Essays in financial econometrics and asset pricing

par

Kokouvi Tewou

Département de sciences économiques

Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en sciences économiques

Mars, 2020

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Essays in financial econometrics and asset pricing

présentée par :

Kokouvi Tewou

a été évaluée par un jury composé des personnes suivantes :

Marine Carrasco,	president du jury
Ilze Kalnina,	directrice de recherche
Bénoît Perron,	co-directeur de recherche
René Garcia,	membre du jury
Bruno Feunou,	examinateur externe
Ian MacDonald,	représentant du doyen de la FAS

Thèse acceptée le 27 Avril 2020

à ma femme Prisca, ma fille Zoey, mes parents Alphonse et Jeanne

Remerciements

Acknowledgments

Je remercie infiniment Ilze Kalnina, ma directrice de recherche pour sa très grande disponibilité ainsi que ses conseils et encouragements tout au long de cette thèse. J'ai beaucoup appris de sa rigueur dans la recherche et je suis très heureux d'avoir travaillé avec elle sur un sujet de recherche aussi passionnant.

J'aimerais aussi dire un grand merci a Christian Dorion, mon co-auteur auprès de qui j'ai beaucoup appris. Je le remercie aussi pour son soutien financier au cours de mon année de marché de travail. Je suis aussi reconnaissant à Benoit Perron et à Marine Carrasco qui ont bien voulu accepter de faire partir de mon comité de thèse depuis ma deuxième année de thèse. Leurs conseils et aides multiformes m'ont été très précieux. Je remercie aussi William McCausland pour son soutien durant mon année de marché de travail. Mes remerciements vont également aux enseignants, chercheurs et au personnel administratif du département des sciences économiques et du CIREQ. Cette thèse a été réalisée avec le soutien financier du département et du CIREQ.

Je tiens aussi à remercier mes amis et collègues avec qui j'ai partagé cette étape importante de mon parcours académique. Merci à Karim, Ada, Soumouaga, David, Jonathan, Louphou, Magnim, Felicien Marius, Hervé, Alex, Désiré, Doukhali ainsi qu'à tous les autres collègues du CIREQ et du département de sciences économiques. Je remercie également toutes les personnes dont j'ai fait la connaissance depuis le debut mon séjour à Montréal et qui m'ont beaucoup soutenu et aidé. Enfin j'ai une pensée pour mes parents, ma femme et mes frères et soeurs pour le soutien et les encouragements tout au long de mes études.

Résumé

Cette thèse est organisée en trois chapitres. Dans le premier chapitre, qui est co-écrit avec Ilze Kalnina, nous proposons un test statistique pour évaluer l'adéquation de la volatilité idiosyncratique comme mesure du risque idioyncratique. Nous proposons un test statistique qui est basé sur l'idée qu'un bon proxy du risque idiosyncratique devrait être non correélé à travers les actifs financiers. Nous démontrons que l'estimation de la volatililité est sujet à des erreurs qui rendent le test non standard. Nous proposons un modèle à facteurs qui permet de réduire sinon éliminer les corrélations dans la volatilité idiosyncratique, avec comme ultime but d' aboutir à un facteur qui satisfait mieux aux critères souhaités du risque idiosyncratique.

Dans le deuxième chapitre de ma thèse, qui est co-écrit avec Christian Dorion et Pierre Chaigneau, nous proposons une méthodologie pour étudier l'importance des risques d'ordres supérieurs dans la valorisation des actifs financiers. A la suite de Kraus and Litzenberger (1976) et Harvey and Siddique (2000a), beaucoup d'études ont analysé l'aversion aux risques de skewness et kurtosis de façon inconditionnelle. Dans ce chapitre, nous proposons une méthodogie qui permet de faire une analyse conditionnelle assez précise de l'aversion au risques d'ordres superieurs. Notre étude complémente la littérature dans la mesure ou nous étudions aussi la valuation des risques d'ordre plus élevé que la kurtosis à savoir l'hyperskewness et l'hyperkurtosis qui sont théoriquement valorisés dans certaines fonction d'utilité comme le CRRA.

Dans le dernier chapitre de ma thése, j'étudie la structure à terme de la prime de risque pour le risque de co-skewness, un risque qui mesure l'asymmétrie systématique dans les actions individuelles. Nous y proposons une méthode assez générale qui permet de faire une analyze mutli-horizon contrairement à la plupart des études existantes.

Mots-clés

Risque systématique, risque idiosyncratique, modele à facteurs, volatilité idiosyncratique, risques d'ordres supérieurs, structure à terme, co-skewness, options, prime de risque.

Abstract

This thesis is organized in three chapters. In the first chapter (which is co-authored with Ilze Kalnina), we propose a statistical test to assess the adequacy of the most popular measure of idiosyncratic risk, which is the idiosyncratic volatility. Our test statistic exploits the idea that a “good" measure of the idiosyncratic risk should be uncorrelated in the cross-section. Using in-fill asymptotics, we study the theoretical properties of the test and find that it has a non-standard behaviour due to various biases induced by the latency of the idiosyncratic volatility. Moreover, we propose a regression model that can be used to reduce if not eliminate the cross-sectional dependences in assets idiosyncratic volatilities.

The second chapter of my thesis is the fruit of a colaboration with Christian Dorion and Pierre Chaigneau. In this chapter, we study the relevance of higher-order risk aversion in asset pricing. The evidence in Kraus and Litzenberger (1976) and Harvey and Siddique (2000a) has spurred the literature on the estimation of the risk premiums attached to skewness and kurtosis risk in addition to the standard variance risk. However, most of these studies focus on the estimation of unconditional premiums or average premiums. In this chapter, we propose a methodology that allows to accurately estimate the time-varying higher-order risk aversions using options prices. Our study complements the literature as we also study the higher-order risks beyond the kurtosis such as hyperskewness and hyperkurtosis risks which are valued by a CRRA investor. .

In my third chapter, I study the term-structure of price of co-skewness risk. Co-Skewness risk captures the portion of the stock returns asymmetry that arises as a result of market returns asymmetry. I propose a general methodology that allows to study the multi-horizon pricing of this risk in contrast to many existing studies.

Keywords

Systematic risk, idiosyncratic risk, factor model, idiosyncratic volatility, higher-order risk, term structure, co-skewness, options, risk premium.

Table of contents

List of Tables

List of Figures

Chapter 1

Cross-Sectional Dependence in Idiosyncratic Volatility*

1.1 Introduction

In a panel of assets, returns are generally cross-sectionally dependent. This dependence is usually modelled using the exposure of assets to some common return factors, such as the Fama-French factors. In this return factor model (R-FM), the total volatility of an asset return can be decomposed into two parts: a component due to the exposure to the common return factors (the systematic volatility), and a residual component termed the idiosyncratic volatility (IdioVol). These two components of the volatility of returns are the most popular measures of the systematic risk and idiosyncratic risk of an asset.

Idiosyncratic volatility is important in economics and finance for several reasons. For example, when arbitrageurs exploit the mispricing of an individual asset, they are exposed to the idiosyncratic risk of the asset and not the systematic risk (see, e.g., Campbell, Lettau, Malkiel, and Xu (2001)).[1] Also, idiosyncratic volatility measures the exposure to the

*This chapter is co-authored with Ilze Kalnina. We benefited from discussions with Marine Carrasco, Yoosoon Chang, Valentina Corradi, Russell Davidson, Jean-Marie Dufour, Prosper Dovonon, Kirill Evdokimov, Sílvia Gonçalves, Peter Hansen, Jean Jacod, Dennis Kristensen, Joon Park, Benoit Perron, and Dacheng Xiu. We thank seminar participants at University of Amsterdam, Bank of Canada, Concordia, HEC Montreal, Indiana, LSE, McGill, NC State, Pennsylvania, Surrey, Toulouse, UCL, Warwick, Western Ontario, as well as participants of various conferences.

[1] A stock is said to be mispriced with respect to a given model if the expected value of the return on the stock is not consistent with the model.

idiosyncratic risk in imperfectly diversified portfolios. A recent observation is that the IdioVols seem to be strongly correlated in the cross-section of stocks.[2] Herskovic, Kelly, Lustig, and Nieuwerburgh (2016) argue this is due to a common IdioVol factor, which they relate to household risk. Moreover, cross-sectional dependence in IdioVols is important for option pricing, see Gourier (2016a).

This paper provides an econometric framework for studying the cross-sectional dependence in the idiosyncratic volatilities using high frequency data. We show that naive estimators, such as covariances and correlations of estimated IdioVols used by several empirical studies, are substantially biased. The bias arises due to the use of error-laden estimates of IdioVols. We provide the bias-corrected estimators.

To study idiosyncratic volatilities, we introduce the idiosyncratic volatility factor model (IdioVol-FM). Just like a return factor model, R-FM, such as the Fama-French model, decomposes returns into common and idiosyncratic returns, the IdioVol-FM decomposes the IdioVols into systematic and residual (non-systematic) components. The IdioVol factors may or may not be related to the return factors. The IdioVol factors can include the volatility of the return factors, or, more generally, (possibly non-linear) transformations of the spot covariance matrices of any observable variables, such as the average variance and average correlation factors of Chen and Petkova (2012). We propose bias-corrected estimators of the components of the IdioVol-FM model.

We provide the asymptotic theory for this model. For example, it allows us to test whether the residual (non-systematic) components of the IdioVols exhibit cross-sectional dependence. This allows us to identify the network of dependencies in the residual IdioVols across stocks.

Our bias-corrected estimators and inference results are an application of a new asymptotic theory that we develop for general estimators of quadratic covariation of vector-valued transformations of spot covariance matrices. This theoretical contribution is of its own interest. An example of alternative applications is the study of cross-sectional dependence of asset betas. Two features make the development of this asymptotic theory difficult. First, preliminary estimation of volatility results in first-order biases even in the special case of quadratic variation of the volatility one stock without any transformations, as in Vetter (2015). Second, we consider general nonlinear functionals in multivariate settings, which substantially

[2] See Connor, Korajczyk, and Linton (2006), Duarte, Kamara, Siegel, and Sun (2014), Herskovic, Kelly, Lustig, and Nieuwerburgh (2016), and Christoffersen, Fournier, and Jacobs (2018).

complicates the analysis.

Throughout the paper, we use factors that are specified by the researcher. An example of our Return Factor Model is the so-called Fama-French factor model, which has three observable factors, or the CAPM, which has one observable factor (the market portfolio return). An example of our IdioVol factors is the market volatility, which can be estimated from the market index. Thus, our setup is different from settings such as PCA where factors are identified from the cross-section of the assets studied. The treatment of the latter case adds an additional layer of complexity to the model and is beyond the scope of the current paper.

We apply our methodology to high-frequency data on the 30 Dow Jones Industrial Average components. We study the IdioVols with respect to two models for asset returns: the CAPM and the three-factor Fama-French model.[3] In both cases, the average pairwise correlation between the IdioVols is high (0.55). We verify that this dependence cannot be explained by the missing return factors. This confirms the recent findings of Herskovic, Kelly, Lustig, and Nieuwerburgh (2016) who use low frequency (daily and monthly) return data. We then consider the IdioVol-FM. We use two sets of IdioVol factors: the market volatility alone and the market volatility together with volatilities of nine industry ETFs. With the market volatility as the only IdioVol factor, the average pairwise correlation between residual (non-systematic) IdioVols is substantially lower (0.25) than between the total IdioVols. With the additional nine industry ETF volatilities as IdioVol factors, average correlation between the residual IdioVols decreases further (to 0.18). However, neither of the two sets of the IdioVol factors can fully explain the cross-sectional dependence in the IdioVols. We map out the network of dependencies in residual IdioVols across all stocks.

This paper analyzes cross-sectional dependence in idiosyncratic volatilities. This should *not* be confused with the analysis of cross-sectional dependence in total and idiosyncratic *returns*. A growing number of papers study the latter question using high frequency data. These date back to the analysis of realized covariances and their transformations, see, e.g., Barndorff-Nielsen and Shephard (2004) and Andersen, Bollerslev, Diebold, and Wu (2006). A continuous-time factor model for asset returns with observable return factors was first studied in Mykland and Zhang (2006). Various return factor models with observable factors have been studied by, among others, Bollerslev and Todorov (2010), Fan, Furger, and Xiu (2016), Li,

[3]The high frequency Fama-French factors are provided by Aït-Sahalia, Kalnina, and Xiu (2019).

Todorov, and Tauchen; Li, Todorov, and Tauchen (2017a; 2017b), and Aït-Sahalia, Kalnina, and Xiu (2019). Emerging literature also studies the cross-sectional dependence in returns using high-frequency data and latent return factors, see Ait-Sahalia and Xiu; Ait-Sahalia and Xiu (2019; 2017) and Pelger; Pelger (2019a; 2019b). Importantly, the models in the above papers are silent on the cross-sectional dependence structure in the IdioVols.

The Realized Beta GARCH model of Hansen, Lunde, and Voev (2014) imposes a structure on the cross-sectional dependence in IdioVols. This structure is tightly linked with the return factor model parameters, whereas our stochastic volatility framework allows separate specification of the return factors and the IdioVol factors.[4]

Our inference theory is related to several results in the existing literature. First, as mentioned above, we generalize the result of Vetter (2015). Jacod and Rosenbaum; Jacod and Rosenbaum (2013; 2015), Li, Todorov, and Tauchen (2016) and Li, Liu, and Xiu (2019) estimate integrated functionals of volatilities, which includes idiosyncratic volatilities. The latter problem is simpler than the problem of the current paper in the sense that $\sqrt{n}$-consistent estimation is possible, and no first-order bias terms due to preliminary estimation of volatilities arise. The need for a first-order bias correction due to preliminary estimation of volatility has also been observed in the literature on the estimation of the leverage effect, see Aït-Sahalia, Fan, and Li (2013), Aït-Sahalia, Fan, Laeven, Wang, and Yang (2017), Kalnina and Xiu (2017) and Wang and Mykland (2014). The biases due to preliminary estimation of volatility can be made theoretically negligible when an additional, long-span, asymptotic approximation is used. This requires the assumption that the frequency of observations is high enough compared to the time span, see, e.g., Corradi and Distaso (2006), Bandi and Renò (2012), Li and Patton (2018), and Kanaya and Kristensen (2016).

In the empirical section, we define a network of dependencies using (functions of) quadratic covariations of IdioVols. This approach can be compared with the network connectedness measures of Diebold and Yilmaz (2014). The latter measures are based on forecast error variance decompositions from vector autoregressions. They capture co-movements in forecast errors. In contrast, we assume a general semimartingale setting, and our framework captures realized co-movements in idiosyncratic volatilities, while accounting for the measurement errors in these volatilities.

[4]In the Beta GARCH model, the IdioVol of a stock is a product of its own (total) volatility, and one minus the square of the correlation between the stock return and the market return.

The remainder of the paper is organized as follows. Section 1.2 introduces the model and the quantities of interest. Section 3.4 describes the identification and estimation. Section 1.4 presents the asymptotic properties of our estimators. Section 1.6 uses high-frequency stock return data to study the cross-sectional dependence in IdioVols using our framework. Section 1.5 contains Monte Carlo simulations. The Appendix contains all proofs and additional figures.

1.2 Model and Quantities of Interest

We first describe a general factor model for the returns (R-FM), which allows us to define the idiosyncratic volatility. We then introduce the idiosyncratic volatility factor model (IdioVol-FM). In this framework, we proceed to define the cross-sectional measures of dependence between the total IdioVols, as well as the residual IdioVols, which take into account the dependence induced by the IdioVol factors.

We start by introducing some notation. Suppose we have (log) prices on d_S assets such as stocks and on d_F observable factors. We stack them into the d-dimensional process $Y_t = (S_{1,t}, \ldots, S_{d_S,t}, F_{1,t}, \ldots, F_{d_F,t})^\top$ where $d = d_S + d_F$. The observable factors $F_1, \ldots F_{d_F}$ are used in the R-FM model below. We assume that all observable variables jointly follow an Itô semimartingale, i.e., Y_t follows

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \tag{1.1}$$

where W is a d^W-dimensional Brownian motion ($d^W \geq d$), σ_s is a $d \times d^W$ stochastic volatility process, and J_t denotes a finite variation jump process. The reader can find the full list of assumptions in Section 1.4.1. We also assume that the spot covariance matrix process $C_t = \sigma_t \sigma_t^\top$ of Y_t is a continuous Itô semimartingale,[5]

$$C_t = C_0 + \int_0^t \widetilde{b}_s ds + \int_0^t \widetilde{\sigma}_s dW_s. \tag{1.2}$$

We denote $C_t = (C_{ab,t})_{1 \leq a,b \leq d}$. For convenience, we also use the alternative notation $C_{UV,t}$ to refer to the spot covariance between two elements U and V of Y, and $C_{U,t}$ to refer to $C_{UU,t}$.

We assume a standard continuous-time factor model for the asset returns.

[5]Note that assuming that Y and C are driven by the same d^W-dimensional Brownian motion W is without loss of generality provided that d' is large enough, see, e.g., equation (8.12) of Aït-Sahalia and Jacod (2014).

Definition (Factor Model for Returns, R-FM). *For all $0 \leq t \leq T$ and $j = 1, \ldots, d_S$,*[6]

$$\begin{aligned} dS_{j,t} &= \beta_{j,t}^{\top} dF_t^c + \tilde{\beta}_{j,t}^{\top} dF_t^d + dZ_{j,t} \quad \textit{with} \\ [Z_j, F]_t &= 0. \end{aligned} \tag{1.3}$$

In the above, $dZ_{j,t}$ is the idiosyncratic return of stock j. The superscripts c and d indicate the continuous and jump part of the processes, so that $\beta_{j,t}$ and $\tilde{\beta}_{j,t}$ are the continuous and jump factor loadings. For example, the k-th component of $\beta_{j,t}$ corresponds to the time-varying loading of the continuous part of the return on stock j to the continuous part of the return on the k-th factor. We set $\beta_t = (\beta_{1,t}, \ldots, \beta_{d_S,t})^{\top}$ and $Z_t = (Z_{1,t}, \ldots, Z_{d_S,t})^{\top}$.

We do not need the return factors F_t to be the same across assets to identify the model, but without loss of generality, we keep this structure as it is standard in empirical finance. These return factors are assumed to be observable, which is also standard. For example, in the empirical application, we use two sets of return factors: the market portfolio and the three Fama-French factors, which are constructed in Aït-Sahalia, Kalnina, and Xiu (2019).

A continuous-time factor model for returns with observable factors was originally studied in Mykland and Zhang (2006) in the case of one factor and in the absence of jumps. A burgeoning literature uses related models to study the cross-sectional dependence of total and/or idiosyncratic returns, see Section 1.1 for details. This literature does not consider the cross-sectional dependence in the IdioVols. Below, we use the R-FM to define the IdioVol, and proceed to study the cross-sectional dependence of IdioVols using the IdioVol Factor Model.

We define the idiosyncratic Volatility (IdioVol) as the spot volatility of the $Z_{j,t}$ process and denote it by C_{Zj}.Notice that the R-FM in (1.3) implies that the factor loadings β_t as well as IdioVol are functions of the total spot covariance matrix C_t. In particular, the vector of factor loadings satisfies

$$\beta_{jt} = (C_{F,t})^{-1} C_{FS_j,t}, \tag{1.4}$$

[6]The quadratic covariation of two vector-valued Itô semimartingales X and Y, over the time span $[0, T]$, is defined as

$$[X, Y]_T = p\text{-}\lim_{M \to \infty} \sum_{s=0}^{M-1} (X_{t_{s+1}} - X_{t_s})(Y_{t_{s+1}} - Y_{t_s})^{\top},$$

for any sequence $t_0 < t_1 < \ldots < t_M = T$ with $\sup_s \{t_{s+1} - t_s\} \to 0$ as $M \to \infty$, where p-lim stands for the probability limit.

for $j = 1, \ldots, d_S$, where $C_{F,t}$ denotes the spot covariance matrix of the factors F, which is the lower $d_F \times d_F$ sub-matrix of C_t; and $C_{FSj,t}$ denotes the covariance of the factors and the j^{th} stock, which is a vector consisting of the last d_F elements of the j^{th} column of C_t. The IdioVol of stock j is also a function of the total spot covariance matrix C_t,

$$\underbrace{C_{Zj,t}}_{\text{IdioVol of stock j}} = \underbrace{C_{Yj,t}}_{\text{total volatility of stock j}} - (C_{FSj,t})^\top (C_{F,t})^{-1} C_{FSj,t}. \tag{1.5}$$

By the Itô lemma, (1.4) and (1.5) imply that factor loadings and IdioVols are also Itô semimartingales with their characteristics related to those of C_t.

We now introduce the Idiosyncratic Volatility Factor model (IdioVol-FM). In the IdioVol-FM, the cross-sectional dependence in the IdioVol shocks can be potentially explained by certain IdioVol factors. We assume the IdioVol factors are known functions of the matrix C_t. In the empirical application, we use the market volatility as the IdioVol factor, which has been used in Herskovic, Kelly, Lustig, and Nieuwerburgh (2016) and Gourier (2016a); we discuss other possibilities below. We allow the IdioVol factors to be any known functions of C_t as long as they satisfy a certain polynomial growth condition in the sense of being in the class $\mathcal{G}(p)$ below,

$$\begin{aligned} \mathcal{G}(p) = \{H : H &\text{ is three-times continuously differentiable and for some } K > 0, \\ &\|\partial^j H(x)\| \le K(1 + \|x\|)^{p-j}, j = 0, 1, 2, 3\}, \quad \text{for some } \ p \ge 3. \end{aligned} \tag{1.6}$$

Definition (Idiosyncratic Volatility Factor Model, IdioVol-FM). *For all $0 \le t \le T$ and $j = 1, \ldots, d_S$, the idiosyncratic volatility C_{Zj} follows,*

$$\begin{aligned} dC_{Zj,t} &= \gamma_{Zj}^\top d\Pi_t + dC_{Zj,t}^{resid} \quad \textit{with} \\ [C_{Zj}^{resid}, \Pi]_t &= 0, \end{aligned} \tag{1.7}$$

where $\Pi_t = (\Pi_{1t}, \ldots, \Pi_{d_\Pi t})$ is a $\mathbb{R}^{d_\Pi}$-valued vector of IdioVol factors, which satisfy $\Pi_{kt} = \Pi_k(C_t)$ with the function $\Pi_k(\cdot)$ belonging to $\mathcal{G}(p)$ for $k = 1, \ldots, d_\Pi$.

We call the residual term $C_{Zj,t}^{resid}$ the residual IdioVol of asset j. Our assumptions imply that the components of the IdioVol-FM, $C_{Zj,t}, \Pi_t$ and $C_{Zj,t}^{resid}$, are continuous Itô semimartingales. We remark that both the dependent variable and the regressors in our IdioVol-FM are not directly observable and have to be estimated, and our asymptotic theory takes that into account. As will see in Section 3.4, this preliminary estimation implies that the naive estimators of all the dependence measures defined below are biased. One of the contributions of this paper is to quantify this bias and provide the bias-corrected estimators for all the quantities of interest.

The class of IdioVol factors permitted by our theory is rather wide as it includes general non-linear transforms of the spot covolatility process C_t. For example, IdioVol factors can be linear combinations of the total volatilities of stocks, see, e.g., the average variance factor of Chen and Petkova (2012). Other examples of IdioVol factors are linear combinations of the IdioVols, such as the equally-weighted average of the IdioVols, which Herskovic, Kelly, Lustig, and Nieuwerburgh (2016) denote by the "CIV". The IdioVol factors can also be the volatilities of any other observable processes.

Having specified our econometric framework, we now provide the definitions of some natural measures of dependence for (residual) IdioVols. Their estimation is discussed in Section 3.4.

Before considering the effect of IdioVol factors by using the IdioVol-FM decomposition, one may be interested in quantifying the dependence between the IdioVols of two stocks j and s. A natural measure of dependence is the quadratic-covariation based correlation between the two IdioVol processes,

$$Corr\left(C_{Zj}, C_{Zs}\right) = \frac{[C_{Zj}, C_{Zs}]_T}{\sqrt{[C_{Zj}, C_{Zj}]_T}\sqrt{[C_{Zs}, C_{Zs}]_T}}. \tag{1.8}$$

Alternatively, one may consider the quadratic covariation $[C_{Zj}, C_{Zs}]_T$ without any normalization. In Section 1.4.4, we use the estimator of the latter quantity to test for the presence of cross-sectional dependence in IdioVols.

To measure the residual cross-sectional dependence between the IdioVols of two stocks after accounting for the effect of the IdioVol factors, we use again the quadratic-covariation based correlation,

$$Corr\left(C_{Zj}^{resid}, C_{Zs}^{resid}\right) = \frac{[C_{Zj}^{resid}, C_{Zs}^{resid}]_T}{\sqrt{[C_{Zj}^{resid}, C_{Zj}^{resid}]_T}\sqrt{[C_{Zs}^{resid}, C_{Zs}^{resid}]_T}}. \tag{1.9}$$

In Section 1.4.4, we use the quadratic covariation between the two residual IdioVol processes $[C_{Zj}^{resid}, C_{Zs}^{resid}]_T$ without normalization for testing purposes.

We want to capture how well the IdioVol factors explain the time variation of IdioVol of the j^{th} asset. For this purpose, we use the quadratic-covariation based analog of the coefficient of determination. For $j = 1, \dots, d_S$,

$$R_{Zj}^{2,IdioVol\text{-}FM} = \frac{\gamma_{Zj}^\top [\Pi, \Pi]_T \gamma_{Zj}}{[C_{Zj}, C_{Zj}]_T}. \tag{1.10}$$

It is interesting to compare the correlation measure between IdioVols in equation (1.8) with the correlation between the residual parts of IdioVols in (1.9). We consider their difference,

$$Corr\left(C_{Zj}, C_{Zs}\right) - Corr\left(C_{Zj}^{resid}, C_{Zs}^{resid}\right) \tag{1.11}$$

to see how much of the dependence between IdioVols can be attributed to the IdioVol factors. In practice, if we compare assets that are known to have positive covolatilities (typically, stocks have that property), another useful measure of the common part in the overall covariation between IdioVols is the following quantity,

$$Q_{Zj,Zs}^{IdioVol\text{-}FM} = \frac{\gamma_{Zj}^\top [\Pi, \Pi]_T \gamma_{Zs}}{[C_{Zj}, C_{Zs}]_T}. \tag{1.12}$$

This measure is bounded by 1 if the covariations between residual IdioVols are nonnegative and smaller than the covariations between IdioVols, which is what we find for every pair in our empirical application with high-frequency observations on stock returns.

We remark that our framework can be compared with the following null hypothesis studied in Li, Todorov, and Tauchen (2016), $H_0: \; C_{Zj,t} = a_{Zj} + \gamma_{Zj}^\top \Pi_t, \; 0 \leq t \leq T$. This H_0 implies that the IdioVol is a deterministic function of the factors, which does not allow for an error term. In particular, this null hypothesis implies $R_{Zj}^{2,IdioVol\text{-}FM} = 1$.

1.3 Estimation

We now discuss the problem of the identification and estimation of the quantities of interest introduced in Section 1.2. We do so by showing that this problem is a special case of a more general problem, which is of its own interest, and solving the latter problem.

For the identification, the strategy is to show that each of the quantities of interest introduced in Section 1.2,

$$
\begin{aligned}
&[C_{Zj}, C_{Zs}]_T,\ Corr\left(C_{Zj}, C_{Zs}\right),\ \gamma_{Z_j},\ [C_{Zj}^{resid},\ C_{Zj}^{resid}]_T, \\
&Corr\left(C_{Zj}^{resid}, C_{Zs}^{resid}\right),\ Q_{Zj,Zs}^{IdioVol\text{-}FM},\ \text{and}\ R_{Zj}^{2,IdioVol\text{-}FM},
\end{aligned}
\tag{1.13}
$$

for $j, s = 1, \ldots, d_S$, can be written as

$$
\varphi\left([H_1(C), G_1(C)]_T, \ldots, [H_\kappa(C), G_\kappa(C)]_T\right),
\tag{1.14}
$$

where φ as well as H_r and G_r, for $r = 1, \ldots, \kappa$, are known real-valued functions. Each element in (1.14) is of the form $[H(C), G(C)]_T$, i.e., it is a quadratic covariation between functions of C_t. $[H(C), G(C)]_T$ is observable from continuous-record observations on Y in (1.1), which means it can be estimated from (discrete) high-frequency observations on Y.

While the identification is relatively simple, the estimation problem has to address the biases due to preliminary estimation of (idiosyncratic) volatility. To this end, we introduce two estimators of $[H(C), G(C)]_T$. Section 1.4 derives the joint asymptotic distribution of several objects of this type, $[H_r(C), G_r(C)]_T$ for $r = 1, \ldots, \kappa$. The asymptotic distribution of the general estimand in (1.14), and hence of every quantity of interest in equation (1.13), follows by the Delta method.

We start by discussing the identification of the first estimand in (1.13), which is the quadratic covariation between j^{th} and s^{th} IdioVol, $[C_{Zj}, C_{Zs}]_T$. It can be written as $[H(C), G(C)]_T$ if we choose $H(C_t) = C_{Zj,t}$ and $G(C_t) = C_{Zs,t}$. By (1.5), both $C_{Zj,t}$ and $C_{Zs,t}$ are functions of C_t. Next, consider $Corr\left(C_{Zj}, C_{Zs}\right)$ defined in (1.8). By the same argument, its numerator and each of the two components in the denominator can be written as $[H(C), G(C)]_T$ for different functions H and G. Therefore, $Corr\left(C_{Zj}, C_{Zs}\right)$ is itself a known function of three objects of the form $[H(C), G(C)]_T$.

To show that the remaining quantities in (1.13) can also be expressed in terms of objects of the form $[H(C), G(C)]_T$, note that the IdioVol-FM implies

$$
\gamma_{Zj} = \left([\Pi, \Pi]_T\right)^{-1}[\Pi, C_{Zj}]_T \quad \text{and} \quad [C_{Zj}^{resid}, C_{Zs}^{resid}]_T = [C_{Zj}, C_{Zs}]_T - \gamma_{Zj}^{\top}[\Pi, \Pi]_T\gamma_{Zs},
$$

for $j, s = 1, \ldots, d_S$. Since $C_{Zj,t}$, $C_{Zs,t}$ and every element in Π_t are real-valued functions of C_t, the above equalities imply that all quantities of interest in (1.13) can be written as real-valued, known functions of a finite number of quantities of the form $[H(C), G(C)]_T$.

We now turn to the estimation of $[H(C), G(C)]_T$. Suppose we have discrete observations on Y_t over an interval $[0, T]$. Denote by Δ_n the distance between observations. It is well known that we can estimate the spot covariance matrix C_t at time $(i-1)\Delta_n$ with a local truncated realized volatility estimator,

$$\widehat{C}_{i\Delta_n} = \frac{1}{k_n \Delta_n} \sum_{m=0}^{k_n - 1} (\Delta_{i+m}^n Y)(\Delta_{i+m}^n Y)^\top 1_{\{\|\Delta_{i+m}^n Y\| \le \chi \Delta_n^{\varpi}\}}, \tag{1.15}$$

where $\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n}$ and where k_n is the number of observations in a local window.[7] Throughout the paper we set $\widehat{C}_{i\Delta_n} = (\widehat{C}_{ab,i\Delta_n})_{1 \le a,b \le d}$.

We propose two estimators for the general quantity $[H(C), G(C)]_T$. The first is a bias-corrected analog of the definition of quadratic covariation between two Itô processes,

$$\begin{aligned}
\widehat{[H(C), G(C)]}_T^{AN} = \frac{3}{2k_n} \sum_{i=1}^{[T/\Delta_n] - 2k_n + 1} \Bigg(\Big(H(\widehat{C}_{(i+k_n)\Delta_n}) - H(\widehat{C}_{i\Delta_n}) \Big) \Big(G(\widehat{C}_{(i+k_n)\Delta_n}) - G(\widehat{C}_{i\Delta_n}) \Big) \\
- \frac{2}{k_n} \sum_{g,h,a,b=1}^{d} (\partial_{gh} H \partial_{ab} G)(\widehat{C}_{i\Delta_n}) \Big(\widehat{C}_{ga,i\Delta_n} \widehat{C}_{gb,i\Delta_n} + \widehat{C}_{gb,i\Delta_n} \widehat{C}_{ha,i\Delta_n} \Big) \Bigg),
\end{aligned} \tag{1.16}$$

where the factor $3/2$ and the last term correct for the biases arising due to the preliminary estimation of volatility C_t.

Our second estimator is based on the following equality, which follows by the Itô lemma,

$$[H(C), G(C)]_T = \sum_{g,h,a,b=1}^{d} \int_0^T (\partial_{gh} H \partial_{ab} G)(C_t) \overline{C}_t^{gh,ab} dt, \tag{1.17}$$

where $\overline{C}_t^{gh,ab}$ denotes the covariation between the volatility processes $C_{gh,t}$ and $C_{ab,t}$. The quantity is thus a non-linear functional of the spot covariance and spot volatility of volatility matrices. Our second estimator is a bias-corrected version of the sample counterpart of the

[7] It is also possible to define more flexible kernel-based estimators as in Kristensen (2010).

"linearized" expression in (1.17),

$$[\widehat{H(C),G(C)}]_T^{LIN} = \frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} (\partial_{gh} H \partial_{ab} G)(\widehat{C}_{i\Delta_n}) \times$$
$$\Big((\widehat{C}_{gh,(i+k_n)\Delta_n} - \widehat{C}_{gh,i\Delta_n})(\widehat{C}_{ab,(i+k_n)\Delta_n} - \widehat{C}_{ab,i\Delta_n}) - \frac{2}{k_n}(\widehat{C}_{ga,i\Delta_n}\widehat{C}_{gb,i\Delta_n} + \widehat{C}_{gb,i\Delta_n}\widehat{C}_{ha,i\Delta_n}) \Big). \tag{1.18}$$

Consistency for a similar estimator has been established by Jacod and Rosenbaum (2015).[8] We go beyond their result by deriving the asymptotic distribution and proposing a consistent estimator of its asymptotic variance.

If we had observations on $C_{i\Delta_n}$, the estimators of $[H(C),G(C)]_T$ would not need any bias-correction terms. However, due to the replacement of $C_{i\Delta_n}$ by its estimate $\widehat{C}_{i\Delta_n}$, two types of bias-correction terms arise: a multiplicative correction $3/2$, as well as an additive bias-correction term

$$-\frac{3}{k_n^2} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \left(\sum_{g,h,a,b=1}^{d} (\partial_{gh} H \partial_{ab} G)(\widehat{C}_{i\Delta_n}) \Big(\widehat{C}_{ga,i\Delta_n}\widehat{C}_{gb,i\Delta_n} + \widehat{C}_{gb,i\Delta_n}\widehat{C}_{ha,i\Delta_n} \Big) \right). \tag{1.19}$$

We remark that this additive bias correction term is (up to a scale factor) an estimator of the asymptotic covariance between the sampling errors embedded in estimators of $\int_0^T H(C_t)dt$ and $\int_0^T G(C_t)dt$ defined in Jacod and Rosenbaum (2013).

The two estimators are identical when H and G are linear, for example, when estimating the covariation between two volatility processes. In the univariate case $d = 1$, when $H(C) = G(C) = C$, our estimator coincides with the volatility of volatility estimator of Vetter (2015), which was extended to allow for jumps in Jacod and Rosenbaum (2015). Our contribution is the extension of this theory to the multivariate $d > 1$ case with nonlinear functionals.

[8] Jacod and Rosenbaum (2015) derive the probability limit of the following estimator:

$$\frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} (\partial^2_{gh,ab} H)(\widehat{C}_{i\Delta_n}) \Big((\widehat{C}_{(i+k_n)\Delta_n} - \widehat{C}_{i\Delta_n})(\widehat{C}_{(i+k_n)\Delta_n} - \widehat{C}_{i\Delta_n}) - \frac{2}{k_n}(\widehat{C}_{ga,i\Delta_n}\widehat{C}_{gb,i\Delta_n} + \widehat{C}_{gb,i\Delta_n}\widehat{C}_{ha,i\Delta_n}) \Big).$$

1.4 Asymptotic Properties

In this section, we first present the full list of assumptions for our asymptotic results. We then state the asymptotic distribution for the general functionals introduced in the previous section, and develop estimators for the asymptotic variance. Finally, to illustrate the application of the general theory, we describe three statistical tests about the IdioVols, which we later implement in the empirical and Monte Carlo analysis.

1.4.1 Assumptions

Recall that the d-dimensional process Y_t represents the (log) prices of stocks, S_t, and factors F_t.

Assumption 1. *Suppose Y is an Itô semimartingale on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$,*

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_E \delta(s,z)\mu(ds,dz),$$

where W is a d^W-dimensional Brownian motion ($d^W \geq d$) and μ is a Poisson random measure on $\mathbb{R}_+ \times E$, with E an auxiliary Polish space with intensity measure $\nu(dt,dz) = dt \otimes \lambda(dz)$ for some σ-finite measure λ on E. The process b_t is $\mathbb{R}^d$-valued optional, σ_t is $\mathbb{R}^d \times \mathbb{R}^{d^W}$-valued, and $\delta = \delta(w,t,z)$ is a predictable $\mathbb{R}^d$ -valued function on $\Omega \times \mathbb{R}_+ \times E$. Moreover, $\|\delta(w, t\wedge\tau_m(w), z)\| \wedge 1 \leq \Gamma_m(z)$, for all (w,t,z), where (τ_m) is a localizing sequence of stopping times and, for some $r \in [0,1]$, the function Γ_m on E satisfies $\int_E \Gamma_m(z)^r \lambda(dz) < \infty$. The spot volatility matrix of Y is then defined as $C_t = \sigma_t \sigma_t^\top$. We assume that C_t is a continuous Itô semimartingale,[9]

$$C_t = C_0 + \int_0^t \widetilde{b}_s ds + \int_0^t \widetilde{\sigma}_s dW_s. \tag{1.20}$$

where $\widetilde{b}$ is $\mathbb{R}^d \times \mathbb{R}^d$-valued optional.

With the above notation, the elements of the spot volatility of volatility matrix and spot

[9] Note that $\widetilde{\sigma}_s = (\widetilde{\sigma}_s^{gh,m})$ is $(d \times d \times d^W)$-dimensional and $\widetilde{\sigma}_s dW_s$ is $(d \times d)$-dimensional with $(\widetilde{\sigma}_s dW_s)^{gh} = \sum_{m=1}^{d^W} \widetilde{\sigma}_s^{gh,m} dW_s^m$.

covariation of the continuous martingale parts of X and c are defined as follows,

$$\overline{C}_t^{gh,ab} = \sum_{m=1}^{d^W} \widetilde{\sigma}_t^{gh,m} \widetilde{\sigma}_t^{ab,m}, \quad \overline{C}_t^{\prime g,ab} = \sum_{m=1}^{d^W} \sigma_t^{gm} \widetilde{\sigma}_t^{ab,m}. \tag{1.21}$$

We assume the following for the process $\widetilde{\sigma}_t$:

Assumption 2. *$\widetilde{\sigma}_t$ is a continuous Itô semimartingale with its characteristics satisfying the same requirements as that of C_t.*

Assumption 1 is very general and nests most of the multivariate continuous-time models used in economics and finance. It allows for potential stochastic volatility and jumps in returns. Assumption 2 is required to obtain the asymptotic distribution of estimators of the quadratic covariation between functionals of the spot covariance matrix C_t. It is not needed to prove consistency. This assumption also appears in Vetter (2015), Kalnina and Xiu (2017) and Wang and Mykland (2014).

1.4.2 Asymptotic Distribution

We have seen in Section 3.4 that all quantities of interest in (1.13) are functions of multiple objects of the form $[H(C), G(C)]_T$. Therefore, if we can obtain a multivariate asymptotic distribution for a vector with elements of the form $[H(C), G(C)]_T$, the asymptotic distributions for all our estimators follow by the delta method. The current section presents this asymptotic distribution.

