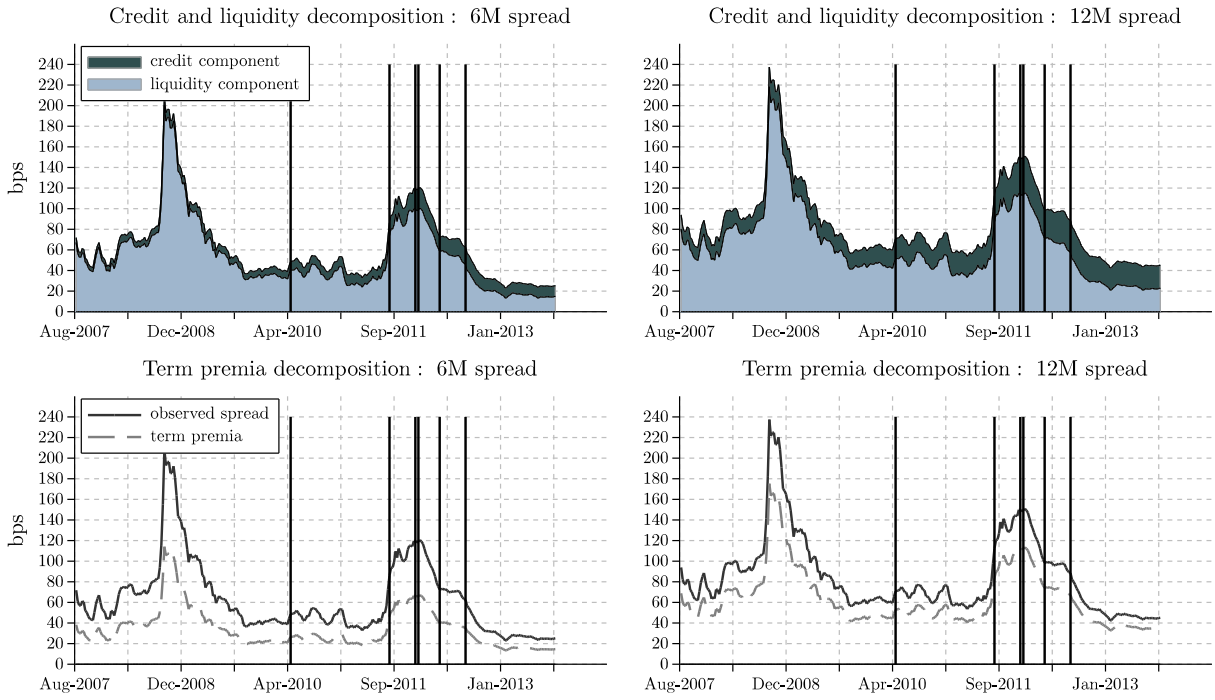
Table 3.4: Descriptive statistics of EURIBOR-OIS components

		Total spread		Risk premium	
		Credit	Liquidity	Credit	Liquidity
average level (in bps)	Spread 3M	5.43	48.22	4.22	15.50
	Spread 6M	9.35	53.70	8.28	28.21
	Spread 9M	14.49	59.42	13.53	38.47
	Spread 12M	20.82	65.36	19.97	47.44
average (% of spread avg)	Spread 3M	10.21	90.73	7.86	28.89
	Spread 6M	14.97	85.94	13.14	44.74
	Spread 9M	19.77	81.09	18.31	52.05
	Spread 12M	24.36	76.46	23.17	55.04

Notes: The modeled spreads are decomposed into four components, along two dimensions: credit vs. liquidity and expected part vs. risk premium. The risk premia are the parts of the spreads that would not exist is investors were risk-neutral. The table shows for instance that for the 9-month maturity, 70% of the EURIBOR-OIS spread correspond to risk premia, a quarter of which ($\simeq 18/(18+52)$) being accounted for by aversion to credit risk.

The second row of Figure 3.3 displays the decomposition of the observed spread into the risk premium and the expected component: the risk premium component and the observed spread have very similar features, and are positively and highly correlated. Together with Table 3.4, we see that the share of the spreads explained by risk premia is increasing with the maturity: for the 3-month spread, credit and liquidity risk premia account respectively for 8% and 29% of the total spread average ; and for the 12-month spread, respectively 23% and 55% (see third and fourth columns of Table 3.4). In times of distress (Lehman collapse or the European debt crisis), the level of risk premia, which are the compensations

Figure 3.3: 6M EURIBOR-OIS spreads decomposition



Notes: Date ranges from August 31, 2007 to September 13, 2013. Units are in basis points. Top panel represents the stacked components of the spread: light grey component is the liquidity component and the dark grey corresponds to the credit component. Bottom panel represents the modeled spread and its term premia. The black vertical axes stand from left to right for: SMP program announcements (first two axis), VLTRO announcement and allotments (next three axis), and Mario Draghi's London speech (last axis).

for exposures to non-diversifiable systematic risk, increases for all maturities.

Figure 3.4 confirms the previous statements by presenting decompositions of the term structure of EURIBOR-OIS spreads at different dates. In particular, the second and third rows show respectively the decomposition of the expected component of spreads and of risk premia. Under the expectation hypothesis (i.e. in the absence of risk premia), the liquidity risk term structure is downward sloping whereas the credit component is smaller and almost constant with respect to maturity. Conversely, looking at the last row of Figure 3.4, both credit and liquidity risk premia are upward sloping with respect to maturity. Note that these features are not specific to the four chosen dates.

In the next section, we exploit the time series and the term structure of the spreads components to analyze the effectiveness of unconventional monetary policies in the Eurozone.

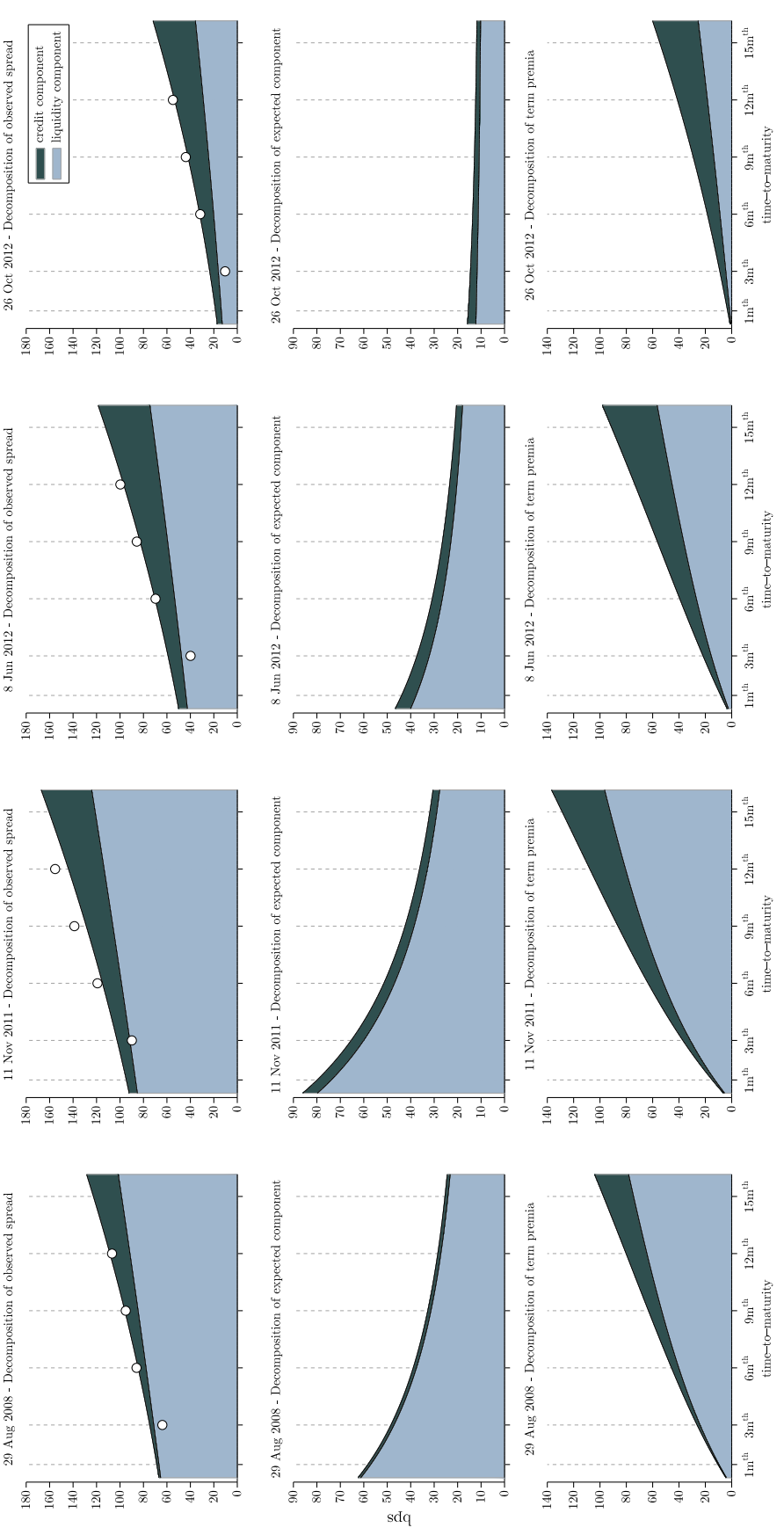
3.6.3 The impact of unconventional monetary policy on interbank risk

The main programs of unconventional monetary policies in the Eurozone can be broadly separated into three periods. The *Securities Market Program* (SMP) consisted in sterilized bond-buying on the secondary market. It was designed to "ensure depth and liquidity in [...] market segments that are dysfunctional" and was implemented in May 2010 and August 2011. Later, on the December 8, 2011, the ECB disclosed the design of *Very Long Term Refinancing Operations* (VLTRO), whereby 3-year maturity open market operations were proposed in the form of reverse repo. Two allotments were granted on December 21, 2011 and on the February 29, 2012, of respectively EUR 489bn and EUR 530bn to 523 and 800 banks. More recently, during August 2012, Mario Draghi announced the setting of *Outright Monetary Transactions* (OMT) in his London speech. Conditionally on fiscal adjustments or precautionary programs enforcement by candidate countries, the ECB is ready to trade in secondary sovereign bond markets with "no ex ante quantitative limits". Whereas this framework has been announced it has not been applied in practice yet.

Interestingly, the EURIBOR-OIS spreads have decreased continuously since the VLTRO announcement in December 2011. This drop has led many commentators (and central bankers) to claim that the ECB unconventional refinancing operations were successful in alleviating interbank market tensions. In particular, according to ECB officials, the non-standard VLTRO operations addressed "only the liquidity side of the [interbank market] problem" (see Draghi [2012]'s interview with the Wall Street Journal, published on February 24, 2012). Our results support this view as the liquidity component of the spreads has slowly faded away since the VLTRO announcement date (see Figure 3.3, first row). A further positive effect can also be attributed to the OMT announcement through liquidity (see the last vertical bar in the charts).

The same pattern can be observed in Figure 3.4. After the SMP and before the VLTRO announcement (second column of charts), liquidity risk still accounts for most of the term structure of interbank spreads with between 90 to 120 basis points depending on the maturity. However, after the VLTRO allotments, liquidity risk represents only 40 to 60 basis points across maturities (see third column of Figure 3.4) and further drops to around 20 basis points for all maturities after the OMT announcement (fourth column). In comparison, looking at both Figures 3.3 and 3.4, those policy measures had only a small impact on the credit components of the spreads: between November 2011 and October 2012, its range goes from [10 bps, 50 bps] to [5 bps, 40 bps]. Even though there is a small drop

Figure 3.4: Decomposition of EURIBOR - OIS term structure



Notes: First to fourth columns correspond respectively to the following dates: August 29, 2008; November 11, 2011; January 8, 2012; and October 26, 2012. First row represents the stacked components of the term structure: the light grey component is the liquidity component and the dark one is the credit component. The white dots are the observed spreads. Second row presents the same components under the expectation hypothesis. The graph at the bottom represents decomposition of the risk premia, that are the differences between the total modeled spreads and the expected components of the spreads (the sum of the last two rows of chart results in the first row.).

in the credit component, the evidence of the effectiveness of unconventional monetary policies on credit risk is far thinner than on liquidity risk.

Turning to the second and third rows of Figure 3.4, it appears that unconventional monetary policies were followed by decreases in both the expected components and the risk premia. Furthermore, we observe that these drops mainly come from the liquidity parts of the spreads, showing that the VLTROs and OMT have had an effect on both decreasing the expectations of credit and liquidity risks, and were successful in alleviating the effects of aversion to this source of risk.

All in all, even if the EURIBOR-OIS spreads have not really reacted to the 2010 SMP program, our results suggest that the more recent unconventional monetary policy measures undertaken within the Eurosystem have contributed to improve bank liquidity positions and to stabilize the credit risk in the Eurozone. The next subsection focuses on this latter aspect by showing how these measures have affected the bank probabilities of default.

3.6.4 Model-implied probabilities of default

Following Doshi, Ericsson, Jacobs, and Turnbull [2013], we present an additional by-product of our framework, which is the computation of model-implied probabilities of default (PDs). In our model, the panel of banks is homogeneous and the probabilities of default are not bank-dependent. Formally, for any bank i, we have:

$$\begin{aligned}\mathbb{P}(d_{t+h}^{(i)}=1|d_t^{(i)}=0,\underline{w_t}) &= 1-\mathbb{P}(d_{t+1}^{(i)}=0,\ldots,d_{t+h}^{(i)}=0|d_t^{(i)}=0,\underline{w_t})\\ &= 1-\mathbb{E}_t^{\mathbb{P}}(\exp(-\lambda_c(x_{c,t+1})-\cdots-\lambda_c(x_{c,t+h})))\\ &= 1-\mathbb{E}_t^{\mathbb{P}}\left(\exp\left(\Lambda_c[x_{c,t+1}^2+\cdots+x_{c,t+h}^2]\right)\right). \qquad (3.18)\end{aligned}$$

The last term of the previous equation is a multi-horizon Laplace transform of $x_{c,t}^2$, which can be computed analytically by means of recursive formulas of the same kind as those presented in System (3.12) (replacing μ^* and Φ^* by μ and Φ, and redefining Λ as the matrix with $(\Lambda_c, 0)$ on its diagonal.). The computation requires an estimate of the default recovery rate θ_c. To the best of our knowledge, the existing literature presents no euro-area figure that can serve as a basis for the calibration of such a parameter. Hence, we set it to 91.25%, which is the recovery rate on unsecured deposits on U.S. banks with at least \$5bn assets (see Kuritzkes, Schuermann, and Weiner [2005]).[7]

7. Christensen, Lopez, and Rudebusch [2014] note that such a recovery rate is high – compared to usual corporate-bond recovery rates – because an unsecured deposit is more senior in the liability structure of a bank than senior unsecured debt.

Figure 3.5 displays the physical (upper plot) and risk-neutral (lower plot) one-year PDs resulting from this computation. Confidence bands are added on the plots; these bands reflect the uncertainty regarding the model parameterization. These confidence bands are obtained by drawing 1000 sets of model parameters from their asymptotic joint distribution. For each set of parameters, we use the quadratic Kalman filter to estimate time series of $(x_{c,t}, x_{\ell,t})$ and compute the implied (time series of) PDs. For each date, the confidence intervals are based on the percentiles of the 1000 simulated PDs. It appears that risk neutral probabilities are very far from their physical counterparts, the deviations being accounted for by sizable credit-risk premia. These findings are in line with those of a large body of empirical studies highlighting the substantial deviations existing between physical and risk-neutral PDs.[8] The existence of credit-risk premia constitutes one of the main explanations for the so-called credit-spread puzzle (see e.g. Amato and Remolona [2003]). This puzzle corresponds to the observation that observed credit spreads tend to be higher than average credit-losses (while they should be equal under some conditions, that notably include the risk-neutrality of investors).

Our estimated physical probabilities of default are roughly comprised between 0.1% and 0.4%. While small, this order of magnitude is however consistent with historical default data of investment-grade issuers. For instance, Moody's [2011] reports that, on average over the period 1983-2010, the one-year default rate of a A-rated financial institutions is of 0.1%. (The median rating of EURIBOR-panel banks is A across the three main rating agencies.) On a longer time-scale, Moody's [2013] indicates that the default rate of A-rated corporates has been of 0.10% (respectively 0.06%) over the period 1920-2013 (respectively 1970-2013).[9] Figure 3.5 illustrates that VLTROs and the OMT announcement (represented by the last four vertical bars on the chart) were effective in reducing bank probabilities of default whether corrected from risk premia or not. However, at the end of the sample, these probabilities remain higher than their mid-2007 value.

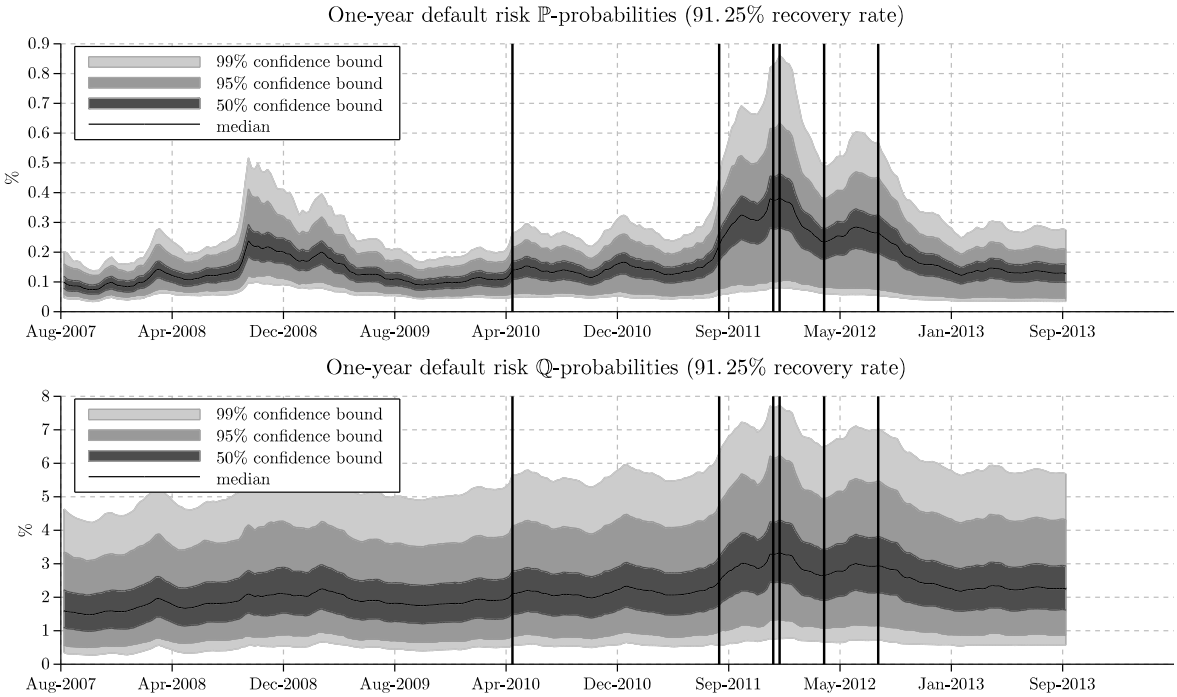
Conclusion

We develop a no-arbitrage two-factor quadratic term structure model for the EURIBOR-OIS spreads across several maturities, from August 2007 to September 2013. To identify credit and liquidity components in the spreads, we exploit credit and liquidity proxies based on CDS prices, market liquidity and funding liquidity measures. Our decomposi-

8. See for instance Monfort and Renne [2014] in the case of sovereign issuers and Elton, Gruber, Agrawal, and Mann [2004] in the case of corporate issuers.

9. For lower-rated investment-grade issuers (Baa using the Moody's rating system, which is equivalent to the BBB rating of S&P), the default rates for these two periods are respectively of 0.27% and of 0.17%.

Figure 3.5: Default probabilities of banks under the physical and pricing measures



Notes: Time ranges from August 31, 2007 to September 13, 2013. The upper plot show the model-implied one-year probability of default of a bank of the panel (banks are assumed to share the same characteristics). This probability is derived using Equation (3.18). The lower chart shows the risk-neutral probability of default, which is obtained by using the same formula replacing the physical dynamics parameters by the risk-neutral ones. Shaded areas are the 50% to 99% confidence bounds of these probabilities. The black vertical axes stand from left to right for: SMP program announcements (first two axis), VLTRO announcement and allotments (next three axis), and Mario Draghi's London speech (last axis).

tion handles potential interdependence between credit and liquidity risks and is consistent across maturities. We find that the liquidity risk generates most of the variance of the spread over the estimation period. The credit risk is less volatile, but represents more than half of the spread level in late 2012. Our decomposition allows us to shed new light on the effects of unconventional monetary policy of the ECB on the interbank risk. We show that whereas the bond-purchase programs of 2010 and 2011 were not followed by decreases in any of the EURIBOR-OIS spread components, the VLTROs and the OMT announcements have had a substantial impact, mainly on the liquidity risk. At the end of the sample, the liquidity risk is at its lowest since the beginning of the financial crisis.

Appendices to Chapter 3

3.A Risk neutral distribution of w_t

The R.N. conditional distribution of w_t given $\underline{w_{t-1}}$ has a p.d.f., with respect to the same historical distribution, which is given by:

$$\exp\left[\Gamma_{t-1}'(X_t - \mu - \Phi X_{t-1}) - \frac{1}{2}\Gamma_{t-1}'\Gamma_{t-1} + g(\underline{r_t})\right].$$

Since this p.d.f. factorizes into a function depending on process X_t and a function depending on process r_t, the independence between these processes is preserved under the R.N. world. The derivation of the dynamics of X_t is standard (see e.g. Ang and Piazzesi [2003]). Moreover, since (d_t', ℓ_t') does not appear in the S.D.F., the conditional distribution of (d_t', ℓ_t') given $(\underline{X_t}, \underline{d_{t-1}}, \underline{\ell_{t-1}})$ are the same in both worlds. Indeed, it is a consequence of the following lemma.

Lemma 3.A.1. *If $w_t = (w_{1,t}', w_{2,t}')'$ and if the S.D.F. $M_{t-1,t}$ is a function of $w_{1,t}$ only, the conditional distribution of $w_{2,t}$ given $(\underline{w_{1,t}}, \underline{w_{2,t-1}})$ is the same in both worlds.*

Proof. We have:

$$f_t^{\mathbb{Q}}(w_{1,t}|\underline{w_{t-1}}) f_t^{\mathbb{Q}}(w_{2,t}|\underline{w_{1,t}}, \underline{w_{2,t-1}}) = M_{t-1,t}(\underline{w_{1,t}}) \exp(-r_{t-1}(\underline{w_{t-1}})) f_t^{\mathbb{P}}(w_{1,t}|\underline{w_{t-1}}) f_t^{\mathbb{P}}(w_{2,t}|\underline{w_{1,t}}, \underline{w_{2,t-1}}).$$

Integrating both sides w.r.t. $w_{2,t}$ gives:

$$f_t^{\mathbb{Q}}(w_{1,t}|\underline{w_{t-1}}) = M_{t-1,t}(\underline{w_{1,t}}) \exp(-r_{t-1}(\underline{w_{t-1}})) f_t^{\mathbb{P}}(w_{1,t}|\underline{w_{t-1}})$$

and the result follows. ■

3.B Pricing recursions and non causality

For any $\tau \in \{0, \ldots, h\}$, we have:

$$\begin{aligned} B(t+\tau, h-\tau) &= \exp(-r_{t+\tau}) \mathbb{E}^{\mathbb{Q}}\Big[B(t+\tau+1, h-\tau-1)(1 - d_{t+\tau+1}^{(j)} - \theta_c d_{t+\tau+1}^{(j)}) \\ &\times \quad (1 - \ell_{t+\tau+1}^{(i)} + \theta_\ell \ell_{t+\tau+1}^{(i)}) \Big| \underline{w_{t+\tau}}, d_{t+\tau}^{(j)} = 0, \ell_{t+\tau}^{(i)} = 0 \Big] \end{aligned}$$

Assumption A: $B(t+\tau+1, h-\tau-1)$ is a function of $(\underline{X_{t+\tau+1}}, \underline{r_{t+\tau+1}})$ but not of $(\underline{d_{t+\tau}^{(j)}}, \underline{\ell_{t+\tau}^{(i)}})$.

We get, using the law of iterated expectations:

$$B(t+\tau,h-\tau)=\exp(-r_{t+\tau})\mathbb{E}^{\mathbb{Q}}\Big[B(t+\tau+1,h-\tau-1)\times \mathbb{E}^{\mathbb{Q}}\Big\{\big(1-(1-\theta_c)d^{(j)}_{t+\tau+1}\big)\big(1-(1-\theta_\ell)\ell^{(i)}_{t+\tau+1}\big)\Big|\underline{w_{t+\tau}},X_{t+\tau+1},r_{t+\tau+1},d^{(j)}_{t+\tau}=0,\ell^{(i)}_{t+\tau}=0\Big\}\Big|\underline{w_{t+\tau}},d^{(j)}_{t+\tau}=0,\ell^{(i)}_{t+\tau}=0\Big]$$

Using the conditional independence of $d^{(j)}_{t+\tau+1}$ and $\ell^{(i)}_{t+\tau+1}$ and the approximations:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\big(1-(1-\theta_c)d^{(j)}_{t+\tau+1}\big)\big|\underline{w_{t+\tau}},X_{t+\tau+1},r_{t+\tau+1},d^{(j)}_{t+\tau}=0,\ell^{(i)}_{t+\tau}=0\right] &= \exp\{-(1-\theta_c)\lambda_{c,t+\tau+1}\}\\ \mathbb{E}^{\mathbb{Q}}\left[\big(1-(1-\theta_\ell)\ell^{(j)}_{t+\tau+1}\big)\big|\underline{w_{t+\tau}},X_{t+\tau+1},r_{t+\tau+1},d^{(j)}_{t+\tau}=0,\ell^{(i)}_{t+\tau}=0\right] &= \exp\{-(1-\theta_\ell)\lambda_{\ell,t+\tau+1}\},\end{aligned}$$

we get:

$$B(t+\tau,h-\tau)=\mathbb{E}^{\mathbb{Q}}\left[B(t+\tau+1,h-\tau-1)\exp(-r_{t+\tau}-\lambda_{t+\tau+1})\big|\underline{w_{t+\tau}},d^{(j)}_{t+\tau}=0,\ell^{(i)}_{t+\tau}=0\right] \tag{3.19}$$

Moreover, since $(d^{(j)}_t,\ell^{(i)}_t)$ does not Granger-cause (r_t,X_t), $B(t+\tau,h-\tau)$ is not a function of $\underline{d_{t+\tau}},\underline{\ell_{t+\tau}}$ and Assumption A is confirmed. In order to use Equation (3.19) recursively backward, starting at $\tau=h-1$, we have to check assumption A for $B(t+h,0)$ which is obviously satisfied since $B(t+h,0)=1$. The recursive use of Equation (3.19) for $\tau=h-1,h-2,\ldots,0$ gives:

$$B(t,h)=\mathbb{E}^{\mathbb{Q}}\left[\exp\{-r_t-\lambda_{t+1}-\ldots-r_{t+h-1}-\lambda_{t+h}\}\big|\underline{w_t},d^{(j)}_t=0,\ell^{(i)}_t=0\right],$$

which depends on $(\underline{r_t},X_t)$, not on $(\underline{d^{(j)}_t},\underline{\ell^{(i)}_t})$, and gives formula (3.6).

3.C Solving for yield/spread loadings in a QTSM

3.C.1 Computing the Laplace transform of $Z_t=\left[X_t',Vec(X_tX_t')\right]'$

Lemma 3.C.1. *If $\varepsilon^*_{t+1}\sim\mathcal{N}(0,I)$, we have*

$$\mathbb{E}_t\left[\exp(\theta'\varepsilon^*_{t+1}+\varepsilon^{*\prime}_{t+1}V\varepsilon^*_{t+1})\right]=\frac{1}{|I-2V|^{1/2}}\exp\left[\frac{1}{2}\theta'(I-2V)^{-1}\theta\right]. \tag{3.20}$$

Proof. It can be shown that

$$\forall u\in\mathbb{R}^n,\quad \int_{\mathbb{R}^n}\exp(-u'Qu+\nu'u)\,\mathrm{d}u=\frac{\pi^{n/2}}{|Q|^{1/2}}\exp\left(\frac{1}{4}\nu'Q^{-1}\nu\right). \tag{3.21}$$

Therefore, we have:

$$\begin{aligned}\mathbb{E}_t\left[\exp(\theta'\varepsilon^*_{t+1}+\varepsilon'^*_{t+1}V\varepsilon^*_{t+1})\right] &= \int_{\mathbb{R}^n}\exp(\theta'\varepsilon+\varepsilon'V\varepsilon)\frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}\varepsilon' I\varepsilon\right)d\varepsilon' \\ &= \frac{1}{(2\pi)^{n/2}}\int_{\mathbb{R}^n}\exp\left[-\varepsilon'\left(\frac{1}{2}I-V\right)\varepsilon+\theta'\varepsilon\right]\,\mathrm{d}u \\ &= \frac{1}{|I-2V|^{1/2}}\exp\left[\frac{1}{2}\theta'(I-2V)^{-1}\theta\right]\end{aligned}$$

■

Let X_t be a random vector of size n following Gaussian VAR(1) dynamics: $X_t = \mu + \Phi X_{t-1} + \Omega\varepsilon_t$, where ε_t are i.i.d. normalized Gaussian vectors, and $\Sigma = \Omega\Omega'$ is the conditional variance-covariance matrix of X_t. We define Z_t as the augmented vector of factors composed of X_t and of its vectorized outer-product, that is: $Z_t = \left[X'_t, Vec(X_tX'_t)\right]'$.