Let $H_1, G_1, \ldots, H_\kappa, G_\kappa$ be some arbitrary elements of $\mathcal{G}(p)$ defined in equation (1.6). We are interested in the asymptotic behavior of vectors

$$\left([H_1(\widehat{C), G_1}(C)]_T^{AN}, \ldots, [H_\kappa(\widehat{C), G_\kappa}(C)]_T^{AN}\right)^\top \text{ and}$$
$$\left([H_1(\widehat{C), G_1}(C)]_T^{LIN}, \ldots, [H_\kappa(\widehat{C), G_\kappa}(C)]_T^{LIN}\right)^\top.$$

The following theorem summarizes the joint asymptotic behavior of the estimators.

Theorem 1. *Let $[H_r(\widehat{C), G_r}(C)]_T$ be either $[H_r(\widehat{C), G_r}(C)]_T^{AN}$ or $[H_r(\widehat{C), G_r}(C)]_T^{LIN}$ defined in (1.16) and (1.18), respectively. Suppose Assumption 1 and Assumption 2 hold. Fix $k_n = \theta \Delta_n^{-1/2}$ for some $\theta \in (0, \infty)$ and set $(8p-1)/4(4p-r) \leq \varpi < \frac{1}{2}$. Then, as $\Delta_n \to 0$,*

$$\Delta_n^{-1/4}\begin{pmatrix}[H_1(\widehat{C), G}_1(C)]_T - [H_1(C), G_1(C)]_T \\ \dots \\ [H_\kappa(\widehat{C), G}_\kappa(C)]_T - [H_\kappa(C), G_\kappa(C)]_T\end{pmatrix} \xrightarrow{L\text{-}s} MN(0, \Sigma_T), \tag{1.22}$$

where $\Sigma_T = \left(\Sigma_T^{r,s}\right)_{1\leq r,s\leq\kappa}$ denotes the asymptotic covariance between the estimators $[H_r(\widehat{C), G}_r(C)]_T$ and $[H_s(\widehat{C), G}_s(C)]_T$. The elements of the matrix Σ_T are

$$\begin{aligned}
\Sigma_T^{r,s} &= \Sigma_T^{r,s,(1)} + \Sigma_T^{r,s,(2)} + \Sigma_T^{r,s,(3)},\\
\Sigma_T^{r,s,(1)} &= \frac{6}{\theta^3}\sum_{g,h,a,b=1}^{d}\sum_{j,k,l,m=1}^{d}\int_0^T \left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_s)\right)\Big[C_t(gh,jk)C_t(ab,lm)\\
&\qquad + C_t(ab,jk)C_t(gh,lm)\Big]dt,\\
\Sigma_T^{r,s,(2)} &= \frac{151\theta}{140}\sum_{g,h,a,b=1}^{d}\sum_{j,k,l,m=1}^{d}\int_0^T \left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_t)\right)\Big[\overline{C}_t^{gh,jk}\overline{C}_t^{ab,lm}\\
&\qquad + \overline{C}_t^{ab,jk}\overline{C}_t^{gh,lm}\Big]dt,\\
\Sigma_T^{r,s,(3)} &= \frac{3}{2\theta}\sum_{g,h,a,b=1}^{d}\sum_{j,k,l,m=1}^{d}\int_0^T \left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_t)\right)\Big[C_t(gh,jk)\overline{C}_t^{ab,lm}\\
&\qquad + C_t(ab,lm)\overline{C}_t^{gh,jk} + C_t(gh,lm)\overline{C}_t^{ab,jk} + C_t(ab,jk)\overline{C}_t^{gh,lm}\Big]dt,
\end{aligned}$$

with

$$C_t(gh,jk) = C_{gj,t}C_{hk,t} + C_{gk,t}C_{hj,t}.$$

The convergence in Theorem 1 is stable in law (denoted L-s, see for example Aldous and Eagleson (1978) and Jacod and Protter (2012)). The limit is mixed gaussian and the precision of the estimators depends on the paths of the spot covariance and the volatility of volatility process. The rate of convergence $\Delta_n^{-1/4}$ has been shown to be the optimal for volatility of volatility estimation (under the assumption of no volatility jumps).

The asymptotic variance of the estimators depends on the tuning parameter θ whose choice may be crucial for the reliability of the inference. We document the sensitivity of the inference theory to the choice of the parameter θ in a Monte Carlo experiment (see Section

1.5).

1.4.3 Estimation of the Asymptotic Covariance Matrix

To provide a consistent estimator for the element $\Sigma_T^{r,s}$ of the asymptotic covariance matrix in Theorem 1, we introduce the following quantities:

$$\widehat{\Omega}_T^{r,s,(1)} = \Delta_n \sum_{g,h,a,b=1}^{d} \sum_{j,k,l,m=1}^{d} \sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s(\widehat{C}_{i\Delta_n}))$$
$$\times \Big[\widetilde{C}_{i\Delta_n}(gh, jk)\widetilde{C}_{i\Delta_n}(ab, lm) + \widetilde{C}_{i\Delta_n}(ab, jk)\widetilde{C}_{i\Delta_n}(gh, lm)\Big],$$

$$\widehat{\Omega}_T^{r,s,(2)} = \sum_{g,h,a,b=1}^{d} \sum_{j,k,l,m=1}^{d} \sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s(\widehat{C}_{i\Delta_n})) \Big[\frac{1}{2}\widehat{\lambda}_i^{n,gh}\widehat{\lambda}_i^{n,jk}\widehat{\lambda}_{i+2k_n}^{n,ab}\widehat{\lambda}_{i+2k_n}^{n,lm}$$
$$+ \frac{1}{2}\widehat{\lambda}_i^{n,ab}\widehat{\lambda}_i^{n,lm}\widehat{\lambda}_{i+2k_n}^{n,gh}\widehat{\lambda}_{i+2k_n}^{n,jk} + \frac{1}{2}\widehat{\lambda}_i^{n,ab}\widehat{\lambda}_i^{n,jk}\widehat{\lambda}_{i+2k_n}^{n,gh}\widehat{\lambda}_{i+2k_n}^{n,lm} + \frac{1}{2}\widehat{\lambda}_i^{n,gh}\widehat{\lambda}_i^{n,lm}\widehat{\lambda}_{i+2k_n}^{n,ab}\widehat{\lambda}_{i+2k_n}^{n,jk}\Big],$$

$$\widehat{\Omega}_T^{r,s,(3)} = \frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{j,k,l,m=1}^{d} \sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s(\widehat{C}_{i\Delta_n}))$$
$$\Big[\widetilde{C}_{i\Delta_n}(gh, jk)\widehat{\lambda}_i^{n,ab}\widehat{\lambda}_i^{n,lm} + \widetilde{C}_{i\Delta_n}(ab, lm)\widehat{\lambda}_i^{n,gh}\widehat{\lambda}_i^{n,jk}$$
$$+ \widetilde{C}_{i\Delta_n}(gh, lm)\widehat{\lambda}_i^{n,ab}\widehat{\lambda}_i^{n,jk} + (\widetilde{C}_{i\Delta_n}(ab, jk)\widehat{\lambda}_i^{n,gh}\widehat{\lambda}_i^{n,lm}\Big],$$

with $\widehat{\lambda}_i^{n,jk} = \widehat{C}_{i+k_n}^{n,jk} - \widehat{C}_i^{n,jk}$ and $\widetilde{C}_{i\Delta_n}(gh, jk) = (\widehat{C}_{gj,i\Delta_n}\widehat{C}_{hk,i\Delta_n} + \widehat{C}_{gk,i\Delta_n}\widehat{C}_{hj,i\Delta_n})$.
The following result holds,

Theorem 2. *Suppose the assumptions of Theorem 1 hold, then, as $\Delta_n \longrightarrow 0$*

$$\frac{6}{\theta^3}\widehat{\Omega}_T^{r,s,(1)} \xrightarrow{\mathbb{P}} \Sigma_T^{r,s,(1)} \tag{1.23}$$
$$\frac{3}{2\theta}[\widehat{\Omega}_T^{r,s,(3)} - \frac{6}{\theta}\widehat{\Omega}_T^{r,s,(1)}] \xrightarrow{\mathbb{P}} \Sigma_T^{r,s,(3)} \tag{1.24}$$
$$\frac{151\theta}{140}\frac{9}{4\theta^2}[\widehat{\Omega}_T^{r,s,(2)} + \frac{4}{\theta^2}\widehat{\Omega}_T^{r,s,(1)} - \frac{4}{3}\widehat{\Omega}_T^{r,s,(3)}] \xrightarrow{\mathbb{P}} \Sigma_T^{r,s,(2)}. \tag{1.25}$$

The estimated matrix $\widehat{\Sigma}_T$ is symmetric but is not guaranteed to be positive semi-definite. By Theorem 1, $\widehat{\Sigma}_T$ is positive semi-definite in large samples. An interesting question is the estimation of the asymptotic variance using subsampling or bootstrap methods, and we leave it for future research.

Remark 1: Results of Jacod and Rosenbaum (2015) and a straightforward extension of

Theorem 1 can be used to show that the rate of convergence in equation (1.23) is $\Delta_n^{-1/2}$, and the rate of convergence in (1.25) is $\Delta_n^{-1/4}$. The rate of convergence in (1.24) can be shown to be $\Delta_n^{-1/4}$.

Remark 2: In the one-dimensional case ($d = 1$), much simpler estimators of $\Sigma_T^{r,s,(2)}$ can be constructed using the quantities $\widehat{\lambda}_i^{n,jk}\widehat{\lambda}_i^{n,lm}\widehat{\lambda}_{i+k_n}^{n,gh}\widehat{\lambda}_{i+k_n}^{n,xy}$ or $\widehat{\lambda}_i^{n,jk}\widehat{\lambda}_i^{n,lm}\widehat{\lambda}_i^{n,gh}\widehat{\lambda}_i^{n,xy}$ as in Vetter (2015). However, in the multidimensional case, the latter quantities do not identify separately the quantity $\overline{C_t}^{jk,lm}\overline{C_t}^{gh,xy}$ since the combination $\overline{C_t}^{jk,lm}\overline{C_t}^{gh,xy} + \overline{C_t}^{jk,gh}\overline{C_t}^{lm,xy} + \overline{C_t}^{jk,xy}\overline{C_t}^{gh,lm}$ shows up in a non-trivial way in the limit of the estimator.

Corollary 3. *For $1 \le r \le \kappa$, let $[H_r(\widehat{C), G}_r(C)]_T$ be either $[H_r(\widehat{C), G}_r(C)]_T^{AN}$ or $[H_r(\widehat{C), G}_r(C)]_T^{LIN}$ defined in (1.18) and (1.16), respectively. Suppose the assumptions of theorem 1 hold. Then,*

$$\Delta_n^{-1/4}\widehat{\Sigma}_T^{-1/2}\begin{pmatrix} [H_1(\widehat{C), G}_1(C)]_T - [H_1(C), G_1(C)]_T \\ \vdots \\ [H_\kappa(\widehat{C), G}_\kappa(C)]_T - [H_\kappa(C), G_\kappa(C)]_T \end{pmatrix} \xrightarrow{L} N(0, I_\kappa). \tag{1.26}$$

In the above, we use the notation L to denote the convergence in distribution and I_κ the identity matrix of order κ. Corollary 3 states the standardized asymptotic distribution, which follows directly from the properties of stable-in-law convergence. Similarly, by the delta method, standardized asymptotic distribution can also be derived for the estimators of the quantities in (1.13). These standardized distributions allow the construction of confidence intervals for all the latent quantities of the form $[H_r(C), G_r(C)]_T$ and, more generally, functions of these quantities.

1.4.4 Tests

As an illustration of application of the general theory, we provide three tests about the dependence of idiosyncratic volatility. Our framework allows to test general hypotheses about the joint dynamics of any subset of the available stocks. The three examples below are stated for one pair of stocks, and correspond to the tests we implement in the empirical and Monte Carlo studies.

First, one can test for the absence of dependence between the IdioVols of the returns on assets j and s,

$$H_0^1 : [C_{Zj}, C_{Zs}]_T = 0 \text{ vs } H_1^1 : [C_{Zj}, C_{Zs}]_T \neq 0.$$

The null hypothesis H_0^1 is rejected whenever the t-test exceeds the $1 - \alpha/2$-quantile of the standard normal distribution, $Z_{1-\alpha/2}$,

$$\Delta_n^{-1/4} \frac{\left| \widehat{[C_{Zj}, C_{Zs}]_T} \right|}{\sqrt{\widehat{\text{AVAR}}\left(C_{Zj}, C_{Zs}\right)}} > Z_{1-\alpha/2}.$$

Second, we can test for all the IdioVol factors Π being irrelevant to explain the dynamics of IdioVol shocks of stock j,

$$H_0^2 : [C_{Zj}, \Pi]_T = 0 \text{ vs } H_1^2 : [C_{Zj}, \Pi]_T \neq 0.$$

Under this null hypothesis, the vector of IdioVol factor loadings equals zero, $\gamma_{Z_j} = 0$. The null hypothesis H_0^2 is rejected when

$$\Delta_n^{-1/4} \left(\widehat{[C_{Zj}, \Pi]_T} \right)^\top \left(\widehat{\text{AVAR}}\left(C_{Zj}, \Pi\right) \right)^{-1} \widehat{[C_{Zj}, \Pi]_T} > \mathcal{X}^2_{d_\Pi, 1-\alpha}, \tag{1.27}$$

where d_Π denotes the number of IdioVol factors, and where $\mathcal{X}^2_{d_\Pi, 1-\alpha}$ is the $(1 - \alpha)$ quantile of the $\mathcal{X}^2_{d_\Pi}$ distribution. One can of course also construct a t-test for irrelevance of any one particular IdioVol factor. The final example is a test for absence of dependence between the residual IdioVols,

$$H_0^3 : [C_{Zj}^{resid}, C_{Zs}^{resid}]_T = 0 \text{ vs } H_1^3 : [C_{Zj}^{resid}, C_{Zs}^{resid}]_T \neq 0.$$

The null can be rejected when the following t-test exceeds the critical value,

$$\Delta_n^{-1/4} \frac{\left| \widehat{[C_{Zj}^{resid}, C_{Zs}^{resid}]_T} \right|}{\sqrt{\widehat{\text{AVAR}}\left(C_{Zj}^{resid}, C_{Zs}^{resid}\right)}} > Z_{\alpha/2}. \tag{1.28}$$

Each of the above estimators

$$[\widehat{C_{Zj}, C_{Zs}}]_T, [\widehat{C_{Zj}, \Pi}]_T, \text{ and } [\widehat{C_{Zj}^{resid}, C_{Zs}^{resid}}]_T$$

can be obtained by choosing appropriate pair(s) of transformations H and G in the general estimator $[H(\widehat{C), G}(C)]_T$, see Section 3.4 for details. Any of the two types of the latter estimator can be used,

$$[H(\widehat{C), G}(C)]_T^{AN} \text{ or } [H(\widehat{C), G}(C)]_T^{LIN}.$$

For the first two tests, the expression for the true asymptotic variance, AVAR, is obtained using Theorem 1 and its estimation follows from Theorem 2. The asymptotic variance in the third test is obtained by applying the delta method to the joint convergence result in Theorem 1. The expression for the estimator of the asymptotic variance, $\widehat{\text{AVAR}}$, follows from Theorem 2. Under R-FM and the assumptions of Theorem 1, Corollary 3 implies that the asymptotic size of the two types of tests for the null hypotheses H_0^1 and H_0^2 is α, and their power approaches 1. The same properties apply for the tests of the null hypotheses H_0^3 with our R-FM and IdioVol-FM representations.

Theoretically, it is possible to test for absence of dependence in the IdioVols at each point in time. In this case the null hypothesis is $H_0^{1\prime} : [C_{Zj}, C_{Zs}]_t = 0 \text{ for all } 0 \leq t \leq T$, which is, in theory, stronger than our $H_0^{1\prime}$. In particular, Theorem 1 can be used to set up Kolmogorov-Smirnov type of tests for $H_0^{\prime 1}$ in the same spirit as Vetter (2015). However, we do not pursue this direction in the current paper for two reasons. First, the testing procedure would be more involved. Second, empirical evidence suggests nonnegative dependence between IdioVols, which means that in practice, it is not too restrictive to assume $[C_{Zj}, C_{Zs}]_t \geq 0\ \forall t$, under which H_0^1 and $H_0^{1\prime}$ are equivalent.

1.5 Monte Carlo

This section investigates the finite sample properties of our estimators and tests. The data generating process (DGP) is similar to that of Li, Todorov, and Tauchen (2013) and is constructed as follows. Denote by Y_1 and Y_2 log-prices of two individual stocks, and by X the log-price of the market portfolio. Recall that the superscript c indicates the continuous part of a process. We assume

$$dX_t = dX_t^c + dJ_{3,t}, \quad dX_t^c = \sqrt{C_{X,t}}dW_t,$$

and, for $j = 1, 2$,

$$dY_{j,t} = \beta_t dX_t^c + d\widetilde{Y}_{j,t}^c + dJ_{j,t}, \quad d\widetilde{Y}_{j,t}^c = \sqrt{C_{Zj,t}}d\widetilde{W}_{j,t}.$$

In the above, C_X is the spot volatility of the market portfolio, $\widetilde{W}_1$, and $\widetilde{W}_2$ are Brownian motions with $\mathrm{Corr}(d\widetilde{W}_{1,t}, d\widetilde{W}_{2,t}) = 0.4$, and W is an independent Brownian motion; J_1, J_2, and J_3 are independent compound Poisson processes with intensity equal to 2 jumps per year and jump size distribution $N(0, 0.02^2)$. The beta process is time-varying and is specified as $\beta_t = 0.5 + 0.1\sin(100t)$.

We next specify the volatility processes. As our building blocks, we first generate four processes $f_1, \ldots, f_4$ as mutually independent Cox-Ingersoll-Ross processes,

$$\begin{aligned} df_{1,t} &= 5(0.09 - f_{1,t})dt + 0.35\sqrt{f_{1,t}}\Big(-0.8dW_t + \sqrt{1-0.8^2}dB_{1,t}\Big), \\ df_{j,t} &= 5(0.09 - f_{j,t})dt + 0.35\sqrt{f_{1,t}}dB_{j,t} \quad , \text{ for } j = 2, 3, 4, \end{aligned}$$

where $B_1, \ldots, B_4$ and independent standard Brownian Motions, which are also independent from the Brownian Motions of the return Factor Model.[10] We use the first process f_1 as the market volatility, i.e., $C_{X,t} = f_{1,t}$. We use the other three processes f_2, f_3, and f_4 to construct three different specifications for the IdioVol processes $C_{Z1,t}$ and $C_{Z2,t}$, see Table 1.4 for details. The common Brownian Motion W_t in the market portfolio price process X_t and its volatility process $C_{X,t} = f_{1,t}$ generates a leverage effect for the market portfolio. The value of the leverage effect is -0.8, which is standard in the literature, see Kalnina and Xiu (2017), Aït-Sahalia, Fan, and Li (2013) and Aït-Sahalia, Fan, Laeven, Wang, and Yang (2017).

We set the time span T equal 1,260 or 2,520 days, which correspond approximately to 5 and 10 business years. These values are standard in the nonparametric leverage effect estimation literature (see Aït-Sahalia, Fan, and Li (2013) and Kalnina and Xiu (2017)), where the rate of convergence is also $\Delta^{-1/4}$. Each day consists of 6.5 trading hours. We consider two different values for the sampling frequency, $\Delta_n = 1$ minute and $\Delta_n = 5$ minutes. We follow

[10] The Feller property is satisfied implying the positiveness of the processes $(f_{j,t})_{1\le j\le 4}$.

Li, Todorov, and Tauchen (2016) and set the truncation threshold u_n in day t at $3\widehat{\sigma}_t\Delta_n^{0.49}$, where $\widehat{\sigma}_t$ is the squared root of the annualized bipower variation of Barndorff-Nielsen and Shephard (2004). We use 10,000 Monte Carlo replications in all the experiments.

We first investigate the finite sample properties of the estimators under Model 3. The considered estimands include:

- the IdioVol factor loading of the first stock, γ_{Z1},
- the contribution of the market volatility to the variation of the IdioVol of the first stock $R_{Z1}^{2,IdioVol\text{-}FM}$,
- the correlation between the idiosyncratic volatilities of stocks 1 and 2, $Corr\,(C_{Zj}, C_{Zs})$,
- the correlation between residual idiosyncratic volatilities, $Corr\left(C_{Zj}^{resid}, C_{Zs}^{resid}\right)$.

The interpretation of simulation results is much simpler when the quantities of interest do not change across simulations. To achieve that, we generate once and keep fixed the paths of the processes $C_{X,t}$ and $(f_{j,t})_{0\le j\le 27}$ and replicate several times the other parts of the DGP. In Table 1.6, we report the bias and the interquartile range (IQR) of the two type of estimators for each quantity using 5 minutes data sampled over 10 years. We choose four different values for the width of the subsamples, which corresponds to $\theta = 1.5, 2, 2.5$ and 3 (recall that the number of observations in a window is $k_n = \theta/\sqrt{\Delta_n}$). It seems that larger values of the parameters produce better results. Next, we investigate how these results change when we increase the sampling frequency. In Table 1.7, we report the results with $\Delta_n = 1$ minute in the same setting. We note a reduction of the bias and IQR at all levels of significance. However, the magnitude of the decrease of the IQR is very small. Finally, we conduct the same experiment using data sampled at one minute over 5 years, see Table 1.8. Despite using more than twice as many observations than in the first experiment, the precision is not as good. In other words, increasing the time span is more effective for precision gain than increasing the sampling frequency. This result is typical for $\Delta_n^{1/4}$-convergent estimators, see, e.g., Kalnina and Xiu (2017).

Next, we study the empirical rejection probabilities of the three statistical tests as outlined in Section 1.4.4. The first null hypothesis is the absence of dependence between the IdioVols (for which we use Model 1), $H_0^1 : [C_{Z1}, C_{Z2}]_T = 0$. The second null hypothesis we test is the absence of dependence between the IdioVol of the first stock and the market volatility (for which we use Model 1), $H_0^2 : [C_{Z1}, C_X]_T = 0$. The third null hypothesis is the absence of dependence in the two residual IdioVols (for which we use Model 2), $H_0^3 : [C^{resid}_{Z1Z1}, C^{resid}_{Z2Z2}]_T = 0$.

The three panels of Table 1.5 contain the empirical rejection probabilities for the three null hypotheses. We present the results for two sampling frequencies ($\Delta_n = 1$ minute and $\Delta_n = 5$ minutes) and the two type of estimators (AN and LIN). We see that the empirical rejection probabilities are reasonably close to the nominal size of the test. Neither type of estimator (AN or LIN) seems to dominate the other. Consistent with the asymptotic theory, the empirical rejection probabilities of the three tests become closer to the nominal size of the test when frequency is higher.

1.6 Empirical Analysis

We apply our methods to study the cross-sectional dependence in IdioVol using high frequency data. One of our main findings is that stocks' idiosyncratic volatilities co-move strongly with the market volatility. This is a quite surprising finding. It is of course well known that the total volatility of stocks moves with the market volatility. However, we stress that we find that the strong effect is still present when considering the idiosyncratic volatilities.

We use full record transaction prices from NYSE TAQ database for 30 constituents of the DJIA index over the time period 2003-2012, see Table 1.1. After removing the non-trading days, our sample contains 2517 days. The selected stocks were the constituents of the DJIA index in 2007. We also use the high-frequency data on nine industry Exchange-Traded Funds, ETFs (Consumer Discretionary, Consumer Staples, Energy, Financial, Health Care, Industrial, Materials, Technology, and Utilities), and the high-frequency size and value Fama-French factors, see Aït-Sahalia, Kalnina, and Xiu (2019). For each day, we consider data from the regular exchange opening hours from time stamped between 9:30 a.m. until 4 p.m. We clean the data following the procedure suggested by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008), remove the overnight returns and then sample at 5 minutes. This sparse sampling has been widely used in the literature because the effect of the microstructure noise and potential asynchronicity of the data is less important at this frequency, see also Liu, Patton, and Sheppard (2015).

The parameter choices for the estimators are as follows. Guided by our Monte Carlo results, we set the length of window to be approximately one week for the estimators in Section 3.4 (this corresponds to $\theta = 2.5$ where $k_n = \theta \Delta_n^{-1/2}$ is the number of observations in a window). The truncation threshold for all estimators is set as in the Monte Carlo study ($3\widehat{\sigma}_t \Delta_n^{0.49}$ where $\widehat{\sigma}_t^2$ is the bipower variation).

Figures ?? and ?? contain plots of the time series of the estimated R^2_{Yj} of the return factor model (R-FM) for each stock.[11] Each plot contains monthly R^2_{Yj} from two return factor models, CAPM and the Fama-French regression with market, size, and value factors. Figures ?? and ?? show that these time series of all stocks follow approximately the same trend with a considerable increase in the contribution around the crisis year 2008. Higher R^2_{Yj} indicates that the systematic risk is relatively more important, which is typical during crises. R^2_{Yj} is consistently higher in the Fama-French regression model compared to the CAPM regression model, albeit not by much. We proceed to investigate the dynamic properties of the panel of idiosyncratic volatilities.

We first investigate the dependence in the (total) idiosyncratic volatilities. Our panel has 435 pairs of stocks. For each pair of stocks, we compute the correlation between the IdioVols, $Corr\left(C_{Zi}, C_{Zj}\right)$. All pairwise correlations are positive in our sample, and their average is 0.55. Figure 1.1 maps the network of dependency in the IdioVol. We simultaneously test 435 hypotheses of no correlation, and Figure 1.1 connects only the assets, for which the null is rejected. We account for multiple testing by controlling the false discovery rate at 5%. Overall, Figure 1.1 shows that the cross-sectional dependence between the IdioVols is very strong.

Could missing factors in the R-FM provide an explanation? Omitted return factors in the R-FM are captured by the idiosyncratic returns, and can therefore induce correlation between the estimated IdioVols, provided these missing return factors have non-negligible volatility of volatility. To investigate this possibility, we consider the correlations between idiosyncratic

[11] For the j^{th} stock, our analog of the coefficient of determination in the R-FM is $R^2_{Yj} = 1 - \frac{\int_0^T c_{Zj,t}dt}{\int_0^T c_{Yj,t}dt}$. We estimate R^2_{Yj} using the general method of Jacod and Rosenbaum (2013). The resulting estimator of R^2_{Yj} requires a choice of a block size for the spot volatility estimation; we choose two hours in practice (the number of observations in a block, say l_n, has to satisfy $l_n^2 \Delta_n \to 0$ and $l_n^3 \Delta_n \to \infty$, so it is of smaller order than the number of observations k_n in our estimators of Section 3.4).

returns, $\text{Corr}(Z_i, Z_j)$.[12] Table 1.2 presents a summary of how estimates $\text{Corr}(Z_i, Z_j)$ are related to the estimates of correlation in IdioVols, $Corr\left(C_{Zi}, C_{Zj}\right)$. In particular, different rows in Table 1.2 display average values of $\widehat{Corr}\left(C_{Zi}, C_{Zj}\right)$ among those pairs, for which $\widehat{Corr}(Z_i, Z_j)$ is below some threshold. For example, the last-but-one row in Table 1.2 indicates that there are 56 pairs of stocks with $\widehat{Corr}(Z_i, Z_j) < 0.01$, and among those stocks, the average correlation between IdioVols, $Corr\left(C_{Zi}, C_{Zj}\right)$, is estimated to be 0.579. We observe that $\widehat{Corr}\left(C_{Zi}, C_{Zj}\right)$ is virtually the same compared to pairs of stocks with high $\text{Corr}(Z_i, Z_j)$. These results suggest that missing return factors cannot explain dependence in IdioVols for all considered stocks. This finding is in line with the empirical analysis of Herskovic, Kelly, Lustig, and Nieuwerburgh (2016) with daily and monthly returns.

To understand the source of the strong cross-sectional dependence in the IdioVols, we consider the Idiosyncratic Volatility Factor Model (IdioVol-FM) of Section 1.2. We first use the market volatility as the only IdioVol factor.[13] Table 1.3 reports the estimates of the IdioVol loading ($\widehat{\gamma}_{Zi}$) and the R^2 of the IdioVol-FM ($R_{Zi}^{2,IdioVol\text{-}FM}$, see equation (1.10)). Table 1.3 uses two different definitions of IdioVol, one defined with respect to CAPM, and a second IdioVol defined with respect to Fama-French three factor model. For every stock, the estimated IdioVol factor loading is positive, suggesting that the idiosyncratic volatility co-moves with the market volatility. Next, Figure 1.2 shows the implications for the cross-section of the one-factor IdioVol-FM when the IdioVol is defined with respect to CAPM. The average pairwise correlations between the residual IdioVols, $\widehat{Corr}\left(C_{Zi}, C_{Zj}\right)$, decrease to 0.25. However, the market volatility cannot explain all cross-sectional dependence in residual IdioVols, as evidenced by the remaining links in Figure 1.2.

Finally, we consider an IdioVol-FM with ten IdioVol factors, market volatility and the volatilities of nine industry ETFs. Figure 1.3 shows the implications for the cross-section of this ten-factor IdioVol-FM when the IdioVol is defined with respect to CAPM. The average

[12] Our measure of correlation between the idiosyncratic returns dZ_i and dZ_j is

$$\text{Corr}(Z_i, Z_j) = \frac{\int_0^T C_{ZiZj,t}dt}{\sqrt{\int_0^T C_{Zi,t}dt}\sqrt{\int_0^T C_{Zj,t}dt}}, \quad i, j = 1, \ldots, d_S, \tag{1.29}$$

where $C_{ZiZj,t}$ is the spot covariation between Z_i and Z_j. Similarly to R^2_{Yj}, we estimate $\text{Corr}(Z_i, Z_j)$ using the method of Jacod and Rosenbaum (2013).

[13] We also considered the volatility of size and value Fama-French factors. However, both these factors turned out to have very low volatility of volatility and therefore did not significantly change the results.

pairwise correlations between the residual IdioVols, $\widehat{Corr}\,(C_{Zi}, C_{Zj})$, decrease further to 0.18. However, significant dependence between the residual IdioVols remains, as evidenced by the remaining links in Figure 1.2. Our results suggest that there is room for considering the construction of additional IdioVol factors based on economic theory, for example, along the lines of the heterogeneous agents model of Herskovic, Kelly, Lustig, and Nieuwerburgh (2016).

1.7 Conclusion

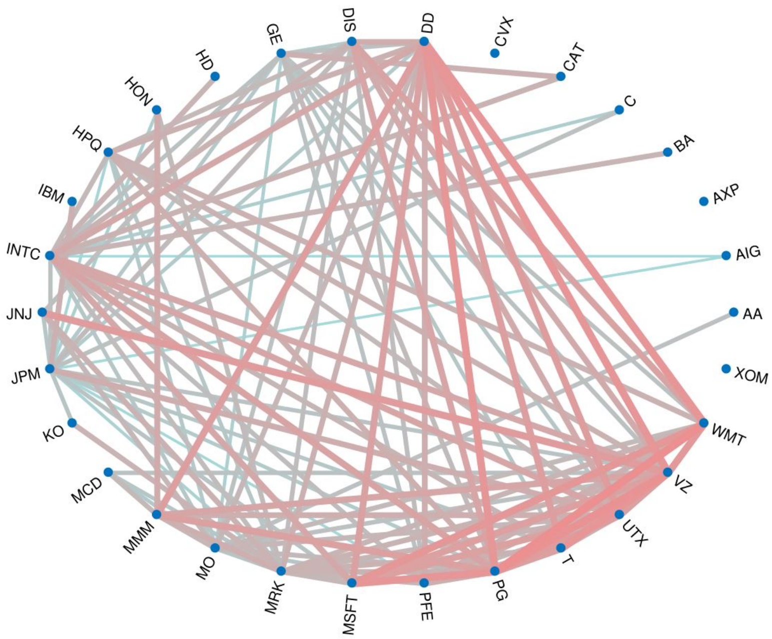
We introduce an econometric framework for analysis of cross-sectional dependence in the IdioVols of assets using high frequency data. First, we provide bias-corrected estimators of standard measures of dependence between IdioVols, as well as the associated asymptotic theory. Second, we study an IdioVol factor model, in which we decompose the variation in IdioVols into two parts: the variation related to the systematic factors such as the market volatility, and the residual variation. We provide the asymptotic theory that allows us to test, for example, whether the residual (non-systematic) components of the IdioVols exhibit cross-sectional dependence.

To provide the bias-corrected estimators and inference results, we develop a new asymptotic theory for general estimators of quadratic covariation of vector-valued (possibly) nonlinear transformations of the spot covariance matrices. This theoretical contribution is of its own interest, and can be applied in other contexts. For example, our results can be used to conduct inference for the cross-sectional dependence in asset betas.

We apply our methodology to the 30 Dow Jones Industrial Average components, and document strong cross-sectional dependence in their idiosyncratic volatilities. We consider two different sets of idiosyncratic volatility factors, and find that neither can fully account for the cross-sectional dependence in idiosyncratic volatilities. We map out the network of dependencies in residual (non-systematic) idiosyncratic volatilities across the stocks.

Figures and Tables of Chapter 1

Figure 1.1: The network of dependencies in total IdioVols.



The color and thickness of each line is proportional to the estimated value of $Corr\left(C_{Zi}, C_{Zj}\right)$, the quadratic-covariation based correlation between the IdioVols, defined in equation (1.8) (red and thick lines indicate high correlation). We simultaneously test 435 null hypotheses of no correlation, and the lines are only plotted when the null is rejected.

Figure 1.2: The network of dependencies in residual IdioVols with a single IdioVol factor: the market variance.

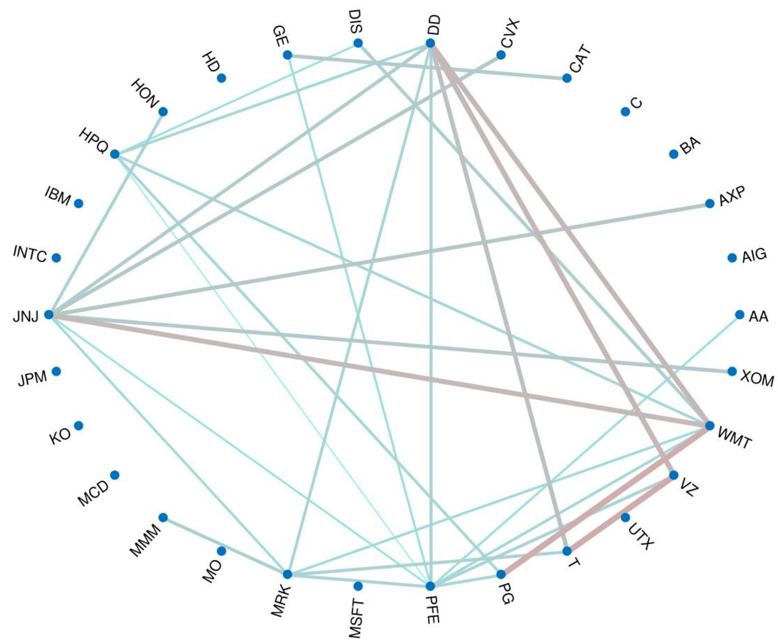
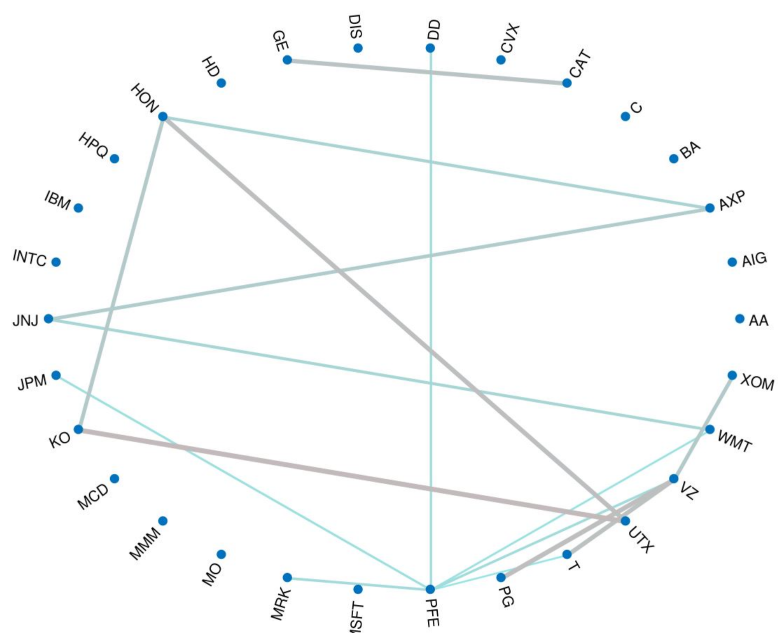
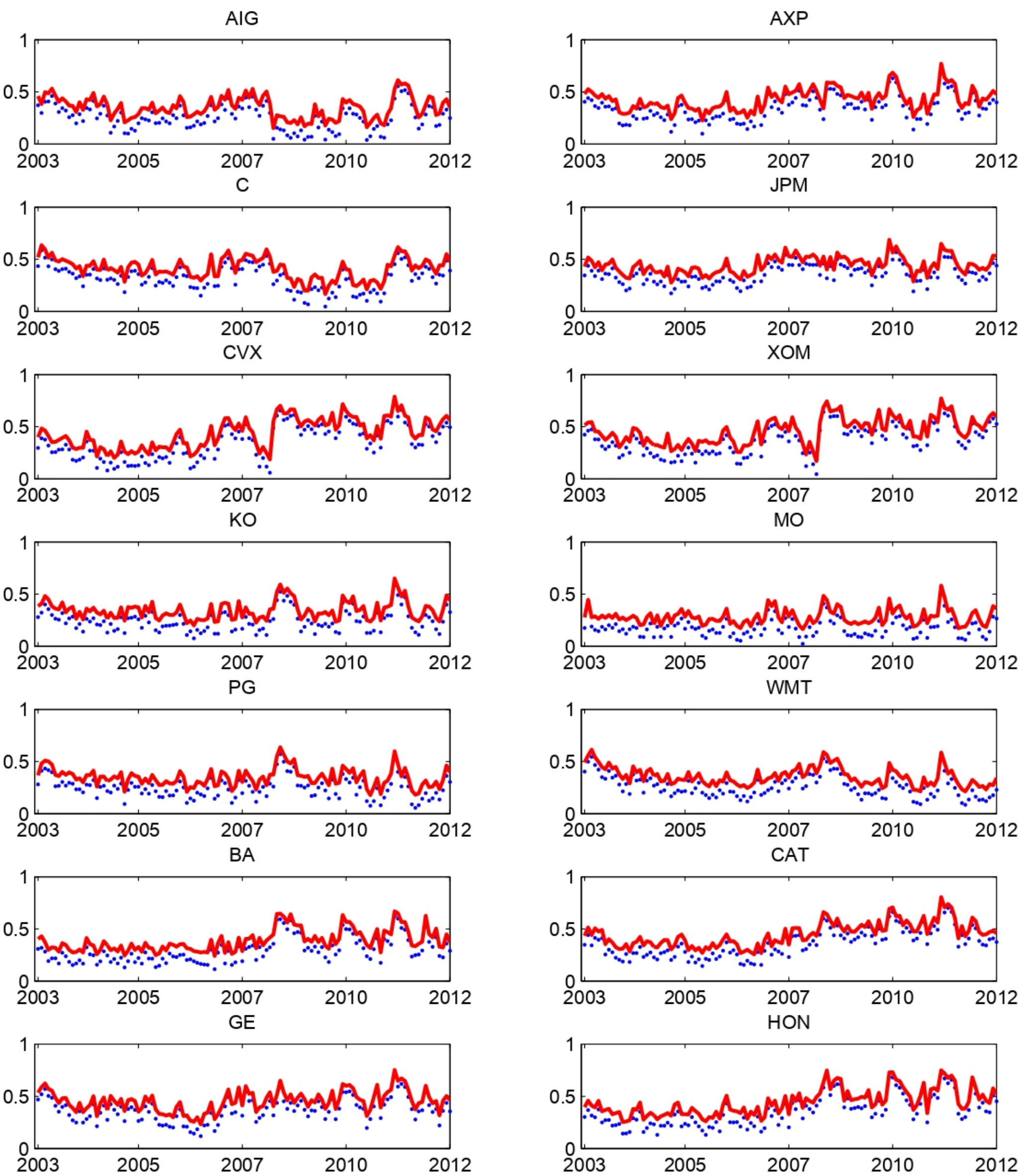


Figure 1.3: The network of dependencies in residual IdioVols with ten IdioVol factors: the market variance and the variances of nine industry ETFs.