Let us consider $u \in \mathbb{R}^n$ and V a square symmetric matrix of size n. The conditional Laplace transform of Z_{t+1} is denoted by φ_t and defined by:

$$\varphi_t(u,V) = \mathbb{E}_t\left\{\exp\left[(u', Vec(V)')\times Z_{t+1}\right]\right\} = \mathbb{E}_t\left\{\exp\left[u'X_{t+1}+X'_{t+1}VX_{t+1}\right]\right\}$$

In the following, we compute the explicit affine form of the conditional Laplace transform of Z_{t+1}. Let us first consider the term in the expectation; substituting $\mu+\Phi X_t+\Omega\varepsilon_{t+1}$ for X_{t+1} leads to:

$$\begin{aligned}\exp\{u'X_{t+1}+X'_{t+1}VX_{t+1}\} &= \exp\left\{u'(\mu+\Phi X_t)+\mu'V\mu+2\mu'V\Phi X_t+X'_t\Phi'V\Phi X_t\right\} \\ &\quad\times\quad \exp\left\{\left[u'\Omega+2(\mu+\Phi X_t)'V\Omega\right]\varepsilon_{t+1}+\varepsilon'_{t+1}\left[\Omega'V\Omega\right]\varepsilon_{t+1}\right\}\end{aligned}$$

Taking the conditional expectation leaves the first part of the previous expression unchanged as everything is known in t. For the second part of the previous expression, we apply Lemma 3.C.1 and algebraic computation leads to:

$$\begin{aligned}&\mathbb{E}_t\left[\exp\left\{\left[u'\Omega+2(\mu+\Phi X_t)'V\Omega\right]\varepsilon_{t+1}+\varepsilon'_{t+1}\left[\Omega'V\Omega\right]\varepsilon_{t+1}\right\}\right] \\ =\;& \exp\Bigg\{-\frac{1}{2}\log\left|I_n-2\Omega'V\Omega\right|+\frac{1}{2}u'\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'u+2u'\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'V\mu \\ &+\;2\mu'V\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'V\mu \\ &+\;\left[2u'\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'V\Phi+4\mu'V\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'V\Phi\right]X_t \\ &+\;X'_t\left[2\Phi'V\Omega(I_n-2\Omega'V\Omega)^{-1}\Omega'V\Phi\right]X_t\Bigg\}.\end{aligned}$$

Putting together the first and the second part in the expectation, we obtain: $\varphi_t(u,V) = \exp\{a_1(u,V)'X_t + X_t'a_2(u,V)X_t + b(u,V)\}$, where:

$$\begin{aligned}
a_1(u,V) &= \Phi'\left[u + 2V\mu + 2V\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'u + 4V\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'V\mu\right] \\
a_2(u,V) &= \Phi'\left[V + 2V\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'V\right]\Phi \\
b(u,V) &= u'\mu + \mu'V\mu - \frac{1}{2}\log\left|I_n - 2\Omega'V\Omega\right| + \frac{1}{2}u'\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'u \\
&+ \quad 2u'\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'V\mu + \quad 2\mu'V\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega'V\mu.
\end{aligned}$$

Then, noticing that:

$$\Omega(I_n - 2\Omega'V\Omega)^{-1}\Omega' = \left[\Omega^{-1'}(I_n - 2\Omega'V\Omega)^{-1}\Omega^{-1}\right]^{-1} = \left[\Sigma^{-1} - 2V\right]^{-1},$$

we can simplify the previous expressions and obtain:

$$\begin{aligned}
a_2(u,V) &= \Phi'V(I_n - 2\Sigma V)^{-1}\Phi \\
a_1(u,V) &= \Phi'\left[(I_n - 2V\Sigma)^{-1}(u + 2V\mu)\right] \\
b(u,V) &= u'(I_n - 2\Sigma V)^{-1}\left(\mu + \frac{1}{2}\Sigma u\right) + \mu'V(I_n - 2\Sigma V)^{-1}\mu - \frac{1}{2}\log\left|I_n - 2\Sigma V\right|.
\end{aligned}$$

3.C.2 Calculation of our model's loadings

Let us denote by λ_t the total intensity, that is: $\lambda_t = (1-\theta_c)\lambda_{c,t} + (1-\theta_\ell)\lambda_{\ell,t}$. We have: $\lambda_t = X_t'\Lambda X_t$ where $\Lambda = diag[(1-\theta_c)\Lambda_c, (1-\theta_\ell)\Lambda_\ell]$. We can then re-express the pricing formula (3.9) as:

$$S(t,h) = -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=1}^{h} X_{t+i}'\Lambda X_{t+i}\right\}\right]\right),$$

which is the log of the multihorizon Laplace transform of a quadratic combination of Gaussian variables. Let us postulate that: $S(t,h) = \theta_{0,h} + \theta_{1,h}'X_t + X_t'\theta_{2,h}X_t$. (We know that the model belongs to the class of quadratic term structure models, and that the spreads at all maturities can be expressed as a quadratic combination of X_t.) Using the

law of iterated expectation, we obtain the following recursion:

$$\begin{aligned} S(t,h) &= -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\mathbb{E}_t^{\mathbb{Q}}\left(\exp\left\{-\sum_{i=1}^{h}X_{t+i}'\Lambda X_{t+i}\right\}\Big|\underline{X_{t+h-1}}\right)\right]\right) \\ &= -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=1}^{h-1}X_{t+i}'\Lambda X_{t+i}\right\}\varphi_{t+h-1}^{\mathbb{Q}}(0,-\Lambda)\right]\right) \\ &= -\frac{1}{h}\log\left(\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left\{-\sum_{i=1}^{h-1}X_{t+i}'\Lambda X_{t+i}\right\}\left[b^{\mathbb{Q}}(0,-\Lambda)+\right.\right.\right. \\ &\qquad\left.\left.\left. a_1^{\mathbb{Q}}(0,-\Lambda)'X_{t+h-1}+X_{t+h-1}'a_2^{\mathbb{Q}}(0,-\Lambda)X_{t+h-1}\right]\right]\right) \\ &\triangleq -\frac{1}{h}\left(\Theta_{0,h}+\Theta_{1,h}'X_t+X_t'\Theta_{2,h}X_t\right). \end{aligned}$$

Eventually, the recursive equations of system (3.12) are obtained by using the closed-form coefficients of the conditional Laplace transform of the previous section, plugging the risk-neutral parameters μ^* and Φ^* (and recalling that we have we have $\Omega = I_2$).

3.D The Quadratic Kalman Filter

The QKF is based on the fact that the measurement equations are quadratic in the latent factor $X_t = (x_{c,t}, x_{l,t})'$ but affine in the augmented vector $Z_t = (X_t', Vec(X_tX_t'))'$. Let Y_t be the set of measured variables, thus $Y_t = \left[S(t,13),\, S(t,26),\, S(t,39),\, S(t,52),\, P_{c,t},\, P_{\ell,t}\right]'$. The measurement equations can be transformed in affine functions of Z_t:

$$\begin{aligned} \begin{pmatrix} S(t,h) \\ P_{c,t} \\ P_{\ell,t} \end{pmatrix} &= \begin{pmatrix} \theta_{0,h} \\ \pi_{c,0} \\ \pi_{l,0} \end{pmatrix} + \begin{pmatrix} \theta_{1,h}^{(c)} & \theta_{1,h}^{(\ell)} & \theta_{2,h}^{(c)} & 0 & 0 & \theta_{2,h}^{(\ell)} \\ \pi_{c,1} & 0 & \pi_{c,2} & 0 & 0 & 0 \\ 0 & \pi_{\ell,1} & 0 & 0 & 0 & \pi_{\ell,2} \end{pmatrix} Z_t + \begin{pmatrix} \sigma_\eta \eta_{t,h} \\ \sigma_{\nu_c}\nu_{c,t} \\ \sigma_{\nu_l}\nu_{l,t} \end{pmatrix}, \\ \Longrightarrow Y_t &\triangleq A + \widetilde{B}Z_t + D\zeta_t \end{aligned}$$

Approximating the conditional distribution of Z_{t+1} given Z_t by a Gaussian distribution and considering the augmented state-space model based on Z_t, we can use the exact same method as is presented in Chapter 2. In order to get the global likelihood maximum, the estimation is achieved in two steps. The *Artificial Bee Colony* stochastic algorithm (see Karaboga and Basturk [2007]) is used to find the potential maxima areas of parameters. The results are then used as starting values for a usual simplex maximization algorithm and the best estimate is selected.

3.E Identifiability and estimation constraints

3.E.1 Parameter contraints

For interpretation purposes, the fluctuations of credit and liquidity proxies are required to correlate positively with the associated intensities. Formally, this is obtained by imposing that, for most values of the factors $x_{c,t}$ and $x_{\ell,t}$, the intensities and the proxies are monotonously increasing with respect to the corresponding factor ($x_{c,t}$ or $x_{\ell,t}$). Recall that the intensity functions are purely quadratic, i.e. of the form $\Lambda_i x_{i,t}^2$, with Λ_c and Λ_ℓ strictly positive to ensure that the intensities are always non-negative. We therefore impose that each $x_{i,t}$ is positive most of the time.

$$\forall i = \{c, \ell\}, \qquad \mathbb{P}\left(x_{i,t} < 0\right) = \alpha \iff \mathbb{E}(x_{i,t}) = -q_{\mathcal{N}(0,1)}(\alpha)\sqrt{\mathbb{V}(x_{i,t})},$$

where $\mathbb{E}(\bullet)$ and $\mathbb{V}(\bullet)$ are the unconditional expectation and variance operators (under the physical measure), and $q_{\mathcal{N}(0,1)}(\alpha)$ is the level-α quantile of the normalized Gaussian distribution and α is typically a small number. We impose the same thing for the proxies, namely:

$$\forall i = \{c, \ell\}, \qquad \mathbb{P}\left(x_{i,t} < -\frac{\pi_{i,1}}{2\pi_{i,2}}\right) = \alpha \iff \mathbb{E}(x_{i,t}) = -\frac{\pi_{i,1}}{2\pi_{i,2}} - q_{\mathcal{N}(0,1)}(\alpha)\sqrt{\mathbb{V}(x_{i,t})}.$$

We impose $\pi_{i,1} > 0$ and $\pi_{i,2} > 0$, which implies that the constraints on the proxies are over-verified. Then, since our factors are jointly Gaussian, the first two unconditional moments are easily computed:

$$\begin{aligned} \mathbb{E}(X_t) &= (I_2 - \Phi)^{-1}(\mu_c,\, \mu_\ell)' \\ Vec[\mathbb{V}(X_t)] &= (I_4 - \Phi \otimes \Phi)^{-1} Vec(I_2) =: (\upsilon_c, \upsilon_{c\ell}, \upsilon_{c\ell}, \upsilon_\ell)', \end{aligned}$$

Eventually, we get the condition:

$$(\mu_c, \mu_\ell)' = (I_2 - \Phi)\left[-q_{\mathcal{N}(0,1)}(\alpha)\,(\sqrt{\upsilon_c}, \sqrt{\upsilon_\ell})'\right]. \tag{3.22}$$

In the estimation, we set $\alpha = 0.025$. We also control the accuracy of the fit of the proxies, and impose that both $\sigma_{\nu_c}^2$ and $\sigma_{\nu_l}^2$ equal 0.1 (a tenth of the proxies' variance).

3.E.2 Identifiability

In order to see if our model parameters are identifiable, we consider an affine transformation $\widetilde{X}_t$ of X_t, i.e. $\widetilde{X}_t = m + MX_t$, and we check that, if we have an observationally equivalent model when X_t is replaced by $\widetilde{X}_t$, then we necessarily have $m = 0$ and $M = I_2$

(i.e. $\widetilde{X}_t \equiv X_t$).

As the proxies are respectively functions of only one component of X_t, M has to be diagonal. Hence the alternative factors can be written: $\widetilde{x}_{i,t} = M_i x_{i,t} + m_i$ for $i = \{c, \ell\}$. The conditional variance of $\widetilde{X}_t$ must be equal to I_2, thus $M = I_2$. At that stage, we have that $\widetilde{X}_t = m + X_t$. $\widetilde{X}_t$ therefore follows a $VAR(1)$ with the same autoregressive matrix than X_t. Since Equation (3.22) also has to apply for $\widetilde{X}_t$, the latter necessarily features the same dynamics as X_t; therefore $m = 0$ and $\widetilde{X}_t \equiv X_t$.

Chapter 4

Recursive Compound Autoregressive Processes

Abstract

This Chapter develops a new general class of multivariate affine processes called *recursive affine*. We define these processes considering a multivariate process partitioned in different blocks. The conditional Laplace transform of each block given *(i)* the past values of the total vector, and *(ii)* the present values of blocks located above in the total vector, is an exponential-affine combination of the conditioning variables. This recursive definition allows to consider a broad class of distributions where conditional independence between components can be abandoned while the entire process still belongs to the affine class. After providing examples of recursive affine processes (such as stochastic volatility processes), we derive the statistical properties of this new class. Conditional and marginal first two moments are shown to be available in closed-form, leading to simple forecasting formulas. As for pricing, we provide the closed-form change of measure formulas using an exponential-affine stochastic discount factor. Finally we discuss the estimation issues depending on the observability of the process.

Résumé

Comme souligné dans l'introduction de la thèse, les processus dits *affines* sont couramment utilisés dans les modèles de valorisation des actifs car ils permettent d'obtenir les taux d'intérêt des obligations en formule fermée. Ils sont aisément définis à l'aide de leur transformée de Laplace conditionnelle. On dit qu'un processus est affine si sa transformée de Laplace conditionnelle à son passé est une fonction exponentielle-affine de son passé. Il existe un ensemble vaste de processus affines univariés, comme par exemple l'auto-régressif gaussien, l'auto-régressif Poisson, ou l'auto-régressif gamma. Hormis dans le cas purement gaussien, il peut être difficile de construire des processus affines multivariés. Une manière triviale est de considérer plusieurs processus affines univariés et de les supposer indépendants entre eux. Le vecteur composé de ces différents processus est évidemment affine. La plupart du temps, on peut relâcher cette hypothèse pour considérer des processus univariés affines que l'on suppose indépendants conditionnellement à leur passé. Sous réserve que la distribution ainsi produite soit bien définie, cette approche permet d'introduire de la causalité de Granger entre les différentes composantes du processus multivarié tout en conservant la propriété affine du vecteur. En revanche, aucune méthode générale n'a été considérée pour construire des processus multivariés à composantes non conditionnellement indépendantes, tout en restant dans la classe affine.

Dans ce chapitre, on introduit une nouvelle méthode de construction de processus affine multivarié par récurrence. On appelle ces processus *affines récursifs*. De même que pour les processus affines décrits ci-avant, les processus affines récursifs sont définis grâce à leur transformée de Laplace conditionnelle. Supposons un vecteur aléatoire X_t divisé en plusieurs blocs $(X_{i,t})$ de dimension supérieure ou égale à 1. Chaque bloc est un processus récursif affine si sa transformée de Laplace conditionnelle au passé du processus X_{t-1} et aux blocs situés avant lui dans le vecteur, i.e. les $X_{j,t}$ avec $j < i$, est une fonction exponentielle-affine des variables conditionnantes. La dépendance conditionnelle entre les différents blocs est donc définie par récurrence : le second bloc dépend conditionnellement du premier, le troisième bloc dépend conditionnellement des deux précédents, etc. Il est facile de montrer que les processus récursifs affines appartiennent à la classe des processus affines : la transformée de Laplace conditionnelle de n'importe quel bloc étant donné le passé du processus multivarié complet X_{t-1} est une fonction exponentielle-affine de X_{t-1}.

Grâce à la définition précédente, on montre que la classe des processus récursifs affines contient de nombreux exemples. On considère dans un premier temps des processus à volatilité stochastique, en utilisant des combinaisons entre des processus gaussiens, gamma, et Wishart. Plusieurs cas sont présentés, en utilisant des combinaisons différentes des proces-

sus précédents. On peut aussi considérer des processus multivariés à valeurs discrètes, des processus avec changement de régime, ou des processus vectoriels auto-régressifs gamma (voir Chapitre 5).

Étant donné que les processus récursifs affines font partie de la classe des processus affines, ils possèdent les mêmes propriétés. En particulier, les deux premiers moments conditionnels (au passé) des processus récursifs affines sont des fonctions affines du passé du processus. En revanche, la matrice de variance-covariance conditionnelle au passé du processus n'est pas forcément diagonale ou bloc-diagonale, contrairement au cas où les blocs sont conditionnellement indépendants. La possession des deux premiers moments conditionnels en formule fermée permet d'exprimer la dynamique des processus multivariés sous la forme d'un VAR semi-fort, où les chocs sont des différences de martingales de moyenne nulle et de variance unitaire. Sous cette forme, il est facile d'obtenir les conditions de stationarité, les deux premiers moments marginaux du processus, et les formules de prédiction optimale ainsi que la variance des erreurs de prédiction.

Les processus récursifs affines peuvent naturellement être utilisés à des fins de valorisation des actifs. On utilise pour cela une approche de *back modeling*, où l'on spécifie la dynamique des facteurs sous la mesure risque neutre sous la forme d'un processus récursif affine et le facteur d'escompte stochastique comme une fonction exponentielle-affine du processus. Grâce aux transformées d'Esscher, on montre que la transition de la dynamique risque-neutre à la dynamique historique est disponible en formule fermée. De plus, si les coefficients de sensibilité au risque dans le facteur d'escompte stochastique sont constants dans le temps, le processus est aussi récursif affine sous la dynamique historique. Dans ce cas, il est possible d'appliquer des méthodes classiques d'estimation. Si le processus récursif affine est directement observable, l'estimation peut être effectuée par maximum de vraisemblance ou pseudo maximum de vraisemblance. Si certaines composantes du vecteur ne sont pas observables, des méthodes de filtrage peuvent être considérées.

4.1 Introduction

Affine processes are widely used in asset pricing models since they are able to provide simple closed-form formulas in many situations (see for instance Duffie and Kan [1996] or Darolles, Gouriéroux, and Jasiak [2006]). They can be defined as processes which possess an exponential-affine conditional Laplace transform given their past (see general introduction of the Thesis). This class is indeed very broad, containing for instance the autoregressive Gaussian, the autoregressive Poisson, or the autoregressive Gamma univariate processes (see Gouriéroux and Jasiak [2006]). Except in the Gaussian case, building multivariate affine processes can be a challenging task. A first possibility is to assume independence between several univariate affine processes and stack them in a single vector.[1] Obviously such a multivariate process is affine. This independence assumption can often be loosened to impose only conditional independence between univariate components given their past. Provided that the conditional distribution is well-defined, such a process is, again, affine. However, building conditionally dependent multivariate affine processes in a general fashion has not yet been explored.

This paper provides a new class of multivariate affine processes that we call *recursive affine*, which allows for conditional dependence between elements in a recursive way. They can be defined as follows. We consider a vector of random variables that we denote by X_t. This vector is partitioned in several blocks $X_{i,t}$ of dimension one or greater than one. The conditional Laplace transform of $X_{i,t}$ given the past values X_{t-1} and the present values of blocks located above the block i in the vector (namely $\{X_{j,t},\, j < i\}$) is an exponential-affine function of the conditioning variables. Therefore, recursive affine processes are such that the conditional dependence goes from the first to the second block, from the first and second to the third block, and so on. recursive affine processes are shown to be affine as well, that is the conditional Laplace transform of any set of blocks given the past X_{t-1} is an exponential-affine function of X_{t-1}.

We are able to provide a large set of examples of recursive affine processes one can build, such as stochastic volatility processes. We consider for example vectorial autoregressive processes with conditionally Gaussian shocks. The conditional variance-covariance matrix of the shocks can be defined as any semi-positive matrix process, such as a Wishart autoregressive process. A second important example is the vectorial autoregressive gamma

1. This is often the case for the multivariate Cox, Ingersoll, and Ross [1985] process, the continuous-time equivalent of the autoregressive gamma process. In most asset pricing models, authors usually assume that the short-term risk-less interest rate is an affine combination of several independent CIR processes, see for instance Longstaff and Schwartz [1992], Dai and Singleton [2000] or Kim and Singleton [2012].

process (VARG), where components can be cross-correlated. This can help modeling stochastic volatility matrices or formulating positive factor models (see Chapter 5).

Using the properties of affine processes, we are able to derive closed-form formulas for conditional and marginal first two moments. As for every affine process, the conditional first two moments of recursive affine processes given the past are affine functions of the past values of the process. This allows us to formulate the dynamics of the process X_t in a semi-strong VAR representation, where the vector of shocks is a martingale difference with zero mean and unit variance. In particular, the conditional variance-covariance matrix can be time-varying and non-diagonal. This representation is convenient to obtain the second-order stationarity conditions, the first two marginal moments, the optimal forecast, and the variance of forecasting errors in closed-form.

Since affine processes are mostly used for asset pricing purposes, we show that it is possible to obtain closed-form change of measure formulas between the risk-neutral and the physical world using an exponential-affine stochastic discount factor. Considering a back modeling strategy, we define X_t as an affine process in the risk-neutral world. With the exponential-affine specification of the stochastic discount factor, the physical dynamics of X_t are easily obtained using the properties Esscher transforms (see Bertholon, Monfort, and Pegoraro [2008]). When the prices of risk are time-invariant, we obtain that X_t is recursive affine in the physical world as well. When that is the case, estimation strategies are easier to obtain. On the one hand, when the process is entirely observable, pseudo MLE or MLE can be performed; on the other hand, when X_t is only partly observable, Kalman filtering approximations can be readily obtained.

The remainder of the paper is organized as follows. We define the recursive affine processes in Section 4.2, and present several useful examples. In Section 4.3, we present the weak VAR representation of recursive affine processes and provide first two conditional moments, first two marginal moments, and forecasting formulas. Section 4.4 treats the use of recursive affine processes for asset pricing. Finally, we consider the estimation issues in Section 4.5.

4.2 The Class of Recursive Affine Processes

In this section, we define the class of recursive affine processes through their conditional Laplace transforms in the most general form. We show that these processes authorize a recursive structure for the conditional dependence between the components of the process while preserving the affine properties of the total process. We then present useful exam-

ples of recursive affine processes, such as stochastic volatility processes, discrete-valued multivariate processes, and vectorial autoregressive gamma processes.

4.2.1 Definition of Recursive Affine Process

Let us consider a multivariate discrete-time stochastic process $(X_t)_{t\in\mathbb{Z}}$. The vector X_t is partitioned into N subvectors $X_{i,t}$, $i \in \{1,\dots,N\}$, of size n_i. The total size of $X_t = (X_{1,t}^{'},\dots,X_{N,t}^{'})^{'}$ is denoted by $n := \sum_{i=1}^{N} n_i$. We also denote by $\underline{X}_t = (X_t, X_{t-1},\dots)$ the set of information at date t.

Definition 4.2.1. *The process $(X_t)_{t\in\mathbb{Z}}$ is said recursive affine if the conditional Laplace transforms:*

$$\mathbb{E}\left[\exp\left(u_i^{'} X_{i,t}\right) \mid X_{i-1,t},\dots,X_{1,t},\underline{X_{t-1}}\right], \quad i \in \{2,\dots,N\}, \tag{4.1}$$

are of the form:

$$\exp\left[\sum_{j=1}^{i-1} c_{i,j}^{'}(u_i)X_{j,t} + a_i'(u_i)X_{t-1} + b_i(u_i)\right], \quad \textit{for} \quad i \in \{2,\dots,N\}, \tag{4.2}$$

and, for $i = 1$,

$$\mathbb{E}\left[\exp\left(u_1^{'} X_{1,t}\right) \mid \underline{X_{t-1}}\right] = \exp\left[a_1^{'}(u_1)X_{t-1} + b_1(u_1)\right], \tag{4.3}$$

where the functions $a_i(\bullet)$, $b_i(\bullet)$ and $c_{i,j}(\bullet)$ transform vectors of size n_i in vectors of respective size n, 1, and n_j.

This definition thus implies that, for $i \in \{2,\dots,N\}$, the conditional Laplace transform of $X_{i,t}$ given the present values $(X_{i-1,t}^{'},\dots,X_{1,t}^{'})'$ and all the past values $\underline{X_{t-1}}$, are exponential-affine in $X_{i-1,t},\dots,X_{1,t}$ and X_{t-1}. A first important result is that, for any $i \in \{1,\dots,N\}$, the joint conditional Laplace transform of $(X_{1,t}^{'},\dots,X_{i,t}^{'})^{'}$, given $\underline{X_{t-1}}$, is exponential-affine in X_{t-1}. In particular, this is true for $i = N$ and, therefore, the process $(X_t)_{t\in\mathbb{Z}}$ is affine. We have the following result.

Proposition 4.2.1. *For any $i \in \{1,\dots,N\}$, the joint conditional Laplace transform of $(X_{1,t}^{'},\dots,X_{i,t}^{'})^{'}$, given $\underline{X_{t-1}}$, is given by:*

$$\mathbb{E}\left[\exp(u_1'X_{1,t} + \dots + u_i'X_{i,t}) \mid \underline{X_{t-1}}\right] = \exp\left[\widetilde{a}_i'(u_1,\dots,u_i)X_{t-1} + \widetilde{b}_i(u_1,\dots,u_i)\right] \tag{4.4}$$

where the functions $\widetilde{a}_i(\bullet)$ *and* $\widetilde{b}_i(\bullet)$ *are obtained recursively from:*

$$\left\{\begin{array}{rcl} \widetilde{a}_1(u_1) & = & a_1(u_1)\,,\ \widetilde{b}_1(u_1) = b_1(u_1) \\ \widetilde{a}_i(u_1,\ldots,u_i) & = & \widetilde{a}_{i-1}\left[u_1 + c_{i,1}(u_i),\ \ldots,\ u_{i-1} + c_{i,i-1}(u_i)\right] + a_i(u_i) \\ \widetilde{b}_i(u_1,\ldots,u_i) & = & \widetilde{b}_{i-1}\left[u_1 + c_{i,1}(u_i),\ \ldots,\ u_{i-1} + c_{i,i-1}(u_i)\right] + b_i(u_i)\,, \end{array}\right. \tag{4.5}$$

Proof. See Appendix 1. ■

Corollary 4.2.1.1. *The n-dimensional stochastic process* $(X_t)_{t\in\mathbb{Z}}$ *is affine and the conditional Laplace transform of* X_t*, given* $\underline{X}_{t-1}$*, is:*

$$\varphi_{t-1}(u) = \mathbb{E}\left[\exp(u' X_t)\,|\,\underline{X}_{t-1}\right] = \exp\left[\widetilde{a}'_n(u)X_{t-1} + \widetilde{b}_n(u)\right]$$

where $u = (u_1,\ldots,u_N)'$.

Proof. Straightforward consequence of Proposition 4.2.1. ■

The previous results provide a convenient way to specify a general multivariate affine process $(X_t)_{t\in\mathbb{Z}}$ of dimension $n = \sum_{i=1}^{N} n_i$. Indeed, it is possible to decompose this specification into N specifications of conditional distributions of the subvectors $X_{i,t}$. In particular, if $n_i = 1$ for any i $(N = n)$, we see that the specification of a n-variate affine process can be decomposed into the specification of n univariate conditional distributions. The multivariate processes thus obtained possess all the important properties characterizing the affine class (see Darolles, Gouriéroux, and Jasiak [2006]).

4.2.2 Examples

The previous method for specifying multivariate affine processes is extremely flexible. In the following section, we provide an extensive yet not exhaustive list of examples of recursive affine processes.

Stochastic Volatility Gaussian Processes

Let us assume that $X_{1,t}$ is a univariate Extended Autoregressive Gamma process, denoted by $ARG_\nu(\alpha,\beta,\mu)$ with parameters $\nu \geq 0$, $\alpha \geq 0$, $\beta > 0$ and $\mu > 0$ (cf Chapter 5 or Monfort, Pegoraro, Renne, and Roussellet [2014]). The conditional distribution of $X_{1,t}$, given $X_{1,t-1}$, is a non-central Gamma distribution $\gamma_\nu(\alpha + \beta\, X_{1,t-1}, \mu)$. In other words, the conditional distribution of $X_{1,t}$ can be defined by the mixture $\gamma_{\nu+z_t}(\mu)$, where z_t follows a Poisson distribution $\mathcal{P}(\alpha + \beta\, X_{1,t-1})$ conditionally on $X_{1,t-1}$. The conditional Laplace

transform of $X_{1,t}$, given $X_{1,t-1}$, is given by:

$$\forall\, u_1 < \frac{1}{\mu}\,, \quad \mathbb{E}\left[\exp\left(u_1 X_{1,t}\right)|X_{1,t-1}\right] = \exp\left[\frac{u_1\mu}{1-u_1\mu}(\alpha+\beta X_{1,t-1}) - \nu\log(1-u_1\mu)\right]. \quad (4.6)$$

We assume that the conditional distribution of $X_{2,t}$, given $\underline{X_{1,t}}$ and $\underline{X_{2,t-1}}$, is the Gaussian distribution $\mathcal{N}(\rho\, X_{2,t-1}, X_{1,t})$ with the associated conditional Laplace transform:

$$\mathbb{E}\left[\exp\left(u_2\, X_{2,t}\right)\,|\,\underline{X_{1,t}}\,,\underline{X_{2,t-1}}\right] = \exp\left[u_2\,\rho\, X_{2,t-1} + \frac{1}{2}\,u_2^2\, X_{1,t}\right]. \quad (4.7)$$

The process $X_t = (X_{1,t}, X_{2,t})'$ is recursive affine, therefore affine.

Stochastic Volatility VAR Processes

The previous example can easily be extended in order to get a VAR process with stochastic conditional variance-covariance matrix. We define m independent univariate Extended Autoregressive Gamma processes $(X_{1,t},\ldots,X_{m,t})$. We consider m other univariate processes $(X_{m+1,t},\ldots,X_{2m,t})$ and assume that, for any $i\in\{1,\ldots,m\}$:

$$X_{m+i,t}\,|\,X_{m+i-1,t},\ldots,X_{1,t},\underline{X_{t-1}} \sim \mathcal{N}\left(\sum_{j=1}^{i-1}\psi_{i,j}\,X_{m+j,t} + \sum_{k=1}^{m}\varphi_{i,k}\,X_{m+k,t-1},\; X_{i,t}\right).$$

The process $X_t = (X_{1,t},\ldots,X_{2m,t})'$ is recursive affine, therefore affine. A second method would be to assume that, conditionally on $(X_{1,t},\ldots,X_{m,t})$ and $\widetilde{X}_{t-1}$, the stacked vector $\widetilde{X}_t = (X_{m+1,t},\ldots,X_{2m,t})'$ is distributed, as follows:

$$\widetilde{X}_t|X_{1,t},\ldots,X_{m,t},\underline{X_{t-1}} \sim \mathcal{N}\left(\nu + \Phi\,\widetilde{X}_{t-1}, \sum_{j=1}^{m} X_{j,t}\,\Omega_j\right),$$

where the matrices Ω_j are symmetric of size $(m\times m)$, positive semi-definite. These matrices might be degenerated, of any rank greater or equal to 1. The number of Extended ARG processes entering the variance of the Gaussian VAR can also be smaller than m. Another possible stochastic volatility VAR process is presented hereafter.

Processes with Wishart Variance-Covariance Matrices

Let us consider here the case where $X_{1,t} = vec(W_t)$ is a vector of size $\frac{m(m+1)}{2}$, and $(W_t)_{t\in\mathbb{Z}}$ is a size $(m\times m)$ Wishart autoregressive (WAR) process. It is well known that $(X_{1,t})_{t\in\mathbb{Z}}$ is an affine process (see Gouriéroux, Jasiak, and Sufana [2009], Gouriéroux, Monfort, and Sufana [2010], Gouriéroux and Sufana [2010] and Gouriéroux and Sufana [2011]). Let $X_{2,t}$ be a n_2-dimensional vector. We define the conditional distribution of

$X_{2,t}$ given $\underline{X_{t-1}}$ and $X_{1,t}$ by:

$$X_{2,t}|X_{1,t}, \underline{X_{t-1}} \sim \mathcal{N}\left(\nu + \Phi_1 X_{1,t} + \Phi_2 X_{2,t-1}, \Omega W_t \Omega'\right) ,$$

where ν is a vector of size n_2, Φ_1 and Φ_2 are matrices of respective sizes $(n_2 \times \frac{m(m+1)}{2})$ and $(n_2 \times n_2)$, and Ω is a matrix of size $(n_2 \times m)$. We get a conditionally Gaussian VAR(1) process with a stochastic variance-covariance matrix The process $X_t = (X'_{1,t}, X'_{2,t})'$ is recursive affine, therefore affine.

Vector Autoregressive Gamma Processes

A vector Autoregressive Gamma process $X_t = (X_{1,t}, \ldots, X_{n,t})'$ with all $X_{i,t}$ of scalar size, denoted $VARG_\nu$ with $\nu = (\nu_1, \ldots, \nu_n)'$, can be defined as:

$$\forall\, i \in \{1, \ldots, n\}\,, \quad X_{i,t}|X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}} \sim \gamma_{\nu_i}\left(\alpha_i + \sum_{j=1}^{i-1} \beta_{i,j} X_{j,t} + \gamma_i' X_{t-1}, \mu_i\right) ,$$

with $\nu_i \geq 0$, $\alpha_i \geq 0$, $\beta_{i,j} \geq 0$, γ have all entries non-negative, and $\mu_i > 0$. It can be easily shown that the $VARG$ is a recursive affine process, therefore affine. All components of X_t are non-negative. In addition, if $\nu_i = 0$, $X_{i,t}$ can stay at zero for an extended period of time (see Monfort, Pegoraro, Renne, and Roussellet [2014] for further details).