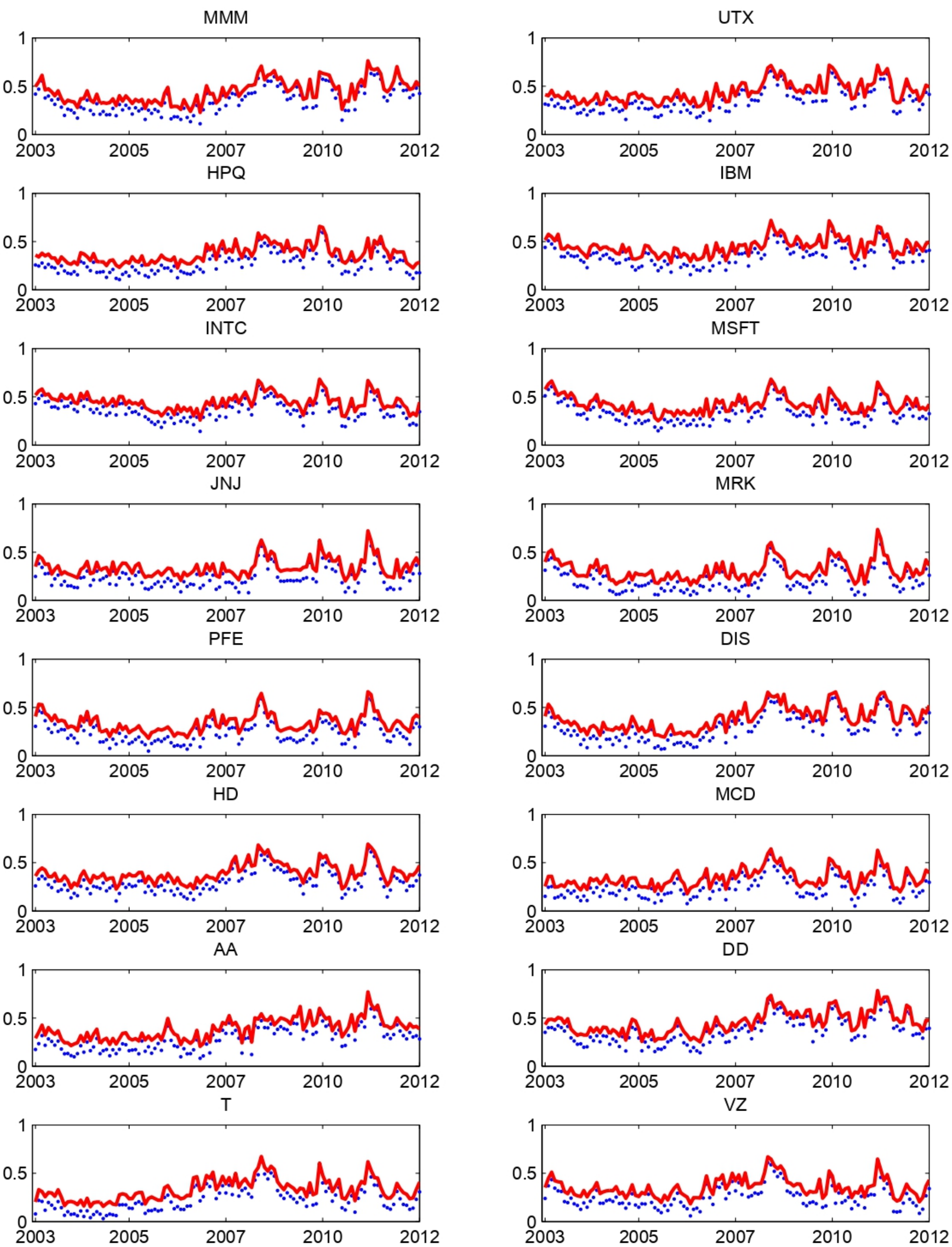
In both figures, the color and thickness of each line is proportional to the estimated value of $Corr\left(C_{Zi}^{resid}, C_{Zj}^{resid}\right)$, the quadratic-covariation based correlation between the IdioVols, defined in (1.9), of each pair of stocks (red and thick lines indicate high correlation). We simultaneously test 435 null hypotheses of no correlation, and the lines are only plotted when the null is rejected.

Figure 1.4: Monthly R^2 of two return factor models.



We plot monthly R^2 of two return factor models ($\widehat{R}^2_{Yj}$): the CAPM (the blue dotted line) and the Fama-French three factor model (the red solid line). Stocks are represented by tickers (see Table 1.1 for full stock names).

Figure 1.5: Monthly R^2 of two return factor models (continued).



We plot monthly R^2 of two return factor models ($\widehat{R}^2_{Yj}$): the CAPM (the blue dotted line) and the Fama-French three factor model (the red solid line). Stocks are represented by tickers (see Table 1.1 for full stock names).

Table 1.1: List of DJIA stocks

Sector	Stock	Ticker
Financial	American International Group, Inc.	AIG
	American Express Company	AXP
	Citigroup Inc.	C
	JPMorgan Chase & Co.	JPM
Energy	Chevron Corp.	CVX
	Exxon Mobil Corp.	XOM
Consumer Staples	Coca Cola Company	KO
	Altria	MO
	The Procter & Gamble Company	PG
	Wal-Mart Stores	WMT
Industrials	Boeing Company	BA
	Caterpillar Inc.	CAT
	General Electric Company	GE
	Honeywell International Inc	HON
	3M Company	MMM
	United Technologies	UTX
Technology	Hewlett-Packard Company	HPQ
	International Bus. Machines	IBM
	Intel Corp.	INTC
	Microsoft Corporation	MSFT
Health Care	Johnson & Johnson	JNJ
	Merck & Co.	MRK
	Pfizer Inc.	PFE
Consumer Discretionary	The Walt Disney Company	DIS
	Home Depot Inc	HD
	McDonald's Corporation	MCD
Materials	Alcoa Inc.	AA
	E.I. du Pont de Nemours & Company	DD
Telecommunications Services	AT&T Inc.	T
	Verizon Communications Inc.	VZ

The table lists the stocks used in the empirical application (for the time period 2003-2012). They are the 30 constituents of DJIA in 2007. The first column provides the Global Industry Classification Standard (GICS) sectors, the second the names of the companies and the third their tickers.

Table 1.2: Idiosyncratic returns correlations

	CAPM			FF3 Model		
$\lvert\widehat{\text{Corr}}(Z_i, Z_j)\rvert$	Pairs	Avg $\lvert\widehat{\text{Corr}}(Z_i, Z_j)\rvert$	Avg $\widehat{Corr}\,(C_{Zi}, C_{Zj})$	Pairs	Avg $\lvert\widehat{\text{Corr}}(Z_i, Z_j)\rvert$	Avg $\widehat{Corr}\,(C_{Zi}, C_{Zj})$
< 0.6	435	0.038	0.510	435	0.038	0.512
< 0.5	434	0.036	0.509	434	0.037	0.512
< 0.4	434	0.036	0.509	434	0.037	0.512
< 0.3	434	0.036	0.509	434	0.037	0.512
< 0.2	431	0.035	0.508	430	0.035	0.511
< 0.1	403	0.028	0.503	404	0.029	0.506
< 0.075	383	0.025	0.500	382	0.026	0.502
< 0.050	315	0.018	0.487	316	0.019	0.489
< 0.025	177	0.006	0.447	178	0.007	0.452
< 0.010	80	0.001	0.415	81	0.002	0.414
< 0.005	43	0.000	0.385	41	0.001	0.409

Each row in this table describes the subset of pairs of stocks with $\lvert\widehat{Corr(Z_i, Z_j)}\rvert$ below a threshold in column one. The table considers two R-FMs: the left panel defines the IdioVol with respect to CAPM, and the right panel defines the IdioVol with respect to the three-factor Fama-French model. In both cases, the market volatility is the only IdioVol factor. Each panel reports three quantities for the given subset of pairs: the number of pairs, average absolute pairwise correlation in idiosyncratic returns, and average pairwise correlation between IdioVols.

Table 1.3: Idiosyncratic volatility factor model regression statistics

	CAPM			FF3 Model		
Stock	$\widehat{\gamma}_z$	$\widehat{R}_Z^{2,IdioVol\text{-}FM}$	p-val	$\widehat{\gamma}_z$	$\widehat{R}_Z^{2,IdioVol\text{-}FM}$	p-val
AIG	1.49	0.02	0.093	1.53	0.02	0.085
AXP	3.02	0.27	0.146	2.98	0.27	0.149
C	3.46	0.108	0.007	3.48	0.11	0.007
JPM	2.44	0.20	0.007	2.46	0.21	0.006
CVX	1.08	0.51	0.030	1.07	0.51	0.030
XOM	0.60	0.48	0.044	0.61	0.49	0.043
KO	0.33	0.58	0.012	0.33	0.58	0.011
MO	0.44	0.35	0.001	0.44	0.35	0.001
PG	0.43	0.63	0.001	0.43	0.63	0.002
WMT	0.45	0.58	0.006	0.45	0.56	0.008
BA	0.47	0.42	0.003	0.48	0.44	0.003
CAT	0.69	0.49	0.009	0.69	0.48	0.009
GE	1.14	0.26	0.003	1.15	0.26	0.002
HON	0.53	0.44	0.014	0.53	0.43	0.014
MMM	0.39	0.55	0.000	0.38	0.54	0.000
UTX	0.50	0.52	0.003	0.50	0.53	0.004
HPQ	0.65	0.33	0.004	0.66	0.34	0.004
IBM	0.35	0.48	0.011	0.35	0.47	0.012
INTC	0.46	0.46	0.003	0.46	0.46	0.003
MSFT	0.68	0.52	0.008	0.67	0.51	0.010
JNJ	0.41	0.68	0.007	0.40	0.67	0.007
MRK	0.54	0.32	0.001	0.54	0.32	0.001
PFE	0.43	0.34	0.002	0.43	0.34	0.001
DIS	0.57	0.48	0.001	0.58	0.49	0.001
HD	0.66	0.45	0.010	0.66	0.45	0.010
MCD	0.29	0.29	0.003	0.29	0.29	0.003
AA	3.03	0.41	0.019	3.04	0.42	0.018
DD	0.61	0.59	0.001	0.61	0.59	0.001
T	0.76	0.45	0.003	0.76	0.44	0.003
VZ	0.54	0.55	0.000	0.54	0.54	0.001

Estimates of the IdioVol factor loading ($\widehat{\gamma}_Z$, see equation (1.7)), and the contribution of the market volatility to the variation in the IdioVols ($\widehat{R}_Z^{2,IdioVol\text{-}FM}$, see equation (1.10)). The table considers two R-FMs: the left panel defines the IdioVol with respect to CAPM, and the right panel defines the IdioVol with respect to the three-factor Fama-French model. In both cases, the market volatility is the only IdioVol factor. P-val is the p-value of the test of the absence of dependence between the IdioVol and the market volatility for a given individual stock, see equation (1.27).

Table 1.4: Monte Carlo factors specification

	$C_{Z1,t}$	$C_{Z2,t}$
Model 1	$0.1 + 1.5f_{2,t}$	$0.1 + 1.5f_{3,t}$
Model 2	$0.1 + 0.6c_{XX,t} + 0.4f_{2,t}$	$0.1 + 0.5c_{XX,t} + 0.5f_{3,t}$
Model 3	$0.1 + 0.45c_{XX,t} + f_{2,t} + 0.4f_{4,t}$	$0.1 + 0.35c_{XX,t} + 0.3f_{3,t} + 0.6f_{4,t}$

floatfootDifferent specifications for the Idiosyncratic Volatility processes $C_{Z1,t}$ and $C_{Z2,t}$.

Table 1.5: Size and power of the different tests.

	$\Delta_n = 5$ minutes						$\Delta_n = 1$ minute					
	$\theta = 1.5$		$\theta = 2.0$		$\theta = 2.5$		$\theta = 1.5$		$\theta = 2.0$		$\theta = 2.5$	
	AN	LIN	AN	LIN	AN	LIN	AN	LIN	AN	LIN	AN	LIN
Panel A : $H_0^1 : [C_{Z1}, C_{Z2}]_T = 0$, Model 1												
$\alpha = 10\%$	9.7	10.6	10.6	12.6	9.7	10.3	10.2	9.7	10.0	10.2	9.8	10.2
$\alpha = 5\%$	4.7	5.1	4.5	5.3	4.8	5.6	5.3	5.3	5.2	5.3	4.9	5.1
$\alpha = 1\%$	0.9	1.1	0.9	1.2	0.9	1.1	1.1	1.1	1.2	1.1	1.0	1.0
Panel B : $H_0^2 : [C_{Z1}, C_X]_T = 0$, Model 1												
$\alpha = 10\%$	12.1	10.2	10.0	10.6	9.8	11.0	11.0	10.4	10.3	10.4	10.4	10.4
$\alpha = 5\%$	6.2	5.0	4.5	5.2	4.6	5.4	5.5	5.4	5.2	5.1	5.2	5.3
$\alpha = 1\%$	1.5	1.0	0.8	1.0	0.9	1.2	1.1	1.1	1.0	0.9	0.8	1.0
Panel C : $H_0^3 : [C^{resid}_{Z1Z1}, C^{resid}_{Z2Z2}]_T = 0$, Model 2												
$\alpha = 10\%$	10.0	10.1	12.1	10.8	9.9	12.6	10.1	10.3	10.6	11.3	10.1	11.4
$\alpha = 5\%$	5.0	6.3	5.1	6.3	5.1	6.7	5.5	5.5	5.3	5.9	5.2	6.0
$\alpha = 1\%$	1.1	1.5	0.8	1.6	1.1	1.4	1.1	1.2	1.3	1.3	1.3	1.5

Panel A contains the empirical rejection probabilities of the test of absence of dependence between IdioVols. Panel B contains the empirical rejection probabilities of the test of absence of dependence between the IdioVol and the market volatility. Panel C contains the empirical rejection probabilities of the test absence of dependence between residual IdioVols. $T = 10$ years. α denotes the nominal size of the test.

Table 1.6: Finite sample properties of our estimators using 10 years of data sampled at 5 minutes

$\widehat{\theta}$	AN 1.5	AN 2	AN 2.5	AN 3	LIN 1.5	LIN 2	LIN 2.5	LIN 3
				Median Bias				
$\widehat{\gamma}_{Z1}$	-0.047	-0.025	-0.011	-0.003	-0.006	0.001	0.009	0.015
$\widehat{R}_{Z1}^{2,IdioVol\text{-}FM}$	0.176	0.130	0.103	0.085	0.181	0.140	0.112	0.092
$\widehat{Corr}\left(C_{Z1},C_{Z2}\right)$	-0.288	-0.212	-0.163	-0.133	-0.249	-0.190	-0.146	-0.120
$\widehat{Corr}\left(C_{Z1}^{resid},C_{Z2}^{resid}\right)$	-0.189	-0.113	-0.064	-0.034	-0.150	-0.091	-0.047	-0.021
				IQR				
$\widehat{\gamma}_{Z1}$	0.222	0.166	0.138	0.121	0.226	0.168	0.139	0.122
$\widehat{R}_{Z1}^{2,IdioVol\text{-}FM}$	0.210	0.188	0.172	0.152	0.181	0.166	0.152	0.140
$\widehat{Corr}\left(C_{Z1},C_{Z2}\right)$	0.404	0.325	0.263	0.223	0.338	0.283	0.237	0.205
$\widehat{Corr}\left(C_{Z1}^{resid},C_{Z2}^{resid}\right)$	0.456	0.384	0.315	0.272	0.388	0.337	0.285	0.250

The true values are $\gamma_{Z1}=0.450$, $R_{Z1}^{2,IdioVol\text{-}FM}=0.342$, $Corr\left(C_{Z1},C_{Z2}\right)=0.523$, $Corr\left(C_{Z1}^{resid},C_{Z2}^{resid}\right)=0.424$. Model 3.

Table 1.7: Finite sample properties of our estimators using 10 years of data sampled at 1 minute.

$\widehat{\theta}$	AN 1.5	AN 2	AN 2.5	AN 3	LIN 1.5	LIN 2	LIN 2.5	LIN 3
				Median Bias				
$\widehat{\gamma}_{Z1}$	-0.022	-0.012	-0.003	0.004	-0.003	-0.000	0.006	0.012
$\widehat{R}_{Z1}^{IdioVol\text{-}FM}$	0.107	0.091	0.073	0.056	0.113	0.095	0.075	0.058
$\widehat{Corr}\left(C_{Z1}, C_{Z2}\right)$	-0.147	-0.104	-0.073	-0.048	-0.133	-0.097	-0.067	-0.042
$\widehat{Corr}\left(C_{Z1}^{resid}, C_{Z2}^{resid}\right)$	-0.135	-0.086	-0.058	-0.039	-0.119	-0.078	-0.052	-0.032
				IQR				
$\widehat{\gamma}_{Z1}$	0.156	0.112	0.088	0.075	0.157	0.112	0.088	0.075
$\widehat{R}_{Z1}^{IdioVol\text{-}FM}$	0.201	0.146	0.118	0.100	0.184	0.138	0.113	0.096
$\widehat{Corr}\left(C_{Z1}, C_{Z2}\right)$	0.340	0.238	0.184	0.150	0.309	0.226	0.177	0.145
$\widehat{Corr}\left(C_{Z1}^{resid}, C_{Z2}^{resid}\right)$	0.417	0.291	0.228	0.184	0.378	0.274	0.217	0.177

The true values are $\gamma_{Z1} = 0.450$, $R_{Z1}^{2,IdioVol\text{-}FM} = 0.336$, $Corr\left(C_{Z1}, C_{Z2}\right) = 0.514$, $Corr\left(C_{Z1}^{resid}, C_{Z2}^{resid}\right) = 0.408$. Model 3.

Table 1.8: Finite sample properties of our estimators using 5 years of data sampled at 1 minute.

$\widehat{\theta}$	AN 1.5	AN 2	AN 2.5	AN 3	LIN 1.5	LIN 2	LIN 2.5	LIN 3
				Median Bias				
$\widehat{\gamma}_{Z1}$	-0.019	-0.011	-0.007	0.000	-0.001	-0.001	0.002	0.008
$\widehat{R}_{Z1}^{2,IdioVol\text{-}FM}$	0.115	0.096	0.081	0.069	0.119	0.100	0.084	0.071
$\widehat{Corr}\left(C_{Z1}, C_{Z2}\right)$	-0.168	-0.101	-0.064	-0.038	-0.149	-0.092	-0.057	-0.033
$\widehat{Corr}\left(C_{Z1}^{resid}, C_{Z2}^{resid}\right)$	-0.141	-0.079	-0.035	-0.007	-0.127	-0.067	-0.029	-0.001
				IQR				
$\widehat{\gamma}_{Z1}$	0.215	0.159	0.128	0.110	0.216	0.158	0.129	0.110
$\widehat{R}_{Z1}^{2,IdioVol\text{-}FM}$	0.282	0.204	0.168	0.144	0.260	0.194	0.161	0.139
$\widehat{Corr}\left(C_{Z1}, C_{Z2}\right)$	0.472	0.337	0.263	0.213	0.436	0.319	0.252	0.206
$\widehat{Corr}\left(C_{Z1}^{resid}, C_{Z2}^{resid}\right)$	0.541	0.412	0.324	0.266	0.510	0.391	0.311	0.256

The true values are $\gamma_{Z1} = 0.450$, $R_{Z1}^{2,IdioVol\text{-}FM} = 0.35$, $Corr\left(C_{Z1}, C_{Z2}\right) = 0.517$, $Corr\left(C_{Z1}, C_{Z2}\right) = 0.417$. Model 3.

Appendix for Chapter 1 (A)

Sections A1, A2, and A3 contain all the proofs.

A1 Proof of Theorem 1

We start by introducing some notation. Our notation is similar to that of the proofs of Jacod and Rosenbaum (2015) whenever possible.

A1.1 Notation

Throughout, we denote by K a generic constant, which may change from line to line. When it depends on a parameter p, we use the notation K_p instead. We let by convention $\sum_{i=a}^{a'} = 0$ when $a > a'$.

By the usual localization argument, there exists a π-integrable function J on E and a constant such that the stochastic processes in (1.20) and (1.21) satisfy

$$\|b\|, \|\widetilde{b}\|, \|c\|, \|\widetilde{c}\|, J \leq A, \|\delta(w,t,z)\|^r \leq J(z). \tag{C.1}$$

For any càdlàg bounded process Z, we set

$$\eta_{t,s}(Z) = \sqrt{\mathbb{E}\Big(\sup_{0<u\leq s} \|Z_{t+u} - Z_t\|^2 | \mathcal{F}_i^n\Big)}, \quad \text{and}$$
$$\eta_{i,j}^n(Z) = \sqrt{\mathbb{E}\Big(\sup_{0\leq u\leq j\Delta_n} \|Z_{(i-1)\Delta_n+u} - Z_{(i-1)\Delta_n}\|^2 | \mathcal{F}_i^n\Big)}.$$

For convenience, we decompose Y_t as

$$Y_t = Y_0 + Y_t' + \sum_{s\leq t} \Delta Y_s.$$

where $Y'_t = \int_0^t b'_s ds + \int_0^t \sigma_s dW_s$ and $b'_t = b_t - \int \delta(t,z) 1_{\{\|\delta(t,z)\| \le 1\}} \pi(dz)$.
Let $\widehat{C}_i'^n$ be the local estimator of the spot variance of the unobservable process Y', that is

$$\widehat{C}_i'^n = \frac{1}{k_n \Delta_n} \sum_{u=0}^{k_n-1} (\Delta_{i+u}^n Y')(\Delta_{i+u}^n Y)'^\top = (\widehat{C}_i'^{n,gh})_{1 \le g,h \le d}. \tag{C.2}$$

There is no jump truncation applied in the definition of $\widehat{C}_i'^n$ since the process Y' is continuous. Hence, it is more convenient to work with $\widehat{C}_i'^n$ rather than $\widehat{C}_i^n$ (defined in (1.15)). Let's also define

$$\alpha_i^n = (\Delta_i^n Y')(\Delta_i^n Y')^\top - C_i^n \Delta_n, \;\; \nu_i^n = \widehat{C}_i'^n - C_i^n, \;\text{ and }\; \lambda_i^n = \widehat{C}_{i+k_n}'^n - \widehat{C}_i'^n, \tag{C.3}$$

which satisfy

$$\nu_i^n = \frac{1}{k_n \Delta_n} \sum_{j=0}^{k_n-1} (\alpha_{i+j}^n + (C_{i+j}^n - C_i^n)\Delta_n) \;\text{ and }\; \lambda_i^n = \nu_{i+k_n} - \nu_i^n + \Delta_n (C_{i+k_n}^n - C_i^n). \tag{C.4}$$

The following multidimensional quantities will be used in the sequel

$$\begin{aligned}
&\zeta(1)_i^n = \frac{1}{\Delta_n} \Delta_i^n Y' (\Delta_i^n Y')^\top - C_{i-1}^n, \qquad \zeta(2)_i^n = \Delta_i^n c, \\
&\zeta'(u)_i^n = \mathbb{E}(\zeta(u)_i^n | \mathcal{F}_{i-1}^n), \qquad \zeta''(u)_i^n = \zeta(u)_i^n - \zeta'(u)_i^n, \text{with }\; \zeta^r(u)_i^n = \left(\zeta^r(u)_i^{n,gh}\right)_{1 \le g,h \le d}.
\end{aligned}$$

We also define, for $m \in \{0, \ldots, 2k_n - 1\}$ and $j, l \in \mathbb{Z}$,

$$\varepsilon(1)_m^n = \begin{cases} -1 & if\ 0 \le m < k_n \\ +1 & if\ k_n \le m < 2k_n, \end{cases}, \varepsilon(2)_m^n = \sum_{q=m+1}^{2k_n-1} \varepsilon(1)_q^n = (m+1) \wedge (2k_n - m - 1),$$

For any u, v, m, u', v', we set

$$z_{u,v}^n = \begin{cases} 1/\Delta_n & if\ u = v = 1 \\ 1 & \text{otherwise,} \end{cases}$$

$$\begin{aligned}
\lambda(u,v;m)_{j,l}^n &= \frac{3}{2k_n^3} \sum_{q=0 \vee (j-m)}^{(l-m-1) \vee (2k_n-m-1)} \varepsilon(u)_q^n \varepsilon(u)_{q+m}^n, \qquad \lambda(u,v)_m^n = \lambda(u,v;m)_{0,2k_n}^n, \\
M(u,v;u',v')_n &= z_{u,v}^n z_{u',v'}^n \sum_{m=1}^{2k_n-1} \lambda(u,v)_m^n \lambda(u',v')_m^n.
\end{aligned}$$

Additionally, set

$$\overline{A11}(H, gh, u; G, ab, v)^n_T = \frac{3}{2k_n^3} \sum_{i=2k_n}^{[T/\Delta_n]} \Big(\sum_{j=0}^{2k_n-1} \varepsilon(u)^n_j \varepsilon(v)^n_j \Big) (\partial_{gh} H \partial_{ab} G)(C^n_{i-2k_n}) \zeta(u)^{n,gh}_i \zeta(v)^{n,ab}_i$$
$$= \lambda(u,v)^n_0 \sum_{i=2k_n}^{[T/\Delta_n]} (\partial_{gh} H \partial_{ab} G)(C^n_{i-2k_n}) \zeta(u)^{n,gh}_i \zeta(v)^{n,ab}_i, \tag{C.5}$$

and

$$\overline{A12}(H, gh, u; G, ab, v)^n_T = \frac{3}{2k_n^3} \sum_{i=2k_n}^{[T/\Delta_n]} (\partial_{gh} H \partial_{ab} G)(C^n_{i-2k_n}) \sum_{m=1}^{(i-1)\wedge(2k_n-1)} \sum_{j=0}^{(2k_n-m-1)} \varepsilon(u)^n_j \varepsilon(v)^n_{j+m}$$
$$\times \zeta_{gh}(u)^n_{i-m} \zeta_{ab}(v)^n_i. \tag{C.6}$$

A1.2 Auxiliary Results

We provide some useful theorems and lemmas here, which are used to prove Theorem 1. These theorems and lemmas are proved in Appendix A3 below.

Theorem 1. *Let $[H(\widehat{C), G}(C)]_T^{LIN'}$ and $[H(\widehat{C), G}(C)]_T^{AN'}$ be the infeasible estimators obtained by replacing $\widehat{C}^n_i$ by $\widehat{C}'^n_i$ in the definition of $[H(\widehat{C), G}(C)]_T^{LIN}$ and $[H(\widehat{C), G}(C)]_T^{AN}$ in (1.18) and (1.16). As long as $(8p-1)/4(4p-r) \le \varpi < \frac{1}{2}$, we have*

$$\Delta_n^{-1/4}\Big([H(\widehat{C), G}(C)]_T^{LIN} - [H(\widehat{C), G}(C)]_T^{LIN'} \Big) \xrightarrow{\mathbb{P}} 0$$
$$\text{and} \quad \Delta_n^{-1/4}\Big([H(\widehat{C), G}(C)]_T^{AN} - [H(\widehat{C), G}(C)]_T^{AN'} \Big) \xrightarrow{\mathbb{P}} 0. \tag{C.7}$$

Theorem 1 allows, in particular, to focus on the derivation of the asymptotic distributions of $[H(\widehat{C), G}(C)]_T^{LIN'}$ and $[H(\widehat{C), G}(C)]_T^{AN'}$. The next theorem connects the two estimators that we have introduced. To state the theorem, define

$$[H(\widehat{C), G}(C)]_T^{A} = \frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \Bigg(\Big(\partial_{gh} H \partial_{ab} G \Big)(C^n_i) \Big[(\widehat{C}'^{n,gh}_{i+k_n} - \widehat{C}'^{n,gh}_i)(\widehat{C}'^{n,ab}_{i+k_n} - \widehat{C}'^{n,ab}_i)$$
$$- \frac{2}{k_n} (\widehat{C}'^{n,ga}_i \widehat{C}'^{n,hb}_i + \widehat{C}'^{n,gb}_i \widehat{C}'^{n,ha}_i) \Big] \Bigg).$$

with $C^n_i = C_{(i-1)\Delta_n}$, and the superscript A stands for "approximated". For simplicity, we do not index the above quantity by a prime although it depends on $\widehat{C}'^n_i$ instead of $\widehat{C}^n_i$.

Theorem 2. *Under the assumptions of Theorem 1, we have*

$$\Delta_n^{-1/4}\Big([H(\widehat{C),G}(C)]_T^{LIN'} - [H(\widehat{C),G}(C)]_T^{A}\Big) \xrightarrow{\mathbb{P}} 0 \quad and$$
$$\Delta_n^{-1/4}\Big([H(\widehat{C),G}(C)]_T^{AN'} - [H(\widehat{C),G}(C)]_T^{A}\Big) \xrightarrow{\mathbb{P}} 0. \tag{C.8}$$

Theorem 2 shows that the two estimators $[H(\widehat{C),G}(C)]_T^{LIN'}$ and $[H(\widehat{C),G}(C)]_T^{AN'}$ can be approximated by a certain quantity with an error of approximation of order smaller than $\Delta_n^{-1/4}$.

Now, we decompose the approximated estimator as follows

$$[H(\widehat{C),G}(C)]_T^{(A)} = [H(\widehat{C),G}(C)]_T^{(A1)} - [H(\widehat{C),G}(C)]_T^{(A2)}, \tag{C.9}$$

with

$$[H(\widehat{C),G}(C)]_T^{(A1)} = \frac{3}{2k_n}\sum_{g,h,a,b=1}^{d}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\big(\partial_{gh}H\partial_{ab}G\big)(C_{i-1}^n)(\widehat{C}_{i+k_n}^{\prime n,gh} - \widehat{C}_i^{\prime n,gh})(\widehat{C}_{i+k_n}^{\prime n,ab} - \widehat{C}_i^{\prime n,ab}),$$

and

$$[H(\widehat{C),G}(C)]_T^{(A2)} = \frac{3}{k_n^2}\sum_{g,h,a,b=1}^{d}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\big(\partial_{gh}H\partial_{ab}G\big)(\widehat{C}_i^{\prime n})(\widehat{C}_i^{\prime n,ga}\widehat{C}_i^{\prime n,hb} + \widehat{C}_i^{\prime n,gb}\widehat{C}_i^{\prime n,ha}).$$

The following theorem holds:

Theorem 3. *Under the assumptions of Theorem 1, we have*

$$\frac{1}{\Delta_n^{1/4}}\Bigg([H(\widehat{C),G}(C)]_T^{(A1)} - \sum_{g,h,a,b=1}^{d}\sum_{u,v=1}^{2}\overline{A11}(H,gh,u;G,ab,v)_T^n + \overline{A12}(H,gh,u;G,ab,v)_T^n$$
$$+ \overline{A12}(G,ab,v;H,gh,u)_T^n\Bigg) \xRightarrow{\mathbb{P}} 0.$$

Lemma 1. *For any càdlàg bounded process Z, for all $t,s>0$, $j,k\geq 0$, set $\eta_{t,s}=\eta_{t,s}(Z)$. Then,*

$$\Delta_n\mathbb{E}\Bigg(\sum_{i=1}^{[t/\Delta_n]}\eta_{i,k_n}\Bigg) \longrightarrow 0, \qquad \Delta_n\mathbb{E}\Bigg(\sum_{i=1}^{[t/\Delta_n]}\eta_{i,2k_n}\Bigg) \longrightarrow 0,$$
$$\mathbb{E}\Bigg(\eta_{i+j,k}|\mathcal{F}_i^n\Bigg) \leq \eta_{i,j+k} \quad and \quad \Delta_n\mathbb{E}\Bigg(\sum_{i=1}^{[t/\Delta_n]}\eta_{i,4k_n}\Bigg) \longrightarrow 0.$$

Lemma 2. *Let Z be a continuous Itô process with drift b_t^Z and spot variance process C_t^Z, and set*

$\eta_{t,s} = \eta_{t,s}(b^Z, c^Z)$. *Then, the following bounds hold:*

$$
\begin{aligned}
&\left|\mathbb{E}(Z_t \middle| \mathcal{F}_0) - t b_0^Z\right| \leq K t \eta_{0,t} \\
&\left|\mathbb{E}(Z_t^j Z_t^k - t C_0^{Z,jk} \middle| \mathcal{F}_0)\right| \leq K t^{3/2} (\sqrt{\Delta_n} + \eta_{0,t}) \\
&\left|\mathbb{E}\big((Z_t^j Z_t^k - t C_0^{Z,jk})(C_t^{Z,lm} - C_0^{Z,lm}) \middle| \mathcal{F}_0\big)\right| \leq K t^2 \\
&\left|\mathbb{E}(Z_t^j Z_t^k Z_t^l Z_t^m \middle| \mathcal{F}_0) - \Delta_n^2 (C_0^{Z,jk} C_0^{Z,lm} + C_0^{Z,jl} C_0^{Z,km} + C_0^{Z,jm} C_0^{Z,kl})\right| \leq K t^{5/2} \\
&\left|\mathbb{E}(Z_t^j Z_t^k Z_t^l \middle| \mathcal{F}_0)\right| \leq K t^2 \\
&\left|\mathbb{E}(\prod_{l=1}^{6} Z_t^{j_l} \middle| \mathcal{F}_0) - \frac{\Delta_n^3}{6} \sum_{l<l'} \sum_{k<k'} \sum_{m<m'} C_0^{Z,j_l j_{l'}} C_0^{Z,j_k j_{k'}} C_0^{Z,j_m j_{m'}}\right| \leq K t^{7/2} \\
&\mathbb{E}\Big(\sup_{w \in [0,s]} \left\|Z_{t+w} - Z_t\right\|^q \Big| \mathcal{F}_t\Big) \leq K_q s^{q/2}, \text{ and } \left\|\mathbb{E}\Big(Z_{t+s} - Z_t\Big) \Big| \mathcal{F}_t\right\| \leq K s. \qquad \text{(C.10)}
\end{aligned}
$$

$$\text{(C.11)}$$

Lemma 3. *Let ζ_i^n be a r-dimensional $\mathcal{F}_i^n$-measurable process satisfying $\|\mathbb{E}(\zeta_i^n | \mathcal{F}_{i-1}^n)\| \leq L'$ and $\mathbb{E}\Big(\|\zeta_i^n\|^q \Big| \mathcal{F}_{i-1}^n\Big) \leq L_q$. Also, let φ_i^n be a real-valued $\mathcal{F}_i^n$-measurable process with $\mathbb{E}\Big(\|\varphi_{i+j-1}^n\|^q \Big| \mathcal{F}_{i-1}^n\Big) \leq L^q$ for $q \geq 2$ and $1 \leq j \leq 2k_n - 1$. Then,*

$$\mathbb{E}\left(\left\| \sum_{j=1}^{2k_n-1} \varphi_{i+j-1}^n \zeta_{i+j}^n \right\|^q \middle| \mathcal{F}_{i-1}^n\right) \leq K_q L^q \Big(L_q k_n^{q/2} + L'^q k_n^q\Big).$$

Lemma 4. *Under the assumptions of Theorem 1, we have:*

$$
\begin{aligned}
&\left|\mathbb{E}\Big(\lambda_i^{n,jk} \lambda_i^{n,lm} \lambda_{i+2k_n}^{n,gh} \lambda_{i+2k_n}^{n,ab} \Big| \mathcal{F}_i^n\Big) - \frac{4}{k_n^2}\Big(C_i^{n,ga} C_i^{n,hb} + C_i^{n,gb} C_i^{n,ha})(C_i^{n,jl} C_i^{n,km} + C_i^{n,jm} C_i^{n,kl}\Big)\right. \\
&- \frac{4\Delta_n}{3}\Big(C_i^{n,jl} C_i^{n,km} + C_i^{n,jm} C_i^{n,kl}\Big)\overline{C}_i^{n,gh,ab} - \frac{4\Delta_n}{3}\Big(C_i^{n,ga} C_i^{n,hb} - C_i^{n,gb} C_i^{n,ha}\Big)\overline{C}_i^{n,jk,lm} \\
&\left. - \frac{4(k_n\Delta_n)^2}{9} \overline{C}_i^{n,gh,ab} \overline{C}_i^{n,jk,lm}\right| \leq K \Delta_n (\Delta_n^{1/8} + \eta_{i,4k_n}^n\Big).
\end{aligned}
$$

Lemma 5. *Under the assumptions of Theorem 1, we have:*

$$\left|\mathbb{E}\Big(\nu_i^{n,jk} \nu_i^{n,lm} \nu_i^{n,gh} \Big| \mathcal{F}_i^n\Big)\right| \leq K \Delta_n^{3/4} \Big(\Delta_n^{1/4} + \eta_{i,k_n}^n\Big), \qquad \text{(C.12)}$$

$$\left|\mathbb{E}\Big(\nu_i^{n,jk} \nu_i^{n,lm} \Big(c_{i+k_n}^{n,gh} - c_i^{n,gh}\Big) \Big| \mathcal{F}_i^n\Big)\right| \leq K \Delta_n^{3/4} \Big(\Delta_n^{1/4} + \eta_{i,k_n}^n\Big), \qquad \text{(C.13)}$$

$$\left|\mathbb{E}\Big(\nu_i^{n,jk} \Big(c_{i+k_n}^{n,lm} - c_i^{n,lm}\Big) \Big(c_{i+k_n}^{n,gh} - c_i^{n,gh}\Big) \Big| \mathcal{F}_i^n\Big)\right| \leq K \Delta_n^{3/4} \Big(\Delta_n^{1/4} + \eta_{i,k_n}^n\Big), \qquad \text{(C.14)}$$

$$\left|\mathbb{E}\Big(\nu_i^{n,jk}\lambda_i^{n,lm}\lambda_i^{n,gh}\Big|\mathcal{F}_i^n\Big)\right| \leq K\Delta_n^{3/4}\Big(\Delta_n^{1/4}+\eta_{i,2k_n}^n\Big), \tag{C.15}$$

$$\left|\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\lambda_i^{n,gh}\Big|\mathcal{F}_i^n\Big)\right| \leq K\Delta_n^{3/4}\Big(\Delta_n^{1/4}+\eta_{i,2k_n}^n\Big). \tag{C.16}$$

Lemma 6. *Under the assumptions of Theorem 1, we have:*

$$\frac{1}{\Delta_n^{1/4}}\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\rho_{gh}(u,v)_i^n\zeta_{ab}^{'}(v)_i^n \xrightarrow{\mathbb{P}} 0, \quad \forall \quad (u,v) \tag{C.17}$$

$$\frac{1}{\Delta_n^{1/4}}\Big(\overline{A11}(H,gh,u;G,ab,v)-\int_0^T(\partial_{gh}H\partial_{ab}G)(C_t)\overline{C}_t^{gh,ab}dt\Big) \xrightarrow{\mathbb{P}} 0 \;\; when \;\; (u,v)=(2,2) \tag{C.18}$$

$$\frac{1}{\Delta_n^{1/4}}\Big(\overline{A11}(H,gh,u;G,ab,v)-\frac{3}{\theta^2}\int_0^T(\partial_{gh}H\partial_{ab}G)(C_t)(C_t^{ga}C_t^{hb}+C_t^{gb}C_t^{ha})dt\Big) \xrightarrow{\mathbb{P}} 0 \tag{C.19}$$

when $(u,v)=(1,1)$,

$$\frac{1}{\Delta_n^{1/4}}\overline{A11}(H,gh,u;G,ab,v) \xrightarrow{\mathbb{P}} 0 \;\; when \;\; (u,v)=(1,2),(2,1) \tag{C.20}$$

A1.3 Return to the Proof of Theorem 1

We now continue the proof of Theorem 1. By Theorem 3, we have

$$\frac{1}{\Delta_n^{1/4}}\Bigg([H(\widehat{C),G}(C)]_T^{(A1)}-\sum_{g,h,a,b=1}^{d}\sum_{u,v=1}^{2}\overline{A11}(H,gh,u;G,ab,v)_T^n+\overline{A12}(H,gh,u;G,ab,v)_T^n$$
$$+\overline{A12}(G,ab,v;H,gh,u)_T^n\Bigg) \xrightarrow{\mathbb{P}} 0.$$

Recalling the definition of $\overline{A12}(H,gh,u;G,ab,v)_T^n$ from C.6, Lemma 6 implies that

$$\frac{1}{\Delta_n^{1/4}}\Bigg([H(\widehat{C),G}(C)]_T^{(A)}-[H(C),G(C)]_T-\frac{3}{2k_n^3}\sum_{g,h,a,b}^{d}\sum_{u,v=1}^{2}\sum_{i=2k_n}^{[T/\Delta_n]}$$
$$\Big[(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\rho_{gh}(u,v)_i^n\zeta_{ab}^{''}(v)_i^n+(\partial_{ab}H\partial_{gh}G)(C_{i-2k_n}^n)\rho_{ab}(v,u)_i^n\zeta_{gh}^{''}(v)_i^n\Big]\Bigg) \xrightarrow{\mathbb{P}} 0. \tag{C.21}$$

Next, define

$$\xi(H,gh,u;G,ab,v)_i^n=\frac{1}{\Delta_n^{1/4}}(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\rho_{gh}(u,v)_i^n\zeta_{ab}''(v)_i^n,$$
$$Z(H,gh,u;G,ab,v)_t^n=\Delta_n^{1/4}\sum_{i=2k_n}^{[t/\Delta_n]}\xi(H,gh,u;G,ab,v)_i^n.$$

Notice that (C.21) implies

$$\frac{1}{\Delta_n^{1/4}}\Big([H(\widehat{C), G}(C)]_T^{(A)} - [H(C), G(C)]_T\Big) \stackrel{\mathcal{L}}{=} \sum_{g,h,a,b=1}^{d} \sum_{u,v=1}^{2} \frac{1}{\Delta_n^{1/4}}\Big(Z(H, gh, u; G, ab, v)_T^n$$
$$+ Z(H, ab, v; G, gh, u)_T^n\Big). \tag{C.22}$$

Next, observe that to derive the asymptotic distribution of $\Big([H_1(\widehat{C), G_1}(C)]_T^{(A)}, \ldots, [H_\kappa(\widehat{C), G_\kappa}(C)]_T^{(A)}\Big)$, it suffices to study the joint asymptotic behavior of the family of processes $\frac{1}{\Delta_n^{1/4}} Z(H, gh, u; G, ab, v)_T^n$. Notice that $\xi(H, gh, u; G, ab, v)_i^n$ are martingale increments relative to the discrete filtration $(\mathcal{F}_i^n)$. Therefore, to obtain the joint asymptotic distribution of $\frac{1}{\Delta_n^{1/4}} Z(H, gh, u; G, ab, v)_T^n$, it is enough to prove the following three properties:

$$A\Big((H, gh, u; G, ab, v), (H', g'h', u'; G', a'b', v')\Big)_t^n = \sum_{i=2k_n}^{[t/\Delta_n]} \mathbb{E}(\xi(H, gh, u; G, ab, v)_i^n \xi(H', g'h', u'; G', a'b', v')_i^n | \mathcal{F}_{i-1}^n)$$
$$\stackrel{\mathbb{P}}{\Longrightarrow} A\Big((H, gh, u; G, ab, v), (H', g'h', u'; G', a'b', v')\Big)_t, \tag{C.23}$$

$$\sum_{i=2k_n}^{[t/\Delta_n]} \mathbb{E}\Big(\Big|\xi(H, gh, u; G, ab, v)_i^n\Big|^4 \Big| \mathcal{F}_{i-1}^n) \stackrel{\mathbb{P}}{\Longrightarrow} 0, \text{ and} \tag{C.24}$$

$$B(N; H, gh, u; G, ab, v)_t^n := \sum_{i=2k_n}^{[t/\Delta_n]} \mathbb{E}\Big(\xi(H, gh, u; G, ab, v)_i^n \Delta_i^n N | \mathcal{F}_{i-1}^n\Big) \stackrel{\mathbb{P}}{\Longrightarrow} 0, \tag{C.25}$$

for all $t > 0$, all $(H, gh, u; G, ab, v), (H', g'h', u'; G', a'b', v')$ and all martingales N which are either bounded and orthogonal to W, or equal to one component W^j.