Discrete Value Processes

Multivariate processes with some scalar components taking discrete values can also be obtained easily. Consider for instance a univariate autoregressive gamma process $X_{1,t}$, and the conditional distribution of $X_{2,t}$ given $X_{1,t}$ and the past is a Poisson distribution with parameter $\alpha + \beta X_{1,t} + \gamma' X_{t-1}$. Its conditional Laplace transform is given by:

$$\mathbb{E}\left[\exp(u X_{2,t})\middle|X_{1,t}, \underline{X_{t-1}}\right] = \exp[(\alpha + \beta X_{1,t} + \gamma' X_{t-1})(\exp(u) - 1)] ,$$

which is exponential-affine in $X_{1,t}$ and X_{t-1}. Another possibility might be to give this form to the parameter of the Poisson distribution appearing in the specification of a compound Poisson process (see Darolles, Gouriéroux, and Jasiak [2006] for details). We could also define in an obvious way multivariate recursive Poisson processes. Clearly, all these processes are affine.

Processes with Quadratic Terms

Let us consider a Gaussian VAR(1) process:

$$X_{1,t} = \nu + \Phi\, X_{1,t-1} + \varepsilon_t\,,\ \ \varepsilon_t \ \sim\ IIN(0,\Sigma)\,,$$

and let us define $X_{2,t} = vec(X_{1,t}\, X_{1,t}^{'})$. It is well known that the process $(X_{1,t}^{'}, X_{2,t}^{'})^{'}$ is affine (see Cheng and Scaillet [2007], Gouriéroux, Jasiak, and Sufana [2009] and Gouriéroux and Sufana [2011]). We introduce a third vector $X_{3,t}$, with non-negative components such that:

$$X_{3,i,t}\,|\,X_{3,i-1,t},\ldots,X_{3,1,t},\underline{X_{1,t}} \sim \gamma_{\nu_i}\left(\alpha_i + \sum_{j=1}^{i-1}\beta_{i,j}X_{3,j,t} + \gamma_i^{'}\,X_{1,t} + X_{1,t}^{'}\,\Gamma_i X_{1,t}, \mu_i\right) \quad (4.8)$$

where Γ_i is a symmetric semi-definite positive matrix, $\alpha_i \geq 0$, $\beta_{i,j} \geq 0$, , $\mu_i > 0$ and, in order to impose that the term $\alpha_i + \gamma_i^{'}\,X_{1,t} + X_{1,t}^{'}\,\Gamma_i X_{1,t}$ is non-negative, we can impose $\alpha_i > \frac{1}{4}\gamma_i^{'}\,\Gamma_i^{-1}\,\gamma_i$. The process $X_t = (X_{1,t}^{'}, X_{2,t}^{'}, X_{3,t}^{'})^{'}$ is recursive affine. Moreover, if $\nu_i = 0$, the non-negative components $X_{3,i,t}$ may stay at zero for some periods.

Processes with Switching Regimes

Let us assume that $X_{1,t}$ is a homogeneous Markov chain valued in $\{e_1,\ldots,e_L\}$, where e_ℓ, $\ell \in \{1,\ldots,L\}$, is the ℓ^{th} column of the identity matrix I_L. Denoting $\pi_{k,\ell} = \mathbb{P}(X_{1,t} = e_\ell\,|\,X_{1,t-1} = e_k)$ the transition probabilities, we know that $(X_{1,t})_{t\in\mathbb{Z}}$ is an affine process with conditional Laplace transform equal given by:

$$\begin{aligned}
\mathbb{E}\left[\exp(u'X_{1,t})\middle|X_{1,t-1}\right] &= \exp\left[a^{'}(u)X_{1,t-1}\right],\\
\text{with}\quad a^{'}(u) &= [a_1(u),\ldots,a_L(u)]\\
\text{and}\quad a_k(u) &= \log\left(\sum_{\ell=1}^{L}\exp(u^{'}e_\ell)\,\pi_{k,\ell}\right),\quad k\in\{1,\ldots,L\}
\end{aligned}$$

It is possible to incorporate $X_{1,t}$ into a recursive affine process in many ways. For instance, we can introduce the m-dimensional process $(X_{2,t})_{t\in\mathbb{Z}}$ such that the conditional distribution of $X_{2,t}$ given $X_{1,t}$ and $\underline{X_{t-1}}$ is $\mathcal{N}\left(A\,X_{1,t} + \Phi\,X_{2,t-1}, \Sigma(X_{1,t})\right)$ and $X_t = (X_{1,t}^{'}, X_{2,t}^{'})$ is recursive affine, since the Laplace transform of the previous distribution is given by:

$$\mathbb{E}\left[\exp(u'X_{2,t})\middle|X_{1,t},\underline{X_{t-1}}\right] = \exp\left[u^{'}A\,X_{1,t} + u^{'}\Phi X_{2,t-1} + \frac{1}{2}\left(u^{'}\Sigma(e_1)u,\ldots,u^{'}\Sigma(e_L)u\right)\,X_{1,t}\right], \quad (4.9)$$

where A is a size $(m\times L)$ matrix whose rows are the (regime-dependent) constant terms of the VAR dynamics, Φ is the size $(m\times m)$ autoregressive matrix, while $\Sigma(e_\ell)$ is the size

$(m \times m)$ regime-ℓ conditional variance-covariance matrix of $X_{2,t}$. We get a switching level and switching volatility Gaussian process. It is an example, among many others, of the introduction of switching regimes within recursive affine processes.

All the features of the previous examples may be combined and we can consider processes with some components taking real values, other components remaining non-negative and some of them being able to stay at zero, other components taking discrete values, with a Markov chain inducing switching regimes. In summary, the recursive affine class is very wide. In the following sections we develop their specific features.

4.3 Moments, VAR Representations, Stationarity Conditions and Prediction

This section provides moments, VAR representation, stationarity conditions and forecasting formulas for recursive affine processes exploiting the exponential-affine form of their conditional Laplace transform. For sake of notational simplicity, we will first consider the case where the components $X_{i,t}$, for any $i \in \{1, \ldots, N\}$, are univariate, so $N = n$. The general case will be considered in Section 4.3.4.

4.3.1 Conditional Moments and VAR representations

Using Definition 4.2.1, the log-Laplace transform of $X_{i,t}$, given $X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}$ is given by:

$$\psi_t(u_i) = \sum_{j=1}^{i-1} c_{i,j}(u_i) X_{j,t} + a_i'(u_i) X_{t-1} + b_i(u_i)\,. \tag{4.10}$$

The recursive conditional mean and variance of $X_{i,t}$, given $X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}$, are respectively given by:

$$\begin{aligned} \overline{m}_{i,t} &= \sum_{j=1}^{i-1} \left(\frac{\partial c_{i,j}(u_i)}{\partial u_i} \right)_{u_i=0} X_{j,t} + \left(\frac{\partial a_i'(u_i)}{\partial u_i} \right) X_t + \left. \frac{\partial b_i(u_i)}{\partial u_i} \right|_{u_i=0}, \\ \overline{\sigma}^2_{i,t} &= \sum_{j=1}^{i-1} \left(\frac{\partial^2 c_{i,j}(u_i)}{\partial u_i \partial u_i'} \right)_{u_i=0} X_{j,t} + \sum_{j=1}^{n} \left(\frac{\partial^2 a_{i,j}'(u_i)}{\partial u_i \partial u_i'} \right)_{u_i=0} X_{j,t} + \left. \frac{\partial^2 b_i(u_i)}{\partial u_i \partial u_i'} \right|_{u_i=0}, \end{aligned} \tag{4.11}$$

with obvious derivative notations. Denoting by $\overline{m}_t$ and $\overline{\sigma}^2_t$ the vectors whose components are, respectively, $\overline{m}_{i,t}$ and $\overline{\sigma}^2_{i,t}$, $i \in \{1, \ldots, n\}$, we have:

$$\begin{aligned} \overline{m}_t &= C_m\, X_t + A_m\, X_{t-1} + b_m\,, \\ \overline{\sigma}^2_t &= C_\sigma\, X_t + A_\sigma\, X_{t-1} + b_\sigma\,, \end{aligned} \tag{4.12}$$

where C_m and C_σ are lower triangular matrices with 0 on the main diagonal. Denoting by $\varepsilon_t = X_t - \overline{m}_t$, we have:

$$\begin{aligned} \mathbb{E}\left(\varepsilon_{i,t} \,|\, X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}\right) &= 0 \\ \mathbb{V}\left(\varepsilon_{i,t} \,|\, X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}\right) &= \overline{\sigma}^2_{i,t}\,, \end{aligned}$$

and therefore $\mathbb{E}(\varepsilon_{i,t} \,|\, \underline{X_{t-1}}) = 0$, $\mathbb{V}(\varepsilon_{i,t} \,|\, X_{t-1}) = \mathbb{E}(\overline{\sigma}^2_{i,t} \,|\, \underline{X_{t-1}})$ and $Cov(\varepsilon_{i,t}, \varepsilon_{j,t} \,|\, \underline{X_{t-1}}) = 0$, since $\mathbb{E}(\varepsilon_{i,t}\varepsilon_{j,t} \,|\, \underline{X_{t-1}}) = \mathbb{E}\left[\varepsilon_{i,t}\mathbb{E}(\varepsilon_{j,t} \,|\, X_{j-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}) \,|\, \underline{X_{t-1}}\right] = 0$, if $i < j$. We thus have the recursive VAR representation:

$$X_t = C_m\, X_t + A_m\, X_{t-1} + b_m + \varepsilon_t\,, \tag{4.13}$$

where ε_t is a martingale difference whose conditional variance-covariance matrix given the past is $diag[\mathbb{E}(\overline{\sigma}^2_{i,t} \,|\, X_{t-1})]$. From relation (4.13) we get:

$$X_t = \widetilde{A}_m\, X_{t-1} + \widetilde{b}_m + (I_n - C_m)^{-1}\, \varepsilon_t\,, \tag{4.14}$$

where $\widetilde{A}_m = (I_n - C_m)^{-1} A_m$ and $\widetilde{b}_m = (I_n - C_m)^{-1} b_m$. Since C_m is lower triangular with diagonal terms equal to zero, it is nilpotent, i.e. $C_m^h = 0$ for any $h > n$ and, therefore, $(I_n - C_m)^{-1} = I_n + C_m + \ldots + C_m^{n-1}$. We get the conditional moments:

$$\begin{aligned} m_t &:= \mathbb{E}(X_t \,|\, \underline{X_{t-1}}) = \widetilde{A}_m\, X_{t-1} + \widetilde{b}_m \\ \Sigma_t &:= \mathbb{V}(X_t \,|\, \underline{X_{t-1}}) = (I_n - C_m)^{-1}\, diag[\mathbb{E}(\overline{\sigma}^2_{i,t} \,|\, \underline{X_{t-1}})]\, (I_n - C_m)^{-1'}\,. \end{aligned} \tag{4.15}$$

In addition, from relations (4.12) and (4.15) we have:

$$\begin{aligned} \mathbb{E}(\overline{\sigma}^2_t \,|\, \underline{X_{t-1}}) &= C_\sigma\, m_t + A_\sigma\, X_{t-1} + b_\sigma = (C_\sigma\, \widetilde{A}_m + A_\sigma)\, X_{t-1} + C_\sigma\, \widetilde{b}_m + b_\sigma \\ &= A\, X_{t-1} + b \ \text{(say)} \end{aligned}$$

and therefore:

$$\Sigma_t \;=\; (I_n - C_m)^{-1}\, diag[A\, X_{t-1} + b]\, (I_n - C_m)^{-1'}\,.$$

So, if we denote by $\Omega_t = (I_n - C_m)^{-1}\, diag[A\, X_{t-1} + b]^{1/2}$, we have the semi-strong VAR representation:

$$X_t \;=\; \widetilde{A}_m\, X_{t-1} + \widetilde{b}_m + \Omega_t\, \eta_t\,, \tag{4.16}$$

where η_t is a martingale difference with a unit conditional variance-covariance matrix.

4.3.2 Unconditional Moments and Stationarity Conditions

Let us denote by $\widetilde{m}_t$ and $\widetilde{\Sigma}_t$ the unconditional mean and variance-covariance matrix of X_t. From (4.16) we have:

$$\begin{aligned}
\widetilde{m}_t &= \widetilde{A}_m\, \widetilde{m}_{t-1} + \widetilde{b}_m \\
\widetilde{\Sigma}_t &= \mathbb{V}(m_t) + \mathbb{E}(\Sigma_t) \\
&= \widetilde{A}_m\, \widetilde{\Sigma}_{t-1}\, \widetilde{A}_m' + (I - C_m)^{-1}\, diag[A\, \widetilde{m}_{t-1} + b]\, (I - C_m)^{-1'}\,,
\end{aligned} \tag{4.17}$$

or, written in vectorial form, we get:

$$vec(\widetilde{\Sigma}_t) = \left(\widetilde{A}_m \otimes \widetilde{A}_m\right)\, vec(\widetilde{\Sigma}_{t-1}) + [(I_n - C_m)^{-1} \otimes (I_n - C_m)^{-1}]\, (D\, \widetilde{m}_{t-1} + d)\,, \tag{4.18}$$

where $D\, \widetilde{m}_{t-1} + d = vec\,[diag(A\, \widetilde{m}_{t-1} + b)]$. We get the linear recursive system:

$$\begin{pmatrix} \widetilde{m}_t \\ vec(\widetilde{\Sigma}_t) \end{pmatrix} = \begin{pmatrix} \widetilde{A}_m & 0 \\ [(I_n - C_m)^{-1} \otimes (I_n - C_m)^{-1}]\, D & \widetilde{A}_m \otimes \widetilde{A}_m \end{pmatrix} \begin{pmatrix} \widetilde{m}_{t-1} \\ vec(\widetilde{\Sigma}_{t-1}) \end{pmatrix} + \begin{pmatrix} \widetilde{b}_m \\ [(I_n - C_m)^{-1} \otimes (I_n - C_m)^{-1}]\, d \end{pmatrix}. \tag{4.19}$$

Since the eigenvalues of $\widetilde{A}_m \otimes \widetilde{A}_m$ are all the products of the eigenvalues of $\widetilde{A}_m$, we get the following result:

Proposition 4.3.1. *The system (4.19) is convergent iif the moduli of the eigenvalues of $\widetilde{A}_m$ are all strictly smaller than one.*

Corollary 4.3.1.1. *When they exist, the stationary unconditional moments $\widetilde{m}$ and $\widetilde{\Sigma}$ of*

the process (X_t), obtained as the limits of the system (4.19), are:

$$\begin{aligned} \widetilde{m} &= (I_n - \widetilde{A}_m)^{-1}\,\widetilde{b}_m \\ vec(\widetilde{\Sigma}) &= (I_{n^2} - \widetilde{A}_m \otimes \widetilde{A}_m)^{-1}\,[(I_n - C_m)^{-1} \otimes (I_n - C_m)^{-1}]\,(D\,\widetilde{m} + d)\,. \end{aligned} \tag{4.20}$$

In addition, if we iterate relation (4.14) we get:

$$\begin{aligned} X_{t+h} &= \widetilde{A}_m^h\,X_t + \sum_{i=0}^{h-1} \widetilde{A}_m^i\,\left[\widetilde{b}_m + (I_n - C_m)^{-1}\,\varepsilon_{t+h-i}\right] \\ &= \widetilde{A}_m^h\,X_t + (I_n - \widetilde{A}_m^h)(I_n - \widetilde{A}_m)^{-1}\widetilde{b}_m + \sum_{i=0}^{h-1} \widetilde{A}_m^i\,(I_n - C_m)^{-1}\,\varepsilon_{t+h-i}\,, \end{aligned} \tag{4.21}$$

and thus $Cov(X_{t+h}, X_t) = \widetilde{A}_m^h\,\widetilde{\Sigma}$. Therefore, we obtain the following result:

Proposition 4.3.2. *If all the eigenvalues of $\widetilde{A}_m$ have a modulus strictly smaller than one, the process (X_t) is (asymptotically) second order stationary, with mean $\widetilde{m}$ and auto-covariance function $\widetilde{A}_m^h\,\widetilde{\Sigma}$.*

A particular case occurs when $\forall u_i$, $a_{i,j}(u_i) = 0$ for $j > i$. In this case, the element $X_{i,t}$ depends (conditionally and unconditionally) only on the elements located above him in the vector X_t. In other words this means that $X_{j,t}$ does not Granger cause $X_{i,t}$ (for any $j > i$). The matrix A_m is then lower triangular and, since $(I_n - C_m)$ is lower triangular with diagonal terms equal to 1, the same is true for $(I - C_m)^{-1}$. Finally, $\widetilde{A}_m = (I - C_m)^{-1}\,A_m$ is lower triangular with the same main diagonal elements as A_m. This result implies that the eigenvalues of $\widetilde{A}_m$ are the diagonal terms of A_m thus we have the following:

Corollary 4.3.2.1. *If $a_{i,j}(u_i) = 0$ for $j > i$, $(X_t)_{t\in\mathbb{Z}}$ is second-order stationary iif:*

$$\forall i \in \{1, \ldots, n\}, \quad \left|\frac{\partial a_{i,i}(u_i)}{\partial u_i}\right|_{u_i=0} < 1\,,$$

that is the i^{th} element of the gradient of $a_i(u_i)$, setting $u_i = 0$, has a modulus strictly smaller than 1.

4.3.3 Forecasting

From relation (4.21) we see that the optimal forecast of X_{t+h} at t is:

$$\begin{aligned} X_{t+h|t} &:= \mathbb{E}_t(X_{t+h}) = \widetilde{A}_m^h\,X_t + \sum_{i=0}^{h-1} \widetilde{A}_m^i\,\widetilde{b}_m \\ &= \widetilde{A}_m^h\,X_t + (I_n - \widetilde{A}_m^h)(I_n - \widetilde{A}_m)^{-1}\,\widetilde{b}_m\,, \end{aligned} \tag{4.22}$$

and the conditional variance-covariance matrix of the prediction error is given by:

$$vec\left[\mathbb{V}_t(X_{t+h})\right] = \sum_{i=0}^{h-1} \left(\widetilde{A}_m^i \otimes \widetilde{A}_m^i\right) \left[(I_n - C_m)^{-1} \otimes (I_n - C_m)^{-1}\right] \left(DX_{t+h-i-1|t} + d\right) , \quad (4.23)$$

where $X_{t+h-i-1|t}$ is given in (4.22). In summary, we have explicit forms for both the optimal forecasts and the variance-covariance matrices of the prediction errors.

4.3.4 Extension to the General Case

Consider now the general case where $X_t = (X_{1,t}', \ldots, X_{N,t}')'$, each $X_{i,t}$ is of size $n_i \geqslant 1$. The conditional log-Laplace transform of $X_{i,t}$, given $X_{i-1,t}, \ldots, X_{1,t}$, $\underline{X_{t-1}}$ is:

$$\psi_t(u_i) = \sum_{j=1}^{i-1} c_{i,j}'(u_i) X_{j,t} + a_i'(u_i) X_{t-1} + b_i(u_i) , \quad (4.24)$$

where u_i is of size n_i, and $c_{i,j}(\bullet)$ and $a_i(\bullet)$ transform vectors of size n_i in vectors of respective size n_j and n. If we denote by $\overline{m}_{i,t} = \mathbb{E}(X_{i,t} \,|\, X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}})$, we have again:

$$\overline{m}_{i,t} = \sum_{j=1}^{i-1} \left(\frac{\partial c_{i,j}'(u_i)}{\partial u_i}\right)_{u_i=0} X_{j,t} + \left(\frac{\partial a_i'(u_i)}{\partial u_i}\right)_{u_i=0} X_{t-1} + \left.\frac{\partial b_i(u_i)}{\partial u_i}\right|_{u_i=0} , \quad (4.25)$$

with obvious derivative notations. The Jacobians of $c_{i,j}'(u_i)$, $a_i'(u_i)$ and $b_i(u_i)$ are now respectively a matrix of size $(n_i \times n_j)$, a matrix of size $(n_i \times n)$, and a vector of size n_i. Stacking the $\overline{m}_{i,t}$, for any $i \in \{1, \ldots, N\}$ we get the same equation as (4.12):

$$\overline{m}_t = C_m X_t + A_m X_{t-1} + b_m , \quad (4.26)$$

where now the C_m and A_m are $(n \times n)$ matrices being defined by the blocks of the Jacobians of $c_{i,j}'(u_i)$ and of $a_i'(u_i)$ respectively, setting $u_i = 0$. Note that C_m is block-lower triangular, the diagonal blocks being equal to zero. We have the block-recursive VAR representation equivalent to Equation (4.13):

$$X_t \;\; = \;\; C_m X_t + A_m X_{t-1} + b_m + \varepsilon_t , \quad (4.27)$$

the error term ε_t being a martingale difference whose conditional variance-covariance matrix given $\underline{X_{t-1}}$ is block-diagonal. The i^{th} block of this matrix can be expressed as

$\mathbb{E}(\overline{\Sigma}_{i,t} \,|\, \underline{X_{t-1}})$ where $\overline{\Sigma}_{i,t}$ is defined by:

$$\overline{\Sigma}_{i,t} = \sum_{j=1}^{i-1} \sum_{k=1}^{n_j} \left(\frac{\partial^2 c_{i,j}^{(k)'}(u_i)}{\partial u_i \partial u_i'} \right)_{u_i=0} X_{j,t}^{(k)} + \sum_{j=1}^{N} \sum_{k=1}^{n_j} \left(\frac{\partial^2 a_{i,j}^{(k)'}(u_i)}{\partial u_i \partial u_i'} \right)_{u_i=0} X_{j,t-1}^{(k)} + \left. \frac{\partial^2 b_i(u_i)}{\partial u_i \partial u_i'} \right|_{u_i=0}, \tag{4.28}$$

where the exponent $^{(k)}$ stands for the k^{th} element of the considered vector. From Equation (4.28), it is clear that $\overline{\Sigma}_{i,t}$ is an affine function of the $X_{j,t}$ and $X_{j,t-1}$. The VAR representation of Equation (4.14) remains valid:

$$X_t = \widetilde{A}_m \, X_{t-1} + \widetilde{b}_m + (I - C_m)^{-1} \, \varepsilon_t \, . \tag{4.29}$$

Also, we still have:

$$\begin{aligned} m_t &= \mathbb{E}(X_t \,|\, \underline{X_{t-1}}) = \widetilde{A}_m \, X_{t-1} + \widetilde{b}_m \\ \Sigma_t &= \mathbb{V}(X_t \,|\, \underline{X_{t-1}}) = (I_n - C_m)^{-1} \, bdiag[\mathbb{E}(\overline{\Sigma}_{i,t} \,|\, \underline{X_{t-1}})] \, (I_n - C_m)^{-1'} \, , \end{aligned} \tag{4.30}$$

the matrix $bdiag[\mathbb{E}(\overline{\Sigma}_{i,t} \,|\, \underline{X_{t-1}})]$ being the block-diagonal matrix whose diagonal blocks are $\mathbb{E}(\overline{\Sigma}_{i,t} \,|\, \underline{X_{t-1}})$, which are affine functions of the $X_{j,t-1}$'s (see Equation (4.28)). The stationarity conditions remain the same and, in the particular case where $X_{j,t}$ does not cause $X_{i,t}$ for $j > i$, these conditions boil down to, for any $i \in \{1, \ldots, N\}$, the eigenvalues of the $(n_i \times n_i)$ Jacobian matrix of $a_{i,i}(u_i)$, setting $u_i = 0$, must have a modulus strictly smaller than one.

4.4 Pricing

Affine processes are well-known to be particularly useful for asset pricing purposes. They constitute the basis of the so-called *affine models* (see e.g. Duffie and Kan [1996] or Duffie and Singleton [1997]) where interest rates on assets are obtained as closed-form affine combinations of the components of an affine process. In this section, we present the properties of recursive affine processes for asset pricing considering an exponential-affine stochastic discount factor. Notably, the change between the historical and the pricing measure is available in closed-form.

4.4.1 Basic tools

Any no-arbitrage asset pricing model involves four basic mathematical tools (see general Introduction of the thesis): the nominal short-term interest rate between $t-1$ and t (known at $t-1$), denoted by r_{t-1}, which is specified as a function of $\underline{X_t}$; the family of his-

torical conditional probability density functions $f^{\mathbb{P}}(X_t \,|\, \underline{X_{t-1}})$; the family of risk-neutral conditional probability density functions $f^{\mathbb{Q}}(X_t \,|\, \underline{X_{t-1}})$ and the family of stochastic discount factors (s.d.f.) between $t-1$ and t specified as a function of $\underline{X_t}$ and denoted by $M_{t-1,t}(\underline{X_t})$. Under no-arbitrage assumptions, these tools are related by the following expression:

$$f^{\mathbb{P}}(X_t \,|\, \underline{X_{t-1}}) = f^{\mathbb{Q}}(X_t \,|\, \underline{X_{t-1}})\, M^{-1}_{t-1,t}(\underline{X_{t-1}})\, \exp[-r_{t-1}(\underline{X_{t-1}})]\,. \tag{4.31}$$

We adopt the *back modeling* strategy (see Bertholon, Monfort, and Pegoraro [2008]). We first specify a recursive affine risk-neutral dynamics in order to benefit from all the nice properties of affine processes. In particular we have quasi-explicit forms for the multi-horizon Laplace transform (see Gouriéroux, Monfort, Pegoraro, and Renne [2014]) and multi-horizon truncated Laplace transforms, implying quasi-explicit equations for many asset pricing formulas when the short rate is an affine function of $\underline{X_t}$. Once the risk-neutral dynamics and the short-rate are specified, we see from Equation (4.31) that we can choose to specify either any one-period stochastic discount factor $M_{t-1,t}(\underline{X_t})$ satisfying $\mathbb{E}^{\mathbb{Q}}_{t-1}(M^{-1}_{t-1,t}(\underline{X_t})) = \exp[r_{t-1}(\underline{X_{t-1}})]$, or the historical dynamics. At this stage, key considerations are the interpretability of the s.d.f. and the tractability of the historical dynamics in terms of inference. We follow a large literature recommending an exponential-affine specification of the s.d.f.:

$$M_{t-1,t}(\underline{X_t}) = \exp[-r_{t-1}(\underline{X_{t-1}}) - \lambda'_{t-1} X_t + \psi^{\mathbb{Q}}_{t-1}(\lambda_{t-1})]\,, \tag{4.32}$$

where λ_{t-1} is a vector of risk sensitivity parameters, function of $\underline{X_{t-1}}$, and $\psi^{\mathbb{Q}}_{t-1}(u)$ is the risk-neutral conditional log-Laplace transform of X_t given $\underline{X_{t-1}}$. We immediately see from Equation (4.32) that the constraint $\mathbb{E}^{\mathbb{Q}}_{t-1}(M^{-1}_{t-1,t}(\underline{X_t})) = \exp[r_{t-1}(\underline{X_{t-1}})]$ is automatically satisfied. Let us know examine the properties of the family of historical dynamics which are obtained if we choose a recursive affine risk-neutral dynamics and a s.d.f. given by Equation (4.32).

4.4.2 Recursive historical dynamics and Esscher transforms

From Relations (4.31) and (4.32) we see that:

$$f^{\mathbb{P}}(X_t \,|\, \underline{X_{t-1}}) = f^{\mathbb{Q}}(X_t \,|\, \underline{X_{t-1}})\, \exp[\lambda'_{t-1} X_t - \psi^{\mathbb{Q}}_{t-1}(\lambda_{t-1})]\,. \tag{4.33}$$

In other words, $f^{\mathbb{P}}(X_t \,|\, \underline{X_{t-1}})$ is the conditional Esscher transform of $f^{\mathbb{Q}}(X_t \,|\, \underline{X_{t-1}})$ associated with the parameter vector λ_{t-1}. From Equation (4.33) we also see that the

conditional Log-Laplace transform in the historical world is given by:

$$\psi_{t-1}^{\mathbb{P}}(u) = \psi_{t-1}^{\mathbb{Q}}(u + \lambda_{t-1}) - \psi_{t-1}^{\mathbb{Q}}(\lambda_{t-1})\,. \tag{4.34}$$

An important issue is now the tractability of the historical dynamics. In the next proposition we show that the historical dynamics has a recursive structure in which each conditional distribution is easily derived from the corresponding risk-neutral one.

Proposition 4.4.1. *Let us suppose that the risk-neutral dynamics of the n-dimensional stochastic process $X_t = (X_{1,t}', \ldots, X_{N,t}')'$ is recursive affine and that the one-period stochastic discount factor is given by Equation (4.32). The prices of risk λ_{t-1} are expressed as a block vector $\lambda_{t-1} = (\lambda_{1,t-1}', \ldots, \lambda_{N,t-1}')'$. Then, the historical conditional distribution of $X_{N-i,t}$, given $(X_{N-i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}})$ is the conditional Esscher transform of the corresponding risk-neutral distributions associated with the parameters $\widetilde{\lambda}_{N-i,t-1}$, obtained from $\lambda_{i,t-1}$'s by the backward recursion:*

$$\begin{cases} \widetilde{\lambda}_{N,t-1} &= \lambda_{N,t-1} \\ \widetilde{\lambda}_{N-i,t-1} &= \lambda_{N-i,t-1} + \displaystyle\sum_{k=1}^{i} c_{N-k+1,n-i}(\widetilde{\lambda}_{N-k+1,t-1})\,, \quad i \in \{1, \ldots, N-1\}\,. \end{cases} \tag{4.35}$$

Proof. See Appendix 2. ■

From proposition 4.4.1 it is clear that if the components $X_{1,t}', \ldots, X_{N,t}'$ are conditionally independent in the risk-neutral world given $\underline{X_{t-1}}$, the same is true in the historical world. In this case, we obtain $\widetilde{\lambda}_{N-i,t-1} = \lambda_{N-i,t-1}$ for any $i \in \{0, \ldots, N-1\}$. Moreover, if up to a certain order $p < N$, we have $\lambda_{N-i,t-1} = 0$ for $i \in \{0, \ldots, p\}$, we have $\widetilde{\lambda}_{N-i,t-1} = 0$ for $i \in \{0, \ldots, p\}$ and thus, the conditional distributions of $X_{N-i,t}$, given $X_{N-i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}$ are the same under both measures, for all $i \in \{0, \ldots, p\}$.

Corollary 4.4.1.1. *If we denote by $\psi_{i,t}^{\mathbb{Q}}(u_i)$ the risk-neutral log-Laplace transform of $X_{i,t}$ given $X_{i-1,t}, \ldots, X_{1,t}$, $\underline{X_{t-1}}$, the corresponding historical log-Laplace transform is:*

$$\psi_{i,t}^{\mathbb{P}}(u_i) = \psi_{i,t}^{\mathbb{Q}}(u_i + \widetilde{\lambda}_{i,t-1}) - \psi_{i,t}^{\mathbb{Q}}(\widetilde{\lambda}_{i,t-1})\,. \tag{4.36}$$

Proof. Straightforward consequence of Proposition 4.4.1. ■

Corollary 4.4.1.2. *If $\widetilde{\lambda}_{t-1}$ does not depend on $\underline{X_{t-1}}$, then the historical dynamics of (X_t) is recursive affine.*

Proof. If $\widetilde{\lambda}_{t-1} = \widetilde{\lambda}$, it is immediately seen from Equation 4.36 that $\psi_{i,t}^{\mathbb{P}}(u_i)$ is affine in $X_{i-1,t}, \ldots, X_{1,t}, X_{t-1}$. ■

The previous results show that a recursive structure of the historical dynamics is easily derived from the risk-neutral one. Moreover, as shown below, the results about recursive conditional moments and VAR representations are easily transposed to the historical dynamics.

4.4.3 Moments and VAR representation of the historical dynamics

Let us consider the risk-neutral conditional log-Laplace transform of $X_{i,t}$, given $X_{i-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}$, introduced in Equation (4.24):

$$\psi^{\mathbb{Q}}_{i,t}(u_i) = \sum_{j=1}^{i-1} c'_{i,j}(u_i) X_{j,t} + a'_i(u_i) X_{t-1} + b_i(u_i)\,, \tag{4.37}$$

and let us denote by $\overline{m}^{\mathbb{P}}_{i,t}$ and $\overline{\Sigma}^{\mathbb{P}}_{i,t}$ the corresponding recursive historical conditional mean and variance-covariance matrix. Then, we have the following proposition:

Proposition 4.4.2.

$$\overline{m}^{\mathbb{P}}_{i,t} = \left(\frac{\partial \psi^{\mathbb{Q}}_{i,t}(u_i)}{\partial u_i}\right)_{u_i=\widetilde{\lambda}_{i,t-1}} \quad \textit{and} \quad \overline{\Sigma}^{\mathbb{P}}_{i,t} = \left(\frac{\partial^2 \psi^{\mathbb{Q}}_{i,t}(u_i)}{\partial u_i \partial u'_i}\right)_{u_i=\widetilde{\lambda}_{i,t-1}}.$$

Proof. We know that $\overline{m}^{\mathbb{P}}_{i,t}$ is the gradient of $\psi^{\mathbb{P}}_{i,t}(u_i)$ setting $u_i = 0$ and $\overline{\Sigma}^{\mathbb{P}}_{i,t}$ is the matrix of second order derivatives of $\psi^{\mathbb{P}}_{i,t}(u_i)$, setting $u_i = 0$. The result follows directly from the expression of $\psi^{\mathbb{P}}_{i,t}(u_i)$ given in Corollary 4.4.1.1. ■

Proposition 4.4.2 has several interesting consequences. Considering the case $n_i = 1$ for the sake of notational simplicity (the results in the general case are easily obtained), we have:

- (*i*) the expressions of $\overline{m}^{\mathbb{P}}_t$ and $\overline{\Sigma}^{\mathbb{P}}_t$ are given by the same equations as in Equation (4.12) in which the matrices C_m, A_m, b_m, C_σ, A_σ and b_σ are replaced by the equivalent matrices computed from the first-order and second-order derivatives of the (risk-neutral) functions $c_{i,j}(\bullet)$ and $a_i(\bullet)$ evaluated in $\widetilde{\lambda}_{i,t-1}$.
- (*ii*) the recursive form of Equation (4.13) and the VAR of Equation (4.16) are still valid representations, the only difference being that the new matrices C_m, A_m, b_m, A, b, $\widetilde{A}_m$ and $\widetilde{b}_m$ may depend on $\underline{X_{t-1}}$ if $\lambda_{i,t-1}$ depends on X_{t-1}.
- (*iii*) if the $\lambda_{i,t-1}$'s do not depend on $\underline{X_{t-1}}$, Proposition 4.3.2, Corollary 4.3.2.1 and the forecasting formulas (4.22) and (4.23) remain valid using the new matrices.

4.5 Inference

Let us denote by θ the set of unknown parameters. The estimation methods of θ primarily depend on the observability of the vector X_t and on the availability of other observations.

4.5.1 The observable process case

If the process $(X_t)_{t\in\mathbb{Z}}$ is entirely observable, and if the p.d.f.'s $f_t^{\mathbb{P}}(X_{i,t} \,|\, X_{j-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}; \theta)$ are tractable, then the log-likelihood function $\mathcal{L}_T(\theta)$ given by:

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \left[\sum_{i=2}^{N} \left(\log f_t^{\mathbb{P}}(X_{i,t} \,|\, X_{j-1,t}, \ldots, X_{1,t}, \underline{X_{t-1}}; \theta) \right) + \log f_t^{\mathbb{P}}(X_{1,t} \,|\, \underline{X_{t-1}}; \theta) \right] \quad (4.38)$$

is available in closed-form, and the maximum likelihood estimator (MLE) can be computed. If the aforementioned p.d.f.'s are untractable, we can use the pseudo maximum likelihood of order two (see Gouriéroux, Monfort, and Trognon [1984]) since the exact expressions of the recursive conditional mean and variance-covariance matrices in the historical world, namely $\overline{m}_{i,t}^{\mathbb{P}}(\theta)$ and $\overline{\Sigma}_{i,t}^{\mathbb{P}}(\theta)$, are easily deduced from the risk-neutral conditional log-Laplace transforms $\psi_{i,t}^{\mathbb{Q}}(u_i)$ (see Proposition 4.4.2). The pseudo maximum likelihood method thus consists in minimizing the following criterion with respect to θ:

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left\{ \log \left| \overline{\Sigma}_{i,t}^{\mathbb{P}}(\theta) \right| + \left[X_{i,t} - \overline{m}_{i,t}^{\mathbb{P}}(\theta) \right]' \left(\overline{\Sigma}_{i,t}^{\mathbb{P}}(\theta) \right)^{-1} \left[X_{i,t} - \overline{m}_{i,t}^{\mathbb{P}}(\theta) \right] \right\}. \quad (4.39)$$

4.5.2 The (partially) latent process case

In some cases the maximum likelihood method can still be used when X_t is partially observable. For instance, if $X_t = (X_{1,t}', X_{2,t}')'$, where $X_{2,t}$ is observable and $X_{1,t}$ is a latent Markov chain (with a transition probability matrix possibly depending on $\underline{X_{2,t-1}}$) it may be possible to compute the MLE using the Kitagawa-Hamilton algorithm. In general, it is always possible to consider the equivalent of the VAR representation (4.16) in the historical world as a set of transition equations and to add measurement equations which depend on the particular asset pricing model we consider. Since the risk-neutral dynamics is affine, we could for instance obtain interest rate formulas (based on risk-less or defaultable bonds) as affine functions of X_t and the introduction of additive measurement errors would provide a set of affine measurement equations. Finally, we obtain a state-space model. The estimation of this state-space model would depend in particular on the quantitative or qualitative nature of the components of X_t. A mixture of (extended) Kalman filter,

Kitagawa-Hamilton filter or inversion techniques (see Chen and Scott [1993], Pearson and Sun [1994], Monfort and Renne [2013], Monfort and Renne [2014]) can provide a solution to the estimation problem.

4.6 Conclusion

This chapter develops a new class of multivariate affine processes that we call *recursive affine*. We define the dependence between the components of the entire vector in a block recursive fashion. More specifically, recursive affine processes are defined through their conditional Laplace transform: a process is said to be recursive affine if the conditional Laplace transform of each block composing the total process, given the past values of the process and the present values of blocks located above the considered block in the vector, is an exponential-affine function of the conditioning variables. The first block therefore depends only on the past values of the whole process, while the second block depends on the past values of the whole process **and** on the present values of the first block. This definition allows these processes to depart from the conditional independence assumption to build multivariate affine processes.

We show that this class is very wide providing several examples, such as stochastic volatility processes, multivariate discrete-valued processes, or vectorial autoregressive gamma processes. Since recursive affine processes are affine, we derive the usual properties associated with this class of processes. We show that recursive affine processes possess a weak VAR representation, leading to simple closed-form formulas for both conditional and marginal first two moments of the entire process. This representation also allows for a straightforward computation of forecasting formulas.

Last, we show how the recursive affine processes can be used for asset pricing purposes considering an exponential-affine stochastic discount factor. We provide closed-form change of measure formulas using the Esscher transform of the conditional density functions. In the end, we document estimation techniques that can be used to estimate models with recursive affine processes.

Appendices of Chapter 4

4.A Proof of Proposition 4.2.1

The result is obvious for $i = 1$. Let us assume that it is true $i-1$, and let us show that is also true for i. We have:

$$
\begin{aligned}
& \mathbb{E}\left[\exp(u_1' X_{1,t} + \ldots + u_i' X_{i,t}) \,|\, \underline{X_{t-1}}\right] \\
= \; & \mathbb{E}\left\{ \exp\left(\sum_{k=1}^{i-1} u_k' X_{k,t} \right) \mathbb{E}\left[\exp(u_i X_{i,t}) \,|\, X_{1,t}, \ldots, X_{i-1,t}, \underline{X_{t-1}} \right] \,|\, \underline{X_{t-1}} \right\} \\
= \; & \mathbb{E}\left\{ \exp\left(\sum_{k=1}^{i-1} u_k' X_{k,t} \right) \exp\left(\sum_{k=1}^{i-1} c_{i,k}'(u_i) X_{k,t} + a_i'(u_i) X_{t-1} + b_i(u_i) \right) \,|\, \underline{X_{t-1}} \right\} \\
= \; & \exp\left(a_i'(u_i) X_{t-1} + b_i(u_i) \right) \mathbb{E}\left\{ \exp\left(\sum_{k=1}^{i-1} (u_k + c_{i,k}(u_i))' X_{k,t} \right) \,|\, \underline{X_{t-1}} \right\} \\
= \; & \exp\left(a_i'(u_i) X_{t-1} + b_i(u_i) \right) \\
& \qquad \times \exp\left[\widetilde{a}_{i-1}(u_1 + c_{i,1}(u_i), \ldots, u_{i-1} + c_{i,i-1}(u_i))' X_{t-1} \right. \\
& \qquad\qquad \left. + \widetilde{b}_{i-1}(u_1 + c_{i,1}(u_i), \ldots, u_{i-1} + c_{i,i-1}(u_i)) \right] \\
= \; & \exp\left\{ \left[\widetilde{a}_{i-1}(u_1 + c_{i,1}(u_i), \ldots, u_{i-1} + c_{i,i-1}(u_i)) + a_i(u_i) \right]' X_{t-1} \right. \\
& \qquad\qquad \left. + \left[\widetilde{b}_{i-1}(u_1 + c_{i,1}(u_i), \ldots, u_{i-1} + c_{i,i-1}(u_i)) + b_i(u_i) \right] \right\}.
\end{aligned}
$$

The result follows by identification.

4.B Proof of Proposition 4.4.1

Let X be a random variable. The Esscher transform of any density function $f_X(x)$ is given by the family of functions:

$$
f_X^{\lambda}(x) = \frac{\exp(\lambda x) f_X(x)}{\mathbb{E}_X\left[\exp(\lambda X)\right]}
$$

The proof of this proposition is based on the two following lemmas.

Lemma 4.B.1. *Let us consider a random vector $(X_1', X_2')'$ with p.d.f. $f(X_1, X_2)$ and the Esscher transform of $f(X_1, X_2)$ associated with parameters $\lambda = (\lambda_1', \lambda_2')$ and denoted by $f^{\lambda}(X_1, X_2)$. Then, the conditional distribution $f^{\lambda}(X_2 \,|\, X_1)$ is the Esscher transform of $f(X_2 \,|\, X_1)$ associated with parameter the λ_2.*

Proof. We have by definition, using the notation $\propto$ for indicating proportionality:

$$\begin{aligned} & f^{\lambda}(X_1, X_2) \propto f(X_1, X_2) \exp\left(\lambda_1' X_1 + \lambda_2' X_2\right) \\ & \text{or } \; f^{\lambda}(X_1, X_2) \propto f(X_1) \exp(\lambda_1' X_1) f(X_2|X_1) \exp(\lambda_2' X_2) \\ & \text{and } \; f^{\lambda}(X_2|X_1) \propto f(X_2|X_1) \exp(\lambda_2' X_2)\,, \end{aligned} \tag{4.40}$$

which gives the result. ■

Lemma 4.B.2. *Let us consider a n-dimensional random vector $X = (X_1', \ldots, X_n')'$, with p.d.f. $f(X_1, \ldots, X_n)$, and the Esscher transform of $f(X_1, \ldots, X_n)$ associated with parameters $\lambda = (\lambda_1', \ldots, \lambda_n')'$ is denoted by $f^{\lambda}(X_1, \ldots, X_n)$. Let us assume that the Laplace transform associated to the conditional p.d.f. $f(X_i \,|\, X_{i-1}, \ldots, X_1)$ is of the form:*

$$\varphi_i(u) \;=\; \mathbb{E}\left[\exp(u' X_i) \,|\, X_{i-1}, \ldots, X_1\right] = \exp\left[\sum_{j=1}^{i-1} c_{i,j}'(u) X_j + b_i(u)\right]\,, \;\; i \in \{2, \ldots, n\}\,, \tag{4.41}$$

with $\varphi_1(u) = \mathbb{E}\left[\exp(u' X_1)\right] = \exp\left[b_1(u)\right]$. Then, the following results hold:

(a) The marginal p.d.f. $f^{\lambda}(X_1, \ldots, X_{n-i})$ is the Esscher transform of $f(X_1, \ldots, X_{n-i})$ associated with the parameters

$$\lambda_j + \sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\,, \;\; j \in \{1, \ldots, n-i\}\,, \tag{4.42}$$

(with the convention $\sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1}) = 0$, for $i = 0$) where the $\widetilde{\lambda}_j$ are obtained recursively from:

$$\begin{cases} \widetilde{\lambda}_n \;=\; \lambda_n & \\ \widetilde{\lambda}_{n-i} \;=\; \lambda_{n-i} + \sum_{k=1}^{i} c_{n-k+1,n-i}(\widetilde{\lambda}_{n-k+1})\,, \;\; i \in \{1, \ldots, n-1\}\,. \end{cases} \tag{4.43}$$

(b) The conditional distribution $f^{\lambda}(X_{n-i} \,|\, X_{n-i-1}, \ldots, X_1)$ is the Esscher transform of $f(X_{n-i} \,|\, X_{n-i-1}, \ldots, X_1)$ associated with the parameter $\widetilde{\lambda}_{n-i}$ obtained from (4.43).

Proof. Let us consider, first, the computation of $f^{\lambda}(X_1, \ldots, X_{n-i})$, and then the computation of the associated $f^{\lambda}(X_{n-i} \,|\, X_{n-i-1}, \ldots, X_1)$.

– if $i = 0$, (a) is true.

– Let us assume now that a) is true for i:

$$f^{\lambda}(X_1,\dots,X_{n-i}) \propto f(X_1,\dots,X_{n-i}) \exp\left[\sum_{j=1}^{n-i}\left(\lambda_j+\sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\right)' X_j\right]. \tag{4.44}$$

Replacing $f(X_1,\dots,X_{n-i})$ by $f(X_1,\dots,X_{n-i-1})\, f(X_{n-i}\,|\,X_1,\dots,X_{n-i-1})$ on the right-hand side of the previous relation and integrating with respect to X_{n-i} gives:

$$\begin{aligned}
&f^{\lambda}(X_1,\dots,X_{n-i-1})\\
&\propto f(X_1,\dots,X_{n-i-1}) \exp\left[\sum_{j=1}^{n-i-1}\left(\lambda_j+\sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\right)' X_j\right]\\
&\quad\times \int f(X_{n-i}\,|\,X_1,\dots,X_{n-i-1}) \exp\left[\left(\lambda_{n-i}+\sum_{k=1}^{i} c_{n-k+1,n-i}(\widetilde{\lambda}_{n-k+1})\right)' X_{n-i}\right] dX_{n-i}\\
&\propto f(X_1,\dots,X_{n-i-1}) \exp\left[\sum_{j=1}^{n-i-1}\left(\lambda_j+\sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\right)' X_j\right]\\
&\quad\times E\left\{\exp\left[\left(\lambda_{n-i}+\sum_{k=1}^{i} c_{n-k+1,n-i}(\widetilde{\lambda}_{n-k+1})\right)' X_{n-i}\right]\,|\,X_1,\dots,X_{n-i-1}\right\}\\
&\propto f(X_1,\dots,X_{n-i-1}) \exp\left[\sum_{j=1}^{n-i-1}\left(\lambda_j+\sum_{k=1}^{i} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\right) X_j\right]\\
&\quad\times \exp\left[\sum_{j=1}^{n-i-1} c_{n-i,j}\left(\lambda_{n-i}+\sum_{k=1}^{i} c_{n-k+1,n-i}(\widetilde{\lambda}_{n-k+1})\right)' X_j\right]\\
&\propto f(X_1,\dots,X_{n-i-1}) \exp\left[\sum_{j=1}^{n-i-1}\left(\lambda_j+\sum_{k=1}^{i+1} c_{n-k+1,j}(\widetilde{\lambda}_{n-k+1})\right)' X_j\right]
\end{aligned}$$

which is the result for i+1. Finally, Lemma 4.B.1 and the formula for $f^{\lambda}(X_1,\dots,X_{n-i})$ show that $f^{\lambda}(X_{n-i}\,|\,X_1,\dots,X_{n-i-1})$ is the Esscher transform of $f(X_{n-i}\,|\,X_1,\dots,X_{n-i-1})$ associated with the coefficient:

$$\lambda_{n-i}+\sum_{k=1}^{i} c_{n-k+1,n-i}(\widetilde{\lambda}_{n-k+1}) = \widetilde{\lambda}_{n-i}\,. \tag{4.45}$$

■

Now, the proof of Proposition 4.4.1 is a direct consequence of part (b) of Lemma 4.B.2 and of formula (4.43), if we consider the conditional distribution given $\underline{X_{t-1}}$.

Chapter 5

Staying at Zero with Affine Processes

This chapter is based on the article "Staying at Zero with Affine Processes: An Application to Term Structure Modeling" of Monfort, Pegoraro, Renne, and Roussellet [2014].

Abstract

We build an Affine Term Structure Model that provides non-negative yields at any maturity and which is able to accommodate a short-term rate that stays at the zero lower bound (ZLB) for extended periods of time while longer-term rates display high volatility. We introduce these features through a new univariate non-negative affine process called ARG-Zero, and its multivariate affine counterpart (VARG), entailing conditional distributions with a zero-point mass. The affine property of this new class of processes implies both explicit bond pricing and quasi-explicit lift-off probability formulas. We provide an empirical application to Japanese Government Bond (JGB) yields, observed weekly from June 1995 to May 2014 with maturities from six months to ten years. Our four-factor specification is able to closely match yield levels and to capture conditional yield volatilities.

Résumé

L'hypothèse de non-arbitrage des actifs financiers implique que les taux d'intérêt des obligations sans risque ne peuvent théoriquement atteindre des niveaux négatifs. En effet, si les taux d'intérêt devenaient négatifs, les investisseurs préfèreraient conserver de la monnaie et bénéficier d'un taux d'intérêt nul. Ce phénomène est appelé "taux plancher à zéro" (*zero lower bound*). En réalité, les frictions financières peuvent conduire les taux d'intérêt à devenir négatifs, mais il existe toujours une borne inférieure. On considère arbitrairement cette borne basse à zéro dans ce chapitre. À la suite de la récente crise financière, les principales zones monétaires mondiales (États-Unis, zone Euro, Japon, Royaume-Uni) ont vu le taux d'intérêt à court terme de leurs obligations sans risque chuter à des valeurs proches de zéro. Cette situation, résultat d'une politique monétaire la plus accommodante possible, est associée à un fait stylisé majeur : lorsque les taux d'intérêt à court terme tombent au plancher, ils restent à zéro pendant de longues périodes. Le Japon en est un parfait exemple : les taux d'intérêt japonais (JGB) ont atteint le taux plancher à zéro entre 2001 et 2006 et depuis 2008, malgré une courte remontée entre les deux périodes. En conséquence, la partie longue de la courbe des taux d'intérêt décroît car l'anticipation des taux d'intérêt à court terme futurs est plus basse et la totalité de la courbe des taux s'aplatit.

Ce comportement pose problème à la majorité des modèles de taux d'intérêt usuels. Le modèle affine gaussien implique une probabilité strictement positive d'obtenir des taux d'intérêt négatifs ; les modèles quadratiques de structure par terme (QTSM) ou Cox-Ingersoll-Ross produisent des taux d'intérêt positifs, mais traitent le plancher à zéro comme une barrière réflexive, et sont incapables de générer des taux d'intérêt à court terme restant ancrés à zéro pour de longues périodes. Récemment, les modèles de taux *shadow*[1] ont permis de prendre en compte simultanément la positivité des taux et la persistance des taux au plancher. En revanche, ces modèles sortent de la classe des modèles affines de taux d'intérêt et perdent les propriétés associées : les taux d'intérêt à long terme ne peuvent plus être obtenus par une combinaison des facteurs disponible en formule fermée et doivent donc être simulés. Ces modèles sont donc associés à une complexité computationnelle plus élevée, notamment pour la phase d'estimation.

Ce chapitre introduit un modèle affine de taux d'intérêt capable de produire des taux positifs à toutes maturités, un taux d'intérêt à court terme pouvant atteindre et rester au plancher, et dont les taux longs peuvent varier même lorsque le taux court est au plancher et sont exprimés comme des combinaisons affines des facteurs disponibles en

1. Terme non traduit pour désigner le *shadow rate model*.

formule formée. Ce modèle s'appuie sur un nouveau processus univarié affine appelé *auto-régressif Gamma-zéro* (ARG_0). Ce nouveau processus est défini à l'aide de la distribution statique gamma-zéro. Une variable aléatoire est gamma-zéro si elle est distribuée comme une gamma décentrée à paramètre de degré de liberté Poissonien. Lorsque la variable de mixage Poissonienne vaut zéro, la variable gamma est tirée dans une distribution convergeant vers la masse de Dirac en zéro. De cette manière, on obtient une variable aléatoire continue sur $\mathbb{R}_+$ dont la probabilité d'être exactement égale à zéro est non-nulle. On introduit la dynamique auto-régressive en supposant que l'intensité de la variable de mixage Poissonnienne est une fonction affine de la réalisation de la variable gamma-zéro à la période précédente. On obtient un processus univarié continu sur $\mathbb{R}_+$, pouvant atteindre zéro et y rester pour des périodes prolongées. La première section montre que les processus gamma-zéro sont affines ce qui permet de dériver aisément leurs conditions de stationarité, leurs deux premiers moments conditionnels et leurs deux premiers moments marginaux (voir Chapitre 4). De plus, on montre que les probabilités que ces processus atteignent zéro et restent à zéro sont disponibles en formules fermées.

Afin de relier les processus ARG_0 aux processus auto-régressifs gamma, on introduit une classe générale de processus appelés ARG étendus. Cette classe de processus est définie à l'aide de quatre paramètres. En imposant certains de ces paramètres à zéro, on retrouve les processus ARG classiques (sans masse en zéro) et les processus ARG_0 (avec masse en zéro). De la même manière, on montre que les ARG étendus sont des processus affines et on dérive leurs différentes propriétés.

À l'aide de ces processus, on développe un modèle affine de structure par terme des taux d'intérêt. En utilisant la stratégie du *back modeling*, on crée un processus multivarié affine sous $\mathbb{Q}$ en empilant des composantes ARG_0 et ARG dans un même vecteur (voir Chapitre 4). On suppose les différentes composantes conditionnellement indépendantes. En revanche, on impose une forme triangulaire supérieure à la matrice auto-régressive, autorisant de la causalité de Granger des composantes gamma vers les composantes gamma-zéro. Le taux d'intérêt à court terme est défini comme une combinaison linéaire des variables ARG_0 uniquement. Comme toutes ces variables possèdent une masse en zéro, le taux d'intérêt à court terme possède lui aussi une masse en zéro, conformément à son comportement au plancher. En introduisant une constante négative dans la spécification du taux d'intérêt à court terme, on peut aussi considérer une borne basse négative. Le modèle ainsi formulé est affine et les taux d'intérêt à toutes les maturités sont des fonctions affines de l'ensemble des facteurs gamma et gamma-zéro. En conséquence, même lorsque le taux court est à zéro, i.e. lorsque les composantes gamma-zéro sont à zéro, les taux longs peuvent continuer à fluctuer grâce aux composantes gamma. En utilisant un fac-

teur d'escompte stochastique exponentiel-affine, on obtient le même type de dynamique sous la mesure physique, avec des paramètres différents. Comme dans le cas univarié, les probabilités que le taux d'intérêt à court terme atteigne zéro et y reste sont disponibles en formule fermée sous les deux mesures. En revanche, elles peuvent varier dans le temps dans le cas multivarié.

On estime ce modèle avec quatre facteurs latents en considérant des taux d'intérêt sur obligations souveraines japonaises de 1995 à 2014. Afin d'obtenir des résultats empiriques satisfaisants, on inclut dans l'ensemble de variables observables des anticipations de taux futurs issues de sondages, ainsi que des indicateurs de la variance conditionnelle des taux d'intérêt. Le modèle montre une performance satisfaisante à reproduire le comportement des taux d'intérêt, de leur volatilité et des anticipations de taux d'intérêt. À l'aide des estimations, on calcule les probabilités que les taux restent au plancher à différents horizons (2 et 5 ans), sous la mesure historique et risque neutre. Nos calculs montrent que ces probabilités varient substantiellement au cours du temps et la différence entre les probabilités calculées sous les deux mesures peut atteindre les 30 points de pourcentage.

5.1 Introduction

Assuming that storing cash is costless, nominal interest rates cannot turn negative since cash provides a zero interest rate and is always an alternative investment to bonds (see e.g. Black [1995]). In other words, the sole existence of currency implies a zero lower bound (ZLB) on bond yields.[2] Before the outbreak of the 2008 financial crisis, the Bank of Japan was the only large central bank that had brought its policy rates – which drive the short-end of the yield curve – to zero. From 2010 on however, bringing policy rates close to the ZLB has become a common situation for the Fed, the ECB, and the BoE. In all of these currency areas, sharp decreases of short-term interest rates have led the medium- to long-term yields to drop deeply, pushing the entire yield curves to unprecedented low levels.

In this context, reproducing low but non-negative interest rates has become a great concern for the specification of term structure models, and still represents a challenging task.[3] Specifically, to the best of our knowledge, no existing term-structure model is able to simultaneously match the three following characteristics:[4]

(i) consistency with non-negative yields;

(ii) availability of closed-form bond pricing formulas; and

(iii) the ability to accommodate extended periods of zero short-term rates and to evaluate associated of leaving the zero lower bound (lift-off probabilities).

In this paper, we first introduce a new affine process that opens the way to term-structure models consistent with *(i)*, *(ii)*, and *(iii)* simultaneously. This process, which we call Autoregressive Gamma-zero (ARG_0), builds on the original ARG process (see Gouriéroux and Jasiak [2006], Le, Singleton, and Dai [2010] or Creal and Wu [2015]) by extending it to a zero degree-of-freedom parameter. This process has a crucial distinctive feature: its conditional distribution given the past values encompasses a point-mass at zero.[5] This attractive property allows its dynamics to satisfy *(i)* and *(iii)*.[6] We explore the properties

2. In reality, holding cash is not costless since it is subject to theft or physical distraction and is complicated to use for large transactions. These features, along with flight-to-safety phenomena or non-conventional monetary-policy measures, may result in negative interest rates. The framework we develop in the present paper is consistent with the existence of a lower bound which can be negative.

3. Typically, in the widely-used Gaussian no-arbitrage models, the yields of all maturities can take negative values with a strictly positive probability (see e.g. Dai and Singleton [2003], Piazzesi [2010], Diebold and Rudebusch [2013], Duffee [2012] or Gurkaynak and Wright [2012]).

4. While the model proposed by Renne [2012] is consistent with these three points, it can only generate a discrete number of positive yield curves. That is, in Renne's framework, the support of the positive short-term (policy) rate is discrete. Here, we consider short rates whose support is the set of non-negative real numbers (denoted by $\mathbb{R}^+$).

5. This appealing feature is obtained by building on Siegel [1979], who introduces a non-central Chi-squared distribution with zero degree of freedom. This distribution has also a Dirac mass at zero.

6. As noted by Kim [2008], coping with those two features for a short-term interest rate is of utmost importance when building a term-structure model with observed option prices.

of this univariate process, explicitly disclosing its exponential-affine conditional Laplace transform and its first two conditional and unconditional moments. This univariate affine process is then extended to a multivariate affine process which we call Vectorial Autoregressive Gamma (VARG). We adequately exploit these processes to build a multi-factor term-structure model in which the yields at all maturities are non-negative, the short-term interest rate can stay at zero for extended periods of time, and the lift-off probabilities are easily computed under both the historical and pricing measures.

We directly address the issue of point *(ii)*, making closed-form bond-pricing formulas available. Indeed, our short-term interest rate is specified as a linear combination of components that follow VARG processes under both historical and risk-neutral measures. Hence our framework boils down to an Affine Term-Structure Model (ATSM) and the zero-coupon yields for all maturities are explicit affine functions of the factors where the loadings are computable recursively (see e.g. Duffie and Kan [1996] or Darolles, Gouriéroux, and Jasiak [2006]).

The historical and risk-neutral affine property of our term-structure model allows for a great flexibility at the estimation stage. First, assuming the presence of latent factors, the estimation technique is computationally simple using Kalman filtering techniques. Indeed, transition equations of the underlying state-space model are simply given by the VAR representation of our factors' dynamics. Second, it implies that (a) forecasts and (b) conditional variances of yields are affine functions of the factors. Accordingly, this allows us to easily augment the set of measurement equations by relating linear combinations of our latent factors with observable proxies of (a) surveys of professional forecasters and (b) conditional (GARCH-based) yield variances. Including these equations respectively improves (a) the estimation of the factors' physical dynamics (see Kim and Orphanides [2012]) and (b) the consistency of the estimated model with sample moments of order two.[7]

As Japan has been confronted with extremely low interest rates since the mid-90s, its sovereign bond yields constitute a relevant source of data to examine the ability of term-structure models to handle the ZLB.[8] Our estimated model both shows a very good fit of yields and of conditional yield volatilities across maturities. We also find differences

7. Among others, Collin-Dufresne and Goldstein [2002], Adrian and Wu [2009], Andersen and Benzoni [2006], Trolle and Schwartz [2009], Jacobs and Karoui [2009], Almeida, Graveline, and Joslin [2011], Bikbov and Chernov [2011], Creal and Wu [2014] and Christensen, Lopez, and Rudebusch [2014] study the ability of term-structure models to fit conditional volatilities of yields.

8. See e.g. Gorovoi and Linetsky [2004], Ueno, Baba, and Sakurai [2006], Ichiue and Ueno [2007], Kim and Singleton [2012], Christensen and Rudebusch [2013], Kim and Priebsch [2013], Krippner [2013].

between historical and risk-neutral lift-off probabilities. Our model's estimates imply that at the 5-year horizon, the difference between the risk-neutral and historical probabilities of exiting the ZLB can be as large as 35 percentage points.