Using the polynomial growth assumption imposed on H_r and G_r, (C.24) and (C.25) can be proved by a simple extension of (B.105) and (B.106) in Aït-Sahalia and Jacod (2014) to handle multivariate processes.

Next, define

$$V_{ab}^{a'b'}(v, v')_t = \begin{cases} (C_t^{aa'} C_t^{bb'} + C_t^{ab'} C_t^{ba'}) & \text{if} \quad (v, v') = (1, 1) \\ \overline{C}_t^{ab, a'b'} & \text{if} \quad (v, v') = (2, 2) \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\overline{V}_{gh}^{g'h'}(u, u')_t = \begin{cases} (C_t^{gg'} C_t^{hh'} + C_t^{gh'} C_t^{hg'}) & \text{if} \quad (u, u') = (1, 1) \\ \overline{C}_t^{gh, g'h'} & \text{if} \quad (u, u') = (2, 2) \\ 0 & \text{otherwise.} \end{cases}$$

Using again the polynomial growth assumption on H_r and G_r, we can show that,

$$A\Big((H,gh,u;G,ab,v),(H',g'h',u';G',a'b',v')\Big)_t = M(u,v;u',v')\int_0^t (\partial_{gh}H\partial_{ab}G\partial_{g'h'}H\partial_{a'b'}G)(C_s)V_{ab}^{a'b'}(v,v')_s\overline{V}_{gh}^{g'h'}(u,u')_s ds,$$

with

$$M(u,v;u',v') = \begin{cases} 3/\theta^3 & \text{if} \quad (u,v;u',v') = (1,1;1,1) \\ 3/4\theta & \text{if} \quad (u,v;u',v') = (1,2;1,2),(2,1;2,1) \\ 151\theta/280 & \text{if} \quad (u,v;u',v') = (2,2;2,2) \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, we have

$$A\Big((H,gh,u;G,ab,v),(H',g'h',u';G',a'b',v')\Big)_T =$$

$$\begin{cases} \frac{3}{\nu^3}\int_0^T(\partial_{gh}H\partial_{ab}G\partial_{g'h'}H'\partial_{a'b'}G')(C_t)(C_t^{gg'}C_t^{hh'}+C_t^{gh'}C_t^{hg'})(C_t^{aa'}C_t^{bb'}+C_t^{ab'}C_t^{ba'})dt, \\ \qquad \text{if} \ \ (u,v;u',v') = (1,1;1,1) \\ \frac{3}{4\nu}\int_0^T(\partial_{gh}H\partial_{ab}G\partial_{g'h'}H'\partial_{a'b'}G')(C_t)(C_t^{gg'}C_t^{hh'}+C_t^{gh'}C_t^{hg'})\overline{C}_t^{ab,a'b'}dt, \text{if} \ \ (u,v;u',v') = (1,2;1,2) \\ \frac{3}{4\nu}\int_0^T(\partial_{gh}H\partial_{ab}G\partial_{g'h'}H'\partial_{a'b'}G')(C_t)(C_t^{aa'}C_t^{bb'}+C_t^{ab'}C_s^{ba'})\overline{t}_s^{gh,g'h'}dt, \ \text{if} \ \ (u,v;u',v') = (2,1;2,1) \\ \frac{151\nu}{280}\int_0^T(\partial_{gh}H\partial_{ab}G\partial_{g'h'}H'\partial_{a'b'}G')(C_t)\overline{C}_s^{ab,a'b'}\overline{C}_t^{gh,g'h'}dt, \qquad \text{if} \ \ (u,v;u',v') = (2,2;2,2) \\ 0 \qquad \text{otherwise.} \end{cases}$$

Using (C.22), we deduce that the asymptotic covariance between $[H_r(\widehat{C),G_r}(C)]_T^{(A)}$ and $[H_s(\widehat{C),G_s}(C)]_T^{(A)}$ is given by

$$\sum_{g,h,a,b=1}^{d}\sum_{g',h',a',b'=1}^{d}\sum_{u,v,u',v'=1}^{2}\Bigg(A\Big((H_r,gh,u;G_r,ab,v),(H_s,g'h',u';G_s,a'b',v')\Big)_T$$
$$+A\Big((H_r,gh,u;G_r,ab,v),(H_s,a'b',v';G_s,g'h',u')\Big)_T$$
$$+A\Big((H_r,ab,v;G_r,gh,u),(H_s,g'h',u';G_s,a'b',v')\Big)_T$$
$$+A\Big((H_r,ab,v;H_r,gh,u),(H_s,a'b',v';G_s,g'h',u')\Big)_T\Bigg).$$

After some simple calculations, the above expression can be rewritten as

$$\sum_{g,h,a,b=1}^{d}\sum_{j,k,l,m=1}^{d}\Bigg(\frac{6}{\theta^3}\int_0^T\left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_t)\right)\Big[(C_t^{gj}C_t^{hk}+C_t^{gk}C_t^{hj})(C_t^{al}C_t^{bm}+C_t^{am}C_t^{bl})$$
$$+(C_t^{aj}C_t^{bk}+C_t^{ak}C_t^{bj})(C_t^{gl}C_t^{hm}+C_t^{gm}C_t^{hl})\Big]dt$$
$$+\frac{151\theta}{140}\int_0^t\left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_t)\right)\Big[\overline{C}^{gh,jk}\overline{C}^{ab,lm}+\overline{C}^{ab,jk}\overline{C}^{gh,lm}\Big]dt$$
$$+\frac{3}{2\theta}\int_0^t\left(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s(C_t)\right)\Big[(C_t^{gj}C_t^{hk}+C_t^{gk}C_t^{hj})\overline{C}_t^{ab,lm}+(C_t^{al}C_t^{bm}+C_t^{am}C_t^{bl})\overline{C}_t^{gh,jk}$$
$$+(C_t^{gl}C_s^{hm}+C_t^{gm}C_s^{hl})\overline{C}_t^{ab,jk}+(C_t^{aj}C_t^{bk}+C_t^{ak}C_t^{bj})\overline{C}_t^{gh,lm}\Big]dt\Bigg),$$

which completes the proof.

A2 Proof of Theorem 2

Using the polynomial growth assumption on H_r, G_r, H_s and G_s and Theorem 2.2 in Jacod and Rosenbaum (2015), one can show that

$$\frac{6}{\theta^3}\widehat{\Omega}_T^{r,s,(1)}\xrightarrow{\mathbb{P}}\Sigma_T^{r,s,(1)}.$$

Next, by equation (3.27) in Jacod and Rosenbaum (2015), we have

$$\frac{3}{2\theta}[\widehat{\Omega}_T^{r,s,(3)}-\frac{6}{\theta}\widehat{\Omega}_T^{r,s,(1)}]\xrightarrow{\mathbb{P}}\Sigma_T^{r,s,(3)}.$$

Finally, to show that

$$\frac{151\theta}{140}\frac{9}{4\theta^2}[\widehat{\Omega}_T^{r,s,(2)}+\frac{4}{\theta^2}\widehat{\Omega}_T^{r,s,(1)}-\frac{4}{3}\widehat{\Omega}_T^{r,s,(3)}]\xrightarrow{\mathbb{P}}\Sigma_T^{r,s,(2)},$$

we first observe that the approximation error induced by replacing $\widehat{C}_i^n$ by $\widehat{C}_i'^n$ in Theorem 2 is negligible. For $1\le g,h,a,b,j,k,l,m\le d$ and $1\le r,s\le d$, we define

$$\widehat{W}_T^n=\sum_{i=1}^{[T/\Delta_n]-4k_n+1}(\partial_{gh}H_r\partial_{ab}G_r\partial_{gh}H_s\partial_{lm}G_s)(\widehat{C}_i^n)\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_{i+2k_n}^{n,ab}\lambda_{i+2k_n}^{n,lm},$$
$$\widehat{w}(1)_i^n=(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(C_i^n)\mathbb{E}(\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_{i+2k_n}^{n,ab}\lambda_{i+2k_n}^{n,lm}|\mathcal{F}_i^n),$$
$$\widehat{w}(2)_i^n=(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(C_i^n)(\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_{i+2k_n}^{n,ab}\lambda_{i+2k_n}^{n,lm}-\mathbb{E}(\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_{i+2k_n}^{n,ab}\lambda_{i+2k_n}^{n,lm}|\mathcal{F}_i^n)),$$
$$\widehat{w}(3)_i^n=\Big((\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(\widehat{C}_i^n)-(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(C_i^n)\Big)\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_{i+2k_n}^{n,ab}\lambda_{i+2k_n}^{n,lm},$$

$$\widehat{W}(u)_t^n = \sum_{i=1}^{[T/\Delta_n]-4k_n+1} \widehat{w}_i(u),\ u = 1, 2, 3.$$

Now, note that we also have $\widehat{W}_t^n = \widehat{W}(1)_t^n + \widehat{W}(2)_t^n + \widehat{W}(3)_t^n$. By Taylor expansion and using repeatedly the boundedness of C_t, we obtain

$$|\widehat{w}(3)_i^n| \leq (1 + \|\nu_i^n\|^{4(p-1)})\|\nu_i^n\|\|\lambda_i^n\|^2\|\lambda_{i+2k_n}^n\|^2,$$

which implies $\mathbb{E}(|\widehat{w}(3)_i^n|) \leq K\Delta_n^{5/4}$ and hence $\widehat{W}(3)_t^n \xrightarrow{\mathbb{P}} 0$. Using Cauchy-Schwartz inequality and the bound $\mathbb{E}(\|\lambda_i^n\|^q|\mathcal{F}_i^n) \leq K\Delta_n^{q/4}$, we have $\mathbb{E}(|\widehat{w}(2)_i^n|^2) \leq K\Delta_n^2$. Observing furthermore that $\widehat{w}(2)_i^n$ is $\mathcal{F}_{i+4k_n}-$measurable, Lemma B.8 in Aït-Sahalia and Jacod (2014) implies $\widehat{W}(2)_t^n \xrightarrow{\mathbb{P}} 0$.
Next, define

$$\begin{aligned}
w_i^n &= (\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(C_i^n)\Big[\frac{4}{k_n^2\Delta_n}(C_i^{n,ga}C_i^{n,hb} + C_i^{n,gb}C_i^{n,ha})(C_i^{n,jl}C_i^{n,km} + C_i^{n,jm}C_i^{n,kl}) \\
&+ \frac{4}{3}(C_i^{n,jl}C_i^{n,km} + C_i^{n,jm}C_i^{n,kl})\overline{C}_i^{n,gh,ab} + \frac{4}{3}(C_i^{n,ga}C_i^{n,hb} + C_i^{n,gb}C_i^{n,ha})\overline{C}_i^{n,jk,lm} \\
&+ \frac{4(k_n^2\Delta_n)}{9}\overline{C}_i^{n,gh,ab}\overline{C}_i^{n,jk,lm}\Big], \\
W_T^n &= \Delta_n \sum_{i=1}^{[T/\Delta_n]-4k_n+1} w_i^n.
\end{aligned}$$

Using the cadlag property of c and $\overline{C}$, $k_n\sqrt{\Delta_n} \longrightarrow \theta$, and the Riemann integral convergence, we conclude that $W_T^n \xrightarrow{\mathbb{P}} W_T$ where

$$\begin{aligned}
W_T = \int_0^T &(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(C_t)\Big[\frac{4}{\theta^2}(C_t^{ga}C_t^{hb} + C_t^{gb}C_t^{ha})(C_t^{jl}C_t^{km} + C_t^{jm}C_t^{kl}) \\
&+ \frac{4}{3}(C_t^{jl}C_t^{km} + C_t^{jm}C_t^{kl})\overline{C}_t^{gh,ab} + \frac{4}{3}(C_t^{ga}C_i^{hb} + C_t^{gb}C_t^{ha})\overline{C}_t^{jk,lm} + \frac{4\theta^2}{9}\overline{C}_t^{gh,ab}\overline{C}_t^{jk,lm}\Big]dt.
\end{aligned}$$

In addition, by Lemma 4, it holds that

$$\mathbb{E}(|\widehat{W}(1)_T^n - W_T^n|) \leq \Delta_n\mathbb{E}\left(\sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\Delta_n^{1/8} + \eta_{i,4k_n})\right).$$

Hence, by the third result of Lemma 1 we have $\widehat{W}_T^n \xrightarrow{\mathbb{P}} W_t$, from which it follows that

$$\frac{9}{4\theta^2}\Big[\widehat{W}(1)_T^n + \frac{4}{k_n^2}\sum_{i=1}^{[T/\Delta_n]-4k_n+1}(\partial_{gh}H_r\partial_{ab}G_r\partial_{jk}H_s\partial_{lm}G_s)(\widehat{C}_i^n)[C_i^n(jk, lm)C_i^n(gh, ab)]$$

$$
\begin{aligned}
&- \frac{2}{k_n} \sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s)(\widehat{C}_i^n) C_i^n(gh, ab) \lambda_i^{n,jk} \lambda_i^{n,lm} \\
&- \frac{2}{k_n} \sum_{i=1}^{[T/\Delta_n]-4k_n+1} (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s)(\widehat{C}_i^n) C_i^n(jk, lm) \lambda_i^{n,gh} \lambda_i^{n,ab} \Big] \\
&\xrightarrow{\mathbb{P}} \int_0^T (\partial_{gh} H_r \partial_{ab} G_r \partial_{jk} H_s \partial_{lm} G_s)(C_t) \overline{C}_t^{gh,ab} \overline{C}_t^{jk,lm} dt.
\end{aligned}
$$

The result follows from the above convergence, the already invoked symmetry argument, and straightforward calculations.

A3 Proofs of Auxiliary Results

This section is devoted to the proofs of the auxiliary theorems and lemmas (listed in Section A1.2) that were used to prove Theorem 1 and Theorem 2.

A3.1 Proof of Theorem 1

To show this result, let us define the functions

$$
\begin{aligned}
R(x,y) &= \sum_{g,h,a,b=1}^{d} \Big(\partial_{gh} H \partial_{ab} G\Big)(x) \Big(y^{gh} - x^{gh}\Big)\Big(y^{ab} - x^{ab}\Big) \\
S(x,y) &= \Big(H(y) - H(x)\Big)\Big(G(y) - G(x)\Big) \\
U(x) &= \sum_{g,h,a,b=1}^{d} \Big(\partial_{gh} H \partial_{ab} G\Big)(x) \Big(x^{ga} x^{hb} + x^{gb} x^{ha}\Big),
\end{aligned}
$$

for any $\mathbb{R}^d \times \mathbb{R}^d$ matrices x and y. The following decompositions hold,

$$
\begin{aligned}
&[\widehat{H(C), G(C)}]_T^{AN} - [\widehat{H(C), G(C)}]_T^{AN'} \\
&= \frac{3}{2k_n} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \Big[\big(S(\widehat{C}_i^n, \widehat{C}_{i+k_n}^n) - S(\widehat{C}_i'^n, \widehat{C}_{i+k_n}'^n)\big) - \frac{2}{k_n}\big(U(\widehat{C}_i^n) - U(\widehat{C}_i'^n)\big)\Big], \\
&[\widehat{H(C), G(C)}]_T^{LIN} - [\widehat{H(C), G(C)}]_T^{LIN'} \\
&= \frac{3}{2k_n} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \Big[\big(R(\widehat{C}_i^n, \widehat{C}_{i+k_n}^n) - R(\widehat{C}_i'^n, \widehat{C}_{i+k_n}'^n)\big) - \frac{2}{k_n}\big(U(\widehat{C}_i^n) - U(\widehat{C}_i'^n)\big)\Big].
\end{aligned}
$$

By (3.11) in Jacod and Rosenbaum (2015), there exists a sequence of real numbers a_n converging to zero such that

$$
\mathbb{E}(\|\widehat{C}_i^n - \widehat{C}_i'^n\|^q) \le K_q a_n \Delta_n^{(2q-r)\varpi+1-q}, \text{ for any } q > 0. \tag{C.26}
$$

Since H and $G \in \mathcal{G}(p)$, the functions R and S are continuously differentiable and satisfy

$$\begin{aligned} \|\partial J(x,y)\| &\le K(1+\|x\|+\|y\|)^{2p-1} \ \text{ for } \ 1 \le g,h,a,b \le d \ \text{ and } \ J \in \{S,R\}, && \text{(C.27)} \\ \|\partial U(x)\| &\le K(1+\|x\|)^{2p-1}, && \text{(C.28)} \end{aligned}$$

where ∂J (respectively, ∂U) is a vector that collects the first order partial derivatives of the function J (respectively, U) with respect to all the elements of (x,y) (resp x). Using the Taylor expansion, the Jensen's inequality, (C.27) and (C.28), it holds that, for $J \in \{S,R\}$,

$$\begin{aligned} &|J(\widehat{C}_i^n, \widehat{C}_{i+k_n}^n) - J(\widehat{C}_i'^n, \widehat{C}_{i+k_n}'^n)| \le K(1+\|\widehat{C}_i'^n\|^{2p-1} + \|\widehat{C}_{i+k_n}'^n\|^{2p-1}) \\ &\quad \times (\|\widehat{C}_i^n - \widehat{C}_i'^n\| + \|\widehat{C}_{i+k_n}^n - \widehat{C}_{i+k_n}'^n\|) + K\|\widehat{C}_i^n - \widehat{C}_i'^n\|^{2p} + K\|\widehat{C}_{i+k_n}^n - \widehat{C}_{i+k_n}'^n\|^{2p} \ \text{ and} \\ &|U(\widehat{C}_i^n) - U(\widehat{C}_i'^n)| \le K(1+\|\widehat{C}_i'^n\|^{2p-1})(\|\widehat{C}_i^n - \widehat{C}_i'^n\|) + K\|\widehat{C}_i^n - \widehat{C}_i'^n\|^{2p}. \end{aligned}$$

By (3.20) in Jacod and Rosenbaum (2015), we have $\mathbb{E}(\|\widehat{C}_i'^n\|^v) \le K_v$, for any $v \ge 0$. Hence by Hölder inequality, for $\epsilon > 0$ fixed,

$$\begin{aligned} \mathbb{E}(\|\widehat{C}_i'^n\|^{2p-2}\|\widehat{C}_i^n - \widehat{C}_i'^n\|) &\le \Big(\mathbb{E}(\|\widehat{C}_i^n - \widehat{C}_i'^n\|^{(1+\epsilon)})\Big)^{1/1+\epsilon} \Big(\mathbb{E}(\|\widehat{C}_i'^n\|^{(2p-2)(1+\epsilon)/\epsilon})\Big)^{\epsilon/1+\epsilon} \\ &\le K_p \Big(\mathbb{E}(\|\widehat{C}_i^n - \widehat{C}_i'^n\|^{(1+\epsilon)})\Big)^{1/1+\epsilon} \\ &\le K_p a_n \Delta_n^{(2-\frac{1}{1+\epsilon})\varpi + \frac{1}{1+\epsilon} - 1}. \end{aligned}$$

Using the above result and (C.26), the following conditions are sufficient for Theorem 1 to hold:

$$(2 - \frac{r}{1+\epsilon})\varpi + \frac{1}{1+\epsilon} - 1 - \frac{3}{4} \ge 0, \quad (4p-r)\varpi + 1 - 2p - \frac{3}{4} \ge 0, \quad \text{and} \quad (2-r)\varpi + -\frac{3}{4} \ge 0.$$

Using the fact that $0 < \varpi < \frac{1}{2}$, and taking ϵ sufficiently close to zero, we can see that Theorem 1 holds when $(8p-1)/4(4p-r) \le \varpi < \frac{1}{2}$, which completes the proof.

A3.2 Proof of Theorem 2

Note that we have

$$\begin{aligned} [H(\widehat{C}), G(C)]_T^{LIN'} - [H(\widehat{C}), G(C)]_T^{A} &= \frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \psi_i^n(g,h,a,b), \\ [H(\widehat{C}), G(C)]_T^{AN'} - [H(\widehat{C}), G(C)]_T^{A} &= \frac{3}{2k_n} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \Big(\chi_i^n - \sum_{g,h,a,b=1}^{d} (\partial_{gh} H \partial_{ab} G)(C_i^n)\lambda_i^{n,gh}\lambda_i^{n,ab}\Big), \end{aligned}$$

with

$$\psi_i^n(g,h,a,b) = \Big(\big(\partial_{gh} H \partial_{ab} G\big)(\widehat{C}_i'^n) - \big(\partial_{gh} H \partial_{ab} G\big)(C_i^n)\Big)\lambda_i^{n,gh}\lambda_i^{n,ab},$$
$$\chi_i^n = \Big(H(\widehat{C}_{i+k_n}'^n) - H(\widehat{C}_i'^n)\Big)\Big(G(\widehat{C}_{i+k_n}'^n) - G(\widehat{C}_i'^n)\Big).$$

By Taylor expansion, we have

$$\big(\partial_{gh} S \partial_{ab} G\big)(\widehat{C}_i'^n) - \big(\partial_{gh} S \partial_{ab} G\big)(C_i^n) = \sum_{x,y=1}^{d} \Big(\partial^2_{xy,gh} S \partial_{ab} G + \partial^2_{xy,ab} G \partial_{gh} S\Big)(C_i^n)\nu_i^{n,xy}$$
$$+ \frac{1}{2}\sum_{j,k,x,y=1}^{d} \Big(\partial^3_{jk,xy,gh} S \partial_{ab} G + \partial^2_{xy,gh} S \partial^2_{jk,ab} G + \partial^3_{jk,xy,ab} G \partial_{gh} S + \partial^2_{xy,ab} G \partial^2_{jk,gh} S\Big)(\widetilde{c}_i^n)\nu_i^{n,xy}\nu_i^{n,jk}$$

and

$$S(\widehat{C}_{i+k_n}'^n) - S(\widehat{C}_i'^n) = \sum_{gh} \partial_{gh} S(C_i^n)\lambda_i^{n,gh} + \sum_{j,k,g,h} \partial^2_{jk,gh} S(C_i^n)\lambda_i^{n,gh}\nu_i^{n,jk}$$
$$+ \frac{1}{2}\sum_{x,y,g,h} \partial^2_{xy,gh} S(C_i^n)\lambda_i^{n,gh}\lambda_i^{n,xy} + \frac{1}{2}\sum_{x,y,j,k,g,h} \partial^3_{xy,jk,gh} S(CC_i^{n,S})\lambda_i^{n,gh}\nu_i^{n,xy}\nu_i^{n,jk}$$
$$+ \frac{1}{6}\sum_{j,k,x,y,g,h} \partial^3_{jk,xy,gh} S(C_i^{n,S})\lambda_i^{n,jk}\lambda_i^{n,gh}\lambda_i^{n,xy},$$

for $S \in \{H, G\}$, $\widetilde{c}_i^n = \pi C_i^n + (1-\pi)\widehat{C}_i'^n$, $C_i^{n,S} = \pi_S \widehat{C}_i'^n + (1-\pi_S)\widehat{C}_{i+k_n}'^n$, $CC_i^{n,S} = \mu_S C_i^n + (1-\mu_S)\widehat{C}_i'^n$ for $\pi, \pi_H, \mu_H, \pi_G, \mu_G \in [0,1]$. Although $\widetilde{c}_i^n$ and π depend on g, h, a, and b, we do not emphasize this in our notation to simplify the exposition.

Set $\mathcal{F}_i^n = \mathcal{F}_{(i-1)\Delta_n}$. By (4.10) in Jacod and Rosenbaum (2013) we have

$$\mathbb{E}\Big(\Big\|\alpha_i^n\Big\|^q \Big|\mathcal{F}_i^n\Big) \le K_q \Delta_n^q \text{ for all } q \ge 0 \text{ and } \mathbb{E}\Big(\Big|\sum_{j=0}^{k_n-1} \alpha_{i+j}^n\Big|^q \Big|\mathcal{F}_i^n\Big) \le K_q \Delta_n^q k_n^{q/2} \text{ whenever } q \ge 2. \quad \text{(C.29)}$$

Combining (C.29), (C.4), (C.10) with $Z = c$ and the Hölder inequality yields for $q \ge 2$,

$$\mathbb{E}\Big(\Big\|\nu_i^n\Big\|^q \Big|\mathcal{F}_i^n\Big) \le K_q \Delta^{q/4}, \text{ and } \mathbb{E}\Big(\Big\|\lambda_i^n\Big\|^q \Big|\mathcal{F}_i^n\Big) \le K_q \Delta^{q/4}. \quad \text{(C.30)}$$

The bound in the first equation of (C.30) is tighter than that in (4.11) of Jacod and Rosenbaum (2015) due to the absence of volatility jumps. This tighter bound will be useful later in deriving the asymptotic distribution for the approximated estimator. By the boundedness of C_t and the polynomial growth assumption, we have

$$\left|\left(\partial^3_{jk,xy,ab}G\partial_{gh}H+\partial^2_{xy,gh}H\partial^2_{jk,ab}G\right)(\widetilde{c}^n_i)\nu^{n,xy}_i\nu^{n,jk}_i\lambda^{n,gh}_i\lambda^{n,ab}_i\right|\leq K(1+\|\widetilde{c}^n_i\|)^{2(p-2)}\|\nu^n_i\|^2\|\lambda^n_i\|^2.$$

Recalling $\widetilde{c}^n_i=\pi C^n_i+(1-\pi)\widehat{C}'^n_i$ and using the convexity of the function $x^{2(p-2)}$, we can refine the last inequality as follows:

$$\left|\left(\partial^3_{jk,xy,ab}G\partial_{gh}H+\partial^2_{xy,gh}H\partial^2_{jk,ab}G\right)(\widetilde{c}^n_i)\nu^{n,xy}_i\nu^{n,jk}_i\lambda^{n,gh}_i\lambda^{n,ab}_i\right|\leq K\left(1+\|\nu^n_i\|^{2(p-2)}\right)\|\nu^n_i\|^2\|\lambda^n_i\|^2. \tag{C.31}$$

Using the Taylor expansion, the polynomial growth assumption and using similar idea as for (C.31), we have

$$\begin{aligned}
&\chi^n_i-\sum_{g,h,a,b}(\partial_{gh}H\partial_{ab}G)(C^n_i)\lambda^{n,gh}_i\lambda^{n,ab}_i=\\
&\sum_{g,h,a,b,j,k}(\partial_{gh}H\partial^2_{jk,xy}G+\partial_{gh}G\partial^2_{jk,xy}H)(C^n_i)(\lambda^{n,gh}_i+\frac{1}{2}\nu^{n,gh}_i)\lambda^{n,ab}_i\lambda^{n,jk}_i+\varphi^n_i,\quad\text{and}\\
&\sum_{g,h,a,b}\left(\partial_{gh}H\partial_{ab}G\right)(\widehat{C}'^n_i)-\left(\partial_{gh}H\partial_{ab}G\right)(C^n_i)=\\
&\sum_{g,h,a,b,x,y}(\partial_{gh}H\partial^2_{ab,xy}G+\partial_{ab}G\partial^2_{gh,xy}G)(C^n_i)(\nu^{n,xy}_i)\lambda^{n,gh}_i\lambda^{n,ab}_i+\delta^n_i
\end{aligned}$$

with $\mathbb{E}(|\varphi^n_i|\big|\mathcal{F}^n_i)\leq K\Delta_n$ and $\mathbb{E}(|\delta^n_i|\big|\mathcal{F}^n_i)\leq K\Delta_n$ which follow by the Cauchy-Schwartz inequality together with (C.30). Given that $k_n=\theta(\Delta_n)^{-1/2}$, the previous inequalities imply

$$\frac{3\Delta_n^{-1/4}}{2k_n}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\varphi^n_i\overset{\mathbb{P}}{\Longrightarrow}0\quad\text{and}\quad\frac{3\Delta_n^{-1/4}}{2k_n}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\delta^n_i\overset{\mathbb{P}}{\Longrightarrow}0.$$

Therefore, it suffices to show that

$$\frac{3\Delta_n^{-1/4}}{2k_n}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\sum_{g,h,a,b,j,k}(\partial_{gh}H\partial^2_{jk,ab}G+\partial_{gh}H\partial^2_{jk,ab}G)(C^n_i)\lambda^{n,gh}_i\lambda^{n,ab}_i\lambda^{n,jk}_i\xrightarrow{\mathbb{P}}0, \tag{C.32}$$

$$\frac{3\Delta_n^{-1/4}}{2k_n}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\sum_{g,h,a,b,j,k}(\partial_{gh}H\partial^2_{jk,ab}G+\partial_{gh}H\partial^2_{jk,ab}G)(C^n_i)\nu^{n,gh}_i\lambda^{n,ab}_i\lambda^{n,jk}_i\xrightarrow{\mathbb{P}}0. \tag{C.33}$$

These results hold by the bounds in 5.

A3.3 Proof of Theorem 3

First, we decompose the approximated estimator as

$$[H(\widehat{C), G}(C)]_T^{(A)} = [H(\widehat{C), G}(C)]_T^{(A1)} - [H(\widehat{C), G}(C)]_T^{(A2)}, \tag{C.34}$$

with

$$[H(\widehat{C), G}(C)]_T^{(A1)} = \frac{3}{2k_n} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \big(\partial_{gh} H \partial_{ab} G\big)(C_{i-1}^n)(\widehat{C}'^{n,gh}_{i+k_n} - \widehat{C}'^{n,gh}_{i})(\widehat{C}'^{n,ab}_{i+k_n} - \widehat{C}'^{n,ab}_{i}),$$

and

$$[H(\widehat{C), G}(C)]_T^{(A2)} = \frac{3}{k_n^2} \sum_{g,h,a,b=1}^{d} \sum_{i=1}^{[T/\Delta_n]-2k_n+1} \big(\partial_{gh} H \partial_{ab} G\big)(\widehat{C}'^{n}_i)(\widehat{C}'^{n,ga}_i \widehat{C}'^{n,hb}_i + \widehat{C}'^{n,gb}_i \widehat{C}'^{n,ha}_i).$$

In this section, we use the notation $C_{i-1}^n = C_{(i-1)\Delta_n}$ and $\mathcal{F}_i = \mathcal{F}_{(i-1)\Delta_n}$ to simplify the exposition. Given the polynomial growth assumption satisfied by H and G and the fact that $k_n = \theta(\Delta_n)^{-1/2}$, by Theorem 2.2 in Jacod and Rosenbaum (2015) we have

$$\frac{1}{\sqrt{\Delta_n}} \left([H(\widehat{C), G}(C)]_T^{(A2)} - \frac{3}{\theta^2} \sum_{g,h,a,b=1}^{d} \int_0^T \big(\partial_{gh} H \partial_{ab} G\big)(C_t)(c_t^{ga} c_t^{hb} + c_i^{gb} c_t^{ha}) dt \right) = O_p(1),$$

which yields

$$\frac{1}{\Delta_n^{1/4}} \left([H(\widehat{C), G}(C)]_T^{(A2)} - \frac{3}{\theta^2} \sum_{g,h,a,b=1}^{d} \int_0^T \big(\partial_{gh} H \partial_{ab} G\big)(C_t)(c_t^{ga} c_t^{hb} + c_i^{gb} c_t^{ha}) dt \right) \xrightarrow{\mathbb{P}} 0.$$

Using the multivariate quantities defined in Section A1.1, we can show that the following decompositions hold:

$$\widehat{C}'^{n}_i = C_{i-1}^n + \frac{1}{k_n} \sum_{j=0}^{k_n-1} \sum_{u=1}^{2} \overline{\varepsilon}(u)_j^n \zeta(u)_{i+j}^n, \qquad \widehat{C}'^{n}_{i+k_n} - \widehat{C}'^{n}_i = \frac{1}{k_n} \sum_{j=0}^{2k_n-1} \sum_{u=1}^{2} \varepsilon(u)_j^n \zeta(u)_{i+j}^n,$$

$$\begin{aligned}
\lambda_i^{n,gh} \lambda_i^{n,ab} = \frac{1}{k_n^2} \sum_{u=1}^{2} \sum_{v=1}^{2} \Bigg(& \sum_{j=0}^{2k_n-1} \varepsilon(u)_j^n \varepsilon(v)_j^n \zeta(u)_{i+j}^{n,gh} \zeta(v)_{i+j}^{n,ab} \\
& + \sum_{j=0}^{2k_n-2} \sum_{q=j+1}^{2k_n-1} \varepsilon(u)_j^n \varepsilon(v)_q^n \zeta(u)_{i+j}^{n,gh} \zeta(v)_{i+q}^{n,ab} + \sum_{j=1}^{2k_n-1} \sum_{q=0}^{j-1} \varepsilon(u)_j^n \varepsilon(v)_q^n \zeta(u)_{i+j}^{n,gh} \zeta(v)_{i+q}^{n,ab} \Bigg).
\end{aligned}$$

Changing the order of the summation in the last term yields

$$\lambda_i^{n,gh}\lambda_i^{n,ab} = \frac{1}{k_n^2}\sum_{u=1}^{2}\sum_{v=1}^{2}\Bigg(\sum_{j=0}^{2k_n-1}\varepsilon(u)_j^n\varepsilon(v)_j^n\zeta(u)_{i+j}^{n,gh}\zeta(v)_{i+j}^{n,ab} \\ + \sum_{j=0}^{2k_n-2}\sum_{q=j+1}^{2k_n-1}\varepsilon(u)_j^n\varepsilon(v)_q^n\zeta(u)_{i+j}^{n,gh}\zeta(v)_{i+q}^{n,ab} + \sum_{j=0}^{2k_n-2}\sum_{q=j+1}^{2k_n-1}\varepsilon(v)_j^n\varepsilon(u)_q^n\zeta(v)_{i+j}^{n,ab}\zeta(u)_{i+q}^{n,gh}\Bigg).$$

Therefore, we can further rewrite $[H(\widehat{C),G}(C)]_T^{(A1)}$ as

$$[H(\widehat{C),G}(C)]_T^{(A1)} = [H(\widehat{C),G}(C)]_T^{(A11)} + [H(\widehat{C),G}(C)]_T^{(A12)} + [H(\widehat{C),G}(C)]_T^{(A13)}, \text{with}$$

$$[H(\widehat{C),G}(C)]_T^{(A1w)} = \sum_{g,h,a,b=1}^{d}\sum_{u,v=1}^{2}\widehat{A1w}(H,gh,u;G,ab,v)_T^n, \quad w=1,2,3,$$

and

$$\widehat{A11}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\sum_{j=0}^{2k_n-1}(\partial_{gh}H\partial_{ab}G)(C_{i-1}^n)\varepsilon(u)_j^n\varepsilon(v)_j^n\zeta(u)_{i+j}^{n,gh}\zeta(v)_{i+j}^{n,ab},$$

$$\widehat{A12}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\sum_{j=0}^{2k_n-2}\sum_{q=j+1}^{2k_n-1}(\partial_{gh}H\partial_{ab}G)(C_{i-1}^n)\varepsilon(u)_j^n\varepsilon(v)_q^n\zeta(u)_{i+j}^{n,gh}\zeta(v)_{i+q}^{n,ab},$$

$$\widehat{A13}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=1}^{[T/\Delta_n]-2k_n+1}\sum_{j=0}^{2k_n-2}\sum_{q=j+1}^{2k_n-1}(\partial_{gh}H\partial_{ab}G)(C_{i-1}^n)\varepsilon(v)_j^n\varepsilon(u)_q^n\zeta(v)_{i+j}^{n,ab}\zeta(u)_{i+q}^{n,gh},$$

where we clearly have $\widehat{A13}(H,gh,u;G,ab,v)_T^n = \widehat{A12}(G,ab,v;H,gh,u)_T^n$. By a change of the order of the summation,

$$\widehat{A11}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=1}^{[T/\Delta_n]}\sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)}(\partial_{gh}H\partial_{ab}G) \\ \times (C_{i-j-1}^n)\varepsilon(u)_j^n\varepsilon(v)_j^n\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab},$$

$$\widehat{A12}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=2}^{[T/\Delta_n]}\sum_{m=1}^{(i-1)\wedge(2k_n-1)}\sum_{j=0\vee(i+2k_n-1-m-[T/\Delta_n])}^{(2k_n-m-1)\wedge(i-m-1)}(\partial_{gh}H\partial_{ab}G)(C_{i-1-j-m}^n) \\ \times \varepsilon(u)_j^n\varepsilon(v)_{j+m}^n\zeta_{gh}(u)_{i-m}^n\zeta_{ab}(v)_i^n.$$

Now, set

$$\widetilde{A11}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=2k_n}^{[T/\Delta_n]}\sum_{j=0}^{2k_n-1}(\partial_{gh}H\partial_{ab}G)(C_{i-j-1}^n)\varepsilon(u)_j^n\varepsilon(v)_j^n\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab},$$

$$\widetilde{A12}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=2k_n}^{[T/\Delta_n]}\sum_{m=1}^{(i-1)\wedge(2k_n-1)}\sum_{j=0}^{(2k_n-m-1)}(\partial_{gh}H\partial_{ab}G)(C_{i-j-1-m}^n)\varepsilon(u)_j^n\varepsilon(v)_{j+m}^n$$

$$\times\zeta_{gh}(u)_{i-m}^n\zeta_{ab}(v)_i^n,$$

and

$$\overline{A11}(H,gh,u;G,ab,v)_T^n = \frac{3}{2k_n^3}\sum_{i=2k_n}^{[T/\Delta_n]}\Big(\sum_{j=0}^{2k_n-1}\varepsilon(u)_j^n\varepsilon(v)_j^n\Big)(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}$$

$$= \lambda(u,v)_0^n\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab},$$

$$\overline{A12}(H,gh,u;G,ab,v)_T^n$$

$$= \frac{3}{2k_n^3}\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\sum_{m=1}^{(i-1)\wedge(2k_n-1)}\sum_{j=0}^{(2k_n-m-1)}\varepsilon(u)_j^n\varepsilon(v)_{j+m}^n\zeta_{gh}(u)_{i-m}^n\zeta_{ab}(v)_i^n$$

$$= \sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\rho_{gh}(u,v)_i^n\zeta_{ab}(v)_i^n,$$

with

$$\rho_{gh}(u,v)_i^n = \sum_{m=1}^{2k_n-1}\lambda(u,v)_m^n\zeta_{gh}(u)_{i-m}^n.$$

We show below that the following results hold:

$$\frac{1}{\Delta_n^{1/4}}\Big(\widehat{A1w}(H,gh,u;G,ab,v)_T^n - \widetilde{A1w}(H,gh,u;G,ab,v)_T^n\Big) \xrightarrow{\mathbb{P}} 0 \tag{C.35}$$

$$\frac{1}{\Delta_n^{1/4}}\Big(\widetilde{A1w}(H,gh,u;G,ab,v)_T^n - \overline{A1w}(H,gh,u;G,ab,v)_T^n\Big) \xrightarrow{\mathbb{P}} 0 \tag{C.36}$$

for all (H,gh,u,G,ab,v) and $w=1,2$.

Proof of (C.35) for $w = 1$

To prove this result, first, notice that the $\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}$ are scaled by random variables rather that constant real numbers. Next, observe that we can write

$$
\begin{aligned}
\widehat{A11} - \widetilde{A11} &= \widetilde{\widehat{A11}}(1) + \widetilde{\widehat{A11}}(2) + \widetilde{\widehat{A11}}(3) \quad \text{ with} \\
\widetilde{\widehat{A11}}(1) &= \sum_{i=1}^{(2k_n-1)\wedge[T/\Delta_n]} \left(\frac{3}{2k_n^3} \sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right) \zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}, \\
\widetilde{\widehat{A11}}(2) &= \sum_{i=[T/\Delta_n]-2k_n+2}^{[T/\Delta_n]} \frac{3}{2k_n^3} \left(\sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right. \\
&\qquad \left. - \sum_{j=0}^{(2k_n-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right) \zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}, \\
\widetilde{\widehat{A11}}(3) &= \sum_{i=2k_n}^{[T/\Delta_n]-2k_n+1} \frac{3}{2k_n^3} \left(\sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right. \\
&\qquad \left. - \sum_{j=0}^{(2k_n-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right) \zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}.
\end{aligned}
$$

It is easy to see that $\widetilde{\widehat{A12}}(3) = 0$. Using (C.10) with $Z = c$ and (C.29), we obtain

$$
\mathbb{E}(\|\zeta(1)_i^n\|^q|\mathcal{F}^n_{i-1}) \leq K_q, \;\; \mathbb{E}(\|\zeta(2)_i^n\|^q|\mathcal{F}^n_{i-1}) \leq K_q\Delta_n^{q/2}. \tag{C.37}
$$

The polynomial growth assumption on H and G and the boundedness of C_t imply that $|(\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})| \leq K$.
Hence, the random quantities $\left(\frac{3}{2k_n^3}\sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j\right)$ and $\frac{3}{2k_n^3}\sum_{j=0}^{(2k_n-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j$ are $\mathcal{F}^n_{i-1}-$ measurable and are bounded by $\widetilde{\lambda}^n_{u,v}$ defined as

$$
\widetilde{\lambda}^n_{u,v} = \begin{cases} K & \text{if } (u,v) = (2,2) \\ K/k_n & \text{if } (u,v) = (1,2),(2,1) \\ K/k_n^2 & \text{if } (u,v) = (1,1). \end{cases}
$$

Similarly, the quantity

$$
\frac{3}{2k_n^3}\left(\sum_{j=0\vee(i+2k_n-1-[T/\Delta_n])}^{(2k_n-1)\wedge(i-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j - \sum_{j=0}^{(2k_n-1)} (\partial_{gh}H\partial_{ab}G)(C^n_{i-j-1})\varepsilon(u)^n_j\varepsilon(v)^n_j \right),
$$

is $\mathcal{F}^n_{i-1}-$ measurable and bounded by $2\widetilde{\lambda}^n_{u,v}$. Note also that, by (C.37) and the Cauchy Schwartz inequality, we have

$$\begin{aligned}\mathbb{E}(|\zeta(u)^{n,gh}_i\zeta(v)^{n,ab}_i|\big|\mathcal{F}^n_{i-1}) &\leq \mathbb{E}(\|\zeta(u)^n_i\|^2|\mathcal{F}^n_{i-1})^{1/2}\mathbb{E}(\|\zeta(v)^n_i\|^2|\mathcal{F}^n_{i-1})^{1/2}\\ &\leq \begin{cases} K\Delta_n & \text{if } (u,v)=(2,2)\\ K\Delta_n^{1/2} & \text{if } (u,v)=(1,2),(2,1)\\ K & \text{if } (u,v)=(1,1).\end{cases}\end{aligned}$$

The above bounds, together with the fact that $k_n = \theta\Delta_n^{-1/2}$, imply $\mathbb{E}(|\widetilde{\widehat{A11}}(1)|) \leq K\Delta_n^{1/2}$ and $\mathbb{E}(|\widetilde{\widehat{A11}}(2)|) \leq K\Delta_n^{1/2}$ for all (u,v). These two results together imply $\widetilde{\widehat{A11}}(1) = o(\Delta_n^{-1/4})$ and $\widetilde{\widehat{A11}}(2) = o(\Delta_n^{-1/4})$, which yields the result.

Proof of (C.35) for $w=2$

First, observe that $\widehat{A12}-\widetilde{A12} = \widetilde{\widehat{A12}}(1)+\widetilde{\widehat{A12}}(2)$, with

$$\begin{aligned}\widetilde{\widehat{A12}}(1) = &\sum_{i=2}^{(2k_n-1)\wedge[T/\Delta_n]}\Bigg(\sum_{m=1}^{(i-1)}\frac{3}{2k_n^3}\Big(\sum_{j=0\vee(i+2k_n-1-m-[T/\Delta_n])}^{(2k_n-m-1)\wedge(i-m-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-1-j-m})\varepsilon(u)^n_j\varepsilon(v)^n_{j+m}\Big)\\ &\times\zeta_{gh}(u)^n_{i-m}\Bigg)\zeta_{ab}(v)^n_i,\\ \widetilde{\widehat{A12}}(2) = &\sum_{i=[T/\Delta_n]-2k_n+2}^{[T/\Delta_n]}\Bigg(\sum_{m=1}^{(i-1)\wedge(2k_n-1)}\Big(\frac{3}{2k_n^3}\sum_{j=0\vee(i+2k_n-1-m-[T/\Delta_n])}^{(2k_n-m-1)\wedge(i-m-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-1-j-m})\varepsilon(u)^n_j\\ &\times\varepsilon(v)^n_{j+m}\Big)-\sum_{j=0}^{(2k_n-m-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-1-j-m})\varepsilon(u)^n_j\varepsilon(v)^n_{j+m}\Big)\zeta_{gh}(u)^n_{i-m}\Bigg)\zeta_{ab}(v)^n_i.\end{aligned}$$

Notice that the quantity

$$\kappa^{m,n}_i = \frac{3}{2k_n^3}\Big(\sum_{j=0\vee(i+2k_n-1-m-[T/\Delta_n])}^{(2k_n-m-1)\wedge(i-m-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-1-j-m})\varepsilon(u)^n_j\varepsilon(v)^n_{j+m}\Big)$$

is $\mathcal{F}^n_{i-m-1}$ measurable and bounded by $\widetilde{\lambda}^n_{u,v}$. Let

$$\kappa^n_i = \sum_{m=1}^{(i-1)}\frac{3}{2k_n^3}\Big(\sum_{j=0\vee(i+2k_n-1-m-[T/\Delta_n])}^{(2k_n-m-1)\wedge(i-m-1)}(\partial_{gh}H\partial_{ab}G)(C^n_{i-1-j-m})\varepsilon(u)^n_j\varepsilon(v)^n_{j+m}\Big)\zeta_{gh}(u)^n_{i-m}.$$

It follows that κ_i^n is $\mathcal{F}_{i-1}^n$-measurable and we have

$$\begin{aligned}
\mathbb{E}(|\kappa_i^{m,n}|^z \big| \mathcal{F}_0) &\leq (\widetilde{\lambda}_{u,v}^n)^z, \\
|\mathbb{E}(\zeta(u)_{i-m}^n|\mathcal{F}_{i-m-1})| &\leq \begin{cases} K\sqrt{\Delta_n} & \text{if } u=1 \\ K\Delta_n & \text{if } u=2 \end{cases}, \\
\mathbb{E}(\|\zeta(u)_{i-m}^n\|^z|\mathcal{F}_{i-m-1}) &\leq \begin{cases} K_z & \text{if } u=1 \\ K_z\Delta_n^{z/2} & \text{if } u=2 \end{cases}.
\end{aligned}$$

Using Lemma 3, we deduce that for $z \geq 2$,

$$\mathbb{E}(|\kappa_i^n|^z) \leq \begin{cases} K_z(\widetilde{\lambda}_{u,v}^n)^z k_n^{z/2} & \text{if } u=1 \\ K_z(\widetilde{\lambda}_{u,v}^n)^z/k_n^{z/2} & \text{if } u=2 \end{cases} \leq \begin{cases} K_z/k_n^{-3z/2} & \text{if } v=1 \\ K_z k_n^{-z/2} & \text{if } v=2 \end{cases}.$$

Using the above result, we obtain $\frac{1}{\Delta_n^{1/4}}\widetilde{\widehat{A12}}(1) \overset{\mathbb{P}}{\Rightarrow} 0$. A similar argument yields $\frac{1}{\Delta_n^{1/4}}\widetilde{\widehat{A12}}(2) \overset{\mathbb{P}}{\Rightarrow} 0$, which completes the proof of (C.35) for $w=2$.

Proof of (C.36) for $w=1$

Define

$$\Theta(u,v)_0^{(C),i,n} = \frac{3}{2k_n^3}\sum_{j=0}^{2k_n-1}\Big((\partial_{gh}H\partial_{ab}G)(C_{i-j-1}^n) - (\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\Big)\varepsilon(u)_j^n\varepsilon(v)_j^n.$$

By Taylor expansion, the polynomial growth assumption on H and G and using (C.10) with $Z=c$, we have

$$\begin{aligned}
&\Big|\mathbb{E}\Big((\partial_{gh}H\partial_{ab}G)(C_{i-j-1}^n) - (\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)\big|\mathcal{F}_{i-2k_n}^n\Big)\Big| \leq K(k_n\Delta_n) \leq K\sqrt{\Delta_n} \\
&\mathbb{E}(|(\partial_{gh}H\partial_{ab}G)(C_{i-j-1}^n) - (\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}^n)|^q|\mathcal{F}_{i-2k_n}^n)| \leq K(k_n\Delta_n)^{q/2} \leq K\Delta_n^{q/4},
\end{aligned}$$

for $q \geq 2$ and for $j = 0,\ldots,2k_n-1$. Next, observe that $\Theta(u,v)_0^{(C),i,n}$ is $\mathcal{F}_{i-1}^n$ -measurable and satisfies $|\Theta(u,v)_0^{(C),i,n}| \leq \widetilde{\lambda}_{u,v}^n$, $|\mathbb{E}\Big(\Theta(u,v)_0^{(C),i,n}|\mathcal{F}_{i-2k_n}^n\Big)| \leq K\Delta_n^{1/2}\widetilde{\lambda}_{u,v}^n$ and $\mathbb{E}\Big(|\Theta(u,v)_0^{(C),i,n}|^q\big|\mathcal{F}_{i-2k_n}^n\Big) \leq K_q\Delta_n^{q/4}(\widetilde{\lambda}_{u,v}^n)^q$ where the latter follows from the Hölder inequality. We aim to prove that

$$\widehat{E} = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]}\Theta(u,v)_0^{(C),i,n}\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}\right]$$

converges to zero in probability for any H, G, g, h, a, and b with $u, v = 1, 2$.

To show this result, we first introduce the following quantities:

$$\widehat{E}(1) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]} \Theta(u,v)_0^{(C),i,n}\mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n)\right]$$

$$\widehat{E}(2) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]} \Theta(u,v)_0^{(C),i,n}\big(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab} - \mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n)\big)\right],$$

with $\widehat{E} = \widehat{E}(1) + \widehat{E}(2)$. By Cauchy-Schwartz inequality, we have

$$\mathbb{E}(|\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|^q) \leq (\widehat{\lambda}_{u,v}^n)^{q/2}, \text{where } \widehat{\lambda}_{u,v}^n = \begin{cases} K & \text{if } (u,v) = (1,1) \\ K\Delta_n & \text{if } (u,v) = (1,2), (2,1) \\ K\Delta_n^2 & \text{if } (u,v) = (2,2) \end{cases}$$

Since $\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}$ is $\mathcal{F}_i^n$-measurable,

the martingale property of $\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab} - \mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n)$ implies, for all (u, v),

$$\mathbb{E}(|\widehat{E}(2)|^2) \leq K\Delta_n^{-3/2}(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n)^2\widehat{\lambda}_{u,v}^n \leq K\Delta_n.$$

The latter inequality implies $\widehat{E}(2) \stackrel{\mathbb{P}}{\Rightarrow} 0$ for all (u, v). It remains to show that $\widehat{E}(1) \stackrel{\mathbb{P}}{\Rightarrow} 0$.

Here, we recall some bounds under Assumption 2,

$$|\mathbb{E}(\zeta(1)_i^{n,gh}\zeta(2)_i^{n,ab}|\mathcal{F}_{i-1}^n)| \leq K\Delta_n, \tag{C.38}$$

$$|\mathbb{E}(\zeta(1)_i^{n,gh}\zeta(1)_i^{n,ab}|\mathcal{F}_{i-1}^n) - \big(C_{i-1}^{n,ga}C_{i-1}^{n,hb} + C_{i-1}^{n,gb}C_{i-1}^{n,ha}\big)| \leq K\Delta_n^{1/2}, \tag{C.39}$$

$$|\mathbb{E}(\zeta(2)_i^{n,gh}\zeta(2)_i^{n,ab}|\mathcal{F}_{i-1}^n - \overline{C}_{i-1}^{n,gh,ab}\Delta_n)| \leq K\Delta_n^{3/2}(\sqrt{\Delta_n} + \eta_i^n). \tag{C.40}$$

Case $(u, v) \in \{(1, 2), (2, 1)\}$. By (C.38) we have

$$\mathbb{E}(|\widehat{E}(1)|) \leq K\frac{T}{\Delta_n}\frac{1}{\Delta_n^{1/4}}(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n\Delta_n) \leq K\Delta_n^{1/2} \quad \text{so} \quad \widehat{E}(1) \stackrel{\mathbb{P}}{\Rightarrow} 0.$$

Case $(u, v) \in \{(1, 1), (2, 2)\}$. Set

$$\widehat{E}'(1) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]} \Theta(u,v)_0^{(C),i,n}V_{i-2k_n}^n\right]$$

$$\widehat{E}''(1) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]} \Theta(u,v)_0^{(C),i,n}\big(V_{i-1}^n - V_{i-2k_n}^n\big)\right]$$

$$\widehat{E}'''(1) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=2k_n}^{[T/\Delta_n]} \Theta(u,v)_0^{(C),i,n}\Big(\mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n) - V_{i-1}^n\Big)\right]$$

where

$$V_{i-1}^n = \begin{cases} C_{i-1}^{n,ga}C_{i-1}^{n,hb} + C_{i-1}^{n,gb}C_{i-1}^{n,ha} & \text{if } (u,v) = (2,2) \\ \overline{C}_{i-1}^{n,gh,ab}\Delta_n & \text{if } (u,v) = (1,1) \\ 0 & \text{otherwise} \end{cases}$$

Note that we have $\widehat{E}(1) = \widehat{E}'(1) + \widehat{E}''(1) + \widehat{E}'''(1)$. Using (C.39) and (C.40), it can be shown that

$$\mathbb{E}(|\widehat{E}'''(1)|) \leq \begin{cases} K\frac{1}{\Delta_n^{5/4}}(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n)\Delta_n^{1/2} & \text{if } (u,v) = (1,1) \\ K\frac{1}{\Delta_n^{5/4}}(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n)\Delta_n^{3/2} & \text{if } (u,v) = (2,2) \end{cases} \leq K\Delta_n^{1/2} \quad \text{in all cases.}$$

Next, we prove $\widehat{E}'(1) \stackrel{\mathbb{P}}{\Rightarrow} 0$. To this end, write

$$\widehat{E}'(1) = \frac{1}{\Delta_n^{1/4}}\left[\sum_{i=1}^{[T/\Delta_n]-2k_n+1} \Theta(u,v)_0^{(C),i-1+2k_n,n}V_{(i-1)\Delta_n}\right].$$

Using the $\mathcal{F}_{i+2k_n-2}^n$-measurability of the last sum, we are able to show

$$\frac{1}{\Delta_n^{1/4}}\left[\sum_{i=1}^{[T/\Delta_n]-2k_n+1} |\mathbb{E}(\Theta(u,v)_0^{(C),i-1+2k_n,n}V_{(i-1)\Delta_n}|\mathcal{F}_{i-1}^n)|\right] \stackrel{\mathbb{P}}{\Rightarrow} 0 \quad \text{and}$$
$$\frac{2k_n-2}{\Delta_n^{1/2}}\left[\sum_{i=1}^{[T/\Delta_n]-2k_n+1} \mathbb{E}\Big(|\Theta(u,v)_0^{(C),i-1+2k_n,n}V_{(i-1)\Delta_n})|^2\Big)\right] \Rightarrow 0.$$

The first result readily follows from the inequality

$$|\mathbb{E}(\Theta(u,v)_0^{(C),i-1+2k_n,n}V_{(i-1)\Delta_n}|\mathcal{F}_{i-1}^n)| \leq \begin{cases} K\Delta_n^{1/2}\widetilde{\lambda}_{u,v}^n & \text{if } (u,v) = (1,1) \\ K\Delta_n^{1/2}\widetilde{\lambda}_{u,v}^n\Delta_n & \text{if } (u,v) = (2,2) \end{cases} \leq K\Delta_n^{3/2} \quad \text{in all cases,}$$

while the second is a direct consequence of

$$\mathbb{E}(|\Theta(u,v)_0^{(C),i-1+2k_n,n}V_{(i-1)\Delta_n}|^2) \leq \begin{cases} K\Delta_n^{1/2}(\widetilde{\lambda}_{u,v}^n)^2 & \text{if } (u,v) = (1,1) \\ K\Delta_n^{1/2}(\widetilde{\lambda}_{u,v}^n)^2\Delta_n^2 & \text{if } (u,v) = (2,2) \end{cases} \leq K\Delta_n^{5/2} \quad \text{in all cases.}$$

Finally, to prove that $\widehat{E}''(1) \stackrel{\mathbb{P}}{\Longrightarrow} 0$, we use the fact that

$$\begin{aligned}\mathbb{E}(|\Theta(u,v)_0^{(C),i,n}\big(V_{(i-1)\Delta_n} - V_{(i-2k_n)\Delta_n}\big)|) &\leq \mathbb{E}(|\Theta(u,v)_0^{(C),i,n}|^2)^{1/2}\mathbb{E}(|V_{(i-1)\Delta_n} - V_{(i-2k_n)\Delta_n}|^2)^{1/2} \\ &\leq \begin{cases} K\Delta_n^{1/2}\widetilde{\lambda}_{u,v}^n & \text{if } (u,v)=(1,1) \\ K\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n\Delta_n\Delta_n^{1/4} & \text{if } (u,v)=(2,2) \end{cases},\end{aligned}$$

which follows from the Cauchy-Schwartz inequality and earlier bounds. In particular, successive conditioning together with Assumption 2 imply that for $(u,v)=(1,1)$ and $(2,2)$,
$\mathbb{E}(|V_{(i-1)\Delta_n} - V_{(i-2k_n)\Delta_n}|^2) \leq \Delta_n^{1/2}$.

Proof of (C.36) for $w = 2$

Our aim here is to show that

$$\widehat{E}(2) = \frac{1}{\Delta_n^{1/4}}\sum_{i=2k_n}^{[T/\Delta_n]}\Bigg(\sum_{m=1}^{2k_n-1}\Big(\frac{3}{2k_n^3}\sum_{j=0}^{2k_n-m-1}\big[(\partial_{gh}H\partial_{ab}G)(c_{i-j-m-1}^n) - (\partial_{gh}H\partial_{ab}G)(c_{i-2k_n}^n)\big]\varepsilon(u)_j^n\varepsilon(v)_{j+m}^n\Big)\times$$
$$\zeta(u)_{i-m}^{n,gh}\Bigg)\zeta(v)_i^{n,ab} \stackrel{\mathbb{P}}{\Longrightarrow} 0.$$

For this purpose, we introduce some new notation. For any $0 \leq m \leq 2k_n - 1$, set

$$\begin{aligned}\Theta(u,v)_m^{(C),i,n} &= \frac{3}{2k_n^3}\sum_{j=0}^{2k_n-m-1}\big[(\partial_{gh}H\partial_{ab}G)(c_{i-j-m-1}^n) - (\partial_{gh}H\partial_{ab}G)(c_{i-2k_n}^n)\big]\varepsilon(u)_j^n\varepsilon(v)_{j+m}^n \\ \rho(u,v)^{(C),i,n,gh} &= \sum_{m=1}^{2k_n-1}\Theta(u,v)_m^{(C),i,n}\zeta(u)_{i-m}^{n,gh}.\end{aligned}$$

It is easy to see that $\Theta(u,v)_m^{(C),i,n}$ is $\mathcal{F}_{i-m-1}^n$ measurable and satisfies, by Hölder inequality,

$$|\Theta(u,v)_m^{(C),i,n}| \leq \widetilde{\lambda}_{u,v}^n \ \text{ and } \ \mathbb{E}\Big(|\Theta(u,v)_m^{(C),i,n}|^q\big|\mathcal{F}_{i-2k_n}^n\Big) \leq K_q\Delta_n^{q/4}(\widetilde{\lambda}_{u,v}^n)^q.$$

Lemma 3 implies that for $q \geq 2$,

$$\mathbb{E}(|\rho(u,v)^{(C),i,n,gh}|^q) \leq \begin{cases} K_q(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n)^q k_n^{q/2} & \text{if } u=1 \\ K_q(\Delta_n^{1/4}\widetilde{\lambda}_{u,v}^n)^q/k_n^{q/2} & \text{if } u=2 \end{cases} \leq \begin{cases} K_q/k_n^{2q} & \text{if } v=1 \\ K_q k_n^q & \text{if } v=2 \end{cases}. \tag{C.41}$$

Set

$$\widehat{E}'(2) = \frac{1}{\Delta_n^{1/4}} \sum_{i=2k_n}^{[T/\Delta_n]} \rho(u,v)^{(C),i,n,gh} \mathbb{E}(\zeta(v)_i^{n,ab} | \mathcal{F}_{i-1}^n),$$
$$\widehat{E}''(2) = \frac{1}{\Delta_n^{1/4}} \sum_{i=2k_n}^{[T/\Delta_n]} \rho(u,v)^{(C),i,n,gh} (\zeta(v)_i^{n,ab} - \mathbb{E}(\zeta(v)_i^{n,ab} | \mathcal{F}_{i-1}^n)).$$

The martingale increments property implies $\mathbb{E}(|\widehat{E}''(2)|^2) \leq K\Delta_n^{1/2}$ in all the cases, which in turn implies $\widehat{E}''(2) \stackrel{\mathbb{P}}{\Longrightarrow} 0$. Next, using the bounds on $\rho(u,v)^{(C),i,n,gh}$, we obtain that $\widehat{E}'(2) \stackrel{\mathbb{P}}{\Longrightarrow} 0$.

We refer to Jacod and Rosenbaum (2015) for the proofs of Lemma 1 and Lemma 2.

A3.4 Proof of Lemma 3

Set

$$\xi_i^n = \varphi_{i-1}^n \zeta_i^n, \quad \xi_i^{'n} = \mathbb{E}(\xi_i | \mathcal{F}_{i-1}^n) = \mathbb{E}(\varphi_{i-1}^n \zeta_i^n | \mathcal{F}_{i-1}^n) = \varphi_{i-1}^n \mathbb{E}(\zeta_i^n | \mathcal{F}_{i-1}^n), \text{ and } \xi_i^{''n} = \xi_i^n - \xi_i^{'n}.$$

Given that $\|\mathbb{E}(\zeta_i^n | \mathcal{F}_{i-1}^n)\| \leq L'$, we have $\|\xi_i^{'n}\| \leq L' |\varphi_{i-1}^n|$. By the convexity of the function x^q, which holds for $q \geq 2$, we have

$$\| \sum_{j=1}^{2k_n-1} \xi_{i+j}^n \|^q \leq K \Big(\| \sum_{j=1}^{2k_n-1} \xi_{i+j}^{'n} \|^q + \| \sum_{j=1}^{2k_n-1} \xi_{i+j}^{''n} \|^q \Big).$$

Therefore, on the one hand we have

$$\| \sum_{j=1}^{2k_n-1} \xi_{i+j}^{'n} \|^q \leq K k_n^{q-1} \sum_{j=1}^{2k_n-1} \| \xi_{i+j}^{'n} \|^q \leq K k_n^{q-1} L'^q \sum_{j=1}^{2k_n-1} |\varphi_{i+j-1}^n|^q,$$

which by $\mathbb{E}\Big(\|\varphi_{i+j-1}^n\|^q \Big| \mathcal{F}_{i-1}^n \Big) \leq L^q$, satisfies

$$\mathbb{E}(\| \sum_{j=1}^{2k_n-1} \xi_{i+j}^{'n} \|^q | \mathcal{F}_{i-1}^n) \leq K L'^q k_n^{q-1} \sum_{j=1}^{2k_n-1} \mathbb{E}(|\varphi_{i+j-1}^n|^q | \mathcal{F}_{i-1}^n) \leq K L'^q k_n^q L^q.$$

On the other hand, we have $\mathbb{E}(\|\xi_{i+j}^{''n}\|^q | \mathcal{F}_{i-1}^n) \leq \mathbb{E}(\|\xi_{i+j}^n\|^q | \mathcal{F}_{i-1}^n) \leq L_q L^q$ and $\mathbb{E}(\xi_{i+j}^{''n} | \mathcal{F}_{i-1}^n) = 0$, where the first inequality is a consequence of $\mathbb{E}(\|\xi_{i+j}^{'n}\|^q | \mathcal{F}_{i-1}^n) \leq \mathbb{E}(\|\xi_{i+j}^n\|^q | \mathcal{F}_{i-1}^n) \leq L_q L^q$, which follows from the Jensen's inequality and the law of iterated expectation. Hence, by Lemma B.2 of Aït-Sahalia and

Jacod (2014) we have

$$\mathbb{E}(\|\sum_{j=1}^{2k_n-1}\xi''^{n}_{i+j}\|^q|\mathcal{F}^n_{i-1})\leq K_qL^qL_qk_n^{q/2}.$$

To see the latter, we first prove that the required condition $\mathbb{E}(\|\xi^n_i\|^q|\mathcal{F}^n_{i-1})\leq L_qL^q)$ in the Lemma B.2 of Aït-Sahalia and Jacod (2014) can be replaced by $\mathbb{E}(\|\xi^n_{i+j}\|^q|\mathcal{F}^n_{i-1})\leq L_qL^q)$ for $1\leq j\leq 2k_n-1$ without altering the result.

A3.5 Proof of Lemma 4

We use the terminology "successive conditioning" to refer to either of the following two equalities,

$$\begin{aligned}
x_1y_1-x_0y_0 &= x_0(y_1-y_0)+y_0(x_1-x_0)+(x_1-x_0)(y_1-y_0),\\
x_1y_1z_1-x_0y_0z_0 &= x_0y_0(z_1-z_0)+x_0z_0(y_1-y_0)+y_0z_0(x_1-x_0)+x_0(y_0-y_1)(z_0-z_1)\\
&\quad +y_0(x_0-x_1)(z_0-z_1)+z_0(x_0-x_1)(y_0-y_1)+(x_1-x_0)(y_1-y_0)(z_1-z_0),
\end{aligned}$$

which hold for any real numbers x_0,y_0,z_0,x_1,y_1, and z_1.

To prove Lemma 4, we first note that $\lambda_i^{n,jk}\lambda_i^{n,lm}$ is $\mathcal{F}^n_{i+2k_n}$-measurable. Therefore, by the law of iterated expectations, we have

$$\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\lambda^{n,gh}_{i+2k_n}\lambda^{n,ab}_{i+2k_n}|\mathcal{F}^n_i\Big)=\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\mathbb{E}(\lambda^{n,gh}_{i+2k_n}\lambda^{n,ab}_{i+2k_n}|\mathcal{F}^n_{i+2k_n})|\mathcal{F}^n_i\Big).$$

By equation (3.27) in Jacod and Rosenbaum (2015), we have

$$\begin{aligned}
&|\mathbb{E}(\lambda^{n,gh}_{i+2k_n}\lambda^{n,ab}_{i+2k_n}|\mathcal{F}^n_{i+2k_n})-\frac{2}{k_n}(C^{n,ga}_{i+2k_n}C^{n,hb}_{i+2k_n}+C^{n,gb}_{i+2k_n}C^{n,ha}_{i+2k_n})-\frac{2k_n\Delta_n}{3}\overline{C}^{n,gh,ab}_{i+2k_n}|\\
&\quad\leq K\sqrt{\Delta_n}(\Delta_n^{1/8}+\eta^n_{i+2k_n,2k_n}),\quad\text{and}\\
&|\mathbb{E}(\lambda_i^{n,jk}\lambda_i^{n,lm}|\mathcal{F}^n_i)-\frac{2}{k_n}(C_i^{n,jl}C_i^{n,km}+C_i^{n,jm}C_i^{n,kl})-\frac{2k_n\Delta_n}{3}\overline{C}_i^{n,jk,lm}|\leq K\sqrt{\Delta_n}(\Delta_n^{1/8}+\eta^n_{i,2k_n}).
\end{aligned}$$

From the above, it follows that

$$\begin{aligned}
&|\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\Big[\mathbb{E}(\lambda^{n,gh}_{i+2k_n}\lambda^{n,ab}_{i+2k_n}\Big|\mathcal{F}^n_{i+2k_n})-\frac{2}{k_n}(C^{n,ga}_{i+2k_n}C^{n,hb}_{i+2k_n}+C^{n,gb}_{i+2k_n}C^{n,ha}_{i+2k_n})-\frac{2k_n\Delta_n}{3}\overline{C}^{n,gh,ab}_{i+2k_n}\Big]\Big|\mathcal{F}^n_i\Big)|\\
&\quad\leq\sqrt{\Delta_n}\mathbb{E}(|\lambda_i^{n,jk}||\lambda_i^{n,lm}|(\Delta_n^{1/8}+\eta^n_{i+2k_n,2k_n})|\Big|\mathcal{F}^n_i)\leq K\sqrt{\Delta_n}\Delta_n^{1/8}\mathbb{E}(|\lambda_i^{n,jk}||\lambda_i^{n,lm}|\Big|\mathcal{F}^n_i)\\
&\quad+K\sqrt{\Delta_n}\mathbb{E}(|\lambda_i^{n,jk}||\lambda_i^{n,lm}|\eta^n_{i+2k_n,2k_n}|\Big|\mathcal{F}^n_i)\leq K\Delta_n(\Delta_n^{1/8}+\eta^n_{i,4k_n}),
\end{aligned}$$

where the last inequality follows from Lemma 1.

Now, using (C.10) successively with $Z = c$ and $Z = \overline{C}$ (recall that the latter holds under Assumption 2), together with the successive conditioning, we also have

$$
\begin{aligned}
&|\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\Big[\frac{2}{k_n}(C_{i+2k_n}^{n,ga}C_{i+2k_n}^{n,hb}+C_{i+2k_n}^{n,gb}C_{i+2k_n}^{n,ha})+\frac{2k_n\Delta_n}{3}\overline{C}_{i+2k_n}^{n,gh,ab}-\frac{2}{k_n}(C_i^{n,ga}C_i^{n,hb}+C_i^{n,gb}C_i^{n,ha}) \\
&-\frac{2k_n\Delta_n}{3}\overline{C}_i^{n,gh,ab}\Big]\Big|\mathcal{F}_i^n\Big)| \le K\Delta_n\Delta_n^{1/4}, \\
&|\mathbb{E}\Big(\lambda_i^{n,jk}\lambda_i^{n,lm}\Big[\frac{2}{k_n}(C_i^{n,ga}C_i^{n,hb}+C_i^{n,gb}C_i^{n,ha}) \\
&+\frac{2k_n\Delta_n}{3}\overline{C}_i^{n,gh,ab}\Big]-\Big[\frac{2}{k_n}(C_i^{n,jl}C_i^{n,km}+C_i^{n,jm}C_i^{n,kl})+\frac{2k_n\Delta_n}{3}\overline{C}_i^{n,jk,lm}\Big] \\
&\times\Big[\frac{2}{k_n}(C_i^{n,ga}C_i^{n,hb}+C_i^{n,gb}C_i^{n,ha})+\frac{2k_n\Delta_n}{3}\overline{C}_i^{n,gh,ab}\Big]\Big|\mathcal{F}_i^n\Big)| \le K\Delta_n(\Delta_n^{1/8}+\eta_{i,2k_n}^n).
\end{aligned}
$$

The result derives from the last inequality.

Proof of (C.12) in Lemma 5

We start by obtaining some useful bounds for some important quantities. First, using the second statement in Lemma 2 applied to $Z = Y'$, we have

$$|\mathbb{E}(\alpha_i^{n,jk}|\mathcal{F}_i^n)| \le K\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta_{i,1}^n). \tag{C.42}$$

Second, by repeated application of the Cauchy-Schwartz inequality and making use of the third and last statements in Lemma 2 as well as (C.10) with $Z = c$, it can be shown that

$$\Big|\mathbb{E}(\alpha_i^{n,jk}\alpha_i^{n,lm}|\mathcal{F}_i^n)-\Delta_n^2\Big(C_i^{n,jl}C_i^{n,km}+C_i^{n,jm}C_i^{n,kl}\Big)\Big| \le K\Delta_n^{5/2}. \tag{C.43}$$

Next, by successive conditioning and using the bound in (C.10) for $Z = c$ as well as (C.42) and (C.43) , we have for $0 \le u \le k_n - 1$,

$$\Big|\mathbb{E}(\alpha_{i+u}^{n,jk}|\mathcal{F}_i^n)\Big| \le K\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta_{i,u}^n), \tag{C.44}$$

$$\Big|\mathbb{E}(\alpha_{i+u}^{n,jk}\alpha_{i+u}^{n,lm}|\mathcal{F}_i^n)-\Delta_n^2\Big(C_i^{n,jl}C_i^{n,km}+C_i^{n,jm}C_i^{n,kl}\Big)\Big| \le K\Delta_n^{5/2}. \tag{C.45}$$

To show (C.12), we first observe that $\nu_i^{n,jk}\nu_i^{n,lm}\nu_i^{n,gh}$ can be decomposed as

$$\nu_i^{n,jk}\nu_i^{n,lm}\nu_i^{n,gh} = \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,u}^{n,lm}\zeta_{i,u}^{n,gh} + \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-2}\sum_{v=u+1}^{k_n-1}\Big[\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm}\zeta_{i,v}^{n,gh} + \zeta_{i,u}^{n,gh}\zeta_{i,v}^{n,jk}\zeta_{i,v}^{n,lm}$$
$$+ \zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,gh}\zeta_{i,v}^{n,jk}\Big] + \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-2}\sum_{v=u+1}^{k_n-1}[\zeta_{i,u}^{n,jk}\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,gh} + \zeta_{i,u}^{n,gh}\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm} + \zeta_{i,u}^{n,lm}\zeta_{i,u}^{n,gh}\zeta_{i,v}^{n,jk}\Big]$$
$$+ \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-3}\sum_{v=u+1}^{k_n-2}\sum_{w=v+1}^{k_n-1}\Big[\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm}\zeta_{i,w}^{n,gh} + \zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,gh}\zeta_{i,w}^{n,lm} + \zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,jk}\zeta_{i,w}^{n,gh} + \zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,gh}\zeta_{i,w}^{n,jk}$$
$$+ \zeta_{i,u}^{n,gh}\zeta_{i,v}^{n,lm}\zeta_{i,w}^{n,jk} + \zeta_{i,u}^{n,gh}\zeta_{i,v}^{n,jk}\zeta_{i,w}^{n,lm}\Big],$$

with $\zeta_{i,u}^n = \alpha_{i+u}^n + (C_{i+u}^n - C_i^n)\Delta_n$, which satisfies $\mathbb{E}(\|\zeta_{i,u}^n\|^q|\mathcal{F}_i^n) \le K\Delta_n^q$ for $q \ge 2$.
Set

$$\xi_i^n(1) = \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,u}^{n,lm}\zeta_{i,u}^{n,gh}, \quad \xi_i^n(2) = \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-2}\sum_{v=u+1}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm}\zeta_{i,v}^{n,gh}$$
$$\xi_i^n(3) = \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-2}\sum_{v=u+1}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,gh} \quad \text{and} \quad \xi_i^n(4) = \frac{1}{k_n^3\Delta_n^3}\sum_{u=0}^{k_n-3}\sum_{v=u+1}^{k_n-2}\sum_{w=v+1}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm}\zeta_{i,w}^{n,gh}.$$

The following bounds can be established,

$$|\mathbb{E}(\xi_i^n(1)|\mathcal{F}_i^n)| \le K\Delta_n \tag{C.46}$$
$$|\mathbb{E}(\xi_i^n(2)|\mathcal{F}_i^n)| \le K\Delta_n \tag{C.47}$$
$$|\mathbb{E}(\xi_i^n(3)|\mathcal{F}_i^n)| \le K\Delta_n \tag{C.48}$$
$$|\mathbb{E}(\xi_i^n(4)|\mathcal{F}_i^n)| \le K\Delta_n^{3/4}(\Delta_n^{1/4} + \eta_{i,k_n}). \tag{C.49}$$

Proof of (C.46)

The result readily follows from an application of the Cauchy Schwartz inequality coupled with the bound $\mathbb{E}(\|\zeta_{i+u}^n\|^q|\mathcal{F}_i^n) \le K_q\Delta_n^q$ for $q \ge 2$.