The present article relates to the small but fast-growing literature that develops and investigates ZLB-consistent models. Three main approaches stand out: shadow-rate models, quadratic term-structure models (QTSM) and models involving square-root (CIR) processes. The shadow-rate model was introduced by Black [1995] and has been adopted by several recent contributions (see e.g. Ueno, Baba, and Sakurai [2006], Ichiue and Ueno [2007], Ichiue and Ueno [2013], Kim and Singleton [2012], Krippner [2012, 2013], Bauer and Rudebusch [2013], Christensen and Rudebusch [2013], Kim and Priebsch [2013] and Wu and Xia [2013]). In this model, the short-term rate is defined as the maximum between zero and the so-called shadow rate and ZLB periods occur when the latter turns negative. Typically, if the shadow rate follows a Gaussian process, the model can generate prolonged periods of ZLB, making it consistent with features *(i)* and *(iii)*. However, there are no closed-form formulas available for bond prices (this inadequately adresses point *(ii)*) and one has to resort to simulation or approximation techniques to estimate the model (see Kim and Priebsch [2013] or Wu and Xia [2013]). By contrast, QTSM and models based on square-root processes provide closed-form bond pricing formulas and positive yields (seminal contributions are those of Ahn, Dittmar, and Gallant [2002], Leippold and Wu [2002], Cox, Ingersoll, and Ross [1985], Pearson and Sun [1994] and Dai and Singleton [2000]). Nevertheless, these models treat the ZLB as a reflecting barrier. In that case, the probability of having an unchanged short-term rate for two subsequent periods is zero, which makes them inconsistent with feature *(iii)*.[9]

The remainder of the paper is organized as follows. Section 5.2 introduces the non-negative ARG_0 process and highlights its ability to stay at zero. Section 5.3 presents the associated affine term-structure model and derives tractable lift-off probability formulas. Section 5.4 describes the estimation strategy and presents the empirical results. Section 5.5 examines the distributions of future lift-off dates. Section 5.6 concludes and an Appendix gathers proofs and technical results.

9. More precisely, in the case of the CIR process, zero is either a reflecting barrier or an absorbing state [see Karlin and Taylor [1981] and Longstaff [1989]].

5.2 Non-negative affine processes with zero lower bound spells

In this section we introduce the univariate Gamma-zero distribution and extend it to the dynamic case with a new class of processes that we call Autoregressive Gamma-Zero (see Section 5.2.1). A multivariate generalization will be considered in Section 5.3. Like the continuous-time Cox, Ingersoll, and Ross [1985] process – or its discrete-time counterpart, the Autoregressive Gamma process of Gouriéroux and Jasiak [2006] – it cannot take negative values. However, the Autoregressive Gamma-zero can reach the zero value with a strictly positive probability, stay at this lower bound for an extended period of time and become positive again. We present its main properties in Section 5.2.2 and a generalization to the Extended Autoregressive Gamma process is developed in Section 5.2.3.

5.2.1 The ARG_0 process and the zero lower bound

Let us first recall that a Gamma distribution $\gamma_\nu(\mu)$ is a positive distribution defined by a shape (or degree of freedom) parameter $\nu > 0$ and a scale parameter $\mu > 0$. Its probability density function (p.d.f.) is given by:

$$f_X(x\,;\,\nu,\mu) = \frac{\exp(-x/\mu)\,x^{\nu-1}}{\Gamma(\nu)\,\mu^\nu}\,\mathbb{1}_{\{x>0\}}\,.$$

Note that $\gamma_\nu(\mu)$ converges in distribution to the Dirac distribution at zero when ν goes to zero. A non-central Gamma distribution can be defined as an extension of the gamma distribution. Consider a Poisson random variable Z of positive parameter λ, the non-central Gamma distribution $\gamma_\nu(\lambda,\mu)$ is a mixture of $\gamma_{\nu+Z}(\mu)$ distributions (Z being the mixing variable), defined on the set of strictly positive real numbers (denoted by $\mathbb{R}^{++}$), where ν, λ and μ are strictly positive. Remarkably, although its p.d.f. is complicated, its Laplace transform is extremely simple. Indeed, if $X \sim \gamma_\nu(\lambda,\mu)$, we have:

$$\varphi_X(u) = \mathbb{E}[\exp(uX)] = \exp\left[-\nu\log(1-u\,\mu) + \lambda\frac{u\,\mu}{1-u\,\mu}\right], \qquad \text{for} \quad u < \frac{1}{\mu}\,.$$

This distribution can be adapted to the case $\nu = 0$ if $\gamma_0(\mu)$ is considered as the Dirac distribution at zero. We obtain, by definition, a Gamma-zero distribution featuring a point mass at zero.

Definition 5.2.1. *Let X be a non-negative random variable. X follows a Gamma-zero distribution with parameters $\lambda > 0$ and $\mu > 0$, denoted $X \sim \gamma_0(\lambda,\mu)$, if its conditional*

distribution given $Z \sim \mathcal{P}(\lambda)$ *is:*

$$X \,|\, Z \sim \gamma_Z(\mu)\,. \tag{5.1}$$

The p.d.f. and the Laplace transform of X*, respectively* $f_X(x\,;\,\lambda,\mu)$ *and* $\varphi_X(u\,;\,\lambda,\mu)$*, are given by:*

$$\begin{aligned} f_X(x\,;\,\lambda,\mu) &= \sum_{z=1}^{+\infty}\left[\frac{\exp(-x/\mu)\,x^{z-1}}{(z-1)!\,\mu^z} \times \frac{\exp(-\lambda)\lambda^z}{z\,!}\right]\mathbb{1}_{\{x>0\}} + \exp(-\lambda)\mathbb{1}_{\{x=0\}} \\ \varphi_X(u\,;\,\lambda,\mu) &= \exp\left[\lambda\frac{u\mu}{(1-u\mu)}\right] \quad \textit{for} \quad u < \frac{1}{\mu}\,. \end{aligned} \tag{5.2}$$

(Note that the p.d.f. is with respect to the sum of the Lebesgue measure on $\mathbb{R}^{++}$ *and the unit mass at zero.)*

Again, despite the complexity of the density function of Equation (5.2), the Laplace transform of the Gamma-zero distribution is very easy to manipulate. Also, Equation (5.2) sheds light on a key feature of the Gamma-zero distribution: it has a point-mass located at $x = 0$, and $\mathbb{P}(X = 0) = \exp(-\lambda)$. It is extremely easy to simulate in $\gamma_0(\lambda, \mu)$ by first simulating Z in $\mathcal{P}(\lambda)$ and then X in $\gamma_z(\mu)$, where z is the result of the first simulation. As Z equals zero with a strictly positive probability, X also equals zero with a strictly positive probability. [10]

We now turn to the dynamic case, where (X_t) is a discrete-time random process that we call *Autoregressive Gamma-zero* (ARG-*Zero*) process, denoted by $\mathrm{ARG}_0(\alpha, \beta,\, \mu)$ (where $\alpha \geq 0$, $\beta \geq 0$, $\mu > 0$).

Definition 5.2.2. *The random process* (X_t) *is a* $ARG_0(\alpha, \beta, \mu)$ *process of order one if the conditional distribution of* X_{t+1}*, given* $\underline{X_t} = (X_t, X_{t-1}, \ldots)$*, is the Gamma-zero distribution:*

$$(X_{t+1}|\underline{X_t}) \sim \gamma_0(\alpha + \beta X_t, \mu) \quad \textit{for} \quad \alpha \geq 0,\, \mu > 0,\, \beta > 0\,.$$

The conditional probability density function $f\,(x_{t+1} \,|\, X_t; \alpha, \beta, \mu)$ *and the conditional Laplace*

10. Observe also that $\gamma_0(\lambda, \mu)$ is an infinitely divisible distribution given that $\varphi_X(u; \lambda, \mu) = [\varphi_X(u; \lambda/n, \mu)]^n$ (see Filipovic and Zabczyk [2002]).

transform $\varphi_{X,t}(u\,;\,\alpha,\beta,\mu)$ *of the* $ARG_0(\alpha,\beta,\mu)$ *process are respectively given by:*

$$\begin{aligned} f\left(x_{t+1}|X_t;\alpha,\beta,\mu\right) &= \sum_{z=1}^{+\infty}\left[\frac{\exp(-x_{t+1}/\mu)x_{t+1}^{z-1}}{(z-1)!\mu^z}\times\frac{\exp[-(\alpha+\beta X_t)](\alpha+\beta X_t)^z}{z!}\right]\mathbb{1}_{\{x_{t+1}>0\}} \\ &\quad + \quad \exp(-\alpha-\beta\,X_t)\mathbb{1}_{\{x_{t+1}=0\}}\,; \end{aligned} \tag{5.3}$$

$$\begin{aligned} \varphi_{X,t}(u\,;\,\alpha,\beta,\mu) &:= \mathbb{E}\left[\exp\left(uX_{t+1}\right)|\underline{X_t}\right] \\ &= \exp\left[\frac{u\mu}{1-u\mu}(\alpha+\beta\,X_t)\right], \quad \textit{for} \quad u<\frac{1}{\mu}\,. \end{aligned} \tag{5.4}$$

As for the static Gamma-zero distribution, the second element of Equation (5.3) emphasizes the zero-point mass of the ARG_0 process. The conditional probability of the process reaching zero at date $t+1$ is time-varying and given by $\exp(-\alpha-\beta\,X_t)$. Note that there are two main differences between this family of processes and the ARG processes introduced in Gouriéroux and Jasiak [2006]. First, in our case we take the shape parameter equal to 0, which allows the presence of the zero-point mass. Second, we introduce a positive intercept α in the Poisson intensity parameter, preventing the zero lower bound from being an absorbing state. Indeed, when $X_t=0$, the value X_{t+1} equals zero with probability $\mathbb{P}\left(X_{t+1}=0\,|\,X_t=0\right)=\exp(-\alpha)<1$. (In the multivariate case, this probability will depend on the information available at date t).

It is also readily seen from relation (5.4) that (X_t) is a discrete-time affine, or $Car(1)$, process (see Darolles, Gouriéroux, and Jasiak [2006]) since $\varphi_{X,t}(u;\alpha,\beta,\mu)$ is exponential-affine in X_t. This class of processes is particularly useful for building term structure models of interest rates, allowing for simple computation of moments and closed-form tractable pricing formulas. In the next sections, we use the fact that recursive formulas are available for the computation of multi-horizon Laplace transforms defined as:

$$\varphi_{t,h}(u_1,\ldots,u_h)=\mathbb{E}_t\left[\exp(u_1\,X_{t+1}+\ldots+u_h\,X_{t+h})\right]\,.$$

We illustrate the aforementioned properties of the $\text{ARG}_0(\alpha,\beta,\mu)$ process and its relevance for interest rate modeling in a ZLB setting with a simple simulation exercise. Let us denote by r_t the risk-free rate between t and $t+1$ (known in t) and let us assume that its dynamics is given by the following univariate ARG_0 process:

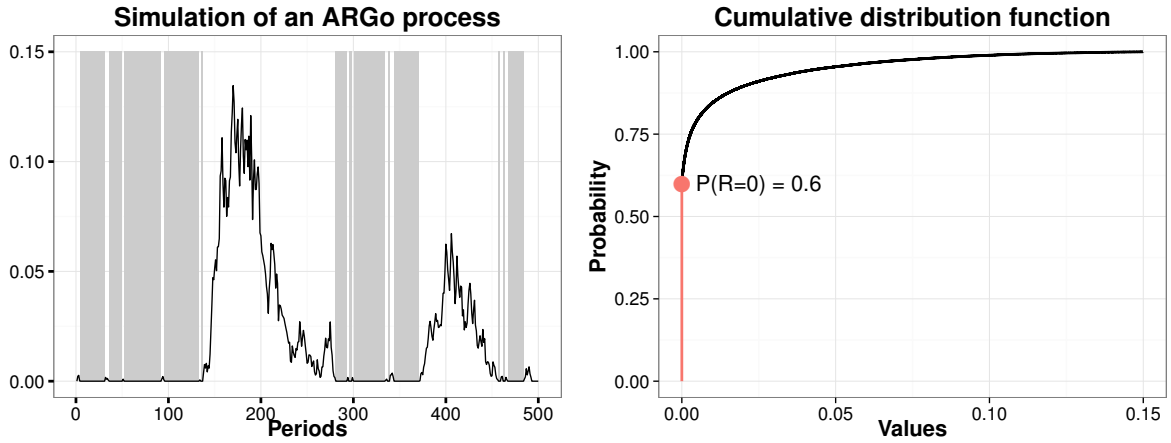
$$\left(r_t\,|\,\underline{r_{t-1}}\right)\sim\gamma_0(\alpha+\beta r_{t-1},\mu)\,, \tag{5.5}$$

where α and β are positive scalars. A model for the short-term rate dynamics described by relation (5.5) can accommodate both protracted periods of zero short-term rates and periods of fluctuations. We simulate this process for 500 periods with parameters cali-

brated as $\alpha = 0.1$, $\beta = 990$ and $\mu = 0.001$. These parameters are such that the marginal mean and standard deviation of process (r_t) are about 0.01 and 0.001, respectively. For such parameters, the conditional probability of staying at the zero lower bound is around 0.9. Figure (5.1) presents the simulated trajectory (left panel) and the computation of the marginal cumulative distribution function (right panel).

As expected, several episodes of prolonged zero lower bound are observed among the 500 simulated values. The grey-shaded areas emphasize the large persistence of the process, which hardly takes off from zero for the first 150 periods. Over the sample, the simulated process hits zero for about 250 periods, that is half of the sample length. The right panel of Figure (5.1) shows that the unconditional probability of the process to be at zero is 0.6. When it is not at zero, the process experiences persistent spikes of between 100 to 150 periods. This behavior of the ARG_0 process appears particularly appealing to model the dynamics of short-term interest rates in a zero lower bound environment.

Figure 5.1: Simulation of an ARG_0 process: a short-term rate with zero lower bound spells



Notes: This figure displays on the left panel the simulated path of a short-term rate dynamics defined by the following conditional distribution: $r_t|r_{t-1} \sim \gamma_0(0.1 + 990 r_{t-1}, 0.001)$. The grey zones correspond to periods where the simulated short rate hits zero. On the right panel we have the associated marginal cumulative distribution function.

5.2.2 Moments, stationarity and lift-off probablities of ARG_0 processes

The exponential-affine form of the Laplace transform given in Equation (5.4) allows for an easy derivation of the properties of $\text{ARG}_0(\alpha, \beta, \mu)$ processes. In this subsection, we show that ARG_0 processes possess simple closed-form formulas for conditional and unconditional moments, stationarity conditions, and especially for calculating conditional

probabilities of reaching zero, staying at zero or leaving zero (lift-off).

First, note that the affine property of the ARG_0 process implies that all conditional cumulants are affine functions of the lagged value of the process. Their derivation is made simple by the use of the log-Laplace transform. Proposition (5.2.1) and associated corollaries derive the first two conditional and unconditional moments of an ARG_0 process.

Proposition 5.2.1. *Let (X_t) be an $ARG_0(\alpha, \beta, \mu)$ process. We use the notation $\rho := \beta\,\mu$. The conditional mean $\mathbb{E}_t(X_{t+1})$ and variance $\mathbb{V}_t(X_{t+1})$ of X_{t+1} given its past are respectively given by:*

$$\mathbb{E}_t(X_{t+1}) = \alpha\mu + \rho X_t \qquad \textit{and} \qquad \mathbb{V}_t(X_{t+1}) = 2\mu^2\alpha + 2\mu\rho X_t = 2\mu\,\mathbb{E}_t(X_{t+1})\,. \tag{5.6}$$

Corollary 5.2.1.1. *(X_t) has the following weak AR(1) representation:*

$$X_{t+1} = \alpha\mu + \rho X_t + \varepsilon_{t+1}\,, \tag{5.7}$$

where (ε_t) is a conditionally heteroskedastic martingale difference, whose conditional variance is $\mathbb{V}(\varepsilon_{t+1} \,|\, \underline{\varepsilon_t}) = 2\mu^2\alpha + 2\mu\rho X_t$.

Corollary 5.2.1.2. *(X_t) is stationary if and only if $\rho < 1$ and, in this case, its unconditional mean and variance are respectively given by:*

$$\mathbb{E}(X_t) = \frac{\alpha\mu}{1-\rho} \qquad \textit{and} \qquad \mathbb{V}(X_t) = \frac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)}\,. \tag{5.8}$$

Proof. See Appendix 5.A. ■

In particular, from the conditional moments given in Proposition 5.2.1, we derive simple expressions for a weak AR(1) representation that helps calculating the unconditional first-two moments of the process. Two key features of the ARG_0 are worth noticing. First, the time-varying conditional variance is proportional to the conditional mean and, thus, it linearly shrinks with the level of X_t. This implies that, in a low-level environment, the ARG_0 process shows low conditional volatility, a typical feature of interest-rates during zero lower bound periods (see Filipovic, Larsson, and Trolle [2013]). Note also that the conditional variance of the ARG_0 process is bounded from below by $2\mu^2\alpha$ when X_t reaches zero. Second, the closed-from availability of the first-two conditional and unconditional moments implies that simple estimation procedures can be used such as the generalized method of moments, or pseudo-maximum likelihood techniques.

We concentrate now on conditional probabilities of an ARG_0 process to reach zero, to stay at zero for more than a certain number of periods, or to *lift-off* in exactly h periods.

To investigate this sojourn time in state zero and the associated lift-off probability, the following lemma proves useful.

Lemma 5.2.1. *Let Z be a random variable valued in $\mathbb{R}^+$ and $\varphi_Z(u)$ is its Laplace transform. Then, we have:*

$$\mathbb{P}_Z\{0\} = \lim_{u\to-\infty} \varphi_Z(u)\,. \tag{5.9}$$

Proof. See Appendix 5.B. ■

This Lemma makes the computation of the conditional probabilities of hitting zero very simple. The main formulas are presented in the following proposition.

Proposition 5.2.2. *Let (X_t) be an $ARG_0(\alpha,\beta,\mu)$ process and let us denote by $\varphi_{t,h}(u_1,\ldots,u_h) = \mathbb{E}_t[\exp(u_1X_{t+1}+\ldots+u_hX_{t+h})]$ its multi-horizon conditional Laplace transform. Then, the following properties hold:*

$$(i)\qquad \mathbb{P}(X_{t+h}=0\,|\,X_t) = \lim_{u\to-\infty}\varphi_{t,h}(0,\ldots,0,u)$$
$$= \exp\left\{-(1-\rho)\left[\frac{\rho^h}{\mu(1-\rho^h)}X_t + \alpha\sum_{k=0}^{h-1}\frac{\rho^k}{1-\rho^{k+1}}\right]\right\},$$

$$(ii)\qquad \mathbb{P}\big[X_{t+1}=0,\ldots,X_{t+h}=0\,\big|\,X_t\big] = \lim_{u\to-\infty}\varphi_{t,h}(u,\ldots,u)$$
$$= \exp(-\alpha\,h-\beta\,X_t)\,,$$

$$(iii)\qquad \mathbb{P}\big[X_{t+1}=0,\ldots,X_{t+h}=0,X_{t+h+1}>0\,\big|\,X_t)\big] = \exp\left[-\alpha\,h-\beta\,X_t\right]\left[1-\exp(-\alpha)\right].$$

Proof. See Appendix 5.B. ■

Corollary 5.2.2.1. *If $X_t=0$, the probability to stay in state 0 for the next $(h-1)$ periods **only** is $(1-p)p^{h-1}$ with $p=\exp(-\alpha)$, and the average sojourn time in zero is given by:*

$$(1-p)\sum_{h=1}^{+\infty} hp^{h-1} = \frac{1}{1-p} = [1-\exp(-\alpha)]^{-1}\,.$$

When $\alpha=0$, this average sojourn time is $+\infty$ and the zero lower bound becomes an absorbing state.

Proposition 5.2.2 is key for calculating lift-off probabilities in economic applications. Corollary 5.2.2.1 stresses the role of the α parameter: the average sojourn time in zero is entirely controlled by α for univariate ARG_0 processes. From an economic point of view, if the short-term interest rate is modeled by an ARG_0 process, α quantifies the average persistence of zero lower bound regimes.

5.2.3 The Extended $\text{ARG}_\nu(\alpha, \beta, \mu)$ process

The $\text{ARG}_0(\alpha, \beta, \mu)$ process described in the previous section and the $\text{ARG}_\nu(\beta, \mu)$ process of Gouriéroux and Jasiak [2006] are nested in a general class of Extended $\text{ARG}_\nu(\alpha, \beta, \mu)$ processes characterized by a degree of freedom parameter $\nu \geq 0$ and a parameter $\alpha \geq 0$. Combining the definitions of Sections 5.2.1 and 5.2.2, we obtain the following:

Definition 5.2.3. *The univariate random process (X_t) is an Extended $ARG_\nu(\alpha, \beta, \mu)$ process of order one if the conditional distribution of X_{t+1}, given $\underline{X_t} = (X_t, X_{t-1}, \ldots)$, is a non-centered Gamma distribution such that:*

$$(X_{t+1}|X_t) \sim \gamma_\nu(\alpha + \beta X_t, \mu)\,, \qquad \text{for} \qquad \alpha \geq 0,\ \nu \geq 0,\ \mu > 0,\ \beta > 0\,.$$

The conditional probability density function $f\left(x_{t+1} \,|\, X_t; \nu, \alpha, \beta, \mu\right)$ and the conditional Laplace transform $\varphi_{X,t}(u\,; \nu, \alpha, \beta, \mu)$ of the Extended $ARG_\nu(\alpha, \beta, \mu)$ process are respectively given by:

$$\begin{aligned} f\left(x_{t+1}|X_t; \nu, \alpha, \beta, \mu\right) &= \sum_{z=1}^{+\infty} \left[\frac{\exp(-x_{t+1}/\mu)\, x_{t+1}^{\nu+z-1}}{\Gamma(\nu+z)\mu^{\nu+z}} \times \frac{\exp[-(\alpha+\beta X_t)]\,(\alpha+\beta X_t)^z}{z!} \right] \mathbb{1}_{\{x_{t+1}>0\}} \\ &\quad + \quad \exp(-\alpha - \beta\, X_t) \mathbb{1}_{\{x_{t+1}=0,\, \nu=0\}}\,, \\ \varphi_{X,t}(u\,; \nu, \alpha, \beta, \mu) &:= \mathbb{E}\left[\exp\left(u X_{t+1}\right) |\underline{X_t}\right] \\ &= \exp\left[\frac{u\mu}{1-u\mu}\beta X_t + \alpha \frac{u\mu}{1-u\mu} - \nu \log(1-u\mu) \right], \ \text{for} \quad u < \frac{1}{\mu}\,. \end{aligned} \tag{5.10}$$

Note that the difference with the ARG_0 process, in terms of conditional Laplace transform, is the additional term $[-\nu \log(1 - u\mu)]$ in the exponential. However, a process with Extended ARG dynamics and $\nu > 0$ does not experience prolonged periods of zero. In line with Proposition 5.2.1, and following the same steps as in Appendix 5.A, we derive the conditional and unconditional first two moments of an Extended ARG process.

Proposition 5.2.3. *Let (X_t) be an Extended $ARG_\nu(\alpha, \beta, \mu)$ process and $\rho := \beta\mu$. The conditional mean and variance of X_{t+1} are respectively given by:*

$$\mathbb{E}_t(X_{t+1}) = \mu\,(\nu + \alpha) + \rho X_t \qquad \text{and} \qquad \mathbb{V}_t(X_{t+1}) = \mu^2(\nu + 2\alpha) + 2\mu\rho X_t\,. \tag{5.11}$$

Corollary 5.2.3.1. *(X_t) is stationary if and only if $\rho < 1$ and, in this case, its unconditional mean and variance are respectively given by:*

$$\mathbb{E}(X_t) = \frac{(\alpha+\nu)\mu}{1-\rho} \qquad \text{and} \qquad \mathbb{V}(X_t) = \frac{2\alpha\mu^2 + \mu^2\nu(1+\rho)}{(1-\rho)(1-\rho^2)}\,.$$

Setting $\nu = 0$, we get the $\text{ARG}_0(\alpha, \beta, \mu)$ family presented in Section 5.2.1 and, assuming

$\alpha = 0$ with $\nu > 0$, we obtain the classical $\text{ARG}_\nu(\beta, \mu)$ family. It is also worth noting from relation (5.10) that, using the extension to random coefficients models, in particular regime-switching models (see Gouriéroux, Monfort, Pegoraro, and Renne [2014]), it would be possible to make the parameters α and ν exogenously random and affine, or linearly dependent on X_t, while staying in the class of affine processes for the augmented process.

In the following sections, we use the previous univariate distributions to construct our multivariate non-negative affine term-structure model where the state vector is composed (under both the risk-neutral and historical probability) of conditionally independent factors with Gamma-zero and Extended Gamma distributions. This assumption of conditional independence characterizing the so-called Vector Autoregressive Gamma process (VARG, say) makes the zero-coupon bond pricing model specification simple while guaranteeing at the same time enough flexibility to match relevant ZLB-linked interest rates stylized facts (see Section 5.4). [11]

5.3 The Non-Negative Affine Term Structure Model

5.3.1 The VARG risk-neutral state dynamics and the affine yield curve formula

In this section we introduce the multivariate non-negative affine term-structure model (NATSM) by directly specifying the risk-neutral ($\mathbb{Q}$) dynamics of the n-dimensional latent state vector $X_t = \left(X_t^{(1)\prime}, X_t^{(2)\prime}\right)'$, where $\dim\left(X_t^{(1)}\right) = n_1$, $\dim\left(X_t^{(2)}\right) = n_2$, and $n = n_1 + n_2$. We also denote by r_t the unobservable short-term rate between t and $t+1$, known at date t. More specifically, we assume that the risk-neutral dynamics of X_t is a Vector ARG (or VARG) process.

Assumption 1. *The risk-neutral distribution of X_{t+1}, conditionally on $\underline{X_t}$, is given by the product of the following conditional distributions:*

$$\left(X_{j,t+1} \,|\, \underline{X_t}\right) \overset{\mathbb{Q}}{\sim} \gamma_{\nu_j}\left(\alpha_j^{\mathbb{Q}} + \beta_j^{\mathbb{Q}\prime} X_t, \mu_j^{\mathbb{Q}}\right), \quad j \in \{1, \ldots, n\}, \tag{5.12}$$

where $\nu_j = 0$ for any $j \in \{1, \ldots, n_1\}$, while $\nu_j \geq 0$ if $j \in \{n_1 + 1, \ldots, n\}$; $\alpha_j^{\mathbb{Q}} \geq 0$, $\mu_j^{\mathbb{Q}} > 0$ and $\beta_j^{\mathbb{Q}}$ is an n-dimensional vector of positive components.
In other words, conditionally on X_t, the n_1 components of $X_{t+1}^{(1)}$ follow independent Gamma-zero distributions, while the n_2 components of $X_{t+1}^{(2)}$ follow independent Non-central Gamma distributions.

11. A general specification of the VARG process with conditional dependence is proposed in Chapter 4 (see also Monfort, Pegoraro, Renne, and Roussellet [2014]).

Given the conditional (given $\underline{X_t}$) independence between the scalar elements of X_{t+1}, the risk-neutral conditional Laplace transform of X_{t+1} given X_t is immediately obtained:

Proposition 5.3.1. *The risk-neutral Laplace transform of X_{t+1}, conditionally on $\underline{X_t}$, is given by:*

$$\varphi_t^{\mathbb{Q}}(u) \;=\; \mathbb{E}^{\mathbb{Q}}\left[\exp\left(\sum_{j=1}^{n} X_{j,t+1}\right)\,\middle|\,\underline{X_t}\right] = \exp\left[\sum_{j=1}^{n} a_j^{\mathbb{Q}}(u_j)'X_t + b_j^{\mathbb{Q}}(u_j)\right] \tag{5.13}$$

where, for any $j \in \{1,\dots,n\}$, we have:

$$a_j^{\mathbb{Q}}(u_j) \;=\; \frac{u_j\,\mu_j^{\mathbb{Q}}}{1-u_j\,\mu_j^{\mathbb{Q}}}\,\beta_j^{\mathbb{Q}} \qquad \textit{and} \qquad b_j^{\mathbb{Q}}(u_j) = \frac{u_j\,\mu_j^{\mathbb{Q}}}{1-u_j\,\mu_j^{\mathbb{Q}}}\,\alpha_j^{\mathbb{Q}} - \nu_j\,\log(1-\,u_j\,\mu_j^{\mathbb{Q}})\,. \tag{5.14}$$

The process (X_t) is therefore a discrete-time affine (Car(1)) process.

Corollary 5.3.1.1. *The process (X_t) is $\mathbb{Q}$-stationary if and only if, for all $j \in \{1,\dots,n\}$, we have $\rho_j^{\mathbb{Q}} = \beta_{j,j}^{\mathbb{Q}}\,\mu_j^{\mathbb{Q}} < 1$.*

Proof. See Chapter 4. ■

Assumption 2. *The nominal short rate process (r_t) is given by the linear combination of the first n_1 components of X_t only, that is:*

$$r_t = \sum_{j=1}^{n_1} \delta_j X_{j,t} = \delta' X_t\,, \tag{5.15}$$

where $\delta = [(\delta_j)'_{j=\{1,\dots,n_1\}}, 0_{n_2}]'$ has the first n_1 entries strictly positive, the remaining ones being equal to zero. [12]

It is straightforward to see that the short-term interest rate still possesses the zero-point mass property given that it is a linear combination of conditionally independent variables following Gamma-zero distributions. Besides, observe that a non-zero short rate lower bound is allowed (as, for instance, in Priebsch [2013]) by simply adding $r_{min} \neq 0$ (say) on the right hand side of Equation (5.15).

In matrix form, the conditional Laplace transform presented in Proposition 5.3.1, can be written as:

$$\varphi_t^{\mathbb{Q}}(u) \;=\; \exp\left[\widetilde{a}^{\mathbb{Q}}(u)'X_t + \widetilde{b}^{\mathbb{Q}}(u)\right],$$

12. Note that δ_j and μ_j cannot be both identified. In the application, we impose that $\mu_j^{\mathbb{Q}} = 1$ for all j to ensure identification constraints.

where:

$$\begin{aligned}
\widetilde{a}^{\mathbb{Q}}(u) &= \beta^{\mathbb{Q}}\left(\frac{u\odot\mu^{\mathbb{Q}}}{1-u\odot\mu^{\mathbb{Q}}}\right)\\
\widetilde{b}^{\mathbb{Q}}(u) &= \alpha^{\mathbb{Q}\prime}\left(\frac{u\odot\mu^{\mathbb{Q}}}{1-u\odot\mu^{\mathbb{Q}}}\right)-\nu'\log\left(1-u\odot\mu^{\mathbb{Q}}\right)\\
\mu^{\mathbb{Q}} &= (\mu_1^{\mathbb{Q}},\ldots,\mu_n^{\mathbb{Q}})',\ \ \beta^{\mathbb{Q}}=(\beta_1^{\mathbb{Q}},\ldots,\beta_n^{\mathbb{Q}}),\\
\alpha^{\mathbb{Q}} &= (\alpha_1^{\mathbb{Q}},\ldots,\alpha_n^{\mathbb{Q}})',\ \ \nu=(0,\ldots,0,\nu_{n_1+1},\ldots,\nu_n)',
\end{aligned}$$

and where $\odot$ denotes the element-wise product and where, with abuse of notations, the division and log operators work element-wise when applied to vectors.