Proof of (C.47)

Using the law of iterated expectation, we have, for $u < v$,

$$\mathbb{E}(\zeta_{i+u}^{n,jk}\zeta_{i+v}^{n,lm}\zeta_{i+v}^{n,gh}|\mathcal{F}_i^n) = \mathbb{E}(\zeta_{i+u}^{n,jk}\mathbb{E}(\zeta_{i+v}^{n,lm}\zeta_{i+v}^{n,gh}|\mathcal{F}_{i+u+1}^n)\big|\mathcal{F}_i^n). \tag{C.50}$$

By successive conditioning, (C.43), and the Cauchy-Schwartz inequality, we also have

$$\begin{aligned}&|\mathbb{E}(\zeta_{i,v}^{n,lm}\zeta_{i,v}^{n,gh}|\mathcal{F}_{i+u+1}^n) - \Delta_n^2(C_{i+u+1}^{n,lg}C_{i+u+1}^{n,mh} + C_{i+u+1}^{n,lh}C_{i+u+1}^{n,mg})\\ &- \Delta_n^2(C_{i+u+1}^{n,gh} - C_i^{n,gh})(C_{i+u+1}^{n,lm} - C_i^{n,lm})| \leq K\Delta_n^{5/2}.\end{aligned}$$

Given that $\mathbb{E}(|\zeta_{i+u}^{n,jk}|^q\big|\mathcal{F}_i^n) \leq \Delta_n^q$, the approximation error involved in replacing $\mathbb{E}(\zeta_{i+v}^{n,lm}\zeta_{i+v}^{n,gh}|\mathcal{F}_{i+u+1}^n)$ by
$\Delta_n^2(C_{i+u+1}^{n,lg}C_{i+u+1}^{n,mh} + C_{i+u+1}^{n,lh}C_{i+u+1}^{n,mg}) + \Delta_n^2(C_{i+u+1}^{n,gh} - C_i^{n,gh})(C_{i+u+1}^{n,lm} - C_i^{n,lm})$ in (C.50) is smaller than $\Delta_n^{7/2}$.
We can also easily show that

$$|\mathbb{E}(\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lm} - C_{i+u}^{n,lm})|\mathcal{F}_i^n)| \leq K\Delta_n^{3/2}(\sqrt{\Delta_n} + \eta_{i,k_n}^n). \tag{C.51}$$

Since $(C_{i+u}^n - C_i^n)$ is $\mathcal{F}_{i+u}^n$-measurable, we use the successive conditioning, the Cauchy-Schwartz inequality, (C.42), (C.43), and the fifth statement in Lemma 2 applied to $Z = c$ to obtain

$$\begin{aligned}|\mathbb{E}(\alpha_{i+u}^{n,gh}(C_{i+u}^{n,lm} - C_i^{n,lm})(C_{i+u}^{n,jk} - C_i^{n,jk})|\mathcal{F}_i^n)| &\leq K\Delta_n^{5/2}\\ |\mathbb{E}(\alpha_{i+u}^{n,jk}\alpha_{i+u}^{n,lm}(C_{i+u}^{n,gh} - C_i^{n,gh})|\mathcal{F}_i^n)| &\leq K\Delta_n^{5/2}\\ |\mathbb{E}\big((C_{i+u}^{n,lm} - C_i^{n,lm})(C_{i+u}^{n,jk} - C_i^{n,jk})(C_{i+u}^{n,gh} - C_i^{n,gh})\big)|\mathcal{F}_i^n)| &\leq K\Delta_n.\end{aligned} \tag{C.52}$$

The following inequalities can be established using (C.42), the successive conditioning together with (C.10) for $Z = c$,

$$\begin{aligned}\Big|\mathbb{E}(\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lg}C_{i+u+1}^{n,mh} + C_{i+u+1}^{n,lh}C_{i+u+1}^{n,mg})|\mathcal{F}_i^n)\Big| &\leq K\Delta_n^{3/2}\\ \Big|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_i^{n,jk})\big(C_{i+u+1}^{n,lg}C_{i+u+1}^{n,mh} + C_{i+u+1}^{n,lh}C_{i+u+1}^{n,mg}\big)|\mathcal{F}_i^n\Big)\Big| &\leq K\Delta_n^{1/2}\\ \Big|\mathbb{E}(\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,gh} - C_i^{n,gh})(C_{i+u+1}^{n,lm} - C_i^{n,lm})|\mathcal{F}_i^n)\Big| &\leq K\Delta_n^{3/2}(\sqrt{\Delta_n} + \eta_{i,k_n}^n).\end{aligned}$$

The last three inequalities together yield $|\mathbb{E}(\xi_i^n(2)|\mathcal{F}_i^n)| \leq K\Delta_n$.

Proof of (C.48)

First, note that, for $u < v$, we have

$$\mathbb{E}(\zeta_{i+u}^{n,jk}\zeta_{i+u}^{n,lm}\zeta_{i+v}^{n,gh}|\mathcal{F}_i^n) = \mathbb{E}(\zeta_{i+u}^{n,jk}\zeta_{i+u}^{n,lm}\mathbb{E}(\zeta_{i+v}^{n,gh}|\mathcal{F}_{i+u+1}^n)\big|\mathcal{F}_i^n). \tag{C.53}$$

By successive conditioning and (C.42), we have

$$|\mathbb{E}(\alpha_{i+w}^{n,gh}|\mathcal{F}_{i+v+1}^n)| \leq K\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta_{i+v+1,w-v}). \tag{C.54}$$

Using the first statement of Lemma applied to $Z=c$, it can be shown that

$$\begin{aligned} & |\mathbb{E}\big((C_{i+w}^{n,gh}-C_{i+v+1}^{n,gh}))|\mathcal{F}_i^n\big)-\Delta_n(w-v-1)\widetilde{b}_{i+v+1}^{n,gh}| \\ \leq \quad & K(w-v-1)\Delta_n\eta_{i+v+1,w-v} \leq K\Delta_n^{1/2}\eta_{i+v+1,w-v}. \end{aligned}$$

The last two inequalities together imply

$$\left|\mathbb{E}\Big(\zeta_{i+w}^{n,gh}|\mathcal{F}_{i+v+1}^n\Big)-(C_{i+v+1}^{n,gh}-C_i^{n,gh})\Delta_n-\Delta_n^2(w-v-1)\widetilde{b}_{i+v+1}^{n,gh}\right| \leq K\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta_{i+v+1,w-v}). \tag{C.55}$$

Since $\mathbb{E}(|\zeta_{i,u}^{n,jk}|^q|\mathcal{F}_i^n)\leq\Delta_n^q$, the error induced by replacing $\mathbb{E}(\zeta_{i+v}^{n,gh}|\mathcal{F}_{i+u+1}^n)$ by $(C_{i+v+1}^{n,gh}-C_i^{n,gh})\Delta_n+\Delta_n^2(w-v-1)\widetilde{b}_{i+v+1}^{n,gh}$ in (C.53) is smaller that $\Delta_n^{7/2}$.
Using Cauchy Schwartz inequality, successive conditioning, (C.52), (C.10) for $Z=c$ and the boundedness of $\widetilde{b}_t$ and C_t we obtain

$$\begin{aligned} \left|\mathbb{E}\Big(\alpha_{i+u}^{n,jk}\alpha_{i+u}^{n,lm}(C_{i+u+1}^{n,jk}-C_i^{n,gh})|\mathcal{F}_{i+u}^n\Big)\right| &\leq K\Delta_n^{5/2} \\ \left|\mathbb{E}\Big(\alpha_{i+u}^{n,jk}\alpha_{i+u}^{n,lm}\widetilde{b}_{i+u+1}^{n,gh}|\mathcal{F}_{i+u}^n\Big)\right| &\leq K\Delta_n^2 \\ \left|\mathbb{E}\Big(\alpha_{i+u}^{n,jk}(C_{i+u}^{n,lm}-C_i^{n,lm})(C_{i+u+1}^{n,gh}-C_i^{n,gh})|\mathcal{F}_i^n\Big)\right| &\leq K\Delta_n^{1/4}\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta_{i,k_n}^n) \\ \left|\mathbb{E}\Big(\alpha_{i+u}^{n,jk}(C_{i+u}^{n,lm}-C_i^{n,lm})\widetilde{b}_{i+u+1}^{n,gh}|\mathcal{F}_i^n\Big)\right| &\leq \Delta_n^{5/4} \\ \left|\mathbb{E}\Big((C_{i+u}^{n,jk}-C_i^{n,gh})(C_{i+u}^{n,lm}-C_i^{n,lm})\widetilde{b}_{i+u+1}^{n,gh}|\mathcal{F}_i^n\Big)\right| &\leq K\Delta_n^{1/2} \\ \left|\mathbb{E}\Big((C_{i+u}^{n,jk}-C_i^{n,jk})(C_{i+u}^{n,lm}-C_i^{n,lm})(C_{i+u+1}^{n,gh}-C_i^{n,gh})|\mathcal{F}_i^n\Big)\right| &\leq K\Delta_n. \end{aligned}$$

The above inequalities together yield $|\mathbb{E}(\xi_i^n(3)|\mathcal{F}_i^n)|\leq K\Delta_n$.

Proof of (C.49)

We first observe that $\xi_i^n(4)$ can be rewritten as

$$\xi_i^n(4)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\zeta_{i+u}^{n,jk}\zeta_{i+v}^{n,lm}\zeta_{i+w}^{n,gh},$$

where

$$\zeta_{i+u}^{n,jk}\zeta_{i+v}^{n,lm}\zeta_{i+w}^{n,gh} = \Big[\alpha_{i+u}^{n,jk}\alpha_{i+v}^{n,lm}\alpha_{i+w}^{n,gh} + \alpha_{i+u}^{n,jk}\Delta_n\alpha_{i+v}^{n,lm}(C_{i+w}^{n,gh} - C_i^{n,gh}) + \alpha_{i+u}^{n,jk}\Delta_n(C_{i+v}^{n,lm} - C_i^{n,lm})\alpha_{i+w}^{n,gh}$$
$$+ \Delta_n^2\alpha_{i+u}^{n,jk}(C_{i+v}^{n,lm} - C_i^{n,lm})(C_{i+w}^{n,gh} - C_i^{n,gh}) + \Delta_n(C_{i+u}^{n,jk} - C_i^{n,jk})\alpha_{i+v}^{n,lm}\alpha_{i+w}^{n,gh}$$
$$+ \Delta_n^2(C_{i+u}^{n,jk} - C_i^{n,jk})\alpha_{i+v}^{n,lm}(C_{i+w}^{n,gh} - C_i^{n,gh}) + \Delta_n^2(C_{i+u}^{n,jk} - C_i^{n,jk})(C_{i+v}^{n,lm} - C_i^{n,lm})\alpha_{i+w}^{n,gh}$$
$$+ \Delta_n^3(C_{i+u}^{n,jk} - C_i^{n,jk})(C_{i+v}^{n,lm} - C_i^{n,lm})(C_{i+w}^{n,gh} - C_i^{n,gh})\Big].$$

Based on the above decomposition, we set

$$\xi_i^n(4) = \sum_{j=1}^{8}\chi(j),$$

with $\chi(j)$ defined below. We aim to show that $|\mathbb{E}(\chi(j)\big|\mathcal{F}_i^n)| \leq K\Delta_n^{3/4}(\Delta_n^{1/4} + \eta_{i,k_n}^n)$, $j = 1,\ldots,8$. First, set

$$\chi(1) = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\alpha_{i+v}^{n,lm}\alpha_{i+w}^{n,gh}.$$

Upon changing the order of the summation, we have

$$\chi(1) = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\alpha_{i+v}^{n,lm}\alpha_{i+w}^{n,gh}.$$

Define also

$$\chi'(1) = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\alpha_{i+v}^{n,lm}\mathbb{E}(\alpha_{i+w}^{n,gh}|\mathcal{F}_{i+v+1}^n).$$

Note that $\mathbb{E}(\chi(1)|\mathcal{F}_i^n) = \mathbb{E}(\chi'(1)|\mathcal{F}_i^n)$.

By Lemma 3, we have for $q \geq 2$,

$$\mathbb{E}\Big(\Big\|\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big\|^q\Big|\mathcal{F}_i^n\Big) \leq K_q\Delta_n^{3q/4}.$$

The Cauchy-Schwartz inequality yields

$$\mathbb{E}\Bigg(\Big|\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\alpha_{i+v}^{n,lm}\mathbb{E}(\alpha_{i+w}^{n,gh}|\mathcal{F}_{i+v+1}^n)\Big|\Bigg|\mathcal{F}_i^n\Bigg) \leq Kk_n^2\Big[\mathbb{E}\Big(\Big|\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big|^4\Big|\mathcal{F}_i^n\Big)\Big]^{1/4}$$

$$\times\Big[\mathbb{E}\Big(\Big|\alpha^{n,lm}_{i+v}\Big|^4\Big|\mathcal{F}^n_i\Big)\Big]^{1/4}\times\Big[\mathbb{E}\Big(\Big|\mathbb{E}(\alpha^{n,gh}_{i+w}|\mathcal{F}^n_{i+v+1})\Big|^2\Big|\mathcal{F}^n_i\Big)\Big]^{1/2}\leq K\Delta_n k_n^2\Delta_n^{3/4}\Delta_n^{3/2}(\sqrt{\Delta_n}+\eta^n_{i,k_n}),$$

where the last iteration is obtained using (C.54) as well as the inequality $(a+b)^{1/2}\leq a^{1/2}+b^{1/2}$, which holds for positive real numbers a and b, and the third statement in Lemma 1. It follows that

$$|\mathbb{E}\Big(\chi(1)\big|\mathcal{F}^n_i\Big)|\leq K\Delta_n^{3/4}(\sqrt{\Delta_n}+\eta^n_{i,k_n}).$$

Next, we introduce

$$\chi(2)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\Delta_n(C^{n,jk}_{i+u}-C^{n,jk}_i)\Big)\alpha^{n,lm}_{i+v}\alpha^{n,gh}_{i+w},$$
$$\chi(3)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha^{n,jk}_{i+v}\Big)\Delta_n(C^{n,lm}_{i+u}-C^{n,lm}_i)\alpha^{n,gh}_{i+w},$$
$$\chi(4)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\Delta_n(C^{n,jk}_{i+u}-C^{n,jk}_i)\Big)\Delta_n(C^{n,lm}_{i+u}-C^{n,lm}_i)\alpha^{n,gh}_{i+w}.$$

Given that for $q\geq 2$, we have

$$\mathbb{E}\Big(\Big\|\sum_{u=0}^{v-1}\Delta_n(C^{n,jk}_{i+u}-C^{n,jk}_i)\Big\|^q\Big|\mathcal{F}^n_i\Big)\leq K_q\Delta_n^{3q/4}\quad\text{and}\quad\mathbb{E}(\|C^{n,jk}_{i+u}-C^{n,jk}_i\|^q\big|\mathcal{F}^n_i)\leq K_q\Delta_n^{q/4}.$$

Similar steps to $\chi(1)$ lead to

$$|\mathbb{E}(\chi(2)\big|\mathcal{F}^n_i)|\leq K\Delta_n^{3/4}(\sqrt{\Delta_n}+\eta^n_{i,k_n})\quad\text{and}\quad|\mathbb{E}(\chi(j)\big|\mathcal{F}^n_i)|\leq K\Delta_n(\sqrt{\Delta_n}+\eta^n_{i,k_n})\quad\text{for}\quad j=3,4.$$

Define

$$\chi(5)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha^{n,jk}_{i+u}\Big)\alpha^{n,lm}_{i+v}\Delta_n(C^{n,gh}_{i+w}-C^{n,gh}_i)$$
$$\chi'(5)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha^{n,jk}_{i+u}\Big)\alpha^{n,lm}_{i+v}\Delta_n\mathbb{E}\big((C^{n,gh}_{i+w}-C^{n,gh}_i)\big|\mathcal{F}^n_{i+v+1}\big)$$
$$\chi(6)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\Delta_n(C^{n,jk}_{i+u}-C^{n,jk}_i)\Big)\alpha^{n,lm}_{i+v}\Delta_n(C^{n,gh}_{i+w}-C^{n,gh}_i)$$
$$\chi(7)=\frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha^{n,jk}_{i+u}\Big)\Delta_n(C^{n,lm}_{i+v}-C^{n,lm}_i)\Delta_n(C^{n,gh}_{i+w}-C^{n,gh}_i),$$

where we have $\mathbb{E}(\chi(5)|\mathcal{F}_i^n) = \mathbb{E}(\chi'(5)|\mathcal{F}_i^n)$. Recalling (C.55), we further decompose $\chi'(5)$ as,

$$\chi'(5) = \sum_{j=1}^{5} \chi(5)[j],$$

with

$$\begin{aligned}
\chi\prime(5)[1] =& \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \sum_{v=0}^{w-1} \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \alpha_{i+v}^{n,lm} \Big(\mathbb{E}\Big(C_{i+w}^{n,gh} - C_i^{n,gh} | \mathcal{F}_{i+v+1}^n \Big) \\
& - (C_{i+v+1}^{n,gh} - C_i^{n,gh})\Delta_n - \widetilde{b}_{i+v+1}^{n,gh} \Delta_n^2 (w-v-1) \Big) \\
\chi\prime(5)[2] =& \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \sum_{v=0}^{w-1} \Delta_n (C_{i+v}^{n,gh} - C_i^{n,gh}) \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \alpha_{i+v}^{n,lm} \\
\chi\prime(5)[3] =& \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \sum_{v=0}^{w-1} \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \Delta_n (C_{i+v+1}^{n,gh} - C_{i+v}^{n,gh}) \alpha_{i+v}^{n,lm} \\
\chi\prime(5)[4] =& \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \sum_{v=0}^{w-1} \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \Delta_n^2 (w-v-1) (\widetilde{b}_{i+v+1}^{n,gh} - \widetilde{b}_{i+v}^{n,gh}) \alpha_{i+v}^{n,lm} \\
\chi\prime(5)[5] =& \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \sum_{v=0}^{w-1} \Delta_n^2 (w-v-1) \widetilde{b}_{i+v}^{n,gh} \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \alpha_{i+v}^{n,lm}.
\end{aligned}$$

Using (C.55), (C.54), (C.51) and following the same strategy proof as for $\chi(1)$, it can be shown that

$$|\mathbb{E}\Big(\chi\prime(5)[j] \big| \mathcal{F}_i^n\Big)| \leq K\Delta_n^{3/4}(\sqrt{\Delta_n} + \eta_{i,k_n}^n), \quad \text{for} \quad j = 1, \dots, 5,$$

which in turn implies

$$|\mathbb{E}\Big(\chi(5) \big| \mathcal{F}_i^n\Big)| \leq K\Delta_n^{3/4}(\sqrt{\Delta_n} + \eta_{i,k_n}^n), \quad \text{for} \quad j = 1, \dots, 5.$$

The term $\chi(6)$ can be handled similarly to $\chi(5)$, hence we conclude that

$$|\mathbb{E}\Big(\chi(6) \big| \mathcal{F}_i^n\Big)| \leq K\Delta_n^{3/4}(\sqrt{\Delta_n} + \eta_{i,k_n}^n).$$

Next, we set

$$\chi(7) = \frac{1}{(k_n\Delta_n)^3} \sum_{w=2}^{k_n-1} \Bigg(\sum_{v=0}^{w-1} \Big(\sum_{u=0}^{v-1} \alpha_{i+u}^{n,jk} \Big) \Delta_n (C_{i+v}^{n,lm} - C_i^{n,lm}) \Delta_n (C_{i+w}^{n,gh} - C_i^{n,gh}) \Bigg).$$

Define

$$\chi(7)[1] = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\left(\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\Delta_n(C_{i+v}^{n,lm}-C_i^{n,lm})\Delta_n(C_{i+v+1}^{n,gh}-C_{i+v}^{n,gh})\right)$$
$$\chi(7)[2] = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\left(\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\Delta_n(C_{i+v}^{n,lm}-C_i^{n,lm})\Delta_n(C_{i+v}^{n,gh}-C_i^{n,gh})\right)$$
$$\chi(7)[3] = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\left(\sum_{v=0}^{w-1}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\Delta_n(C_{i+v}^{n,lm}-C_i^{n,lm})\Delta_n^2(w-v-1)(\widetilde{b}_{i+v+1}^{n,gh}-\widetilde{b}_{i+v}^{n,gh})\right)$$
$$\chi(7)[4] = \frac{1}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\left(\sum_{v=0}^{w-1}\Delta_n^2(w-v-1)\widetilde{b}_{i+v}^{n,gh}\Big(\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}\Big)\Delta_n(C_{i+v}^{n,lm}-C_i^{n,lm})\right).$$

It is easy to see that

$$\chi(7) = \sum_{j=1}^{4}\chi(7)[j].$$

Similarly to calculations used for $\chi(1)$, it can be shown that

$$|\mathbb{E}(\chi(7)[j]\big|\mathcal{F}_i^n)| \le K\Delta_n^{1/4}(\Delta_n^{1/4}+\eta_{i,k_n}), \quad \text{for} \quad j=1,\dots,3.$$

To handle the remaining term $\chi(7)[4]$, we decompose it $\chi(7)[4] = \sum_{j=1}^{9}\chi(7)[4][j]$, where

$$\chi(7)[4][1] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lm}-C_{i+u}^{n,lm})(C_{i+u+1}^{n,gh}-C_{i+u}^{n,gh})$$
$$\chi(7)[4][2] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,gh}-C_i^{n,gh})\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lm}-C_{i+u}^{n,lm})$$
$$\chi'(7)[4][2] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,gh}-C_i^{n,gh})\mathbb{E}(\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lm}-C_{i+u}^{n,lm})|\mathcal{F}_{i+u}^n)$$
$$\chi(7)[4][3] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,lm}-C_i^{n,lm})\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,gh}-C_{i+u}^{n,gh})$$
$$\chi(7)[4][4] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,lm}-C_i^{n,lm})(C_{i+u}^{n,gh}-C_i^{n,gh})\alpha_{i+u}^{n,jk}$$
$$\chi(7)[4][5] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,lm}-C_i^{n,lm})\alpha_{i+u}^{n,jk}(C_{i+v}^{n,gh}-C_{i+u+1}^{n,gh})$$
$$\chi'(7)[2][5] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,lm}-C_i^{n,lm})\alpha_{i+u}^{n,jk}\mathbb{E}((C_{i+v}^{n,gh}-C_{i+u+1}^{n,gh}|\mathcal{F}_{i+u}^n)$$

$$\chi(7)[4][6] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,lm} - C_{i+u}^{n,lm})(C_{i+v}^{n,gh} - C_{i+u+1}^{n,gh})$$
$$\chi(7)[4][7] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,gh} - C_{i}^{n,gh})\alpha_{i+u}^{n,jk}(C_{i+v}^{n,lm} - C_{i+u+1}^{n,lm})$$
$$\chi(7)[4][8] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}(C_{i+u+1}^{n,gh} - C_{i+u}^{n,gh})(C_{i+v}^{n,lm} - C_{i+u+1}^{n,lm})$$
$$\chi(7)[4][9] = \frac{\Delta_n^2}{(k_n\Delta_n)^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}\alpha_{i+u}^{n,jk}(C_{i+v}^{n,lm} - C_{i+u+1}^{n,lm})(C_{i+v}^{n,gh} - C_{i+u+1}^{n,gh}).$$

Using arguments similar to those involved for the treatment of $\chi(1)$, it can be shown that

$$|\mathbb{E}(\chi(7)[4][j]\big|\mathcal{F}_i^n)| \le K\Delta_n^{1/4}(\Delta_n^{1/4} + \eta_{i,k_n}), \quad \text{for} \quad j = 1, \ldots, 8,$$

which yields

$$|\mathbb{E}(\chi(7)\big|\mathcal{F}_i^n)| \le K\Delta_n^{1/4}(\Delta_n^{1/4} + \eta_{i,k_n}).$$

Next, define

$$\chi(8) = \frac{1}{k_n^3}\sum_{w=2}^{k_n-1}\sum_{v=0}^{w-1}\sum_{u=0}^{v-1}(C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+v}^{n,lm} - C_{i}^{n,lm})(C_{i+w}^{n,gh} - C_{i}^{n,gh}).$$

This term can be further decomposed into six components. Successive conditioning and existing bounds give

$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+v}^{n,lm} - C_{i+u}^{n,lm})(C_{i+w}^{n,gh} - C_{i+v}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n$$
$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+v}^{n,lm} - C_{i+u}^{n,lm})(C_{i+v}^{n,gh} - C_{i+u}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n^{3/4}(\Delta_n^{1/4} + \eta_{i,k_n})$$
$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+v}^{n,lm} - C_{i+u}^{n,lm})(C_{i+u}^{n,gh} - C_{i}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n$$
$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+u}^{n,lm} - C_{i}^{n,lm})(C_{i+w}^{n,gh} - C_{i+v}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n$$
$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+u}^{n,lm} - C_{i}^{n,lm})(C_{i+v}^{n,gh} - C_{i+u}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n$$
$$|\mathbb{E}\Big((C_{i+u}^{n,jk} - C_{i}^{n,jk})(C_{i+u}^{n,lm} - C_{i}^{n,lm})(C_{i+u}^{n,gh} - C_{i}^{n,gh})\big|\mathcal{F}_i^n\Big)| \le K\Delta_n$$

These bounds can be used to deduce

$$|\mathbb{E}(\chi(8)\big|\mathcal{F}_i^n)| \le K\Delta_n.$$

This completes the proof.

Proof of (C.13) and (C.14) in Lemma 5

Observe that

$$\nu_i^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh}) = \frac{1}{k_n\Delta_n}\sum_{u=0}^{k_n-1}\zeta_{i,u}^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh}),$$

$$\begin{aligned}
\nu_i^{n,jk}\nu_i^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh}) &= \frac{1}{k_n^2\Delta_n^2}\sum_{u=0}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,u}^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh})\\
&+\frac{1}{k_n^2\Delta_n^2}\sum_{u=0}^{k_n-2}\sum_{v=0}^{k_n-1}\zeta_{i,u}^{n,jk}\zeta_{i,v}^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh})+\frac{1}{k_n^2\Delta_n^2}\sum_{u=0}^{k_n-2}\sum_{v=0}^{k_n-1}\zeta_{i,u}^{n,lm}\zeta_{i,v}^{n,jk}(C_{i+k_n}^{n,gh}-C_i^{n,gh}).
\end{aligned}$$

Hence, (C.13) and (C.14) can be proved using the same strategy as for (C.12).

Proof of (C.15) and (C.16) in Lemma 5

Note that we have

$$\begin{aligned}
\lambda_i^{n,jk}\lambda_i^{n,lm}\nu_i^{n,gh} &= \nu_i^{n,gh}\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}+\nu_i^{n,gh}\nu_i^{n,jk}\nu_i^{n,lm}-\nu_i^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}-\nu_i^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}\\
&+\nu_i^{n,gh}\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})-\nu_i^{n,gh}\nu_i^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})+\nu_i^{n,gh}\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})\\
&-\nu_i^{n,gh}\nu_i^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})+\nu_i^{n,gh}(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,lm}-C_i^{n,lm}),
\end{aligned}$$

and

$$\begin{aligned}
\lambda_i^{n,gh}\lambda_i^{n,jk}\lambda_i^{n,lm} &= \nu_{i+k_n}^{n,gh}\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}+\nu_{i+k_n}^{n,gh}\nu_i^{n,jk}\nu_i^{n,lm}-\nu_{i+k_n}^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}-\nu_{i+k_n}^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}\\
&+\nu_{i+k_n}^{n,gh}\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})-\nu_{i+k_n}^{n,gh}\nu_i^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})+\nu_{i+k_n}^{n,gh}\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})\\
&-\nu_{i+k_n}^{n,gh}\nu_i^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})+\nu_{i+k_n}^{n,gh}(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,lm}-C_i^{n,lm})\\
&-\nu_i^{n,gh}\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}-\nu_i^{n,gh}\nu_i^{n,jk}\nu_i^{n,lm}+\nu_i^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}+\nu_i^{n,gh}\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}\\
&-\nu_i^{n,gh}\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})+\nu_i^{n,gh}\nu_i^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})-\nu_i^{n,gh}\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})\\
&+\nu_i^{n,gh}\nu_i^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})-\nu_i^{n,gh}(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,lm}-C_i^{n,lm})\\
&+\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh})+\nu_i^{n,jk}\nu_i^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh})-\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,gh}-C_i^{n,gh})\\
&-\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,gh}-C_i^{n,gh})+\nu_{i+k_n}^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh})
\end{aligned}$$

$$-\nu_i^{n,jk}(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh})+\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,gh}-C_i^{n,gh})$$
$$-\nu_i^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,gh}-C_i^{n,gh})+(C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh}).$$

From (C.4), notice that ν_i^n is $\mathcal{F}_{i+k_n}^n$-measurable and satisfies $\|\mathbb{E}(\nu_i^n|\mathcal{F}_i^n)\|\leq K\Delta_n^{1/2}$.
The law of iterated expectations and existing bounds imply

$$\begin{aligned}
|\mathbb{E}(\nu_i^{n,lm}\nu_{i+k_n}^{n,jk}|\mathcal{F}_i^n)| &\leq K\Delta_n^{3/4},\\
|\mathbb{E}(\nu_i^{n,lm}\nu_i^{n,gh}\nu_{i+k_n}^{n,jk}|\mathcal{F}_i^n)| &\leq K\Delta_n,\\
|\mathbb{E}(\nu_i^{n,lm}(C_{i+k_n}^{n,gh}-C_i^{n,gh})\nu_{i+k_n}^{n,jk}|\mathcal{F}_i^n)| &\leq K\Delta_n,\\
|\mathbb{E}(\nu_{i+k_n}^{n,lm}(C_{i+k_n}^{n,jk}-C_i^{n,jk})|\mathcal{F}_i^n)| &\leq K\Delta_n^{3/4},\\
|\mathbb{E}((C_{i+k_n}^{n,jk}-C_i^{n,jk})(C_{i+k_n}^{n,lm}-C_i^{n,lm})(C_{i+k_n}^{n,gh}-C_i^{n,gh})|\mathcal{F}_i^n)| &\leq K\Delta_n. \qquad \text{(C.56)}
\end{aligned}$$

It can also be readily verified that

$$\begin{aligned}
&|\mathbb{E}(\nu_{i+k_n}^{n,gh}\nu_{i+k_n}^{n,ab}|\mathcal{F}_{i+k_n}^n)-\frac{1}{k_n}(C_{i+k_n}^{n,ga}C_{i+k_n}^{n,hb}+C_{i+k_n}^{n,gb}C_{i+k_n}^{n,ha})-\frac{k_n\Delta_n}{3}\overline{C}_{i+k_n}^{n,gh,ab}|\\
&\leq K\sqrt{\Delta_n}(\Delta_n^{1/8}+\eta_{i+k_n,k_n}^n).
\end{aligned}$$

Hence, for $\varphi_i^{n,gh}\in\{\nu_i^{n,gh},C_{i+k_n}^{n,gh}-C_i^{n,gh}\}$, which satisfies $\mathbb{E}(|\varphi_i^{n,gh}|^q\Big|\mathcal{F}_i^n)\leq K\Delta_n^{q/4}$ and $\mathbb{E}(\varphi_i^{n,gh}|\mathcal{F}_i^n)\leq K\Delta_n^{1/2}$. One can show that

$$\begin{aligned}
&|\mathbb{E}(\varphi_i^{n,gh}\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}|\mathcal{F}_i^n)-\mathbb{E}\Big(\varphi_i^{n,gh}\Big[\frac{1}{k_n}(C_{i+k_n}^{n,jl}C_{i+k_n}^{n,km}+C_{i+k_n}^{n,jm}C_{i+k_n}^{n,kl})-\frac{k_n\Delta_n}{3}\overline{C}_{i+k_n}^{n,jk,lm}\Big]|\mathcal{F}_i^n\Big)|\\
&\leq K\Delta_n^{3/4}(\Delta_n^{1/4}+\eta_{i,2k_n}^n).
\end{aligned}$$

Next, by combining the successive conditioning together with existing bounds, we have

$$\begin{aligned}
|\mathbb{E}(\varphi_i^{n,gh}\overline{C}_{i+k_n}^{n,jk,lm})| &\leq K\Delta_n^{1/4}(\Delta_n^{1/4}+\eta_{i,k_n}^n)\\
|\mathbb{E}(\varphi_i^{n,gh}C_{i+k_n}^{n,jl}C_{i+k_n}^{n,km})| &\leq K\Delta_n^{1/2},
\end{aligned}$$

which together imply

$$|\mathbb{E}(\varphi_i^{n,gh}\nu_{i+k_n}^{n,jk}\nu_{i+k_n}^{n,lm}|\mathcal{F}_i^n)|\leq K\Delta_n^{3/4}(\Delta_n^{1/4}+\eta_{i,2k_n}^n). \qquad \text{(C.57)}$$

It is easy to see that (C.12), (C.56) and (C.57) and the inequality $\eta_{i,k_n}^n\leq\eta_{i,2k_n}^n$ together yield (C.15) and (C.16).

A3.6 Proof of Lemma 6

(C.17) can be proved easily using the bounds of $\rho(u,v)_i^{n,gh}$ in (C.41) . To show (C.18), (C.19) and (C.20), we set

$$\overline{\overline{A11}}(H,gh,u;G,ab,v)=\lambda(u,v)_0^n\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}.$$

Then it holds that

$$\frac{1}{\Delta_n^{1/4}}\Big(\overline{\overline{A11}}(H,gh,u;G,ab,v)-\overline{A11}(H,gh,u;G,ab,v)\Big)\overset{\mathbb{P}}{\Rightarrow}0.$$

The above result is proved following similar steps as for (C.35) in case $w=1$ by replacing $\Theta(u,v)_0^{(C),i,n}$ by $\lambda(u,v)_0^n((\partial_{gh}H\partial_{ab}G)(C_{i-1})-(\partial_{gh}H\partial_{ab}G)(C_{i-2k_n}))$, which has the same bounds as the former. Next, decompose $\overline{\overline{A11}}$ as follows,

$$\begin{aligned}\overline{\overline{A11}}(H,gh,u;G,ab,v)=\lambda(u,v)_0^n\Bigg[&\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})V_{i-1}^n\\&+\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})\Big(\mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n)-V_{i-1}^n\Big)\\&+\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})\Big(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}-\mathbb{E}(\zeta(u)_i^{n,gh}\zeta(v)_i^{n,ab}|\mathcal{F}_{i-1}^n)\Big)\Bigg].\end{aligned}$$

We follow the proof of (C.36) for $w=1$, and we replace $\Theta(u,v)_0^{(C),i,n}$ by $\lambda(u,v)_0^n(\partial_{gh}H\partial_{ab}G)(C_{i-1})$, which satisfies only the condition $|\lambda(u,v)_0^n(\partial_{gh}H\partial_{ab}G)(C_{i-1})|\le\widetilde{\lambda}_{u,v}^n$. This calculation shows that the last two terms in the above decomposition vanish at a rate faster than $\Delta_n^{1/4}$. Therefore,

$$\frac{1}{\Delta_n^{1/4}}\left(\overline{\overline{A11}}(H,gh,u;G,ab,v)-\lambda(u,v)_0^n\Big(\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})V_{i-1}^n\Big)\right)\Rightarrow0.$$

As a consequence, for $(u,v)=(1,2)$ and $(2,1)$,

$$\frac{1}{\Delta_n^{1/4}}\overline{\overline{A11}}(H,gh,u;G,ab,v)\Rightarrow0.$$

The results follow from the following observation,

$$\frac{1}{\Delta_n^{1/4}}\left(\lambda(u,v)_0^n\Big(\sum_{g,h,a,b=1}^{d}\sum_{i=2k_n}^{[T/\Delta_n]}(\partial_{gh}H\partial_{ab}G)(C_{i-1})V_{i-1}^n(u,v)\Big)\right.$$

$$-\frac{3}{\theta^2}\int_0^T (\partial_{gh} H \partial_{ab} G)(C_t)(C_t^{ga} C_t^{hb} + C_t^{gb} C_t^{ha})dt\Bigg) \Rightarrow 0, \quad \text{for } (u,v) = (2,2),$$

$$\frac{1}{\Delta_n^{1/4}}\left(\sum_{g,h,a,b=1}^{d} \lambda(u,v)_0^n \Big(\sum_{i=2k_n}^{[T/\Delta_n]} (\partial_{gh} H \partial_{ab} G)(C_{i-1}) V_{i-1}^n(u,v)\Big) - [H(C), G(C)]_T\right) \Rightarrow 0,$$

for $(u,v) = (1,1)$.

Chapter 2

Are Three Moments Enough? What Options Tell Us about Higher-Order Risk Aversion. *

2.1 Introduction

The Capital Asset Pricing Model (CAPM) of Sharpe (1964a) and Lintner (1965b) has been at the forefront of the empirical financial literature for many decades. The model builds on the assumption that investors are exclusively rewarded for the exposure to the variance risk, an assumption that implies, in particular, that the pricing kernel is linear in the market returns. As the literature expands, numerous papers have convincingly established the limitations of this CAPM model pushing more recent models to incorporate other factors besides market returns (Ross (1976), Fama and French (1992)).[1] Even though many of these models perform well empirically, most of the newly introduced factors lack economic underpinnings. To overcome this shortcoming, Kraus and Litzenberger (1976) and Harvey and Siddique (2000b) derive preferences for the first moments of a distribution, based on preference theory. They propose a nonlinear pricing kernel that assumes an aversion to the variance of the returns distribution, and an appetence for its skewness. Dittmar (2002a) follows the same approach and

*This chapter is co-authored with Christian Dorion and Pierre Chaigneau.

[1] A CAPM anomaly appears when low-risk stocks outperform high-risk stocks.

also considers an aversion to the kurtosis of the returns distribution. These papers show that such pricing kernels derived from first principles can outperform other popular cross-sectional asset pricing models, including the Fama and French (1993) three-factor model.

A limitation of most of the existing nonlinear pricing kernels, however, is that they do not consider the effect of higher-order moments beyond the kurtosis risk. Yet, higher-order risk preferences also imply preferences for higher-order moments of a distribution (Scott and Horvath (1980a), Ekern (1980)).[2] These higher-order risk preferences have been extensively studied in economic theory (e.g., Ekern (1980), Eeckhoudt and Schlesinger (2006), Liu and Meyer (2013)) and more recently in experiments on risky choices (Deck and Schlesinger (2014)).[3] For example, CARA and CRRA utility are averse to all even-order moments, and loving toward all odd-order moments. A priori, it is unclear whether considering only a subset of moments is adequate for asset pricing purposes, and, if so, which subset of moments should be considered. Should the pricing kernel explicitly incorporate the effect of higher-order risk preferences beyond the kurtosis preference?

To contribute to answering this question, we introduce a novel option-valuation approach that allows us to exploit the richness of option data to better understand the extent to which higher moments of the distribution of market returns matter in the pricing kernel. Options constitute a suitable test asset for documenting the pricing of higher-order risks given the nonlinear patterns in their payoffs. As in Harvey and Siddique (2000b)) and Dittmar (2002a) , we use a Taylor series expansion of marginal utility, so that the pricing kernel is a polynomial function of the market return. Whereas Harvey and Siddique (2000b)) and Dittmar (2002a) consider a Taylor series expansion that only incorporates preferences for the first three and four moments, respectively, we also consider higher-order expansions, thus allowing for the effect of higher-order risk preferences.[4] The first pricing kernel considered, denoted by $\text{SDF}^{(2)}$, relies on a first order Taylor expansion of marginal utility, which implies that only variance

[2]They also have other implications for economic behavior, see Lajeri and Lajeri (2004) and Denuit and Rey (2010) for the case of "edginess", which is an appetence for fifth order risk, as implied by a positive fifth derivative of the utility function in an expected utility framework.

[3]Deck and Schlesinger (2014) find that preferences for the fifth and sixth degree risks are weaker than preferences for the first four degrees of risk.

[4]This type of approach still imposes constraints on the set of possible pricing kernels, which is restricted by expected utility theory: a utility function without any preference for a given moment (in the sense that its corresponding derivative is nil) does not have a preference for higher-order moments either. Harvey and Siddique (2000b) and Dittmar (2002a) approaches implicitly assume $u^{(4)} = 0$ and $u^{(5)} = 0$, respectively. Likewise, our $\text{SDF}^{(k)}$ approach described below implicitly assumes $u^{(k+1)} = 0$.

risk is priced. The second pricing kernel considered, $\text{SDF}^{(3)}$, relies on a second order Taylor expansion of marginal utility, which implies that variance and skewness risks are priced, but that higher-order risks are not. As k tends to infinity, $\text{SDF}^{(k)}$ incorporates the effect of all risk preferences , and thus represents the "true" pricing kernel.

We formally prove that, using any one of the approximate $\text{SDF}^{(k)}$, European options prices can be computed in closed-form as long as the physical dynamics is affine. In such case, the risk-neutral dynamics implied by $\text{SDF}^{(k)}$ needs not to be affine. We combine each approximate SDF with the nonnormal component GARCH model of Babaoglu, Christoffersen, Heston, and Jacobs (2018) to obtain weekly estimates of the risk aversion parameters from S&P 500 index option data. We then compare the ability of these pricing kernels to fit panels of index option prices across time. By doing so, we assess the goodness of the fit of each $\text{SDF}^{(k)}$ using the options root mean square of the relative implied volatility (RIVRMSE).

Our weekly in-sample analysis reveals the following. On virtually all weeks between 1996 and 2017, pricing kernels accounting for skewness ($\text{SDF}^{(3)}$) and kurtosis ($\text{SDF}^{(4)}$) fit the data significantly better than $\text{SDF}^{(2)}$. The pricing errors of $\text{SDF}^{(2)}$ ($\text{SDF}^{(3)}$) are, on average, 2.18 (1.18) times larger than the pricing errors of $\text{SDF}^{(4)}$. Further accounting for hyperskewness and hyperkurtosis, $\text{SDF}^{(6)}$ yields errors that are 0.95 times as large as those of $\text{SDF}^{(4)}$, and that are statistically significant on only 12% of the weeks. Finally, whereas, in sample, increasing the complexity of the SDF should always improve the fit, $\text{SDF}^{(9)}$ does not improve over $\text{SDF}^{(8)}$.

Interestingly, the estimates of the aversion to variance, skewness and kurtosis have the sign predicted by the expected utility theory in almost all the weeks and they evolve smoothly over time. In contrast, the estimates of aversion to hyperskewness and hyperkurtosis are characterized by large sign deviations from their expected signs. Additionally, these estimates are noticeably more volatile than that of the first three moments. Together, sign deviations and noisy estimates thus lend little support to hyperskewness and hyperkurtosis aversions being economically relevant. It rather hints towards overfitting. Importantly, the out-of-sample improvements of $\text{SDF}^{(6)}$ over $\text{SDF}^{(4)}$ have little economic significance.

Putting together these results, we conclude $\text{SDF}^{(4)}$ adequately describes the pricing kernel. Hence, despite the fact that higher-order risks may theoretically matter in asset pricing, index option prices suggest that only the first four moments matter. This result corroborates the approach of Dittmar (2002a), who proposed a pricing kernel which incorporates the asset pricing effect of these four moments. Thus, our results can be viewed as providing empirical support for an asset pricing model based on preferences for the first four moments of the distribution of market returns – and for these moments only, which a priori cannot be taken for granted. Further, this support is derived from index option data, which reflect risk preferences and risk perceptions, regardless of the actual manifestation of such risks.

Our paper relates to a vast literature.[5] Much of the existing literature either focuses on estimating the physical and risk-neutral distributions, or on the premium implied by the difference in moments. In the latter case, risk premia are informative about the interaction of prices and quantities of risk. However, one typically cannot disentangle the two. In particular, time-varying risk premiums do not necessarily rule out constant risk aversion; the quantity of risk faced by a CRRA agent, or her risk aversion coefficient could be changing. Our results, however, clearly reject the CRRA hypothesis at any given point in time. We show that the the relative risk aversion parameters do not move in tandem as implied by CRRA preferences. Indeed, our results show that the aversion to variance risk is low in high volatility periods while aversion to skewness and kurtosis are high during these episodes. Given that risk aversion is expected to be countercyclical, our results suggest that skewness and kurtosis aversions are better proxies for the aggregate risk aversion than variance risk aversion.

Finally, our results demonstrate that a stochastic discount factor with priced variance, skewness and kurtosis risks and maturity-independent risk aversion parameters compares favourably to the discrete-time volatility-dependent pricing kernel of Christoffersen, Heston, and Jacobs (2013). Indeed, we find that the volatility-dependent pricing kernel and $\mathrm{SDF}^{(4)}$ are almost indistinguishable in terms of options fit both in sample and out-of-sample. To appraise the qualitative implication of each SDF, we compute the moments premium. Our results show that the the volatility-dependent pricing kernel implies a systematically higher premium than the $\mathrm{SDF}^{(4)}$ for all the moments orders. For example, the normalized volatility risk premium is 1.24 for the the volatility-dependent pricing kernel against 1.09 for $\mathrm{SDF}^{(4)}$on average. Focusing on a single moment, volatility-dependent pricing kernel thus needs to

[5]Among many others, Bakshi, Kapadia, and Madan (2003), Harvey and Siddique (2000b), Dittmar (2002a), Christoffersen, Fournier, Jacobs, and Karoui (2017), Ait-Sahalia and Lo (1998), Jackwerth (2000) Rosenberg and Engle (2002), Bliss and Panigirtzoglou (2004), Chabi-Yo (2002), Christoffersen, Heston, and Jacobs (2013), Babaoglu, Christoffersen, Heston, and Jacobs (2018).

further inflate risk-neutral variance that $\text{SDF}^{(4)}$ to match option prices.

This paper is organized as follows. Section 2 presents our model, discusses the pricing kernel specification. Section 3 outlines the estimation methodoloy. Section 4 presents the data and presents the empirical results. Section 4 concludes.

2.2 Methodology

The time-t value of asset V with terminal payoff X_T paid at time T is given by:

$$V_t = \mathrm{E}_t^{\mathbb{P}}\left[M_{t,T} X_T\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-r_f (T-t)} X_T\right], \tag{2.1}$$

where $M_{t,T}$ is the pricing kernel and r_f is the instantaneous risk-free rate, and $\mathbb{P}$ and $\mathbb{Q}$ denote respectively the physical and risk-neutral measures. Following Black and Scholes (1973b), the bulk of the option-valuation literature relies on risk-neutral valuation. Outside the continuous-time complete-market framework, however, the Radon-Nikodym that equates the $\mathbb{P}$ and $\mathbb{Q}$ expectations in equation (2.1) crucially hinges on prices of risk that must be estimated if one wants to relate the physical and risk-neutral measures.[6] The prevalent paradigm in the option-valuation literature is to specify a model under $\mathbb{P}$ and a Radon-Nikodym such that the resulting dynamics under $\mathbb{Q}$ are affine, allowing for a closed-form solution (see e.g. Duffie, Pan, and Singleton (2000) among others).

Whereas closed-form solutions are crucial for tractability, affine models have been demonstrated to underperform their non-affine counterparts at fitting option prices, sometimes quite significantly so.[7] Our approach allows for closed-form prices without imposing affine restrictions on the $\mathbb{Q}$ model. Indeed, we depart from the standard risk-neutral valuation approach and focus on $\mathrm{E}_t^{\mathbb{P}}\left[M_{t,T} X_T\right]$. To disentangle between quantities and prices of risk, we use a parametric model to describe the distribution of payoffs, and a flexible nonparametric description of the SDF.

[6] In a calibration exercise where one only cares about matching option prices, these prices of risk are "buried" in the risk-neutral parameters. They become relevant only if one wants to relate option prices to the returns of the underlying assets.

[7] See, for instance, Chernov, Gallant, Ghysels, and Tauchen (2003), Hsieh and Ritchken (2005), and Christoffersen, Dorion, Jacobs, and Wang (2010).

2.2.1 Prices of risk: An arbitrarily precise approximation of the SDF

The Euler equation (see e.g. Hansen and Jagannathan (1991))

$$V_t = \mathrm{E}_t^{\mathbb{P}}\left[M_{t,T} X_T\right]$$

relates prices (or returns) to the representative agent's intertemporal marginal rate of substitution, and the asset payoff X_T, where U is the utility function of the representative agent.[8] As discussed in Dittmar (2002a), under mild conditions, the pricing kernel can be reformulated in terms of aggregate wealth, W, rather than consumption. In line with Dittmar (2002a) and many others,[9] we will work under the assumption that the evolution of market index, S, is a good proxy for the wealth portfolio. Hence,

$$M_{t,T} = \frac{\mathrm{U}'(S_T)}{\mathrm{U}'(S_t)}.$$

In particular, the current value of the index portfolio is given by:[10]

$$S_t = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\mathrm{U}'(S_T)}{\mathrm{U}'(S_t)} S_T\right] \tag{2.2}$$

or equivalently

$$1 = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\mathrm{U}'(S_T)}{\mathrm{U}'(S_t)} (1 + R_{t,T})\right] \tag{2.3}$$

where $R_{t,T} = (S_T - S_t)/S_t$ is the total return on the index portfolio between t and T. Eqn. (2.3) represents a no-arbitrage condition that should be fulfilled by any candidate SDF.

Our approach hinges on a simple Taylor series representation of the stochastic discount factor (SDF). Indeed, $\mathrm{U}'(S_T)$ can be described by

$$\mathrm{U}'(S_T) = \sum_{i=0}^{\infty} \frac{1}{i!}\,\mathrm{U}_t^{(i+1)}(S_t)\,(S_T - S_t)^i = \mathrm{U}'(S_t) + \sum_{j=2}^{\infty} \frac{1}{(j-1)!}\,\mathrm{U}_t^{(j)}(S_t)\,(S_T - S_t)^{j-1}\ . \tag{2.4}$$

[8]In this paper, we allow for time-variation in the representative consumer's utility function. Such time-variation could be driven by changing levels of risk aversion. To simplify the notation, we here abstract for the time index.

[9]Virtually, all tests of the CAPM are implicitly making this assumption (Roll's critique).

[10]To simplify the notation, we here abstract from dividends. Of course, in our empirical exercise, total returns include dividends.

Incidentally, each order of derivative in this representation can be related to the aversion to a given moment of the returns distribution. Consequently, we set

$$\mathrm{RRA}_t^{(j)} = (-1)^{j-1} \frac{S_t^{j-1}\,\mathrm{U}_t^{(j)}(S_t)}{\mathrm{U}_t'(S_t)}, \qquad j \geq 2,$$

which implies that

$$\frac{\mathrm{U}'(S_T)}{\mathrm{U}'(S_t)} = 1 + \sum_{j=2}^{\infty} \frac{(-1)^{j-1}}{(j-1)!}\,\mathrm{RRA}_t^{(j)}\, R_{t,T}^{j-1}. \tag{2.5}$$

For instance, $\mathrm{RRA}^{(2)}$ is the familiar Arrow-Pratt measure of relative risk aversion. For $j > 2$, $\mathrm{RRA}^{(j)}$ generalizes relative risk aversion to skewness aversion (prudence, $j = 3$), aversion to kurtosis (temperance, $j = 4$) and so on.[11]

The higher-order risk preferences have been extensively studied in economic theory (e.g., Ekern (1980), Eeckhoudt and Schlesinger (2006), Liu and Meyer (2013)) and more recently in experiments on risky choices (Deck and Schlesinger (2014))). Scott and Horvath (1980b) show that investors should have a negative preference for even moments and a positive preference for odd ones.[12] This implies in particular that the $\mathrm{RRA}_t^{(j)}$, defined above, should be non-negative for $j \geq 0$.

Note that SDF representation in equation (2.5) is *not* an approximation. In practice, however, using this representation would involve truncating the series at some order k, which would then yield an approximation of the SDF. This approximation is at the heart of our analysis:

$$\frac{\mathrm{U}'(S_T)}{\mathrm{U}'(S_t)} \simeq \mathrm{SDF}_{t,T}^{(k)} = \kappa_{1,t,T}^{(k)} + \sum_{j=2}^{k} \kappa_{j,t,T}^{(k)}\,\mathrm{RRA}_t^{(j)}\, R_{t,T}^{j-1}, \tag{2.6}$$

where $\kappa_{j,t,T}^{(k)}$ are normalization constants whose expressions are available in Appendix B.2.

Actually, the SDF in equation (2.6) would not be an approximation under the hypothesis that the representative agent is indifferent to the moments of the returns distribution beyond

[11] Prudence is necessary for decreasing absolute risk aversion (DARA). DARA implies that as an agent becomes more wealthy she invests more (in dollars) in the risky asset (as opposed to the risk-free asset). Temperance is necessary for decreasing absolute prudence, which is in turn necessary for “standard risk aversion" Kimball (1993)). Under standard “risk aversion", the presence of an exogenous 'background risk" with nonpositive mean (including a pure risk) increases the aversion to other independent risks (“risk vulnerability").

[12] This property is actually satisfied by CARA and CRRA utility functions.

the k^{th} moment. For instance, the pricing kernel studied by Harvey and Siddique (2000b) sets $k = 3$, and that studied by Dittmar (2002a) considers $k = 4$. Our methodology allows us to consider increasingly complex approximations of the SDF and assess to what extent the increased complexity improves the fit to options prices both in-sample and out-of-sample.

2.2.2 Quantity of risk: Moments of the payoff's distribution

The previous section presents the essence of our contribution. To disentangle prices and quantities of risk, however, we unavoidably need a model for the physical distributions of future returns.[13] We here follow Babaoglu, Christoffersen, Heston, and Jacobs (2018, hereafter BCHJ) and use their non-Gaussian GARCH model for the returns of asset S:[14]

$$r_{t+1} \equiv \log\left(\frac{S_{t+1}}{S_t}\right) = r_f + \lambda h_{t+1} - \xi_{t+1} + \nu_{r,t+1}, \tag{2.7}$$

$$\nu_{r,t+1} \equiv \eta y_{t+1} - \frac{h_{t+1}}{\eta}, \tag{2.8}$$

where y_{t+1} is an inverse Gaussian (IG) distribution with degree of freedom $\frac{h_{t+1}}{\eta^2}$, such that $\nu_{r,t+1}$ has mean 0 and variance h_{t+1}.[15] The model has the key property that when $\eta \to 0$, the IG distribution converges to the normal distribution. The convexity adjustment ξ_{t+1} is such that

$$\mathrm{E}_t^{\mathbb{P}}\left[e^{r_{t+1}}\right] = e^{r_f + \lambda h_{t+1}}.$$

[13] Recent model-based recovery theorems study conditions on physical transition dynamics and pricing kernels under which the physical belief can be determined from Arrow-Debreu prices alone. Broadly speaking, this literature shows that recovery is possible whenever physical probabilities are induced by a family of path-independent pricing kernels and satisfy further stability conditions (see e.g. Borovicka, Hansen, and Scheinkman (2016)).

[14] Conditionally Gaussian models do not generate enough (negative) skewness and kurtosis at short horizon to fit the market prices of short-dated options. A plethora of distributions have been studied to account for conditional nonnormality, starting with the generalized error distribution (Duan (1999)), but the inverse Gaussian distribution, popularized by Christoffersen, Heston, and Jacobs (2006a), has the notable advantage of allowing for closed-form European prices. Although our notation differs significantly from the one used by BCHJ, the two formulations of the model are observationally equivalent (cf. Appendix B.3). Our notation is closer to that used in other recent papers, which allows for easier comparison with other models.

[15] In particular, $\mathrm{E}^{\mathbb{P}}[y] = \mathrm{Var}^{\mathbb{P}}(y) = \frac{h_{t+1}}{\eta^2}, \mathrm{E}^{\mathbb{P}}[1/y] = \frac{\eta^2}{h_{t+1}} + \frac{\eta^4}{h_{t+1}^2}$. The conditional density of the future stock price is given by $f_{t-1}(S_t) = \frac{h_t/|\eta^3|}{\sqrt{2\pi y_t^3}} S_t \exp\left(-\frac{1}{2}\left[\sqrt{y_t} - \frac{h_t/\eta^2}{\sqrt{y_t}}\right]^2\right)$.

That is, λ parametrizes the daily equity risk premium (ERP), λh_{t+1}. Given that the ERP is inherently related to preferences, sequentially re-estimating risk aversion coefficients without re-estimating λ would be theoretically inconsistent.

As in BCHJ, we assume the following component structure for volatility:[16]

$$h_{t+1} = q_{t+1} + \rho_h(h_t - q_t) + \nu_{h,t}, \tag{2.9}$$

$$q_{t+1} = \sigma^2 + \rho_q(q_t - \sigma^2) + \nu_{q,t}, \tag{2.10}$$

where the innovations $\nu_{h,t}$ and $\nu_{q,t}$ are defined as

$$\nu_{h,t} = a_h h_t \nu_t + \frac{c_h}{\eta} \upsilon_{r,t}, \qquad \nu_{q,t} = a_q h_t \nu_t + \frac{c_q}{\eta} \upsilon_{r,t}, \qquad \text{and} \quad \nu_t = \left(\frac{h_t}{y_t} - \eta^2 - \frac{\eta^4}{h_t} \right),$$

where ν_t has mean 0. Parameters $\rho_h < \rho_q$ capture the persistence of the transient and persistent processes, respectively. Parameters a_h and a_q control the variance of the variance processes, and parameters c_h and c_q the negative correlation between variance processes and returns, commonly referred to as the leverage effect (Black (1976)). By accounting for more than one volatility component, the model is more flexible and its modelling of the term structure of volatility improves compared to a one-factor model.

We use returns to estimate the parameters of the physical variance processes, as well as η, which parametrizes the physical conditional distribution of returns. We maximize the (inverse-Gaussian) likelihood of the $\{y_t\}$ innovations filtered from the S&P 500 returns between January 1967 and December 2017.[17] Table reports the MLE estimates of the physical parameters.

[16] At least two volatility components are required to adequately describe stock market volatility and option prices. Indeed, two-factor volatility processes more effectively capture the time-series properties of volatility by separately accounting for transient and highly persistent volatility shocks. See, among many others, Engle and Lee (1999); Bates (2000); Andersen, Bollerslev, Diebold, and Ebens (2001); Alizadeh, Brandt, and Diebold (2002); Christoffersen, Jacobs, Ornthanalai, and Wang (2008); and Dorion (2016).

[17] Note that returns between July 1962 and December 1966 are used to warm up the volatility processes; however, the likelihood of these returns is disregarded.

2.2.3 Benchmark BCHJ pricing kernel

Although our main goal is to disentangle the contribution of higher-order moments risks to the variation in the index options prices, it is important to compare the performance of our pricing kernels $\mathrm{SDF}_{t,T}^{(k)}$ with that of a benchmark pricing kernel.

For this comparative analysis, we retain the volatility-dependent pricing kernel which has been initially introduced in Christoffersen, Heston, and Jacobs (2013) and further used in BCHJ. The BCHJ's pricing kernel takes the form:

$$\mathrm{SDF}_{t,t+1} = \Big(\frac{S_{t+1}}{S_t}\Big)^{\phi} \exp\left(\delta_0 + \delta_1 h_{t+1} + \xi h_{t+2}\right),$$

where ϕ captures the first-order risk aversion and ξ denotes the volatility risk premium parameter. This pricing kernel was shown to be a discrete-time analogue of the Radon-Nikodym change of measure used in continuous-time models. When $\xi = 0$, this pricing kernel becomes the monotonic power utility pricing kernel used in Heston and Nandi (2000).

Christoffersen, Heston, and Jacobs (2013) shows that when $\xi > 0$, the pricing kernel is U-shaped in the returns and there exists a divergence between the one-day ahead risk neutral and physical volatilities. They additionally show that this pricing kernel substantially outperforms the monotonic power utility pricing kernel, in terms of options fit, when the returns are conditionally gaussian. BCHJ reinforces this evidence by documenting that the over-performance of this pricing kernel over the power utility pricing kernel is even stronger when one assumes that the returns follow an IG distribution. Finally, it is worth emphasizing that the BCHJ pricing kernel does not belong to our sequence of pricing kernels $\mathrm{SDF}_{t,T}^{(k)}$.

2.3 Estimation

Given an affine returns dynamics under the physical measure $\mathbb{P}$ and the parameters governing this dynamics, we obtain closed-form European option prices using

$$V_t = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\ X_T\right], \tag{2.11}$$

where,

$$X_T = \max(S_T - K_n, 0)$$

is the payoff of a call option with strike K_n expiring at time T.[18] The solution, provided in Appendix B.1, hinges on the fact that the characteristic function of the future price distribution is available in closed-form at any T combined with the polynomial form of $\mathrm{SDF}_{t,T}^{(k)}$. Interestingly, the dynamics under the pricing measure need not, and in all likelihood will not, be affine. This approach is particularly appealing as we also show, in Appendix B.5, that risk-neutral moments can also be computed in closed-form.[19]

Obviously, option prices also depend on the set of RRAs entering the SDF. Hence, in our framework, RRAs can simply be estimated from options. Many studies suggest that risk aversion is time-varying (Aït-Sahalia and Lo (1998), Rosenberg and Engle (2002), Bliss and Panigirtzoglou (2004)). Iteratively estimating the RRAs from prices observed during a day, a week or a month will allow us to investigate the matter. Our main analysis is based on weekly estimates.

For each week t, and for each SDF approximation of order k, we maximize the options log-likelihood following:

$$\begin{aligned} &\max_{\lambda,\, \mathrm{RRA}_t^{(2)},\, \ldots,\, \mathrm{RRA}_t^{(k)}} L_{\text{Options}}\left(\lambda,\, \mathrm{RRA}_t^{(2)},\, \ldots,\, \mathrm{RRA}_t^{(k)}\right) \qquad (2.12)\\ &\text{subject to} \qquad (1+R_{f,t})^{-1} = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\right] \\ &\qquad\qquad\qquad\qquad 1 = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\left(1+R_{t,t+1}\right)\right]\,, \end{aligned}$$

where λ, defined in Section 2.2.2, parametrizes the level of the equity risk premium, and

[18]Our study also uses information on puts prices.

[19]Bakshi, Kapadia, and Madan (2003) derive nonparametric estimates of higher-order risk-neutral moments. In practice, using this methodology involves interpolating and extrapolating observed option prices to unobserved strikes. This step is not necessary using our model since we have a physical model augmented of a flexible SDF that delivers closed-form option prices and as a by-product closed-form risk neutral moments. Interestingly, the closed-form formulas reveal how risk-neutral moments relate to physical moments and the risk aversion parameters which could be useful in other applications.

$L_{\text{Options}}\left(\lambda,\, \text{RRA}_t^{(2)},\, \ldots,\, \text{RRA}_t^{(k)}\right)$ denotes the options log-likelihood which is obtained as

$$L_{\text{Options}}\left(\lambda,\, \text{RRA}_t^{(2)},\, \ldots,\, \text{RRA}_t^{(k)}\right) = -\frac{1}{2}\sum_{n=1}^{N_t}\left(\log(2\pi\sigma_e^2) + \frac{e_n^2}{\sigma_e^2}\right),$$

with N_t denoting the options data size and σ_e^2 is identified using the sample deviation of $\{e_n\}_{n=1}^{N_t}$, and

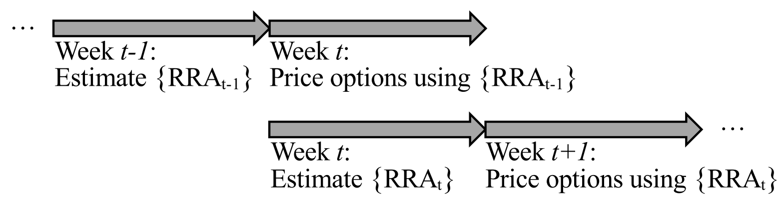
$$e_n = \left(\frac{IV_n^{\text{MODEL}} - IV_n^{\text{MKT}}}{IV_n^{\text{MKT}}}\right).$$

In the above expression IV_n^{MKT} denotes the observed Black and Scholes (1973b) IV of option n and IV_n^{MODEL} is the model predicted IV. We interpret e_n as the relative implied volatility (IV). The likelihood is based on the assumption, that e_n is normally distributed, that is $e_n \sim N(0, \sigma_e^2)$, and uncorrelated in the cross-section (see e.g. Christoffersen, Heston, and Jacobs (2006a)).
Concretely, the first constraint of (2.12) identifies the normalization constants, $\kappa_{j,t,T}^{(k)}$, in equation (2.6). Besides, the second constraint links the ERP parameter and the RRAs together such that the minimization is actually performed on k-dimensional vector (rather than $k+1$). Appendix B.2 provides further details.

We find that the above log-likelihood maximization problem delivers quantitatively similar results to the nonparametric alternative in which one minimizes the sum of the squared relative implied volatility error without imposing the normality assumption on the implied volatility errors. As we show in the sequel, maintaining a distributional assumption on the errors is useful for testing purposes.

Finally, by re-estimating the RRAs weekly, one could argue that some of the time-variation in the estimates is due to fitting noise in prices or compensating for misspecification of the returns dynamics. To alleviate this concern, our focus is on the out-of-sample fit of the pricing model, as illustrated in Figure 2.3.1.

Figure 2.3.1: Sequential Out-of-sample Pricing



2.4 Empirical Results

2.4.1 Data

To estimate the model, we use the returns and prices of options on the S&P 500 index, as a proxy for the market. We obtain the daily index returns from the Center for Research and Security Prices (CRSP). To compute the corresponding excess returns, we use the one-month Treasury bill rate (from Ibbotson Associates) as extracted from Kenneth French's data library. As in BCHJ, our estimation procedure fits the implied volatilities instead of the raw option prices.[20] The implied volatilities of options on the SPX, between January 1996 and December 2017, are obtained from OptionMetrics. We remove in-the-money options and options that violate the usual no-arbitrage conditions. We additionally filter out options with less than one week or more than one year to maturity to mitigate the effect of illiquidity. In the same spirit, we select options with strictly positive volume and open interest. To ease the calculation and interpretation, out-of-money (OTM) put prices are converted into corresponding in-the-money call values by exploiting the call-put parity relationship. This leaves us with 1,378,317 options on SPX. Since our analysis is performed at the weekly frequency, we further group options traded within each week.[21] Our full sample includes 1158 trading weeks. This implies that, on average, each trading week has around 1190 options. We believe that this number is sufficiently large to allow us to accurately pin down the RRA parameters in which we are interested in.

Table 2.3 provides a crisp description of the options data. To highlight the main characteristics of the S&P 500 index option, we sort the data by moneyness and maturity. This table

[20]The usefulness of relying on the implied volatilities in estimating option pricing models can be dated back to Renault (1997).

[21]This choice is not to be confounded with that other papers using weekly observations on one particular day (usually wednesdays) to estimate option pricing models.

reports the percentage of contracts, the average option price, the average Black and Scholes (1973b) implied volatility, the average bid-ask spread in dollars, and the average volume for different maturity and moneyness buckets. Short-term options have between 6 and 45 days to maturity (DTM), and long-term options have between 91 and 365 DTM. Following Israelovr and Kelly (2017), our measure of the moneyness is

$$m_{i,\tau} = \frac{\log\left(K_i/S_i\right)}{\text{VIX}\sqrt{\tau}},$$

where VIX is the CBOE VIX index and τ is the calendar days to maturity. OTM puts are options with moneyness less than -1, ATM options are those options having their moneyness between -1 and 0 and finally OTM calls have their moneyness greater than 0. Our dataset includes around 51% of short-maturity options and 63% of OTM puts. We also observe that OTM puts have a higher implied vol (average of 0.27) than ATM calls (average of 0.16), confirming the well-known volatility smirk.

2.4.2 Returns Maximum Likelihood Estimation

Our estimation methodology is a modified version of the widely used sequential maximum likelihood (see e.g. Broadie, Chernov, and Johannes (2007)). In a typical sequential maximum likelihood, physical parameters are typically estimated using returns data solely. Subsequently, the pricing kernel parameters are estimated setting the physical parameters at their maximum likelihood estimates.

While this approach has the great advantage of easing the interpretation of empirical results, when it comes to comparing the fit of different pricing kernels, it is often limited by the lack of identification of the equity risk premium. To highlight this feature, we plot in Figure 2.1, the returns log-likelihood as a function of equity risk premium (ERP) parameter λ. For 300 equally-spaced values of lambda between -3 and 3, we maximize the log-likelihood over the remaining physical parameters. The red starred point on Figure 2.1 shows the global maximum of the log-likelihood function. The two dotted horizontal lines form the 99% confidence interval around the global maximum. The two dotted vertical lines show the range of λ's that are not rejected at a 1% confidence level by the likelihood ratio (LR) test. As evident from this figure, the range of λ values that are rejected by the LR test is rather wide. A standard solution, widely adopted in the literature, consists in calibrating the λ parameter such that the convexity-adjusted ERP equals the average excess returns.[22] This calibrated value of λ is given by

$$\lambda^{cal} = \frac{\log(\mathrm{E}(r_t^e)))}{\mathrm{Var}(r_t^e)},$$

where r_t^e are the excess returns. The red squared point of Figure 2.1 shows the location of the λ^{cal} and its corresponding log-likelihood. As it clearly appears on this graph, the calibrated value is not rejected by the LR test.

Our estimation technique is a pseudo-sequential maximum likelihood in which the non-ERP parameters are set to their estimated values consistent with λ^{cal}, and the ERP parameter λ is dynamically estimated together with the RRA parameters using option prices as oulined in Section 2.3.[23] Proceeding this way allows to keep the consistency between time-varying risk aversion and time-varying risk compensation (as captured by the equity risk premium).

Table 2.1 presents the Maximum likelihood estimates of the $\mathbb{P}$ parameters with λ set at λ^{cal}. The long run variance parameter σ^2 is calibrated using the sample returns variance, which is close to 17% on an annual basis. As expected, the long-run volatility factor (q_t) is highly persistent with a persistence factor of 0.98. The total volatility factor h_t appears to be less persistent with a persistence factor of 0.74. Overall, our MLE estimates are close to the ones obtained by BCHJ.

2.4.3 Analyzing the in-sample pricing results

We now turn to our paper's main empirical contribution which is the study of the relevance of higher-order risks in capturing the risk premia in the options market.

We estimate the weekly RRA parameters using the methodology outlined in Section 2.3. For the safe of consistency, the RRAs are assumed to be maturity-independent. We set the maximum order of the SDF's Taylor expansion to 10.[24] For each SDF order, we compute the

[22] This solution has been adopted by Christoffersen, Heston, and Jacobs (2013) and BCHJ among others.

[23] We opted for this approach because it is tractable in addition to yielding time-consistent results. A certainly more general approach will be to use a joint maximum likelihood in which all physical parameters are estimated dynamically together with pricing kernel parameters. To the best of our knowledge, the joint ML is usually applied to full sample datasets with time-invariant parameters.

[24] Chung, Johnson, and Schill (2006) include systematic co-moments of orders 3-10 in a standard Fama-MacBeth regression and find that this almost always causes SMB and HML to become insignificant and always

option prices implied by the estimated RRA parameters and compare them with the observed option prices.

Before quantitatively assessing the fit of the different SDFs over the options panel, we want to perform a simple analysis of detecting how often the SDFs appear to be useful in the pricing. Using the in-sample results, we implement the Likelihood Ratio (LR) test of the following two hypotheses: "$H_0^{n,1}$: $SDF^{(n)}$ does not over-perform $SDF^{(n-1)}$ in terms of options fit" and "$H_0^{2,n}$: $SDF^{(n)}$ does not over-perform $SDF^{(n-2)}$ in terms of options fit". Figure 2.3 displays the outcomes of the testing procedures. The purple bars of this figure reports the fraction of weeks for which $H_0^{n,1}$ is rejected at the 1% confidence level and the lighted blue bars depicts the same information for $H_0^{n,2}$. The dotted horizontal line refers to the 1% level. A rejection of $H_0^{n,1}$ can be interpreted as the statistical significance of the $RRA^{(n)}$ implied by $SDF^{(n)}$. Similarly, a rejection of $H_0^{n,2}$ echoes a joint statistical significance of $RRA^{(n)}$ and $RRA^{(n-1)}$ implied by $SDF^{(n)}$.

Focusing on the blue bars, for almost all the weeks, $SDF^{(3)}$ outperforms $SDF^{(2)}$ in term of options fit. This result confirms the usefulness of skewness risk in explaining asset prices, a finding that can be traced back to Kraus and Litzenberger (1976) and Harvey and Siddique (2000b). The kurtosis risk appears to be important in 92% of weeks, corroborating the central finding of Dittmar (2002a). The hyperskewness risk appears to be less valued compared to the skewness and the kurtosis risks. Indeed, there are only 23% of weeks in which, option market participants seem to care about the hyperskewness risk. Surprisingly, the hyperkurtosis risk appears to be more often priced than the hyperskewness risk (hyperkurtosis risk is priced in 50% of the weeks). Figure 2.3 also shows the economic relevance of higher-order risks drastically decline beyond the hyperkurtosis risk, as these risks are valued in less than 12% of the weeks.

To confirm these in-sample results, we report in the first row of Table 2.4 the ratios of the average root mean square of the relative errors in the implied volatility domain (RIVRMSE) for different specifications of the SDFs. Following the forecasting literature, we report the in-sample average RIVRMSE of each specification as a ratio with the average RIVRMSE of the baseline specification, $SDF^{(4)}$ in our study.[25] In parentheses below each ratio, we report

causes their t-statistics to drop dramatically.

[25]This choice of $SDF^{(4)}$ as benchmark is motivated by the current practice in the asset pricing literature, where the first three co-moments are included in almost all Fama-MacBeth regressions.

the Diebold and Mariano (1995) (DM) statistics, based on time-series of RIVRMSEs, comparing the forecast accuracy of the different specification with that of $\text{SDF}^{(4)}$. This statistic accounts for autocorrelation in errors and is normally distributed asymptotically. It is positive (negative) whenever the $\text{SDF}^{(4)}$ outperforms (underperforms) the alternate specification. We report the same ratios for the CRRA and the BCHJ pricing kernel.

As evident from Table 2.4, in-sample, compared to the benchmark $\text{SDF}^{(4)}$, using any SDF beyond the sixth order leads to the same 4% reduction in the RIVRMSE yield by $\text{SDF}^{(6)}$. This confirms our previously established evidence that, on average, higher-order SDFs beyond the 6th order do not seem to be significantly valued by investors in the options market. It is worth noting that the DM t-stats attached to these higher-order SDFs are fairly close to that of $\text{SDF}^{(6)}$. Table 2.4 also shows that a pricing kernel that incorporates hyperskewness risk in addition to the three standard risks (variance, skewness and kurtosis), helps capture an additional 1% of options price anomalies. The improvement, although economically small, appears to be statistically significant according to the DM test.

The top left panel of Figure 2.5 barplots the RIVRMSE for the estimated SDFs. The blue (resp. red) line locates the RIVRMSE for the CRRA (resp. BCHJ) parametric pricing kernels. As depicted on this figure, the RIVRMSE plot is almost flat beyond the sixth-order. The skewness risk seems to be the most useful risk that helps captures the dynamics of option prices. Indeed, adding the skewness risk to the traditional CAPM model (with the variance risk only) decreases the RIVRMSE by 45.5% from 29.2% to 15.9%. The marginal contribution of kurtosis risk is more moderate at a 15.7 %, as adding it to pricing kernel further reduces the RIVRMSE to 13.4%. Figure 2.5 (resp Figure 2.6) dissects the pricing performance of the different SDFs along the maturity (moneyness) dimension. The RIVRMSE bar plot looks almost flat beyond the third order for OTM calls, dwarfing the value-added of the kurtosis in-sample the prices of OTM calls. This evidence is consistent with the well-established intuition that kurtosis risk captures tail risk, which is less valued in OTM calls. It also appears that the non trivial contribution of the skewness risk, that we uncovered earlier, seems to be driven by the OTM puts. This means that skewness risk helps explain the expensiveness of OTM puts compared to OTM calls. This evidence suggests that options market participants's aversion to skewness risk might be partially responsible of the volatility smirk. The skewness risk appears to be also crucial in matching the prices of long maturity-options.

Now, relating these results to the performance of our two parametric pricing kernels, it appears clearly that the CRRA benchmark, even with the time-varying RRA, is outperformed by $\text{SDF}^{(3)}$, $\text{SDF}^{(4)}$ and the BCHJ pricing kernels. This suggests that the restrictions implied by the CRRA pricing kernel are too strong and not supported by the options data. The benchmark $\text{SDF}^{(4)}$ seems to slightly better capture the dynamics of option prices than the BCHJ benchmark. This is evidenced by the 4% improvement in the RIVRMSE depicted in Table 2.4. However, this over-performance does not appear to be significant according to the DM-test.

At this stage, it is worth stressing the value-added of using our methodology. An affine SDF such as BCHJ includes a volatility premium parameter that generates some premium in both variance, skewness and kurtosis risks simultaneously. As a result, using such pricing kernel does not reveal us which higher-order risk seems to be driving the over-performance of non-monotonic pricing kernels over linear pricing kernels such as the CRRA one. Our results, in contrast, allow us to claim that the pricing kernel monotonicity introduced by the single skewness risk is enough to outperform the monotonic CRRA pricing kernel.

2.4.4 Analyzing the out-of-sample pricing results

Our discussion in Section 2.4.3 was based on the in-sample results. Given that the different SDFs, that were compared, differ by their number of parameters, it is natural to wonder if these results and their interpretation are not purely driven by data overfitting or pricing kernel over-parametrization.[26]

To tackle this concern, we implement an out-of-sample analysis in which we use the RRAs estimated h weeks back in the past to price currently traded options, for $h = 1, 2, 3, 4$. Consequently, if our results are not driven by the pricing kernel over-parametrization, we expect the one-week ahead results to quantitatively and qualitatively reflect the in-sample results.

As it appears in the second row of Table 2.4, the key results derived in-sample carry out out-of-sample, at the one week horizon. We uncover that, out-of-sample, $\text{SDF}^{(6)}$ is still the last pricing kernel that matters in options pricing. Out-of-sample, at the one week horizon, $\text{SDF}^{(4)}$

[26]This question is at least relevant for the SDFs of orders 3 to 6 since we found that higher-orders SDFs above the 6th order, even though more parametrized, add little information to options pricing.

and BCHJ become indistinguishable in terms of average fit. This conclusion still applies at the two-week horizon. For horizons longer that 2 weeks, we observe some divergence in the main conclusions. Higher-order risks beyond SDF$^{(4)}$ appear to be completely useless. This implies that these risks can be disregarded if we want to predict options prices at longer horizons. The out-of-sample analysis also reveals that our benchmark SDF$^{(4)}$ becomes slightly outperformed by the BCHJ pricing kernel above the two-week horizon. Successfully pricing options at longer horizons requires some minimum degree of persistence of the estimated parameters. Here, it appears that the degree of persistence of the RRA parameters decline substantially above the two-week horizon.[27]

2.4.5 Analyzing the behaviour of the RRA parameters

This section documents the time-series properties of the estimated RRA parameters.

The estimation technique we employ in this paper relies on a joint estimation of the equity risk premium parameter and the RRA parameters from option prices. Therefore, before addressing the properties of the RRA parameters, we want to highlight the relevance of this methodology.[28]

Figure 2.4 graphs the options root mean square of the relative implied volatility (RIVRMSE) as a function of the ERP parameter λ for one typical week in our sample. For this illustration, we use the week on which the value of the volatility index VIX is the closest to its first quartile.[29] In our sample, this week ranges from September 20, 2004 to September 24, 2004. This figure was produced using 250 values of λ uniformly spaced between -10 and 10. Options traded during this week are priced assuming SDF$^{(4)}$ is the "true" pricing kernel. For each value of λ, we maximize the options log-likelihood over the SDF$^{(4)}$ parameters. The dotted vertical black line shows the λ value that minimizes the RIVRMSE and the dotted horizontal black line shows the corresponding RIVRMSE. As evident from the figure, the RIVRMSE plotted as a function of λ is highly curved and far from being flat as the returns log-likelihood

[27]It is worth stressing that, to ease the comparison, we use a very simple forecasting methodology which consists in predicting future values, simply using past values. Given that some RRA parameters become less persistent beyond the two weeks horizons, it might be appropriate to use some more sophisticated forecasting methodologies such an AR regression or more generally an HAR regression. We leave this for future research.

[28]This procedure presents the additional advantage of yielding a time-varying λ which helps reduce the concerns about affine models being outperformed by non-affine models such as the NGARCH. Indeed, we find that the estimated λ react to changes in market volatility.

[29]We obtain similar results for other quartiles.

reported in Figure 2.1. This evidence reinforces our confidence in our estimation methodology premise that options prices do help better capture the equity risk premium.