Given the exponential-affine form of the risk-neutral conditional Laplace transform of (X_t), it is easy to obtain the following explicit zero-coupon bond pricing formula (see Appendix 5.C for a proof):

Proposition 5.3.2. *If the n-dimensional state vector (X_t) has a risk-neutral dynamics defined by Equation (5.13) and if the short-term interest rate is defined as in Assumption 2, then the price at date t of the zero-coupon bond with residual maturity h, denoted by $P_t(h)$, is given by:*

$$P_t(h)=\exp\left(A_h'X_t+B_h\right), \tag{5.16}$$

where A_h and B_h satisfy the following recursive equations:

$$\begin{aligned}
A_h &= -\delta+\widetilde{a}^{\mathbb{Q}}(A_{h-1})\\
&= -\delta+\beta^{\mathbb{Q}}\left(\frac{A_{h-1}\odot\mu^{\mathbb{Q}}}{1-A_{h-1}\odot\mu^{\mathbb{Q}}}\right) &(5.17)\\
B_h &= B_{h-1}+\widetilde{b}^{\mathbb{Q}}(A_{h-1})\\
&= B_{h-1}+\alpha^{\mathbb{Q}\prime}\left(\frac{A_{h-1}\odot\mu^{\mathbb{Q}}}{1-A_{h-1}\odot\mu^{\mathbb{Q}}}\right)-\nu'\log\left(1-A_{h-1}\odot\mu^{\mathbb{Q}}\right), &(5.18)
\end{aligned}$$

with starting conditions $A_0=0$ and $B_0=0$. The date t continuously-compounded yield associated with a zero-coupon bond maturing in h periods is therefore given by the following non-negative affine function of X_t:

$$\begin{aligned}
R_t(h) &= \overline{A}_h'X_t+\overline{B}_h,\\
\overline{A}_h &= -\frac{1}{h}A_h, \qquad \text{and} \qquad \overline{B}_h=-\frac{1}{h}B_h, \qquad h\geq 1.
\end{aligned} \tag{5.19}$$

The non-negativeness of our NATSM can be easily established from the usual no-arbitrage formula $R_t(h)=-\frac{1}{h}\log\mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(-r_t-\ldots-r_{t-h+1}\right)\right]$ since the short-term rate

is a positive combination of the $X_{i,t}$'s which are all positive.

5.3.2 The VARG historical state dynamics

We have defined the risk-neutral dynamics of X_t in Assumption 1. Let us now determine the historical ($\mathbb{P}$) dynamics of the state vector (X_t). For this, we assume that the one-period stochastic discount factor is based on an exponential-affine change of probability measure $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta' X_{t+1} - \psi_t^{\mathbb{Q}}(\theta)\right]$, where $\psi_t^{\mathbb{Q}}(u) = \log \varphi_t^{\mathbb{Q}}(u)$ denotes the risk-neutral conditional log-Laplace transform of (X_t), and $\theta = (\theta_1, \ldots, \theta_n)'$ denotes the n-dimensional vector of market prices of risk factors. Then, we have:

Proposition 5.3.3. *The historical distribution of X_{t+1}, conditionally on $\underline{X_t}$, is given by the product of the conditional distributions:*

$$(X_{j,t+1} \,|\, \underline{X_t}) \overset{\mathbb{P}}{\sim} \gamma_{\nu_j}\left(\alpha_j^{\mathbb{P}} + \beta_j^{\mathbb{P}'} X_t, \mu_j^{\mathbb{P}}\right), \quad \text{for} \quad j \in \{1, \ldots, n\}, \tag{5.20}$$

where $\alpha_j^{\mathbb{P}} \geq 0$, $\mu_j^{\mathbb{P}} > 0$, and $\beta_j^{\mathbb{P}}$ is an n-dimensional vector of strictly positive components and the historical Laplace transform of X_{t+1}, conditionally to $\underline{X_t}$, is given by:

$$\varphi_t^{\mathbb{P}}(u) = \exp\left[\sum_{j=1}^{n} a_j^{\mathbb{P}}(u_j)' X_t + b_j^{\mathbb{P}}(u_j)\right] \tag{5.21}$$

where, for any $j \in \{1, \ldots, n\}$, we have:

$$\begin{aligned} & a_j^{\mathbb{P}}(u_j) = \frac{u_j\,\mu_j^{\mathbb{P}}}{1 - u_j\,\mu_j^{\mathbb{P}}}\,\beta_j^{\mathbb{P}} \qquad \text{and} \qquad b_j^{\mathbb{P}}(u_j) = \frac{u_j\,\mu_j^{\mathbb{P}}}{1 - u_j\,\mu_j^{\mathbb{P}}}\,\alpha_j^{\mathbb{P}} - \nu_j\,\log(1 - u_j\,\mu_j^{\mathbb{P}}), \\ & \text{with} \qquad \alpha_j^{\mathbb{P}} = \frac{\alpha_j^{\mathbb{Q}}}{1 - \theta_j\,\mu_j^{\mathbb{Q}}}, \qquad \beta_j^{\mathbb{P}} = \frac{1}{1 - \theta_j\,\mu_j^{\mathbb{Q}}}\,\beta_j^{\mathbb{Q}} \qquad \text{and} \qquad \mu_j^{\mathbb{P}} = \frac{\mu_j^{\mathbb{Q}}}{1 - \theta_j\,\mu_j^{\mathbb{Q}}}. \end{aligned} \tag{5.22}$$

Proof. See Appendix 5.D. ∎

Note that the ν_j's are the same in the risk-neutral and the historical worlds. In particular, if $\nu_j = 0$ in the risk-neutral world, it is also true in the historical one, since the negligible sets must be the same for both conditional distributions in order to guarantee the equivalence of the associated probabilities. In line with the notation adopted in the previous section, this historical conditional Laplace transform can be represented in matrix form:

$$\varphi_t^{\mathbb{P}}(u) \quad = \quad \exp\left[\widetilde{a}^{\mathbb{P}}(u)' X_t + \widetilde{b}^{\mathbb{P}}(u)\right]$$

where:

$$\begin{aligned}
\widetilde{a}^{\mathbb{P}}(u) &= \beta^{\mathbb{P}}\left(\frac{u\odot\mu^{\mathbb{P}}}{1-u\odot\mu^{\mathbb{P}}}\right)\\
\widetilde{b}^{\mathbb{P}}(u) &= \alpha^{\mathbb{P}\prime}\left(\frac{u\odot\mu^{\mathbb{P}}}{1-u\odot\mu^{\mathbb{P}}}\right)-\nu'\log\left(1-u\odot\mu^{\mathbb{P}}\right)\\
\mu^{\mathbb{P}} &= (\mu_1^{\mathbb{P}},\ldots,\mu_n^{\mathbb{P}})'\,,\quad \beta^{\mathbb{P}}=(\beta_1^{\mathbb{P}},\ldots,\beta_n^{\mathbb{P}})\,,\quad \text{and}\quad \alpha^{\mathbb{P}}=(\alpha_1^{\mathbb{P}},\ldots,\alpha_n^{\mathbb{P}})'\,.
\end{aligned}$$

5.3.3 Lift-off Probabilities

Let us move now to the problem of investigating the sojourn in state zero of the short rate process (r_t), and the associated lift-off probability. As seen in the previous sections, our multivariate non-negative yield curve model has the convenient property of being affine under both the risk-neutral and historical dynamics. Consequently, our model allows to easily compute multi-horizon Laplace transforms in both worlds and, thus, to explicitly calculate lift-off probabilities.

Let us first remember that, given the exponential-affine nature of the conditional historical Laplace transform of (X_t) (see relation (5.23)), its multi-horizon Laplace transform until $t+k$ is given by:

$$\begin{aligned}
\varphi_{t,k}^{\mathbb{P}}(u_1,\ldots,u_k) &= \mathbb{E}^{\mathbb{P}}\left[\exp\left(u_1'X_{t+1}+\ldots+u_k'X_{t+k}\right)\,\middle|\,X_t\right]\\
&= \exp\left[\mathcal{A}_k'\,X_t+\mathcal{B}_k\right]
\end{aligned}\tag{5.23}$$

where, for any $i\in\{1,\ldots,k\}$, u_i is an n-dimensional vector. The $\mathcal{A}_k$ and $\mathcal{B}_k$ loadings are obtained as the final values $\mathcal{A}_k=\mathcal{A}_k^{(k)}$, $\mathcal{B}_k=\mathcal{B}_k^{(k)}$ of the k-step recursion:

$$\left\{\begin{aligned}
\mathcal{A}_0^{(k)} &= 0 \qquad\text{and}\qquad \mathcal{B}_0^{(k)}=0\,,\\
\mathcal{A}_i^{(k)} &= \widetilde{a}^{\mathbb{P}}\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)=\beta^{\mathbb{P}}\left(\frac{\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)\odot\mu^{\mathbb{P}}}{1-\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)\odot\mu^{\mathbb{P}}}\right),\\
\mathcal{B}_i^{(k)} &= \widetilde{b}^{\mathbb{P}}\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)+\mathcal{B}_{i-1}^{(k)}\\
&= \alpha^{\mathbb{P}\prime}\left(\frac{\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)\odot\mu^{\mathbb{P}}}{1-\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)\odot\mu^{\mathbb{P}}}\right)-\nu'\log\left[1-\left(u_{k+1-i}+\mathcal{A}_{i-1}^{(k)}\right)\odot\mu^{\mathbb{P}}\right]+\mathcal{B}_{i-1}^{(k)}\,.
\end{aligned}\right.\tag{5.24}$$

Proof. See Proposition 3 in Gouriéroux, Monfort, Pegoraro, and Renne [2014]. ■

Given that the yield $R_t(h)$ is an affine function of X_t, it is easily seen that, for any k-dimensional vector v:

$$\begin{aligned}\varphi^{(h)\mathbb{P}}_{R,t,k}(v) &:= \varphi^{(h)\mathbb{P}}_{R,t,k}(v_1,\ldots,v_k) = \mathbb{E}\left[\exp\left(v_1\,R_{t+1}(h)+\ldots+v_k\,R_{t+k}(h)\right)\,|\,X_t\right]\\ &= \varphi^{\mathbb{P}}_{t,k}(v_1\,\overline{A}_h,\ldots,v_k\,\overline{A}_h)\,\exp\left(\overline{B}_h\sum_{j=1}^{k}v_j\right)\,, \qquad (5.25)\end{aligned}$$

where $v_1,\ldots,v_k$ are the scalar entries composing v. Therefore, Equation (5.23) can be used to calculate the yields' multi-horizon conditional Laplace transform. In order to determine lift-off probability formulas, let us introduce the following lemma, generalizing Lemma 5.2.1 to the multivariate framework.

Lemma 5.3.1. *If Z is an n-dimensional random variable valued in $\mathbb{R}^n_+$ and $\varphi_Z(u_1,\ldots,u_n)$ is its Laplace transform, we have:*

$$\mathbb{P}_Z\{0,\ldots,0\} = \lim_{u\to-\infty}\varphi_Z(u,\ldots,u)\,.$$

Proof. Straightforward generalization of the proof of Lemma 5.2.1 using the fact that, here, $Z=0$ is equivalent to $\mathbf{e}'Z=0$ (with $\mathbf{e}=(1,\ldots,1)'$). ■

Then, as far as the lift-off probabilities for the short rate are concerned, we have the following proposition:

Proposition 5.3.4. *Let us consider the short rate process (r_t). The following properties hold:*

$$\begin{aligned}&(i) \quad \mathbb{P}\left[r_{t+k}=0\,|\,X_t\right] = \lim_{u\to-\infty}\varphi^{(1)\mathbb{P}}_{R,t,k}(0,\ldots,0,u)\,;\\ &(ii) \quad \mathbb{P}\left[r_{t+1}=0,\ldots,r_{t+k}=0\,|\,X_t\right] = \lim_{u\to-\infty}\varphi^{(1)\mathbb{P}}_{R,t,k}(u,\ldots,u) = p_{r,t,k} \quad \textit{(say)}\,;\\ &(iii) \quad \mathbb{P}\left[r_{t+1}=0,\ldots,r_{t+k-1}=0,r_{t+k}>0\,|\,X_t\right] = p_{r,t,k-1}-p_{r,t,k}\,,\end{aligned}$$

where $p_{r,t,0}=1$.

The last relation gives the distribution of the first lift-off date. The average sojourn time in state zero is then given by:

$$\sum_{k=1}^{\infty}h\left(p_{r,t,k-1}-p_{r,t,k}\right).$$

In the previous proposition we have introduced explicit formulas concerning the probability of lift-off from the zero lower bound for the short rate process. Using the formula for

truncated Laplace transform in the case of affine processes (see Duffie, Pan, and Singleton [2000] for details), it is possible to provide some tractable formulas if the zero lower bound is replaced by a positive floor $\lambda > 0$ (e.g. $\lambda = 10$ bps). Besides, such formulas are available for interest rates of any maturity. More precisely:

Proposition 5.3.5. *Let us consider the yield process $(R_t(h))$ of maturity h with the multi-horizon conditional Laplace transform given in Equation (5.25). The following properties hold:*

$$(i) \qquad \widetilde{p}^{(h)}_{t,k}(v,\lambda) := \mathbb{P}\left[v' R^{(t+k)}_{t+1}(h) > \lambda \,|\, X_t\right] = \frac{1}{2} + \frac{1}{\pi}\int_0^{+\infty} \frac{Im\left[\varphi^{(h)\mathbb{P}}_{R,t,k}(i\,v\,x)\exp(-i\,\lambda\,x)\right]}{x}\,dx\,;$$

$$(ii) \qquad \mathbb{P}\left[R_{t+k}(h) > \lambda \,|\, X_t\right] = \widetilde{p}^{(h)}_{t,k}(e_k,\lambda)\,;$$

$$(iii) \qquad \mathbb{P}\left(\frac{R_{t+k-m+1}(h) + R_{t+k-m+2}(h) + \ldots + R_{t+k}(h)}{m} > \lambda \,|\, X_t\right) = \widetilde{p}^{(h)}_{t,k}\left(\frac{1}{m} e^{(k)}_{k-m+1}, \lambda\right),$$

where $R^{(t+k)}_{t+1}(h) = (R_{t+1}(h), \ldots, R_{t+k}(h))'$ and $v = (v_1, \ldots, v_k)'$; e_k is the k^{th} column of the (k,k) identity matrix and $e^{(k)}_{k-m+1} = (\underbrace{0, \ldots, 0}_{k-m \text{ times}}, \underbrace{1, \ldots, 1}_{m \text{ times}})'$ denotes here a k-dimensional vector of zeros for the first $k-m$ components and ones for the m others.

Observe that these formulas do not determine the probability that $t+k$ be the first date (between t and $t+k$) at which $R_{t+k}(h) > \lambda$. Nevertheless, this latter information can always be obtained by simulation.

5.4 Empirical analysis of NATSMs

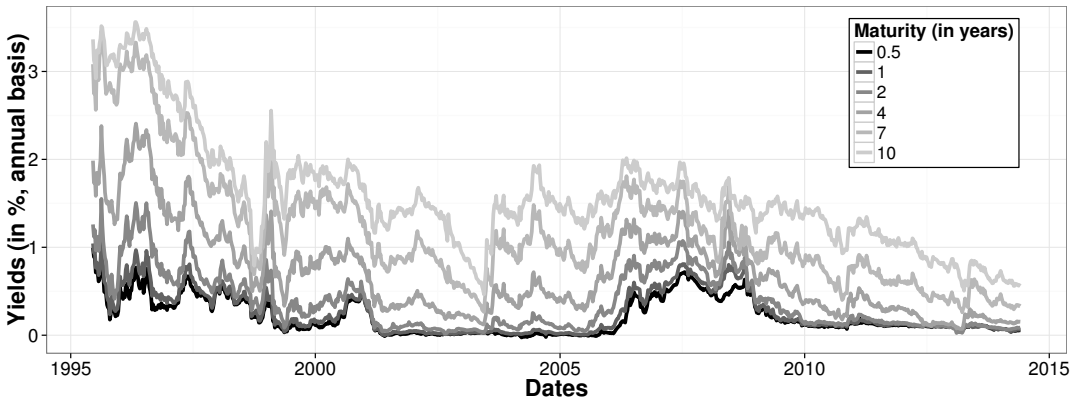
5.4.1 Data and stylized facts

As in Kim and Singleton [2012] and Christensen and Rudebusch [2013], we concentrate on zero-coupon Japanese Government Bond (JGB) yields. The data are weekly (Fridays) and cover the period from June 16, 1995 to May 30, 2014, with residual maturities of six months and one, two, four, seven and ten years.[13] A graphical representation of the yields is provided on Figure 5.2 and descriptive statistics are presented in Table 5.1.

During the first years of our sample, we observe a large decrease in the yields at all maturities. From 1996 to 2001, the 6-month yield stabilizes around 40bps whereas other maturities continue to decrease until 1999, and experience large fluctuations after. From 2001 to 2006, yields literally enter the zero-lower-bound phase, with the 6-month rate

13. The data are extracted from Bloomberg, the tickers of the time series are F10506M, F10501Y, F10502Y, F10504Y, F10507Y, F10510Y.

Figure 5.2: Japanese yields data



Notes: Yields are weekly data from June 16, 1995 to May 30, 2014. Yields are expressed in annualized percentage points figures, with maturity from 6 months (darkest line) to 10 years (lightest line).

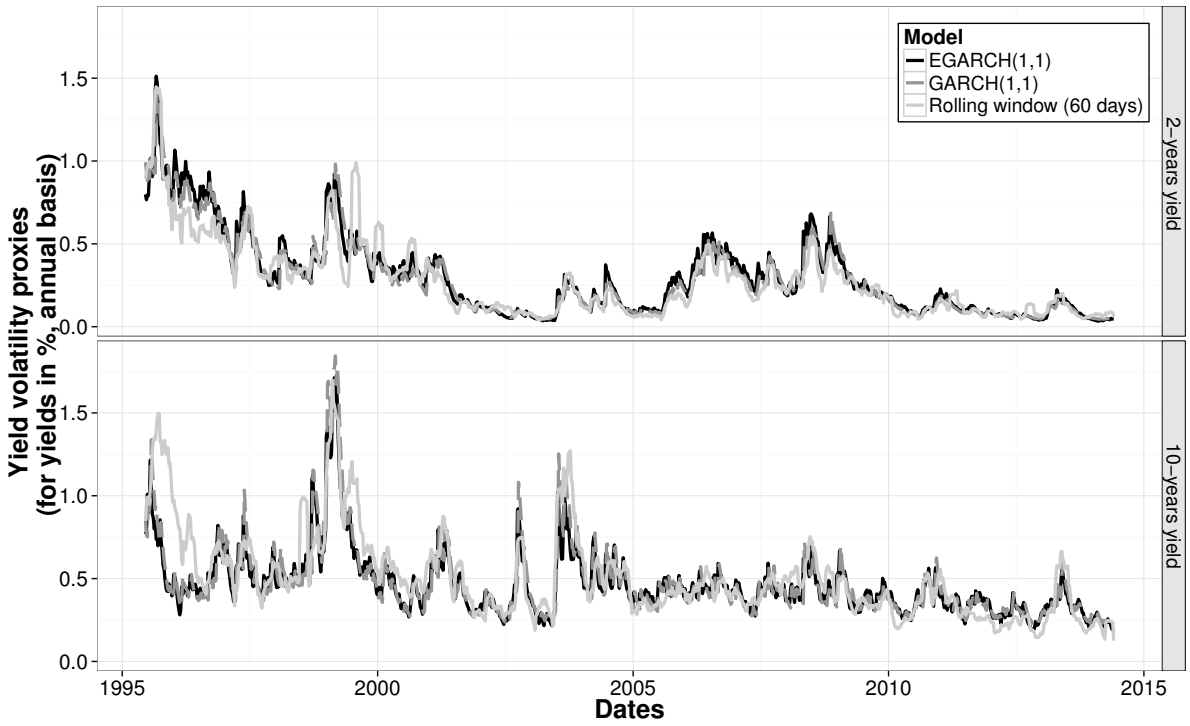
stable at virtually zero.[14] As already noted in Kim and Singleton [2012], during this period, the longer-term yields continued showing large variance. We examine more closely this behavior by computing three different measures of univariate conditional variances. For each yield in the data, we fit a GARCH(1,1) and a EGARCH(1,1) models and extract the associated fitted variances. We also compute a two-month rolling-window variance measure on daily data. All those measures are normalized in the same fashion, taking volatilities expressed in annualized terms. Standard descriptive statistics of those proxies are presented in Table 5.1, and they are represented in Figure 5.3 (for the 2-year and 10-year maturities). This figure illustrates that, for a given maturity, the three variance proxies are close to each other. We hence consider them to be coherent and credible proxies of conditional volatilities of interest rates. This proximity is confirmed by Table 5.2, which presents the correlations between the level of interest rates and the conditional volatility proxies. The correlations between the three volatility proxies exceed 0.75 for all maturities.

Besides, we observe that the behaviors of conditional volatility proxies differ substantially across maturities. For the 2-year maturity (Figure 5.3, top panel), the conditional volatility proxies drop very close to zero when the 6-month rate hits the zero lower bound in 2001. For the longest maturity, the behavior of the three proxies does not show the same decreasing trend (from 1999 to 2004) as for the 2-year yield, even though they experience large spikes in 1995, 1999, and at the end of 2003.

The previous observations allow us to exhibit three important stylized facts: short-term yields can stay at zero for extended periods of time; longer-term yields show sub-

14. Between May 2001 and February 2006, the 6-month yield has mean and standard deviations respectively equal to 1.37bps and 1.42bps.

Figure 5.3: Conditional volatility proxies



Notes: Top and bottom panels respectively present the volatility proxies for the 2-year and the 10-year yields. GARCH and EGARCH conditional volatility models are computed on weekly yield changes whereas the rolling-window volatility is computed on a 2-month window of daily data. We take the square-root of estimated proxies of conditional variance and obtain our conditional volatility proxies. We normalize them to be comparable to annualized yields. We take estimated proxies and normalize them to be comparable to annualized yields.

stantial variations even in a ZLB period and proxies of conditional yield variances show different profiles across the maturity spectrum. A well-specified term structure model should be able to replicate these empirical features.

5.4.2 Estimation Strategy

Since our term-structure model is affine, a natural estimation technique is to use the linear Kalman filter (as in Duan and Simonato [1999] and De Jong [2000]). The model can be easily represented in a (linear) state-space form where the measurement equations are the yield formula (Equation (5.19)) and the transition equations are given by the factor dynamics. Moreover, the affine nature of yields forecasts and of conditional variances naturally provided by our model opens the way to easily introduce new affine measurement equations while preserving the linear specification of the state-space model.

Our first kind of additional measurement equations relate the 2-year and the 10-year conditional variance proxies to their model-implied counterparts (specifically, we retain the

Table 5.1: Mean and standard deviations of yields and volatility proxies

	Maturity	6m	1y	2y	4y	7y	10y
Mean	Yields	0.2142	0.2581	0.3811	0.7091	1.2025	1.5850
	GARCH(1,1)	0.0019	0.0020	0.0029	0.0040	0.0052	0.0047
	EGARCH(1,1)	0.0020	0.0021	0.0030	0.0042	0.0052	0.0046
	rolling-window	0.0022	0.0023	0.0028	0.0040	0.0055	0.0051
Std.	Yields	0.2082	0.2440	0.3347	0.5025	0.6799	0.6576
	GARCH(1,1)	0.0017	0.0019	0.0024	0.0025	0.0024	0.0022
	EGARCH(1,1)	0.0017	0.0019	0.0024	0.0025	0.0021	0.0020
	rolling-window	0.0021	0.0021	0.0023	0.0027	0.0029	0.0026

Notes: Yields are expressed in annualized percentage points. GARCH and EGARCH models are computed on weekly data whereas the rolling-window volatility is computed on a 60-day window of daily data and converted to the weekly frequency keeping only Fridays data. Our volatility proxies are the square roots of the estimated conditional variance proxies; they are normalized to make them homogeneous to annualized yields. 'Mean' and 'Std.' respectively present sample means and standard deviations of our proxies.

EGARCH-based proxy). Recent papers in the term structure literature have highlighted that the estimations of term-structure models based only on yield in levels fail to satisfyingly replicate fluctuations in conditional volatilities (see for instance Jacobs and Karoui [2009] and Cieslak and Povala [2015]).

Second, we augment our state-space model with measurement equations relating the model-implied yields forecasts and survey-based ones. This approach, introduced by Kim and Orphanides [2012], is aimed at handling the persistence problem affecting the estimation of term-structure models (see Kozicki and Tinsley [2001a], Kozicki and Tinsley [2001b] and Jardet, Monfort, and Pegoraro [2013]). More precisely, we use three- and twelve-month-ahead forecasts of the ten-year yield coming from the Consensus Forecasts by Consensus Economics. The latter forecasts are available only from 1999 onward and at the monthly frequency. This missing-data issue is nevertheless easily handled with the Kalman Filter.

In summary, directly fitting survey-based forecasts and conditional variances of yields help to estimate historical and risk-neutral parameters of the factor X_t. This contributes to get reliable model-implied measures of long-horizon interest-rate forecasts and of lift-off probabilities under both measures ($\mathbb{P}$ and $\mathbb{Q}$).

Let us now detail the chosen specifications of the factor dynamics. Preliminary estimations have suggested that the data call for the inclusion of a single factor in the specification

Table 5.2: Correlation between rates and volatility proxies

Maturity	6m			1y			2y		
	level	volatility		level	volatility		level	volatility	
		GARCH	EGARCH		GARCH	EGARCH		GARCH	EGARCH
GARCH	0.63	1		0.68	1		0.74	1	
EGARCH	0.68	0.96	1	0.72	0.98	1	0.78	0.97	1
RW	0.58	0.75	0.76	0.65	0.88	0.89	0.70	0.92	0.92
Maturity	4y			7y			10y		
GARCH	0.72	1		0.54	1		0.31	1	
EGARCH	0.74	0.95	1	0.56	0.95	1	0.35	0.93	1
RW	0.67	0.90	0.90	0.60	0.87	0.86	0.54	0.80	0.80

Notes: Yields are expressed in annualized percentage points. GARCH and EGARCH models are computed on weekly data whereas the rolling-window volatility is computed on a 60-day window of weekly data.

of the short-term rate and four factors seem necessary to get a satisfying fit for both the levels and the conditional variances of yields. Accordingly, we present in the following the estimation of the model where $n_1 = 1$ and $n_2 = 3$. However, since there are causal relationships between the four factors, longer-term yields are combinations of all four factors, which allows them to vary even if $X_t^{(1)}$ and the short rate are equal to zero. Moreover, we set ν at zero for all components; that is, the conditional distributions of the four factors are γ_0.

In the following, we formally present our state-space model. Using the multivariate adaptation of Equation (5.11) and the historical dynamics given in Section 5.3.2, the transition equations can be expressed as follows:

$$\begin{aligned} X_{t+1} &= \underbrace{\mu^{\mathbb{P}} \odot \alpha^{\mathbb{P}} + \mu^{\mathbb{P}} \odot \beta^{\mathbb{P}\prime} X_t}_{\mathbb{E}_t(X_{t+1})} + \underbrace{\left\{ \operatorname{diag}\left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}\prime} X_t \right) \right] \right\}^{1/2}}_{\mathbb{V}_t(X_{t+1})^{1/2}} \varepsilon_{t+1} \\ &= m + M X_t + \Sigma_t^{1/2} \varepsilon_{t+1} \,, \end{aligned} \tag{5.26}$$

where (ε_t) is a martingale difference with zero-mean and identity variance-covariance matrix.

The measurement equations describe the relationship between three types of observable variables and their model-implied (affine) counterparts: the JGB yields described previously, the EGARCH(1,1) conditional variance proxies for the two and ten-year maturities as well as the three and twelve months-ahead surveys of professional forecasters of the

ten-year yield. Observed variables are replicated by the model up to some measurement errors, that we assume to be mutually independent and serially uncorrelated.

We denote by $H = \{26, 52, 104, 208, 364, 520\}$ is the list of available maturities in weeks. The vector of observed yields is denoted by $R_t = [R_t(h)]_{h\in H}$. Besides, $V_t = [V_t(h)]_{h\in\{104,520\}}$ denotes the conditional variance proxies for yield of maturity h, $S_t = [S_t^{(q)}(h)]_{h=520,q\in\{12,52\}}$ denotes the survey of professional forecasters q-periods ahead for the h-maturity yield. The measurement equations for the yields and the survey variables are directly derived from Equation (5.19):

$$R_t(h) = \overline{B}_h + \overline{A}_h' X_t + \sigma_R \eta_{R,h,t}, \quad h \in H \tag{5.27}$$

$$\begin{aligned} S_t^{(q)}(h) &= \overline{B}_h + \overline{A}_h' \mathbb{E}_t^{\mathbb{P}}(X_{t+q}) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)} \\ &= \overline{B}_h + \overline{A}_h' \left(\sum_{i=0}^{q-1} M^i m + M^q X_t \right) + \sigma_{S,h}^{(q)} \eta_{S,h,t}^{(q)}, \quad \text{for } \begin{cases} h = 520, \\ q \in \{13, 52\} \end{cases} \end{aligned} \tag{5.28}$$

where σ_R is the same for all maturities h, and $\eta_{R,h,t}$ and $\eta_{S,h,t}^{(q)}$ are i.i.d. Gaussian white noises.

Eventually, we introduce measurement equations for the volatility proxies based on the conditional variance-covariance matrix of the latent process X_t. As already emphasized, the affine property of the VARG distribution implies that the conditional variance-covariance matrix of X_{t+1} given its own past is affine in X_t. Specifically, the new measurement equations write:

$$\begin{aligned} V_t(h) &= \overline{A}_h' \left\{ \text{diag}\left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(\nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}\prime} X_t \right) \right] \right\} \overline{A}_h + \sigma_{V,h}\, \eta_{V,h,t} \\ &= (\overline{A}_h \odot \overline{A}_h)' \left[\mu^{\mathbb{P}} \odot \mu^{\mathbb{P}} \odot \left(\nu + 2\alpha^{\mathbb{P}} + 2\beta^{\mathbb{P}\prime} X_t \right) \right] + \sigma_{V,h}\, \eta_{V,h,t}\,, \quad h \in \{104, 520\}\,, \end{aligned} \tag{5.29}$$

where $\eta_{V,h,t}$ is a i.i.d. Gaussian white noise. The total vector of observable variables is denoted by $Y_t = [R_t', S_t', V_t']'$. Our vector of observables therefore contains 10 different variables (6 yields, 2 conditional variance proxies and 2 survey-based forecast series). Stacking the transition and measurement equations, we obtain the following state-space model representation:

$$\begin{cases} X_{t+1} &= m + M X_t + \Sigma_t^{1/2} \varepsilon_{t+1} \\ Y_t &= \Gamma_0 + \Gamma_1 X_t + \Omega\, \eta_t \end{cases}. \tag{5.30}$$

where $\eta_t = (\eta'_{R,t}, \eta'_{S,t}, \eta'_{V,t})' \sim \mathcal{IIN}(0, I)$, and Γ_0 and Γ_1 are based on Equations (5.27-5.29).

To estimate the model, we use pseudo-maximum likelihood where an approximation of the likelihood function is derived from the linear Kalman filter. The latter is slightly modified to accommodate the fact that the latent factor X_t is conditionally heteroskedastic. To do so, we run the Kalman filter replacing the real – intractable – log-likelihood derived from conditional Gamma distributions by that obtained from Gaussian distributions, i.e. we approximate ε_{t+1} by a standard Gaussian white noise. The availability of a linear state-space model makes the application of such a procedure very easy (see e.g. Kim and Singleton [2012]). For identification purposes, we impose that $\mu^{\mathbb{P}} = (1, \ldots, 1)'$. In addition, we take a lower-triangular $\beta^{\mathbb{P}}$ matrix, which implies that $X_t^{(2)}$ (that does not enter directly the short-term interest rate specification) Granger-causes $X_t^{(1)} \equiv X_{1,t}$ (that does enter directly the short-term interest rate), but $X_t^{(1)}$ does not Granger-cause $X_t^{(2)}$.

We estimate all risk-neutral parameters and the four market prices of risk in a single step. Historical parameters are then deduced from the estimated parameters. We also estimate the short-term interest rate loading δ_1 (we have $r_t = \delta_1 X_t^{(1)}$) and the measurement-noise standard deviations of the yields. The computation of the parameters standard errors for a preliminary fully-parameterized specification pointed to the non-statistical significativity of some parameters. The latter were further constrained to zero and we eventually end up estimating 16 parameters in an embedded specification. The estimation results are presented in Table 5.3.

5.4.3 Cross-sectional fit

Most of the parameter estimates are highly significantly different from zero. We observe that most of the factors are highly persistent under both measures with the $\rho_i = \mu_i \beta_{i,i}$ parameters being close to one. We present a graphical representation of the filtered factors on Figure 5.4.

The first three factors experience long periods at zero, notably during those periods when the 6-month interest rate is at its lowest level (2001 to 2006). By contrast, Factor 4 experiences large and persistent fluctuations during the whole sample. The interpretation of the factors is facilitated by the analysis of the so-called factor loadings, that describe the affine relationships between the levels and conditional variances of yields on the one hand and the factors on the other one. These loadings are plotted in Figure 5.5. Panel (a) of this figure shows for instance that the first and fourth factor are particularly important

Table 5.3: Parameter estimates

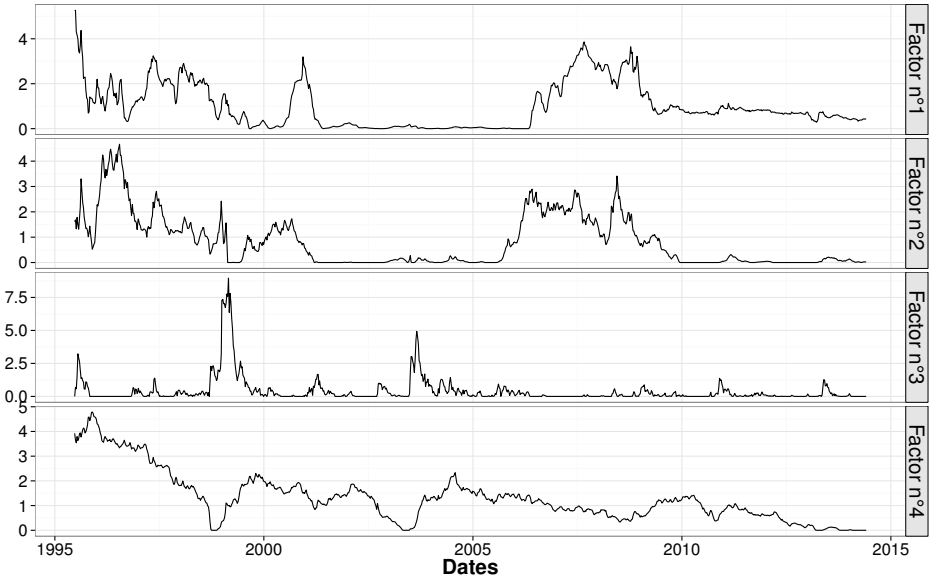
	$\mathbb{Q}$-parameters		$\mathbb{P}$-parameters	
	Estimates	Std.	Estimates	Std.
α_4	3.2455	0.1118	3.2347	0.1113
$\beta_{1,1}$	0.9663	0.0078	0.9794	0.0042
$\beta_{2,2}$	0.9978	0.0005	0.9957	0.0006
$\beta_{3,3}$	0.9486	0.0044	0.9705	0.0023
$\beta_{4,4}$	0.9967	0.0005	0.9933	0.0003
$\beta_{2,1}$	0.0308	0.0041	0.0308	0.0041
$\beta_{3,2}$	0.1094	0.0059	0.1120	0.0061
$\beta_{4,3}$	$3.88{\cdot}10^{-4}$	$2.28{\cdot}10^{-5}$	$3.87{\cdot}10^{-4}$	$2.27{\cdot}10^{-5}$
μ_1	1	–	1.0135	0.0040
μ_2	1	–	0.9980	0.0005
μ_3	1	–	1.0231	0.0023
μ_4	1	–	0.9967	0.0003
Other Parameters				
δ_1	0.0030	0.0003		
θ_1	-0.0133	0.0039	θ_2	0.0020	0.0005
θ_3	-0.0226	0.0022	θ_4	0.0033	0.0003
σ_R	0.0407	0.0003		
σ_V	$3 \cdot 10^{-3}$	–	σ_S	0.15	–

Note: This table reports the estimated parameters of a four-factor model where $r_t = \delta_1 X_{1,t}$. Standard deviations (Std.) are calculated from the outer product of the log-likelihood gradient at the estimated parameter values. The symbol '–' in the standard-deviation column indicates that the parameter has been calibrated. The σ_V's are set to twice the standard deviations of the differences between the GARCH, EGARCH and rolling window variance proxies. The σ_S's are set at the in-sample mean of standard-deviations of forecasts among the professional forecasters. σ_S and σ_R are expressed in percentage points. Last, we impose that the unconditional mean of the short-term interest rate is equal to 100bps.

to account for the fluctuations of short-term and long-term yields, respectively. The influence of the second factor is more evenly spread across the yield curve. Moreover, Panel (b) suggests that changes in the second and third factors have more important impacts on the conditional variances of medium- to long-term yields than the first two factors.

We now turn to the empirical performances of the VARG term-structure model. First, we observe a remarkable cross-sectional fit of the JGB yields with the measurement-noise standard deviations of yields being 4bps and the differences between the oberservable and model-implied yields are nearly indistinguishable to the naked eye (see Figure 5.6).

Figure 5.4: Estimated Factors



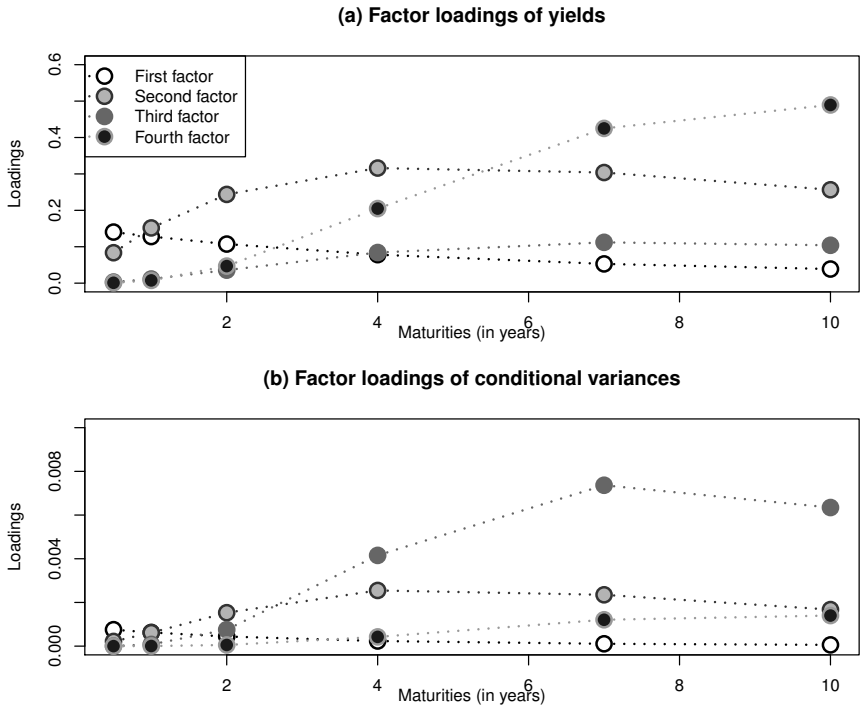
Notes: Factors are filtered estimates from the linear Kalman filter on the full sample (June 1995 to May 2014). The short-term rate r_t is equal to $\delta_1 X_{1,t}$. For all $j > i$ factor $X_{j,t}$ Granger-causes factor $X_{i,t}$.

The top panel of Figure 5.7 presents the fit of the observable conditional variance proxies obtained by our term structure model. The main periods of volatility spikes are well-captured by the model. Also, these plots demonstrate the ability of the model to accommodate different patterns in the volatilities of yields across maturities.

The second panel of Figure 5.7 presents the fit obtained on the survey of professional forecasters equations. For both the 3-month and 1-year horizons, model-implied forecasts of the 10-year yield nicely reproduces the behavior of observed surveys. Note that the standard errors on the survey measurement noise are parameterized with values that are commensurate with the disagreement among forecasters, as measured by the average standard deviations of the professional forecasters declarations (10bps).

On the whole, these results show a great flexibility of our VARG term-structure model, being able not only to closely reproduce both the level and the conditional volatility behavior of yields across maturities, but also to provide expectations under the historical measure that are coherent with survey-based forecasts.

Figure 5.5: Estimated Factors



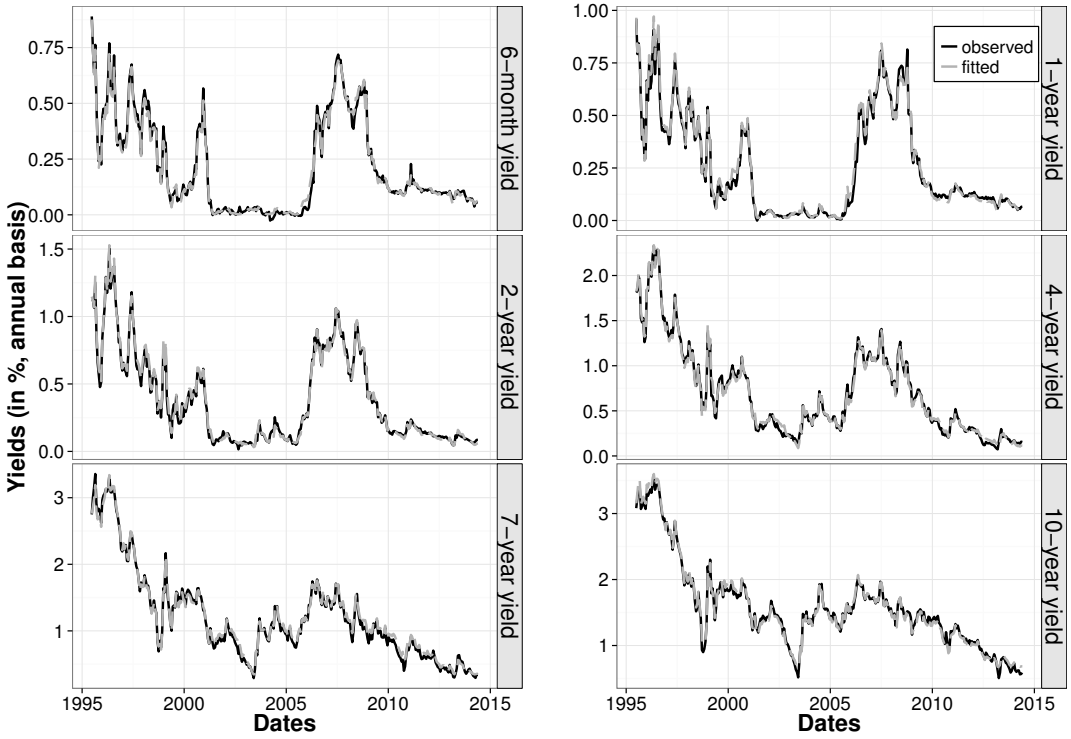
Notes: This figure displays the factor loadings of the levels (Panel (a)) and the conditional variances (Panel (b)) of yields. These levels and variances are affine in factors X_t. Panel (a) relate to Equation (5.27) and Panel (b) to Equation (5.29). The loadings are divided by the sample standard deviations of the estimated factors, that is, using the notations of Equation (5.30), the plotted loadings are of the form $\Gamma_{1,i,j}/\sqrt{Var(X_{j,t})}$, where $\Gamma_{1,i,j}$ is the entry $\{i,j\}$ of matrix Γ_1.

5.5 Lift-off probabilities

As described in previous sections, our specifications entail closed-form and semi closed-form formulas for calculating distribution of future yields. As an application, we compute the model-implied probabilities that the short-term interest rate will remain low for a certain amount of time. We can compute such probabilities under both $\mathbb{P}$ and $\mathbb{Q}$ measures. The discrepancies existing between the $\mathbb{P}$ and $\mathbb{Q}$ probabilities stem from the risk aversion of investors.

We first consider the time-series behavior of such probabilities in Figure 5.8. Specifically, this exercise is based on the probability formula of the short-rate hitting zero in k periods ($r_{t+k} = 0$, see Proposition 5.3.4, *(i)*). We apply the expression for 2- and 5-year-ahead forecasts (i.e. with $k = 104$ and $k = 260$) for both the historical and risk-neutral probabilities (resp. red and black lines of top panels of Figure 5.8). A second exercise exploits the Duffie, Pan, and Singleton [2000] formula to calculate the probabilities of the short-term interest rate being below 25 bps 2- and 5-years ahead, also for both measures (bottom

Figure 5.6: Observed and model-implied yields



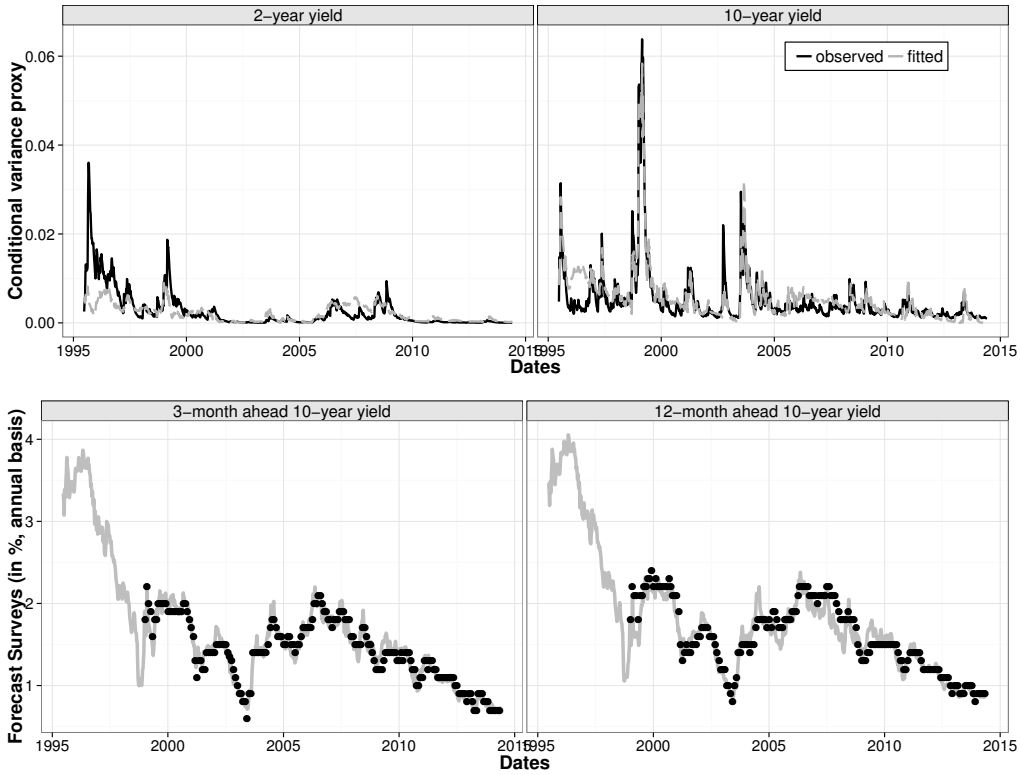
Notes: Yields are observed at the weekly frequency from June 16, 1995 to May 30, 2014. Yields are expressed in annualized percentage points, with maturities from 6 months to 10 years. The black solid lines are the observed yields and the grey dashed lines are the model-implied (or fitted) yields using the term structure framework of Section 5.3 with 4 factors ($n_1 = 1$ and $n_2 = 3$).

panels of Figure 5.8).

Let us focus first on the top-left panel, representing both $\mathbb{P}_t(r_{t+k} = 0)$ and $\mathbb{Q}_t(r_{t+k} = 0)$ for a 2-year horizon. Until 1998, both probabilities are small and begin experiencing fluctuations from that date on. In 1998, coherently with the low level of yields and of with the increase in yield volatilities (see Figure 5.3), we obtain a dramatic increase in the probability to hit the ZLB. After that volatility spike, the probability decreases until 2001. The ZLB period of 2001-2006 corresponds to large increases in both probabilities, reaching levels highest than 75% during 2003. This peak is coherent with a flattening of the yield curve at that date: as short rates stay low and long-term rates begin to drop, agents expect a higher probability of the short rate staying at zero for 2 years on. In some sense, those probabilities are a convenient way to represent information contained in the yield curve. The probability to reach and/or stay at the ZLB increases again at the end of the sample, amid the last financial crisis (from 2009 onwards).

We turn now to the same probabilities for a 5-year horizon (top-right plot of Figure 5.8).

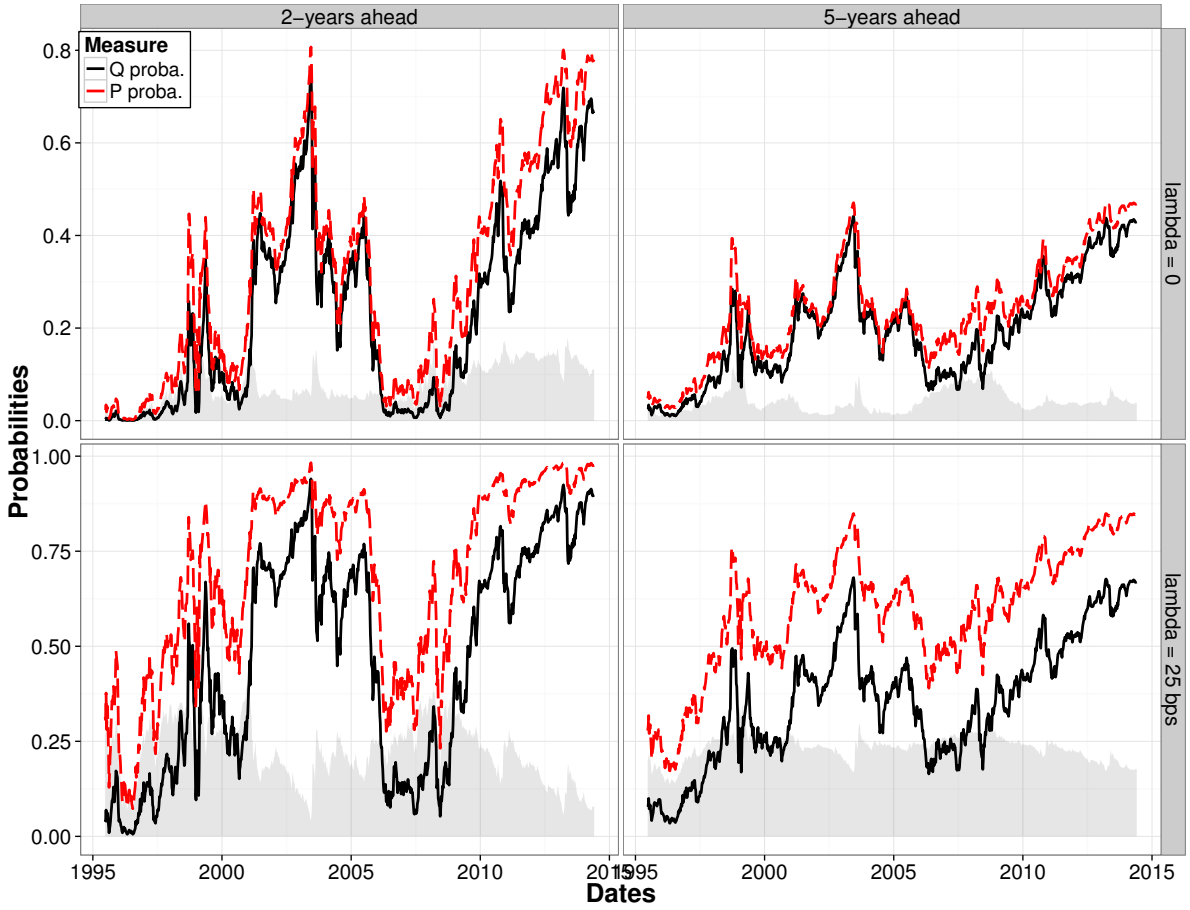
Figure 5.7: Fitted conditional variance proxies and surveys



Notes: The top panel presents the two conditional variance proxies $V_t(h)$ estimated with an EGARCH(1,1) model on 2- and 10-year yields (left and right tiles) of weekly data from June 30, 1995 to May 30, 2014. The black solid lines are the observed variance proxies and the grey dotted lines are the model-implied (or fitted) equivalent. The bottom panel presents the survey of professional forecasters for the 10-year yield, 3- and 12-months ahead. Survey-based data are available at the monthly frequency from 1999 to the end of the sample. The black dots correspond to the observed data, and the grey solid lines are the fitted equivalent.

First and unsurprisingly, the probabilities under both measures are on average lower than for the 2-year horizon. Second, the differences between $\mathbb{P}$ and $\mathbb{Q}$ probabilities under the two measures are smaller than for the previous horizon. This being said, the difference between $\mathbb{P}$ and $\mathbb{Q}$ probabilities are not negligible. In particular, in 2007, in a context of rising short-term rates, $\mathbb{Q}$ probabilities are twice lower than their physical counterparts. This implies that neglecting the existence of risk premia results in a substantial underestimation of the persistency of the ZLB regime. In other words, the short-term interest rate is expected to stay at zero for a longer period under the physical measure. Third, it is interesting to note that over the last five years of data, even if the observed short- to medium-term yields (up to 4 years) are fairly stable (see Figure 5.6), the probability of the short-rate being at zero in years has substantially grown. This phenomenon is consistent with the decrease in longer-term yields, which points to an increase in the perceived expected length of the low-interest rate environment.

Figure 5.8: Time-series of ZLB probabilities: $\mathbb{P}_t(r_{t+k} \leqslant \lambda)$ and $\mathbb{Q}_t(r_{t+k} \leqslant \lambda)$

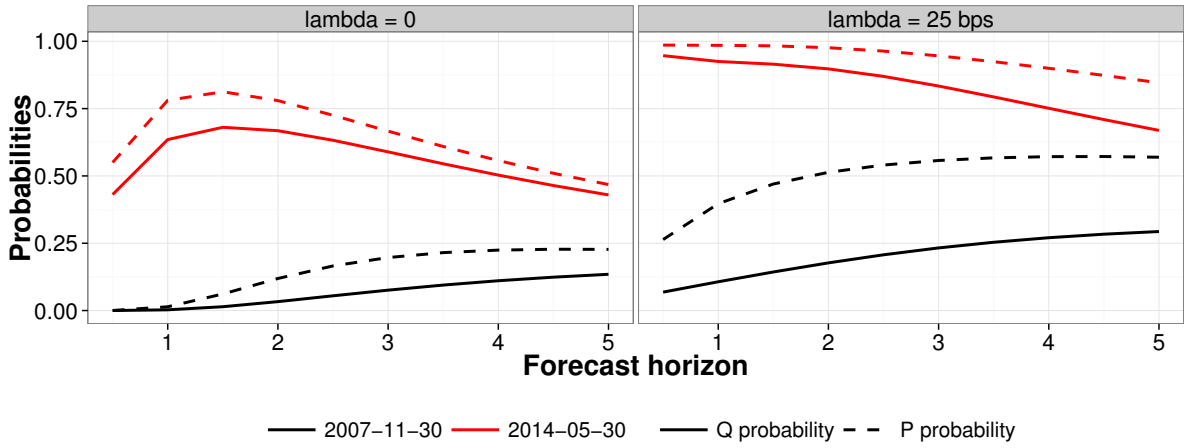


Notes: Probabilities are computed with weekly data from June 16, 1995 to May 30, 2014. The top panels present the probabilities of the short-rate hitting zero in two years (top-left panel) and 5 years (top-right panel). On bottom panels, we represent the probabilities of the short-rate being below 25 bps in 2 years (bottom-left panel) and 5 years (bottom-right panel). Black solid lines are the risk-neutral probabilities whereas red dashed lines are the historical ones; grey-shaded areas are the difference between the two probabilities.

The bottom panels of Figure 5.8 confirm the previous results. Since the threshold is now different from 0 (25 bps), we observe higher probability values under both measures. For instance, the historical and risk-neutral probabilities of going below 25 bps at the 2-year maturity (bottom-left tile) are close to 1 in 2003, and fluctuates between 0.5 and 1 during the ZLB period. The divergence between $\mathbb{P}$ and $\mathbb{Q}$ probabilities are larger than for the upper two plots: in many instances, the physical probabilities of being in a low-rate environment are two to three times larger than the risk-neutral ones.

Figure 5.9 shows conditional $\mathbb{P}$ and $\mathbb{Q}$ probabilities of having low short rates over a richer spectrum of horizons. We consider two dates, the first in late 2007 and the second at

Figure 5.9: Horizon structure of ZLB probabilities: $\mathbb{P}_t(r_{t+k} \leqslant \lambda)$ and $\mathbb{Q}_t(r_{t+k} \leqslant \lambda)$.



Notes: X axis is the horizon k of the short rate r_{t+k} being exactly 0 (left tile), or below 25 bps (right tile). Black and red curves correspond distinguish the date at which these probabilities are evaluated, and respectively correspond to November 30, 2007 and May 30, 2014. Solid and dashed lines represent respectively to $\mathbb{Q}$ and $\mathbb{P}$-probabilities.

the end of the sample (respectively black and red lines, Figure 5.9). For each date, the forecast horizon varies between 6 months and 5 years. Our term-structure model generates different profiles of low-rate probabilities with respect to the forecast horizon: for the earlier date, the horizon structure is globally increasing whereas it is hump-shaped for the latest date of the sample. This illustrates the ability of the model to generate flexible expected paths of future short-term interest rates.

5.6 Conclusion

In this paper, we introduce a new class of Affine Term Structure Models able to provide at the same time non-negative yields at any maturity and a short rate which can at zero for extended periods of time (the ZLB being a non-absorbing state) while the longer-term rates can still vary. These characteristics are obtained by the introduction of a new univariate non-negative affine process called Autoregressive Gamma-zero and its multivariate affine extension (VARG), involving conditional distributions with zero-point masses. The affine nature of our model allows for a great flexibility at the estimation stage. First, a Kalman-filter-based maximum likelihood approach is allowed. Second, the estimation procedure is easily enhanced by explicitly taking into account relevant information like interest rate survey-based forecasts, conditional yield variance proxies. Third, explicit and quasi-explicit formulas are easily derived for calculating the physical and

risk-neutral probabilities of the short-term rate staying at –or close to– zero at different forecast horizon.

We assess the model performances with an application to Japanese government bond yields. Our four-latent-factors VARG term-structure model is able to fit both yield levels and conditional volatilities of yields. We also compute time-varying probabilities of being at the ZLB in the future under the historical and risk-neutral measures. Our results show that differences between such physical and risk-neutral probabilities can be substantial.

The affine framework we develop in this paper can also be used to easily price fixed-income derivatives. We left those different directions for further future research.

Appendices to Chapter 5

5.A Conditional moments of the $ARG_0(\alpha, \beta, \mu)$ process

The conditional cumulant-generating function is $\psi_t(u) = \log(\varphi_t(u)) = \dfrac{u\mu}{1-u\mu}\beta X_t + \dfrac{u\mu}{1-u\mu}\alpha$. Deriving this function with respect to u gives us the conditional expectation and variance of X_{t+1} given X_t:

$$\begin{aligned}
\frac{d}{du}\psi_t(0) &= \frac{\rho(1-u\mu)+\mu(u\rho)}{(1-u\mu)^2}X_t + \frac{\mu\alpha(1-u\mu)+\mu(u\mu\alpha)}{(1-u\mu)^2}\Bigg|_{u=0} \\
&= \frac{\rho}{(1-u\mu)^2}X_t + \frac{\mu\alpha}{(1-u\mu)^2}\Bigg|_{u=0} \\
&= \alpha\mu + \rho X_t \\
\frac{d}{du^2}\psi_t(0) &= \frac{2\mu\rho}{(1-u\mu)^3}X_t + \frac{2\mu^2\alpha}{(1-u\mu)^3}\Bigg|_{u=0} \\
&= 2\mu^2\alpha + 2\mu\rho X_t
\end{aligned}$$

Let us introduce now the following notations: $m_{1,t} = \mathbb{E}(X_t)$ and $m_{2,t} = \mathbb{V}(X_t)$. It easily seen that these unconditional moments are defined by the following system of difference equations:

$$\begin{aligned}
m_{1,t} &= \rho\, m_{1,t-1} + \alpha\mu \\
m_{2,t} &= 2\mu^2\alpha + 2\mu\rho\, m_{1,t-1} + \rho^2\, m_{2,t-1}
\end{aligned}$$

that can be represented in matrix form as:

$$\begin{pmatrix} m_{1,t} \\ m_{2,t} \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 2\mu\rho & \rho^2 \end{pmatrix} \begin{pmatrix} m_{1,t-1} \\ m_{2,t-1} \end{pmatrix} + \begin{pmatrix} \mu\alpha \\ 2\mu^2\alpha \end{pmatrix}. \tag{5.31}$$

This system admits a stationary solution if and only if $\rho < 1$, and it is given by:

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha\mu}{1-\rho} \\ \dfrac{2\alpha\mu^2}{(1-\rho)(1-\rho^2)} \end{pmatrix}. \tag{5.32}$$

m_1 and m_2 are therefore the marginal mean and marginal variance of the stationary $ARG_0(\alpha, \beta, \mu)$ process.

5.B Sojourn time and lift-off probability of the $ARG_0(\alpha, \beta, \mu)$ process

Proof. of Lemma 5.2.1

$$\varphi_X(u) = \int_{\mathbb{R}^+} \exp(ux)\, \mathrm{d}\mathbb{P}_X(x) = \mathbb{P}_X\{0\} + \int_{x>0} \exp(ux)\, \mathrm{d}\mathbb{P}_X(x)$$

Since $x > 0$, $\exp(ux) \to 0$ when $u \to -\infty$, and, using Lebesgue theorem, the integral tends towards 0. ■

Proof. of Proposition 5.2.2

(i) Let us consider an $ARG_0(\alpha, \beta, \mu)$ process X_t and let us study, first, the limit of:

$$\mathbb{E}\left[\exp(uX_{t+h}) \,|\, X_t\right] = \exp\left\{ a^{\circ h}(u)\, X_t + \sum_{k=0}^{h-1} b[a^{\circ k}(u)] \right\},$$

when $u \to -\infty$, in order to calculate $\mathbb{P}(X_{t+h} = 0 \,|\, X_t)$. It can be shown recursively that:

$$\begin{aligned} a^{\circ h}(u) &= \frac{\rho^h u}{1 - u\mu\left[\dfrac{1-\rho^h}{1-\rho}\right]} \\ \sum_{k=0}^{h-1} b[a^{\circ k}(u)] &= (1-\rho)\alpha u\mu \sum_{k=0}^{h-1} \frac{\rho^k}{1 - \rho - u\mu + u\mu\rho^{k+1}}, \end{aligned}$$

and, when $u \to -\infty$, we have:

$$\begin{aligned} \mathbb{P}(X_{t+h} = 0 \,|\, X_t) &= \exp\left[-\frac{\rho^h X_t}{\mu\left(\dfrac{1-\rho^h}{1-\rho}\right)} - (1-\rho)\,\alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1-\rho^{k+1}} \right] \\ &= \exp\left\{ -(1-\rho)\left[\frac{\rho^h X_t}{\mu\,(1-\rho^h)} + \alpha \sum_{k=0}^{h-1} \frac{\rho^k}{1-\rho^{k+1}} \right] \right\}, \end{aligned} \tag{5.33}$$

and the result is proved.

(ii) From Definition 5.2.2 we know that, when X_t follows an $ARG_0(\alpha, \beta, \mu)$ process, $\mathbb{P}(X_{t+1} = 0 \,|\, X_t) = \exp(-\alpha - \beta\, X_t)$. Then, if we denote by $f_h(X_t) = \mathbb{P}(X_{t+h} =$

$0, \ldots, X_{t+1} = 0 \,|\, X_t)$, we can always write:

$$\begin{aligned} f_h(X_t) &= \mathbb{P}(X_{t+h} = 0, \ldots, X_{t+1} = 0 \,|\, X_t) \\ &= \mathbb{P}(X_{t+h} = 0 \,|\, X_{t+h-1} = 0, \ldots, X_{t+1} = 0; X_t)\, f_{h-1}(X_t) \\ &= \mathbb{P}(X_{t+h} = 0 \,|\, X_{t+h-1} = 0)\, f_{h-1}(X_t) \end{aligned}$$

and the result is easily proved by recursion.
(*iii*)

$$\begin{aligned} & \mathbb{P}(X_{t+h} > 0, X_{t+h-1} = 0, \ldots, X_{t+1} = 0 \,|\, X_t) \\ = \; & \mathbb{P}(X_{t+h} > 0 \,|\, X_{t+h-1} = 0, \ldots, X_{t+1} = 0; X_t)\, \mathbb{P}(X_{t+h-1} = 0, \ldots, X_{t+1} = 0 \,|\, X_t) \\ = \; & [1 - \mathbb{P}(X_{t+h} = 0 \,|\, X_{t+h-1} = 0)] \exp\left[-\alpha(h-1) - \beta\, X_t\right] \\ = \; & [1 - \exp(-\alpha)] \exp\left[-\alpha(h-1) - \beta\, X_t\right] . \end{aligned}$$

■

5.C Risk-neutral conditional Laplace transform and yield-to-maturity formula

Proof. of Proposition 5.3.2
Given that, from Assumption 2, we have $r_t = \delta' X_t$, where the first n_1 components are different from zero and the remaining ones are equal to zero, we can write:

$$\begin{aligned} P_t(h) &= \exp\left(A_h + B_h' X_t\right) = \mathbb{E}_t^{\mathbb{Q}}\left[\exp(-r_t) \exp\left(A_{h-1} + B_{h-1}' X_{t+1}\right)\right] \\ &= \exp\left(-r_t + A_{h-1}\right) \mathbb{E}_t^{\mathbb{Q}}\left[\exp\left(B_{h-1}' X_{t+1}\right)\right] \\ &= \exp\left[A_{h-1} + \sum_{j=1}^{n} b_j^{\mathbb{Q}}(B_{j,h-1}) + \left(\sum_{j=1}^{n} a_j^{\mathbb{Q}}(B_{j,h-1}) - \delta\right)' X_t\right] \end{aligned}$$

and the result follows by identification. ■

5.D Historical conditional Laplace transform of the state vector

First of all, the following result holds:

Proposition 5.D.1. *Let us consider a scalar Extended $ARG_\nu(\alpha,\beta,\mu)$ process (X_t) with conditional log-Laplace transform $\psi_t(u) = \frac{\rho u}{1-u\mu}X_t + \frac{u\mu}{1-u\mu}\alpha - \nu\log(1-u\mu)$, with $\rho = \beta\,\mu$. The associated conditional Esscher transform, with parameter $\theta\in\mathbb{R}$, generates the family of probability distributions characterized by the following conditional log-Laplace transform:*

$$\psi_t^*(u) = \frac{u\rho^*}{1-u\mu^*}\,X_t + \frac{u\mu^*}{1-u\mu^*}\,\alpha^* - \nu\,\log(1-u\mu^*)\,, \tag{5.34}$$

which is the log-Laplace transform of an $EARG_\nu(\alpha^*,\beta^*,\mu^*)$ process with

$$\begin{aligned}\rho^* &= \frac{\rho}{(1-\theta\mu)^2}\,,\quad \mu^* = \frac{\mu}{1-\theta\mu}\,,\quad \alpha^* = \frac{\alpha}{1-\theta\mu}\,,\\ \beta^* &:= \frac{\rho^*}{\mu^*} = \frac{\rho}{\mu(1-\theta\mu)} = \frac{\beta}{1-\theta\mu}\,.\end{aligned}$$

Proof. of Proposition 5.3.3
If we consider our change of probability measure $\frac{d\mathbb{P}_{t,t+1}}{d\mathbb{Q}_{t,t+1}} = \exp\left[\theta'\widetilde{X}_{t+1} - \psi_t^{\mathbb{Q}}(\theta)\right]$, where $\mathbb{P}_{t,t+1}$ is the conditionalEsscher transform of $\mathbb{Q}_{t,t+1}$ associated with θ, we have $\psi_{j,t}^{\mathbb{P}}(u_j) = \psi_{j,t}^{\mathbb{Q}}(u_j+\theta_j) - \psi_{j,t}^{\mathbb{Q}}(\theta_j)$ for any $j\in\{1,\ldots,n\}$, and applying Proposition 5.D.1, Proposition 5.3.3 is easily proved. ■

Bibliography

P. Abbassi and T. Linzert. The Effectiveness of Monetary Policy in Steering Money Market Rates During the Recent Financial Crisis. *Journal of Macroeconomics*, 34(4): 945–954, 2012.

T. Adrian and H. Wu. The Term Structure of Inflation Expectations. Staff Reports 362, Federal Reserve Bank of New York, 2009.

D.-H. Ahn, R. F. Dittmar, and A. R. Gallant. Quadratic Term Structure Models: Theory and Evidence. *Review of Financial Studies*, 15(1):243–288, March 2002.

Y. Ait-Sahalia and R. L. Kimmel. Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions. *Journal of Financial Economics*, 98:113–144, 2010.

C. Almeida, J. Graveline, and S. Joslin. Do Interest Rate Options Contain Information About Excess Returns? *Journal of Econometrics*, 164:35–44, 2011.

J. D. Amato and E. M. Remolona. The Credit Spread Puzzle. BIS Quarterly Review, Bank for International Settlements, December 2003.

T. G. Andersen and L. Benzoni. Can Bonds Hedge Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models. Working Papers, Carlson School of Management, University of Minnesota, 2006.

B. D. O. Anderson and J. B. Moore. *Optimal Filtering* . Prentice-Hall, 1979.

M. Andreasen and A. Meldrum. Likelihood Inference in Non-Linear Term Structure Models: The Importance of the Zero Lower Bound. BoE Working Papers, Bank of England, Jan. 2011.

M. Andreasen and A. Meldrum. Dynamic Term Structure Models: The Best Way to Enforce the Zero Lower Bound. Technical report, 2013.

M. Andreasen and A. Meldrum. The SR Approach: A New Estimation Procedure for Non-Linear and Non-Gaussian Dynamic Term Structure Models. *Journal of Econometrics*, 184(2):420–451, February 2015.

A. Ang and M. Piazzesi. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics*, 50 (4):745–787, May 2003.

A. Ang, J. Boivin, S. Dong, and R. Loo-Kung. Monetary Policy Shifts and the Term Structure. *Review of Economic Studies*, 78(2):429–457, 2011.

P. Angelini, A. Nobili, and C. Picillo. The Interbank Market after August 2007: What Has Changed, and Why? *Journal of Money, Credit and Banking*, 43(5):923–958, August 2011.

M. Athans, R. Wishner, and A. Bertolini. Suboptimal State Estimation for Continuous-Time Nonlinear Systems from Discrete Noisy Measurements. *IEEE Transactions on Automatic Control*, 13(5):504–514, 1968.

M. Baadsgaard, J. N. Nielsen, and H. Madsen. Estimating Multivariate Exponential-Affine Term Structure Models from Coupon Bond Prices Using Nonlinear Filtering. Technical report, 2000.

D. Backus, S. Foresi, and C. Telmer. Affine Term Structure Models and the Forward Premium Anomaly. *Journal of Finance*, 56(1):279–304, February 2001.

D. Backus, M. Chernov, and S. E. Zin. Sources of Entropy in Representative Agent Models. *Journal of Finance*, 69(1):51–99, February 2014.

Bank of International Settlements. Zero-Coupon Yield Curves: Technical Documentation. BIS Working Papers 25, Bank for International Settlements, 2005.

Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2002.

M. D. Bauer and G. D. Rudebusch. Monetary Policy Expectations at the Zero Lower Bound. Working Paper Series 2013-18, Federal Reserve Bank of San Francisco, 2013.

D. Beaglehole and M. Tenney. A Nonlinear Equilibrium Model of the Term Structure of Interest Rates: Corrections and Additions. *Journal of Financial Economics*, 32: 345–454, 1992.

A. Berndt, R. Douglas, D. Duffie, M. Ferguson, and D. Schranz. Measuring Default Risk Premia from Default Swap Rates and EDFs. BIS Working Papers 173, Bank for International Settlements, Mar. 2005.

H. Bertholon, A. Monfort, and F. Pegoraro. Econometric Asset Pricing Modelling. *Journal of Financial Econometrics*, 4:407–458, 2008.

R. Bikbov and M. Chernov. Yield Curve and Volatility: Lessons from Eurodollar Futures and Options. *Journal of Financial Econometrics*, 9(1):66–105, 2011.

F. Black. Interest Rates as Options. *Journal of Finance*, 50(5):1371–76, December 1995.

A. N. Bomfim. Counterparty Credit Risk in Interest Rate Swaps During Times of Market Stress. Finance and Economics Discussion Series 2003-09, Board of Governors of the Federal Reserve System (U.S.), 2003.

M. Brandt and D. Chapman. Comparing Multifactor Models of the Term Structure. Wharton school working paper, 2003.

S. J. Brown and P. H. Dybvig. The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates. *Journal of Finance*, 41(3):617–530, July 1986.

M. K. Brunnermeier and L. H. Pedersen. Market Liquidity and Funding Liquidity. *Review of Financial Studies*, 22(6):2201–2238, June 2009.

A. Buraschi, A. Cieslak, and F. Trojani. Correlation Risk and the Term Structure of Interest rates. *Working Paper*, 2008.

C. Burnside. Solving Asset Pricing Models with Gaussian Shocks. *Journal of Economic Dynamics and Control*, 22(3):329–340, March 1998.

J. Y. Campbell, A. Sunderam, and L. M. Viceira. Inflation Bets or Deflation Hedges? The Changing Risk of Nominal Bonds. Working paper, National Bureau of Economic Research, 2013.

S. Carpenter, S. Demiralp, and J. Eisenschmidt. The Effectiveness of Non-Standard Monetary Policy in Addressing Liquidity Risk During the Financial Crisis: The Experiences of the Federal Reserve and the European Central Bank . *Journal of Economic Dynamics and Control*, 43(0):107 – 129, 2014. ISSN 0165-1889.

M. Cecioni, G. Ferrero, and A. Secchi. Unconventional Monetary Policy in Theory and in Practice. Questioni di Economia e Finanza (Occasional Papers) 102, Bank of Italy, Economic Research and International Relations Area, Sept. 2011.

L. Chen, D. Filipovic, and V. Poor. Quadratic Term Structure Models for Risk-Free and Defaultable Rates. *Mathematical Finance*, 14(4):515–536, October 2004.

L. Chen, P. Collin-Dufresne, and R. Goldstein. On the Relation Between the Credit Spread Puzzle and the Equity Premium Puzzle. *Review of Financial Studies*, 22(9): 3367–3409, 2009.

R.-R. Chen and L. Scott. Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income*, 3:14–31, 1993.

R.-R. Chen, X. Cheng, F. J. Fabozzi, and B. Liu. An Explicit, Multi-Factor Credit Default Swap Pricing Model with Correlated Factors. *Journal of Financial and Quantitative Analysis*, 43(01):123–160, March 2008.

R.-R. Chen, X. Cheng, and L. Wu. Dynamic Interactions Between Interest-Rate and Credit Risk: Theory and Evidence on the Credit Default Swap Term Structure. *Review of Finance*, 17(1):403–441, 2011.

P. Cheng and O. Scaillet. Linear-Quadratic Jump-Diffusion Modeling. *Mathematical Finance*, 17(4):575–598, 2007.

J. H. Christensen and G. D. Rudebusch. Modeling Yields at the Zero Lower Bound: Are Shadow Rates the Solution? Working Paper Series 2013-39, Federal Reserve Bank of San Francisco, 2013.

J. H. E. Christensen, J. A. Lopez, and G. D. Rudebusch. Do Central Bank Liquidity Facilities Affect Interbank Lending Rates? *Journal of Business & Economic Statistics*, 32(1):136–151, January 2014.

P. Christoffersen, C. Dorion, K. Jacobs, and L. Karoui. Nonlinear Kalman Filtering in Affine Term Structure Models. *Management Science*, 60(9):2248–2268, 2014.

A. Cieslak and P. Povala. Information in the Term Structure of Yield Curve Volatility. *Journal of Finance*, forthcoming, 2015.

L. Codogno, C. Favero, and A. Missale. Yield spreads on EMU Government Bonds. *Economic Policy*, 18(37):503–532, 2003.

F. Collard and M. Juillard. Accuracy of Stochastic Perturbation Methods: The Case of Asset Pricing Models. *Journal of Economic Dynamics and Control*, 25(6-7):979–999, June 2001.

P. Collin-Dufresne and R. S. Goldstein. Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance*, 57(4): 1685–1730, 08 2002.

P. Collin-Dufresne and B. Solnik. On the Term Structure of Default Premia in the Swap and LIBOR Markets. *Journal of Finance*, 56(3):1095–1115, June 2001.

P. Collin-Dufresne, R. S. Goldstein, and S. Martin. The Determinants of Credit Spread Changes. *Journal of Finance*, 56(6):2177–2207, December 2001.

G. M. Constantinides. A Theory of the Nominal Term Structure of Interest Rates. *Review of Financial Studies*, 5(4):531–52, 1992.

J. C. Cox, J. Ingersoll, Jonathan E, and S. A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–407, March 1985.

D. Creal and C. Wu. Estimation of Affine Term Structure Models With Spanned or Unspanned Stochastic Volatility. *Journal of Econometrics*, 185:60–81, 2015.

D. D. Creal and J. C. Wu. Interest Rate Uncertainty and Economic Fluctuations. Working Papers, Chicago Booth School of Business, 2014.

Q. Dai and K. J. Singleton. Specification Analysis of Affine Term Structure Models. *Journal of Finance*, 55:1943–1978, 2000.

Q. Dai and K. J. Singleton. Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure. *Journal of Financial Economics*, 63(3):415–441, March 2002.

Q. Dai and K. J. Singleton. Term Structure Dynamics in Theory and Reality. *Review of Financial Studies*, 16(3):631–678, July 2003.

S. Darolles, C. Gouriéroux, and J. Jasiak. Structural Laplace Transform and Compound Autoregressive Models. *Journal of Time Series Analysis*, 27(4):477–503, 2006.

F. De Jong. Time Series and Cross-Section Information in Affine Term-Structure Models. *Journal of Business & Economic Statistics*, 18(3):300–314, July 2000.

F. X. Diebold and G. D. Rudebusch. *Yield Curve Modeling and Forecasting*. Introductory Chapters. Princeton University Press, 2013.

H. Doshi, J. Ericsson, K. Jacobs, and S. M. Turnbull. Pricing Credit Default Swaps with Observable Covariates. *Review of Financial Studies*, 2013.

M. Draghi. Q&A: ECB President Mario Draghi, February 23, 2012. *Wall Street Journal*, 2012.

J. Driessen. Is Default Event Risk Priced in Corporate Bonds? *Review of Financial Studies*, 18(1):165–195, 2005.

J.-C. Duan and J.-G. Simonato. Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter. *Review of Quantitative Finance and Accounting*, 13(2): 111–35, September 1999.

S. Dubecq, A. Monfort, J.-P. Renne, and G. Roussellet. Credit and Liquidity in Interbank Rates: A Quadratic Approach. Banque de France Working Paper Series, 2014.

G. R. Duffee. Estimating the Price of Default Risk. *Review of Financial Studies*, 12(1): 197–226, 1999.

G. R. Duffee. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance*, 57(1):405–443, 02 2002.

G. R. Duffee. Bond Pricing and The Macroeconomy. *Hanbook of Economics and Finance (forthcoming)*, 2012.

G. R. Duffee and R. Stanton. Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation. *Journal of Financial Econometrics*, 6(1):108–142, 2008.

D. Duffie and R. Kan. A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6 (4):379–406, 1996.

D. Duffie and D. Lando. Term Structures of Credit Spreads with Incomplete Accounting Information. *Econometrica*, 69(3):633–664, May 2001.

D. Duffie and K. J. Singleton. An Econometric Model of the Term Structure of Interest-Rate Swap Yields. *Journal of Finance*, 52(4):1287–1321, September 1997.

D. Duffie and K. J. Singleton. Modeling Term Structures of Defaultable Bonds. *Review of Financial Studies*, 12(4):687–720, 1999.

D. Duffie, J. Pan, and K. Singleton. Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica*, 68(6):1343–1376, 2000.

J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press, 2012.

J. Eisenschmidt and J. Tapking. Liquidity Risk Premia in Unsecured Interbank Money Markets. Working Paper Series 1025, European Central Bank, March 2009.

E. J. Elton, M. J. Gruber, D. Agrawal, and C. Mann. Factors Affecting the Valuation of Corporate Bonds. *Journal of Banking & Finance*, 28(11):2747–2767, November 2004.

J. Ericsson and O. Renault. Liquidity and Credit Risk. *Journal of Finance*, 61(5):2219–2250, October 2006.

P. Feldhutter and D. Lando. Decomposing Swap Spreads. *Journal of Financial Economics*, 88(2):375–405, May 2008.

D. Filipovic. Separable Term Structures and the Maximal Degree Problem. *Mathematical Finance*, 12(4):341–349, 2002.

D. Filipovic and J. Teichmann. On Finite-Dimensional Term Structure Models. Working paper, Princeton University, 2002.

D. Filipovic and A. B. Trolle. The Term Structure of Interbank Risk. *Journal of Financial Economics*, 109(3):707 – 733, 2013.

D. Filipovic and J. Zabczyk. Markovian Term-Structure Models in Discrete Time. *Annals of Applied Probability*, 12(2):710–729, 2002.

D. Filipovic, M. Larsson, and A. Trolle. Linear-Rational Term Structure Models. Working paper, Swiss Finance Institute at EPFL, 2013.

A. R. Gallant and G. Tauchen. Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions. *Journal of the American Statistical Association*, 93:10–24, 1998.

A. Gelb. *Applied Optimal Estimation*. MIT Press, 1974.

I. Goldstein and A. Pauzner. Demand-Deposits Contracts and the Probability of Bank Runs. *Journal of Finance*, 60(3):1293–1327, 2005.

J.-M. Gonzales-Paramo. The ECB's monetary policy during the crisis. Closing speech at the tenth economic policy conference, malaga, European Central Bank, 2011.

V. Gorovoi and V. Linetsky. Black's Model of Interest Rates as Options, Eigenfunction Expansions and Japanese Interest Rates. *Mathematical Finance*, 14(1):49–78, 2004.

C. Gouriéroux and J. Jasiak. Autoregressive Gamma Processes. *Journal of Forecasting*, 25:129–152, 2006.

C. Gouriéroux and A. Monfort. Quadratic Stochastic Intensity and Prospective Mortality Tables. *Insurance: Mathematics and Economics*, 43(1):174–184, August 2008.

C. Gouriéroux and R. Sufana. Derivative Pricing with Wishart Multivariate Stochastic Volatility. *Journal of Business & Economic Statistics*, 28(3):438–451, 2010.

C. Gouriéroux and R. Sufana. Discrete-Time Wishart Term Structure Models. *Journal of Economic Dynamics and Control*, 35(6):815–824, June 2011.

C. Gouriéroux, A. Monfort, and A. Trognon. Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, 52(3):681–700, 1984.

C. Gouriéroux, A. Monfort, and V. Polimenis. Affine Models for Credit Risk Analysis. *Journal of Financial Econometrics*, 4(3):494–530, 2006.

C. Gouriéroux, J. Jasiak, and R. Sufana. The Wishart Autoregressive Process of Multivariate Stochastic Volatility. *Journal of Econometrics*, 150:167–181, 2009.

C. Gouriéroux, A. Monfort, and R. Sufana. International Money and Stock Market Contingent Claims. *Journal of International Money and Finance*, 29(8):1727–1751, December 2010.

C. Gouriéroux, A. Monfort, F. Pegoraro, and J.-P. Renne. Regime Switching and Bond Pricing. *Journal of Financial Econometrics*, 12(2):237–277, 2014.

R. S. Gurkaynak and J. H. Wright. Macroeconomics and the Term Structure. *Journal of Economic Literature*, 50(2):331–67, June 2012.

F. Gustafsson and G. Hendeby. Some Relations Between Extended and Unscented Kalman Filters. *IEEE Transactions on Signal Processing*, 60(2):545–555, 2012.

J. Gyntelberg and P. D. Wooldridge. Interbank Rate Fixings during the Recent Turmoil. Bis quaterly review, BIS, March 2008.

A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, November 1991.

D. Harville. *Matrix Algebra from a statistician's perspective*. Springer, 1997.

Z. He and W. Xiong. Rollover risk and credit risk. *Journal of Finance*, 67(2):391–429, 2012.

S. Holmes, G. Klein, and D. M. Murray. A Square Root Unscented Kalman Filter for Visual monoSLAM. Technical report, Active Vision Laboratory, Department of Engineering Science of Oxford, 2008.

P. Hordahl and O. Tristani. The Term Structure of Euro-Area Sovereign Bond Yields. Mimeo, 2012.

P. Hordahl and O. Tristani. Macro Factors and Sovereign Bond Spreads: A Quadratic No-Arbitrage Model. Technical report, 2013.

H. Ichiue and Y. Ueno. Equilibrium Yield Curve and the Yield Curve in Low Interest Rate Environment. Working paper series, Bank of Japan, 2007.

H. Ichiue and Y. Ueno. Estimating Term Premia at the Zero Bound: An Analysis of Japanese, U.S., and U.K. Yields. Working paper series, Bank of Japan, 2013.

A. C. Inci and B. Lu. Exchange Rates and Interest Rates: Can Term Structure Models Explain Currency Movements? *Journal of Economic Dynamics and Control*, 28(8): 1595–1624, June 2004.

K. Jacobs and L. Karoui. Conditional Volatility in Affine Term-Structure Models: Evidence from Treasury and Swap Markets. *Journal of Financial Economics*, 91:288–318, 2009.

C. Jardet, A. Monfort, and F. Pegoraro. No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth. *Journal of Banking & Finance*, 37 (2):389–402, 2013.

R. A. Jarrow and F. Yu. Counterparty Risk and the Pricing of Defaultable Securities. *Journal of Finance*, 56(5):1765–1799, October 2001.

R. A. Jarrow, D. Lando, and S. M. Turnbull. A Markov Model for the Term Structure of Credit Risk Spreads. *Review of Financial Studies*, 10(2):481–523, 1997.

R. A. Jarrow, H. Li, S. Liu, and C. Wu. Reduced-Form Valuation of Callable Corporate Bonds: Theory and Evidence. *Journal of Financial Economics*, 95(2):227–248, February 2010.

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. 1970.

X. Jin and J. M. Maheu. Modeling Realized Covariances and Returns. *Journal of Financial Econometrics*, 11(2):335–369, March 2013.

S. Joslin, K. J. Singleton, and H. Zhu. A New Perspective on Gaussian Dynamic Term Structure Models. *Review of Financial Studies*, 24(3):926–970, 2011.

M. Joyce, A. Lasaosa, I. Stevens, and M. Tong. The Financial Market Impact of Quantitative Easing in the United Kingdom. *International Journal of Central Banking*, 7(3): 113–161, September 2011.

S. J. Julier. The Scaled Unscented Transformation. In *Proceedings IEEE American Control Conference*, pages 4555–4559, 2002.

S. J. Julier. The Spherical Simplex Unscented Transformation. In *Proceedings of the American Control Conference*, 2003.

S. J. Julier and J. K. Uhlmann. Unscented Filtering and Nonlinear Estimation. In *Proceedings of the IEEE*, pages 401–422, 2004.

S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A New Approach for Filtering Nonlinear Systems. Technical report, Robotic Research Group, 2000.

R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

R. Kandepu, B. Foss, and L. Imsland. Applying the Unscented Kalman Filter for Nonlinear State Estimation. *Journal of Process Control*, 18(7):753–768, 2008.

D. Karaboga and B. Basturk. A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm. *Journal of Global Optimization*, 39(3):459–471, November 2007.

S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.

D. H. Kim. Zero Bound, Option-Implied PDFs, and Term Structure Models. Finance and Economics Discussion Series 2008-31, Board of Governors of the Federal Reserve System (U.S.), 2008.

D. H. Kim and A. Orphanides. Term Structure Estimation with Survey Data on Interest Rate Forecasts. *Journal of Financial and Quantitative Analysis*, 47(01):241–272, February 2012.

D. H. Kim and M. Priebsch. Estimation of Multi-Factor Shadow-Rate Term Structure Models. Federal reserve board discussion paper series, Federal Reserve Board, 2013.

D. H. Kim and K. J. Singleton. Term Structure Models and the Zero-Lower Bound: A Empirical Investigation of Japanese Yields. *Journal of Econometrics*, 170(1):32–49, September 2012.

S. Kozicki and P. Tinsley. Shifting Endpoints in the Term Structure of Interest Rates. *Journal of Monetary Economics*, 47:613–652, 2001a.

S. Kozicki and P. Tinsley. Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations. *Journal of Economic Dynamics and Control*, 25:149–184, 2001b.

L. Krippner. Modifying Gaussian Term Structure Models when Interest Rates are Near the Zero Lower Bound. Reserve Bank of New Zealand Discussion Paper Series DP2012/02, Reserve Bank of New Zealand, 2012.

L. Krippner. Measuring the Stance of Monetary Policy in Zero Lower Bound Environments. *Economics Letters*, 118(1):135–138, 2013.

D. Kuo, D. Skeie, and J. Vickery. A Comparison of LIBOR to Other Measures of Bank Borrowing Costs. Technical report, FRBNY Working Paper, 2012.

A. Kuritzkes, T. Schuermann, and S. Weiner. Deposit Insurance and Risk Management of the U.S. Banking System: What is the Loss Distribution Faced by the FDIC? *Journal of Financial Services Research*, 27(3):217–242, 2005.

A. Le, K. J. Singleton, and Q. Dai. Discrete-Time Affine-Q Term Structure Models with Generalized Market Prices of Risk. *Review of Financial Studies*, 23(5):2184–2227, 2010.

M. Leippold and L. Wu. Asset Pricing under the Quadratic Class. *Journal of Financial and Quantitative Analysis*, 37(02):271–295, June 2002.

M. Leippold and L. Wu. Design and Estimation of Quadratic Term Structure Models. Technical report, 2007.

H. Li and F. Zhao. Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives. *Journal of Finance*, 61(1):341–378, 02 2006.

J. Liu, F. A. Longstaff, and R. E. Mandell. The Market Price of Risk in Interest Rate Swaps: The Roles of Default and Liquidity Risks. *Journal of Business*, 79(5):2337–2360, September 2006.

F. Longstaff. A Nonlinear General Equilibrium Model of the Term Structure of Interest Rates. *Journal of Finance*, 23:1259–1282, 1989.

F. Longstaff. The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices. *Journal of Business*, 77(3):511–526, July 2004.

F. Longstaff and E. Schwartz. Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *Journal of Finance*, 47(4):1259–82, September 1992.

F. Longstaff, S. Mithal, and E. Neis. Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market. *Journal of Finance*, 60(5):2213–2253, October 2005.

F. Longstaff, J. Pan, L. Pedersen, and K. Singleton. How Sovereign is Sovereign Credit Risk? *Amercian Economic Journal: Macroeconomics*, 3:75–103, April 2011.

J. Lund. Non-Linear Kalman Filtering Techniques for Term-Structure Models. Technical report, 1997.

J. Magnus and H. Neudecker. *The Elimination Matrix: Some Lemmas and Applications.* Society for Industrial and Applied Mathematics. Journal on Algebraic and Discrete Methods, 1980.

J. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley, 1988.

P. S. Maybeck. *Stochastic Models, Estimation, and Control* , volume 2. 1982.

F.-L. Michaud and C. Upper. What Drives Interbank Rates? Evidence from the LIBOR Panel. *BIS Quarterly Review*, March 2008.

A. Monfort and J.-P. Renne. Default, Liquidity and Crises: An Econometric Framework. *Journal of Financial Econometrics*, 11(2):221–262, 2013.

A. Monfort and J.-P. Renne. Decomposing Euro-Area Sovereign Spreads: Credit and Liquidity Risks. *Review of Finance*, pages 1–49, 2014.

A. Monfort, J.-P. Renne, and G. Roussellet. A Quadratic Kalman Filter. *Journal of Econometrics (forthcoming).*

A. Monfort, F. Pegoraro, J.-P. Renne, and G. Roussellet. Staying at Zero with Affine Processes: A New Dynamic Term Structure Model. Technical report, 2014.

Moody's. Defaults and Recoveries for Financial Institution Debt Issuers, 1983-2010. Global credit policy, Moody's Investors Service, 2011.

Moody's. Annual Default Study: Corporate Default and Recovery Rates, 1920-2013. Global credit policy, Moody's Investors Service, 2013.

J. Pan and K. Singleton. Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads. *Journal of Finance*, 63(5):2345–84, 2008.

N. D. Pearson and T.-S. Sun. Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model. *Journal of Finance*, 49(4):1279–1304, September 1994.

M. Piazzesi. Affine Term Structure Models. In *Handbook of Financial Econometrics*, volume 1, chapter 12, pages 389–472. Yacine Ait-Sahalia and Lars Peter Hansen edition, 2010.

M. Priebsch. Computing Arbitrage-Free Yields in Multi-Factor Gaussian Shadow-Rate Term Structure Models. Federal reserve board discussion paper series, Federal Reserve Board, 2013.

M. Realdon. Quadratic Term Structure Models in Discrete Time. Working paper, 2006.

J.-P. Renne. A Model of the Euro-Area Yield Curve with Discrete Policy Rates. Working paper series 395, Banque de France, 2012.

W. Rinnergschwentner, G. Tappeiner, and J. Walde. Multivariate Stochastic Volatility via Wishart Processes: A Comment. *Journal of Business & Economic Statistics*, 30(1): 164–164, September 2011.

P. Schuster and M. Uhrig-Homburg. The Term Structure of Bond Market Liquidity Conditional on the Economic Environment: An Analysis of Government Guaranteed Bonds. Technical report, Karlruhe Institute of Technology, 2012.

K. Schwarz. Mind the Gap: Disentangling Credit and Liquidity in Risk Spreads. Technical report, Columbia University Graduate School of Business, 2009.

A. Siegel. The Noncentral Chi-Squared Distribution with Zero Degrees of Freedom and Testing for Uniformity. *Biometrika*, 66(2):381–386, August 1979.

J. Smith. The Term Structure of Money Market Spreads During the Financial Crisis. *Working paper*, 2010.

H. Tanizaki. *Nonlinear Filters: Estimation and Applications*. 1996.

J. B. Taylor and J. C. Williams. A Black Swan in the Money Market. *American Economic Journal: Macroeconomics*, 1(1):58–83, January 2009.

A. B. Trolle and E. S. Schwartz. A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives. *Review of Financial Studies*, 22(5):2007–2057, May 2009.

Y. Ueno, N. Baba, and Y. Sakurai. The Use of the Black Model of Interest Rates as an Option for Monitoring the JGB Market Expectations. Working paper series, Bank of Japan, 2006.

R. Van Der Merwe and E. A. Wan. The Square-Root Unscented Kalman Filter for State and Parameter-Estimation. In *in International Conference on Acoustics, Speech, and Signal Processing*, pages 3461–3464, 2001.

J. C. Wu and F. D. Xia. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. Working paper series, Chicago Booth School of Business, 2013.