Although the study of risk aversion has attracted a lot of attention in the finance literature, [30] the emphasis has been mainly put on the first-order aversion parameter $\mathrm{RRA}^{(2)}$, and most of the papers assume a CRRA utility function.[31] As we show in the Appendix, assuming a CRRA utility function entails some tight restrictions among the high-orders risk aversion parameters.[32]

The top (resp. bottom) graph of Figure 2.7 depicts the ERP parameter λ (resp. $\mathrm{RRA}^{(2)}$) implied by $\mathrm{SDF}^{(4)}$. Figure 2.8 provides the same information for the CRRA pricing kernel.[33] The red (resp. black) lines show the average of each time-series (resp. zero-level threshold). As evident from these two figures, for both pricing kernels, the two parameters fluctuate substantially over the sample period. The Arrow-Pratt risk aversion measure $\mathrm{RRA}^{(2)}$ appears to be tightly related to λ as the two parameters share almost the same values and move together. While such tight connection between these two parameters is highly expected when estimating $\mathrm{SDF}^{(2)}$, having it satisfied for more general SDFs reinforces the link between risk compensation and risk aversion.

The $\mathrm{RRA}^{(2)}$ implied by the CRRA pricing kernel is positive in all weeks with an average value of 4.13. This average $\mathrm{RRA}^{(2)}$ is more moderate than the 7.36 reported by Rosenberg and Engle (2002) and more close to the value of 4 estimated by Gormsen and Jensen (2018) using the methodology of Bliss and Panigirtzoglou (2004).[34]

The $\mathrm{RRA}^{(2)}$ implied by $\mathrm{SDF}^{(4)}$ takes on positive values in 96% of the weeks with an aver-

[30]The first papers can be traced back to Hansen and Singleton (1982) and Mehra and Prescott (1985) among others.

[31]See for example Rosenberg and Engle (2002), Bliss and Panigirtzoglou (2004), and more recently Gormsen and Jensen (2018).

[32]Our analysis in Section 2.4.3 clearly points to a rejection of this assumption in the options market.

[33]The CRRA pricing kernel was estimated using exactly the same methodology as for our SDFs, except that the higher-order RRA parameters are restricted and the no-arbitrage condition was added as a nonlinear constraint.

[34]Rosenberg and Engle (2002) estimated both a CRRA utility and a nonlinear pricing kernel using Chebyshev polynomials. The paper uses an asymmetric GARCH model with constant mean that was estimated using Maximum Likelihood. This model does not price options in closed-form which limit the amount of data that could be used in the estimation. For instance, their reported estimates are obtained using short-term one-month options from 1991 to 1995. Based on in-sample pricing results, the authors argue that the nonlinear pricing kernel substantially outperforms the CRRA benchmark. The higher-order RRA parameters are not readily available from the expression of the nonlinear pricing kernel. In practice, one would have to compute higher-order derivatives of the SDF to obtain these parameters. Higher-order derivatives calculation is known to be noisy beyond the second order.

age value of 4.37, which is fairly close to the CRRA $\text{RRA}^{(2)}$ average. The fact that the CRRA and the $\text{SDF}^{(4)}$ implied $\text{RRA}^{(2)}$ are, on average, close facilitates to some extent the comparison of higher-order RRAs. The $\text{RRA}^{(2)}$ implied by $\text{SDF}^{(4)}$ was particularly noisy and took on many negative values in the calm period preceding the 2008 financial crisis and close to the end of our sample. Figure 2.9 shows the dynamics of the annualized equity risk premium (ERP) implied by $\text{SDF}^{(4)}$ and the CRRA pricing kernels. Both ERPs fluctuate substantially over time. The average ERP is 10.54% for the CRRA and 10.10% for $\text{SDF}^{(4)}$. The first row of Table 2.8 gives the ERP for the other SDFs. We can observe that, better capturing options market anomalies requires some slightly higher levels of first-order risk aversion. Importantly, the average ERP is 6.23% for $\text{SDF}^{(2)}$ and 7.00% for $\text{SDF}^{(3)}$. The average ERP is thus increasing with the order of the SDF.[35] Here, it is tempting to interpret the increased average ERP as induced by the added risk factors. As such, skewness risk adds a modest ERP of 0.77% while kurtosis risks adds a more substantial 3.1%.

Moving on to higher-order aversion parameters, it is worth stressing that the expected utility framework implies nonnegative RRAs at all orders.[36] Here, we choose to let the options data reveal the true signs of these parameters. This contrasts with Dittmar (2002a) who restricted the parameters when fitting returns on industry-sorted portfolios. It is well known that higher-order RRA parameters have different scales. For example, for a CRRA utility function, we have, for $n \geq 3$,

$$\text{RRA}^{(n)} = \prod_{j=0}^{n-2} (\text{RRA}^{(2)} + j). \tag{2.13}$$

The above formula implies in particular that the RRA's magnitude grows exponentially with the SDF's order. To ease the interpretation, we define some standardized RRAs that are obtained by inverting the CRRA relationship (2.13) at each order.[37] This normalization allows to have all the RRA parameters on the same scale. In the sequel, we refer to these normalized values as being the RRAs. Note that this transformation is such that $\text{RRA}^{(n)}$=$\text{RRA}^{(2)}$ when

[35] A sequential maximum likelihood will imply here a constant ERP at all orders, which is less attractive.

[36] If the investors have a negative preference for even moments and a positive preference for odd ones (as shown by Scott and Horvath (1980)), then the RRAs, which are normalized versions of the absolute aversion parameters, should be positive.

[37] When the estimated RRA is negative, we set the standardized RRA as the negative of the standardized RRA of the negative of the initial estimate, to preserve the signs of the estimate. This normalization is similar to the inversion that allows to compute Black and Scholes implied volatilities from observed option prices.

the true pricing kernel is CRRA. This gives a visual test of the CRRA assumption

To evaluate the robustness of the estimates, we report in Table 2.10 the correlations between the RRA parameters estimated using different SDFs. The observation (n, m) of this table gives the correlation between RRA$^{(n)}$ estimated using SDF$^{(n)}$ and the same RRA$^{(n)}$ estimated using SDF$^{(m)}$. It appears that the RRA$^{(3)}$ inferred from SDF$^{(3)}$ is at least 65% correlated with any version of it estimated from higher-order SDFs. This figure is even higher for SDF$^{(4)}$ at 86%. This correlation is less than 35% for RRA$^{(2)}$ and 25% for RRA$^{(5)}$. This demonstrates that the aversion to skewness and kurtosis are strongly stable when other risk factors are added to the pricing kernel. Figure 2.7 (respectively Figure 2.10) show the RRA$^{(2)}$, RRA$^{(3)}$, and RRA$^{(4)}$ implied by SDF$^{(4)}$ (respectively SDF$^{(6)}$). The dynamics of RRA$^{(2)}$ inferred from SDF$^{(4)}$ becomes more stable at higher orders.

In Panel A of Table 2.5, we report some descriptive statistics for the RRA parameters implied by SDF$^{(4)}$. The skewness aversion parameter RRA$^{(3)}$ takes on positive values in 98% of the weeks compared to 92% for RRA$^{(4)}$. As a result, all the parameters display some positive averages. These statistics are rather satisfactory given that the parameters are freely estimated without any positivity restriction. We also note that the estimates are fairly stable across time. This contrasts with the usual noisy risk premium estimates yield by the Fama-MacBeth regressions. RRA$^{(3)}$ appears to be the most stable time-series as it has the lowest coefficient of variation.[38] All the three RRA parameters appears to be persistent. A simple regression of each RRA on its lagged value yields an adjusted R^2 of 40% for RRA$^{(2)}$, 55% for RRA$^{(3)}$ and 63% for RRA$^{(4)}$. The relative aversion to skewness and kurtosis risks seem to have decreased over time from their very high levels observed at the beginning of the sample during the dot-com bubble.

Panel A of Table 2.6 reports the correlations between the RRA parameters implied by SDF$^{(4)}$. The skewness and the kurtosis aversion parameters tend to move together as evidenced by the positive correlation of 88% between their times-series. Both parameters are negatively correlated with RRA$^{(2)}$. Indeed, RRA$^{(3)}$ has a significantly -0.19 correlation with RRA$^{(2)}$ while the correlation of RRA$^{(4)}$ with RRA$^{(2)}$ is -0.14. This establishes a qualitative distinction with the CRRA utility. The risk aversion parameters do not move in tandem as we use to believe in the CRRA world. This evidence is further corroborated by the correla-

[38]We define the coefficient of variation as the ration between the standard deviation and the absolute values of the average. This statistics is useful when comparing the randomness of variable of different scales

tion between the RRA parameters and the market volatility. $\text{RRA}^{(2)}$ is negatively correlated with the market volatility while $\text{RRA}^{(3)}$ and $\text{RRA}^{(4)}$ are positively correlated with the latter. To further confirm this evidence, we regress each RRA parameter on its lagged value and the market volatility.[39] The results, which are reported in Panel A of Table 2.7, show that $\text{RRA}^{(2)}$ load negatively on the market volatility while $\text{RRA}^{(3)}$ and $\text{RRA}^{(4)}$ load positively on the market volatility. This implies that the aversion to variance risk is low in high volatility periods while aversion to skewness and kurtosis are high during these episodes. Since the market volatility is a countercyclical variable (see Schwert (1989)), we conclude that $\text{RRA}^{(2)}$ is procyclical while $\text{RRA}^{(3)}$ and $\text{RRA}^{(4)}$ are countercyclical. Given that the risk aversion is expected to be a countercyclical variable, our results tend to suggest that skewness and kurtosis aversions are better proxies for the aggregate risk aversion than the first-order risk aversion. In Table 2.11, we run a similar regression using the $\text{RRA}^{(2)}$ implied by the CRRA utility function. The result shows that the first-order risk aversion is also procyclical in the CRRA world. This result is in line with Bliss and Panigirtzoglou (2004) who uncover the same evidence and hypothesize that the results may be driven by the absence of high risk averse investors during periods of high market volatility.

The $\text{RRA}^{(5)}$ implied by $\text{SDF}^{(5)}$ appears to be well-behaved as it has the expected positive sign in 80% of weeks. However, the time-series substantially changes with the addition of the hyperkurtosis risk. Indeed, with a PK that incorporates hyperkurtosis risk, the $\text{RRA}^{(5)}$ times-series take on a positive sign in only 44% of the weeks.

The $\text{RRA}^{(6)}$ time-series is notoriously negative, having positive values in only 26% of the weeks. This sizeable deviation of the sign of the hyperkurtosis aversion from the expected sign doesn't ease the interpretation of the aversion to this risk. As it appears on Figure 2.11, both $\text{RRA}^{(5)}$ and $\text{RRA}^{(6)}$ time-series are noisier than that of the first three RRAs. As a result, even though $\text{SDF}^{(6)}$ outperforms $\text{SDF}^{(4)}$ both in-sample and out-of-sample, it is more reasonable to adopt the latter as the most valid pricing kernel.

2.4.6 Risk Premium Implications

In this section, we investigate the moments risk premium implications of the aversion to the different higher-order moments. Our objective is to learn about the relationship between

[39] Regressing the RRAs on their lagged values allow to alleviate the effect of autocorrelation in the residuals.

risk aversion and moment risk premium. Importantly, we want to answer the following question: what can we learn about risk aversion by simply observing the moments risks premium?

Before tackling this question, we want to formally define the moments risk premium. The first moment risk premium or $\mathrm{RP}^{(1)}$ simply refers to the equity risk premium whose behaviour has been analyzed in Section 2.4.5. For the higher-orders moments, we define a multiplicative risk premium as the ratio of risk-neutral centred moment and its physical counterpart powered by the inverse of the moment order, that is, for $n > 1$:

$$\mathrm{RP}^{(n)} = \left(\frac{\mathbb{E}^{\mathbb{Q}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{Q}}(R_{t\to T})\right)^n\right)}{\mathbb{E}^{\mathbb{P}}\left((R_{t\to T} - \mathbb{E}^{\mathbb{P}}(R_{t\to T}))^n\right)} \right)^{1/n}.$$

Note that the multiplicative moments risk premium are suitably normalized to make them comparable. In the Appendix, we prove that our models deliver closed-form risk neutral moments which eases the calculation of the premium.[40] Table 2.8 reports the sample averages of the one-month ahead moments premium for different pricing kernels. In parentheses below each average premium, we report the Newey-West (1987) t-stat for testing the hypothesis H_0: "$\mathrm{E}(\mathrm{RP}^{(1)})$=0" for the first order moment and H_0: "$\mathrm{E}(\mathrm{RP}^{(n)})$=1" for higher-order moments.

As it can be seen from this table, the null hypothesis H_0 is rejected for all the orders and all the pricing kernels implying the existence of some premium in each moment risks. Here, we observe that $\mathrm{RP}^{(n)}$ is greater than one on average for any $n > 1$ and any pricing kernel. This result implies that moments risk are intensified under the risk neutral moments. Odd moments such are the skewness and the hyperskewness and more negative under $\mathbb{Q}$ while even moments such as variance, kurtosis and hyperkurtosis are more positive under $\mathbb{Q}$ compared to the physical $\mathbb{P}$ measure.

Table 2.8 suggests a striking evidence according to which a simple pricing kernel such $\mathrm{SDF}^{(2)}$, in which only variance risk is priced, can generate high-order moments risk premium up to the sixth order. This evidence implies, in particular that, inferring information on risk aversion from a single higher-order moment may be misleading. Subsequently, each higher-order $\mathrm{RP}^{(n)}$ increases as we add some new factors to the pricing kernel. Indeed, adding skewness and kurtosis to the pricing kernel contribute to significantly increase the $\mathrm{RP}^{(n)}$ and

[40]This was to be expected since we have closed-form option pricing. We prove that risk-neutral moments are functions of physical moments and risk aversion parameters.

thus increase the divergence of $\mathbb{Q}$ moments from the $\mathbb{P}$ moments. The results from Table 2.8 also suggest that hyperskewness and hyperkurtosis have marginal contributions to $\mathrm{RP}^{(n)}$. This last evidence is firstly consistent with our previous result (see Section 2.4.3) that these two risks capture a rather small variation in options prices and secondly in line with the fact that the signs of their corresponding RRA parameters are subject to major deviations from the expected theory predictions (see Section 2.4.5).

Broadly speaking, our results are consistent with Bakshi and Madan (2006) who theoretically shows that the departure between risk-neutral and physical index volatility is connected to the higher-order physical return moments and the parameters of the pricing kernel process. Their theory predicts a positive volatility spreads when investors are risk averse, and when the physical index distribution is negatively skewed and leptokurtic. Empirically, we complete their investigation by looking at the relationship between higher-order risk aversion and their moments risk premium.

Now focusing on the premium implied by $\mathrm{SDF}^{(4)}$, Panel A of Table 2.9 reports the correlation between the estimated $\mathrm{RP}^{(n)}$ on one side, and their correlations with market volatility on the other side. It appears that $\mathrm{RP}^{(2)}$, $\mathrm{RP}^{(3)}$, and $\mathrm{RP}^{(4)}$ are strongly positively correlated, and also co-move with the market volatility. Indeed, the correlation between these three RP's is at least 93%. This results contrasts with the negative correlation that we documented between $\mathrm{RRA}^{(2)}$ and $\mathrm{RRA}^{(3)}$ ($\mathrm{RRA}^{(4)}$) and reinforces our claim that risk aversion should not be directly inferred from single moments risk premium.

Now comparing with the parametric pricing kernels, Table 2.8 additionally shows that the two benchmarks CRRA and BCHJ are also quite successful at capturing risk premium at the different orders. Panel B of Table 2.9 shows that the correlation between $\mathrm{RP}^{(2)}$, $\mathrm{RP}^{(3)}$, and $\mathrm{RP}^{(4)}$ are even stronger that the ones reported for $\mathrm{SDF}^{(4)}$. In particular, $\mathrm{RP}^{(2)}$ and $\mathrm{RP}^{(4)}$ are perfectly correlated (correlation=1). However, we note that the BCHJ implied premium are only barely correlated with the market volatility in contrast with what we observe for $\mathrm{SDF}^{(4)}$. Figures 2.12 and 2.13 plot the $\mathrm{RP}^{(n)}$ time-series for the first six orders for both the BCHJ and the $\mathrm{SDF}^{(4)}$ pricing kernels. As evident from these figures, the BCHJ SDF implies systematically a higher premium than the $\mathrm{SDF}^{(4)}$ for all the moments orders.[41] This might be the underlying reason of the underperformance of the BCHJ utility compared to $\mathrm{SDF}^{(4)}$

[41] The volatility premium parameter in the BCHJ pricing kernel is estimated dynamically each week.

and the CRRA when pricing call options. To further develop this point, note that, the normalized volatility risk premium is 1.24 for the BCHJ against 1.09 for SDF$^{(4)}$on average.[42] This observation demonstrates that it is possible to successfully match option prices without excessively inflating the volatility under the risk neutral measure. [43]

2.4.7 The Shape of the Pricing Kernel

In this section, we study the implications of our pricing results for the shape of the pricing kernel.

In Figure 2.14, we plot the one-month ahead SDFs on three specific weeks.[44] The top left (right) figure displays the shape of the SDFs for the week in which the VIX is at its 25% quartile (50%). The bottom left figure graphs the SDFs on the week in which the VIX is at its 75% quartile. On each graph, the blue line depicts SDF$^{(2)}$ as a function of future index price expressed in terms of returns standard deviations. The red line provides the same information for SDF$^{(3)}$, the green line plots SDF$^{(4)}$ and the black line graphs SDF$^{(5)}$.

As evident from these figures, the SDF$^{(5)}$'s representation is almost confounded with that of SDF$^{(4)}$. This observation is consistent with our early evidence that SDF$^{(5)}$ improves the options fit by only 1%, on average, both in-sample and one-week ahead out-of-sample.

As it was to be expected, SDF$^{(2)}$ is linear function of the wealth. SDF$^{(3)}$ and SDF$^{(4)}$ seem to be non-monotonic as evidenced by their sharp increase in the high positive returns domain. This non-monotonic features appears to be less pronounced on a low-volatility day. It also appears that SDF$^{(4)}$ is less non-monotonic compared to SDF$^{(3)}$. Indeed, a representative agent endowed with SDF$^{(4)}$ seems to have higher marginal utilities in low returns states and

[42] It is interesting to compare our average of the BCHJ volatility risk premium 1.24 to the full sample estimate 1.124 reported by Christoffersen, Heston, and Jacobs (2013)(see Table 4 on page 1989). We hypothesize that this non-negligeable difference might be driven by the difference in our samples. Our sample is far bigger than the ones used in Christoffersen, Heston, and Jacobs (2013)and BCHJ and it has relatively more put options.

[43] It is worth emphasizing that our SDFs are maturity-independent. Bliss and Panigirtzoglou (2004) found that the classical Arrow-Pratt measure of relative risk aversion RRA$^{(2)}$ exhibits some maturity-dependent features. Our framework can be easily extended to incorporate this maturity-dependent feature. Additionally, based on the usual projection argument, we also hypothesize that a maturity-dependent SDF$^{(4)}$ will certainly out-perform the BCHJ pricing kernel. In this paper, we do not pursue this estimation since the theoretical motivation for the maturity-dependent RRA parameters is still not well-understood. Maturity-dependent risk aversion parameters suggest some sort of segmentation of the options markets across horizons which is not always compatible with our maintained assumption of a representative agent in the economy.

[44] The estimated-SDFs are maturity-independent. The maturity level is only used to convert the wealth level into standard deviations returns.

relatively less marginal utilities in high returns states.[45]

Overall, these results confirm the finding of Christoffersen, Heston, and Jacobs (2013) and BCHJ according to which a non-linear SDF is needed to capture the options prices anomalies. Here, we complement their finding by documenting how the shape of SDF is affected by the priced kurtosis risk.

2.4.8 Robustness Analysis

A real challenge for the estimation of the RRAs lies in the fact that the asset pricing factors in which we are interested in are powers of the market returns and are thus expected to be "highly correlated". In table 2.2, we report the correlations between the normalized physical moments of the market returns. The table shows that the correlation between the moments M_2, M_3, M_4, M_5, and M_6 is at least 82% confirming that the risk factors are indeed "highly correlated". As it is well known, a high correlation between asset pricing factors can raise some identification issues. In this paper, we argue that the use of options allows us to alleviate this identification issue. In fact, option prices are highly nonlinear functions of the factors and as such using options presents a better alternative to using stocks returns.

To confirm this intuition, we implement a Monte Carlo exercise whose results are reported in Table 2.12. For one hundred randomly chosen weeks, we generate synthetic option prices using some known pricing kernels reported in the table's rows. $\text{CRRA}^{(n)}$ is the CRRA pricing kernel truncated at order n using the standard Taylor expansion. Recall that the RRAs of the CRRA pricing kernel are functions of the $\text{RRA}^{(2)}$ which we calibrate using the data-based weekly estimates of the CRRA risk aversion parameter $\text{RRA}^{(2)}$. $\text{SDF}^{(n)}$ is the estimated nonparametric pricing kernel of order n whose parameters are calibrated using the weekly estimates of the RRAs. Once the synthetic options prices are generated, we use our estimation methodology to recover the true parameters. The first column of the table provides the average pricing errors measured by the RIVRMSE. The remaining columns report the average of the absolute relative errors on the RRAs (Panel A) and the median of the absolute relative errors on the RRAs (Panel B). It clearly appears that the options pricing errors are small which provides an evidence that the true prices are perfectly matched. As a direct implication, the

[45] Since the average of the SDF is the same for all SDFs, this is was to be expected. Higher values in the left tail of the SDFs will be certainly associated with lower values in the right-tail.

parameters estimates have relatively small errors as measured by the median errors. The $\mathrm{RRA}^{(5)}$ and $\mathrm{RRA}^{(6)}$ parameters have a few outliers that slightly inflate the average errors. Overall, we conclude that the Monte Carlo results are satisfactory for all the pricing kernels.

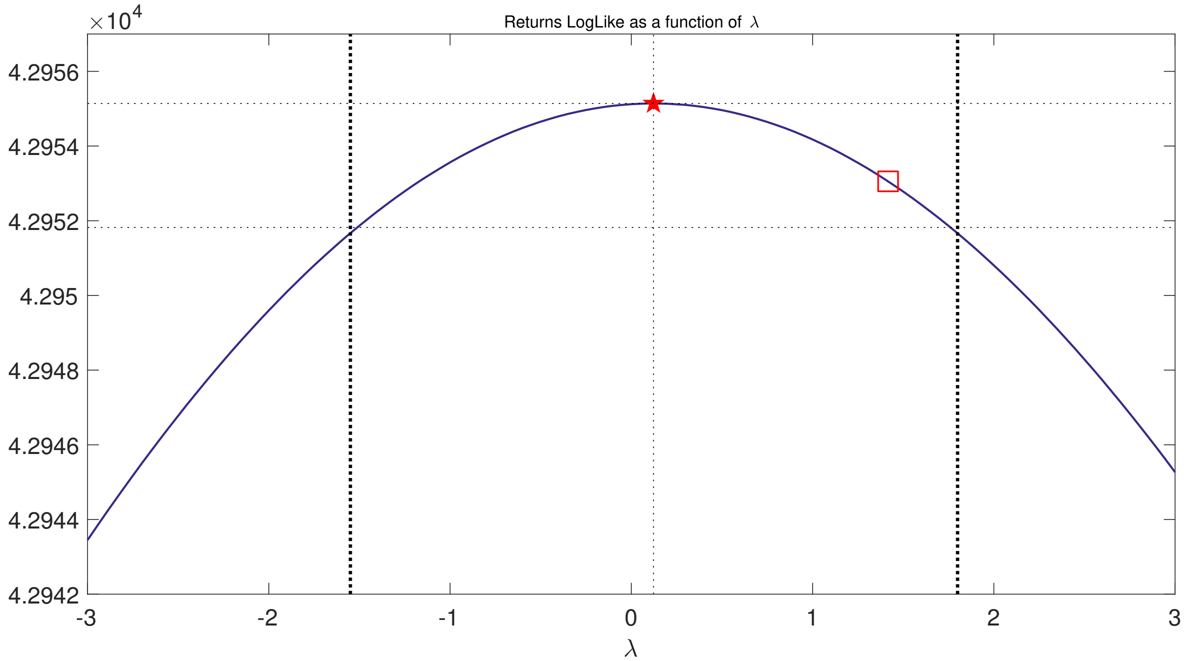
2.5 Conclusion

In this paper, we embed the higher-order risks in an option pricing framework and analyze how the different moments risks contribute to fitting the index option prices. We use a methodology that allows us to accurately pin down the dynamics of the risk aversion parameters. Our results indicate that $\mathrm{SDF}^{(4)}$ is a fairly good approximation of the true SDF, as evidenced by how well it fits options prices. When compared with the BCHJ pricing kernel, which is perhaps the "most successful pricing kernel" in the discrete time literature, we find that both pricing kernel are virtually undistinguishable in terms of options fit.

We also study the higher-order risks beyond the kurtosis such as hyperskewness and hyperkurstosis risks which are valued by a CRRA investor. Empirically, we find that together hyperskewness and hyperkurtosis have a rather small contribution to asset pricing and the estimates of the aversion parameters exhibit substantial deviations from the expected utility theory predictions. Our analysis also reveals that the CRRA pricing kernel is not supported by the options data partially because the variance, skewness and kurtosis risks aversion parameters do not move in tandem as implied by the CRRA utility. Indeed, while skewness and kurtosis risks aversion parameters strongly co-move, the aversion to variance parameter seems to be negatively correlated to the latter.

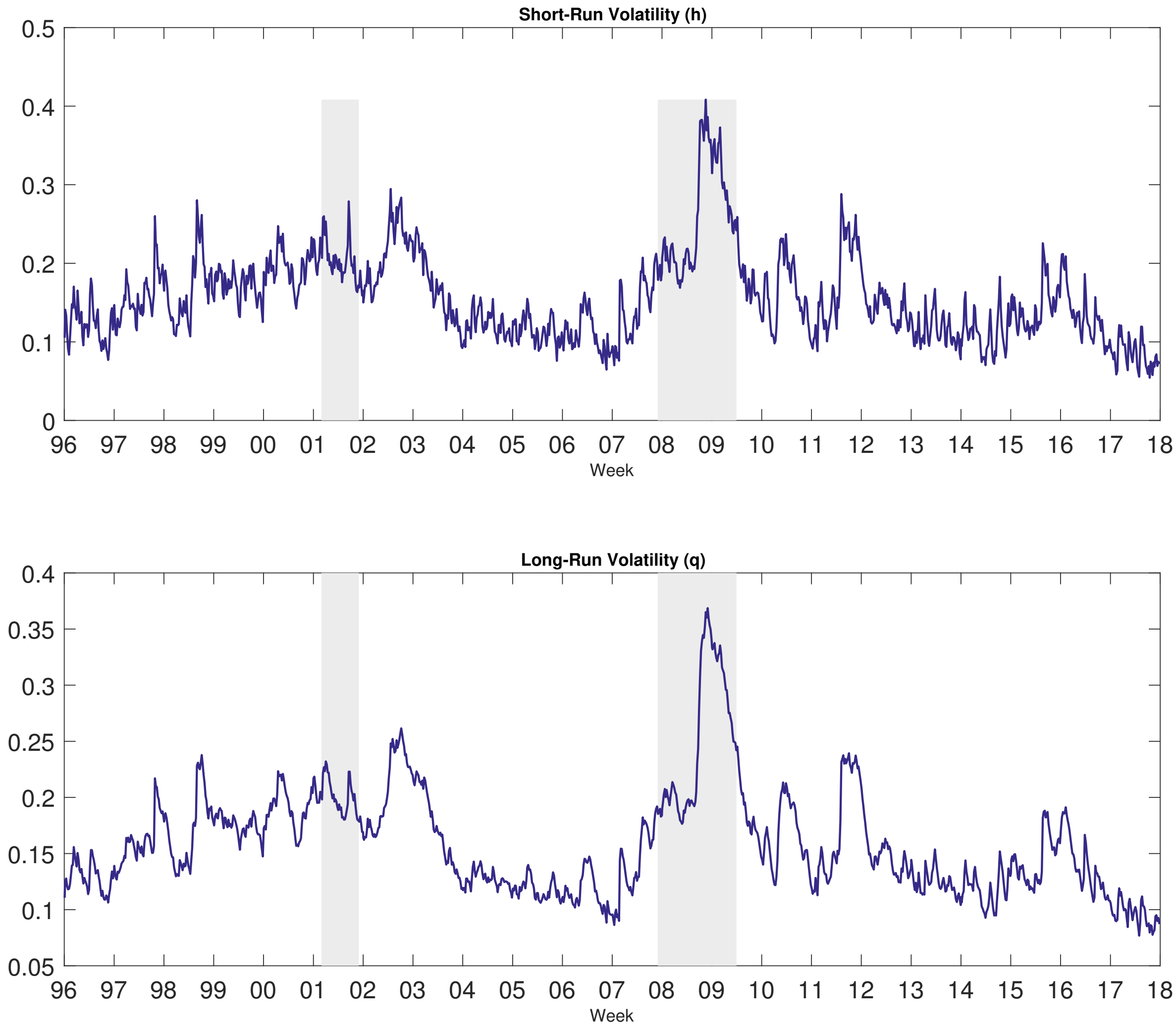
Figures and Tables of Chapter 2

Figure 2.1: Returns Log-likelihood as a function of the Equity Risk Premium Parameter λ



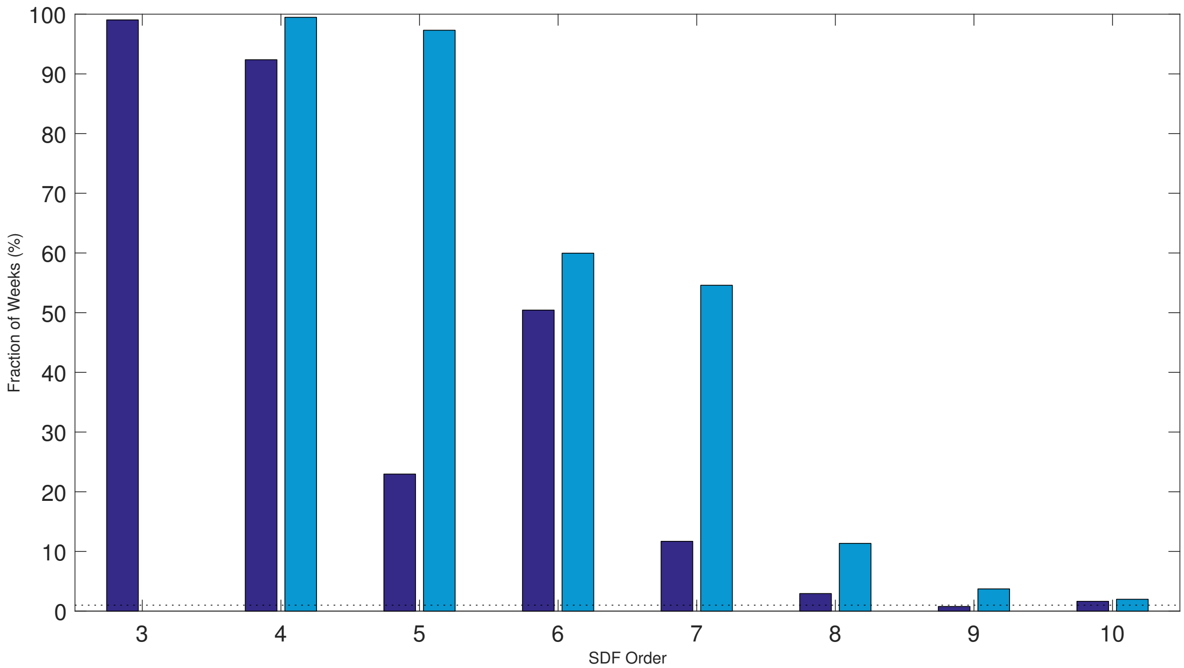
This figure depicts the returns log-likelihood as a function of the equity risk premium (ERP) parameter λ. The values of λ correspond to 300 points uniformly spaced of the range [-3,3]. Next, for each value of λ, we maximize the log-likelihood over the remaining physical parameters. By doing so, we assume that the returns are generated by the Component IG-GARCH(1,1) model. The red starred point shows the global maximum of the log-likelihood function and the red squared point is obtained when the λ parameter is calibrated using the value $\log(\mathrm{E}(r_t^e))) \big/ \mathrm{Var}(r_t^e)$, where r_t^e are the excess returns. This calibrated value is such that the convexity-adjusted ERP equals the average excess returns. In particular, this calibrated λ implies an annualized ERP of 3.91% . The two dotted horizontal lines form the 1% confidence interval around the global maximum. The two dotted vertical lines show the range of λ's that are not rejected at a 1%confidence level by the likelihood ratio test. The lower (resp upper) bound of this range implies an ERP of -4.11% (resp 5.00%) on an annual basis.

Figure 2.2: Time-series of Short-run and Long-run Volatilities



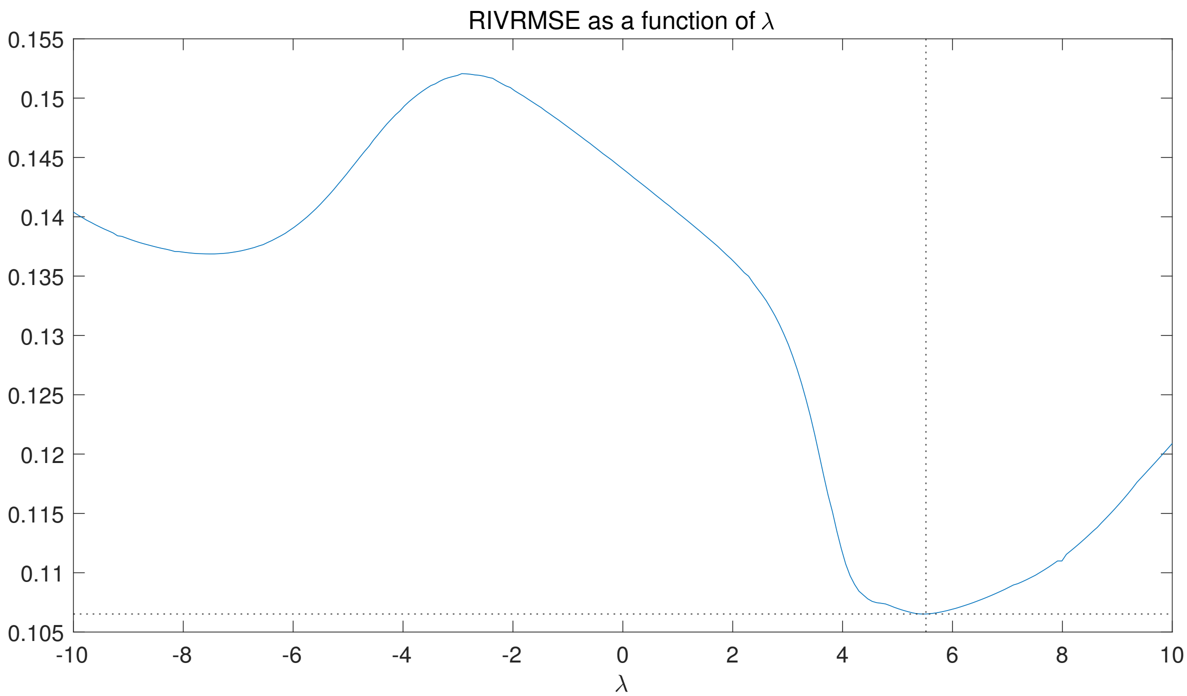
The top (resp. bottom) panel of this figure shows the dynamics of the short-run (resp. long-run) volatility factor implied by the Component IG-GARCH(1,1) physical model. The two volatility factors are filtered using the returns-based maximum likelihood estimates. The ERP parameter λ is calibrated such that the convexity-adjusted model-based ERP equal the average excess returns, that is $\lambda = \log(\mathrm{E}(r_t^e)))\big/\mathrm{Var}(r_t^e)$, where r_t^e.

Figure 2.3: In-sample Likelihood Ratio Test for the Statistical Significance of the RRA parameters



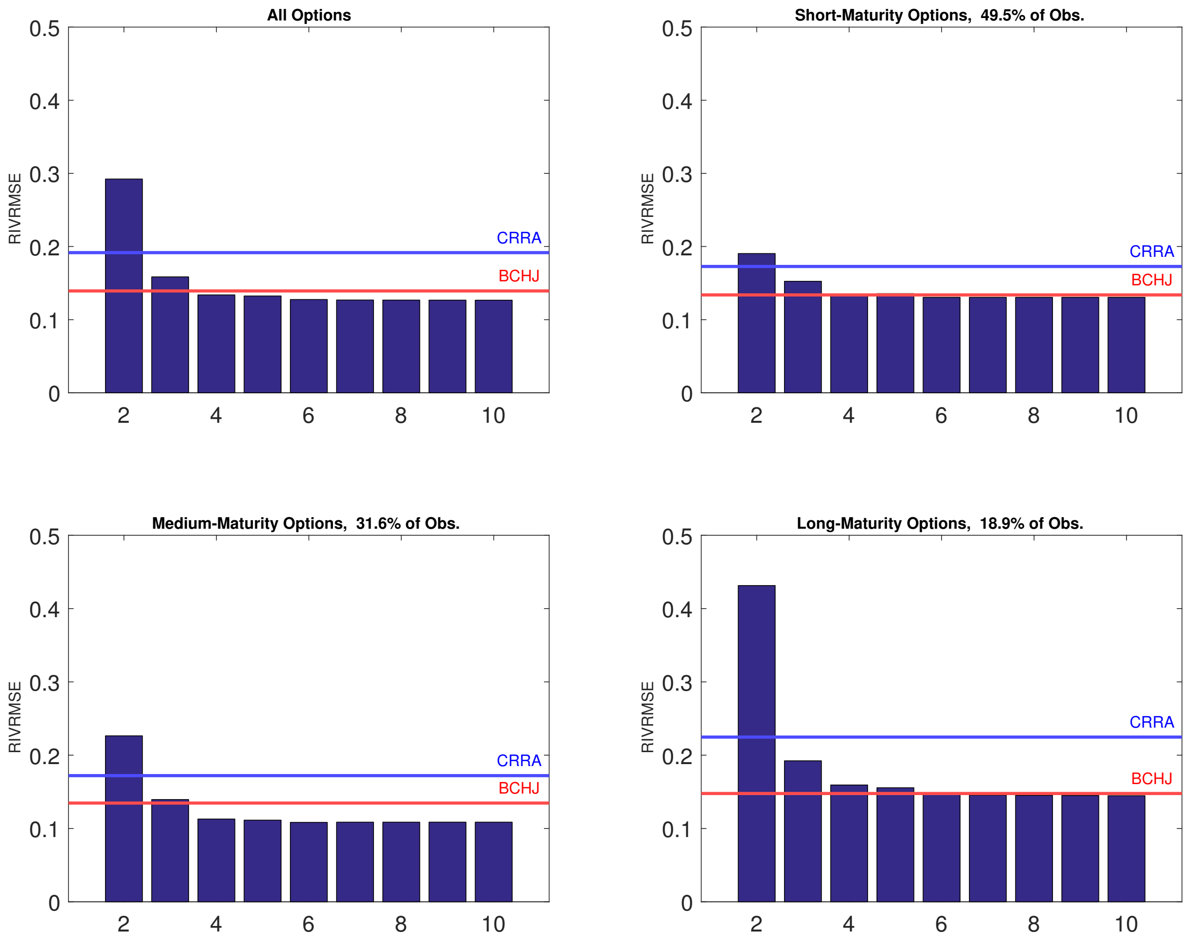
Using the in-sample results, we implement the Likelihood Ratio (LR) test of the following two hypotheses: "$H_0^{n,1}$: SDF$^{(n)}$ does not over-perform SDF$^{(n-1)}$ in terms of options fit" and "$H_0^{2,n}$: SDF$^{(n)}$ does not over-perform SDF$^{(n-2)}$ in terms of options fit". The purple bars reports the fraction of weeks for which $H_0^{n,1}$ is rejected at the 1% confidence level and the lighted blue bars depicts the same information for $H_0^{n,2}$. The dotted horizontal line refers to the 1% level. Here, we interpret a rejection of $H_0^{n,1}$ as the statistical significance of the the RRA$^{(n)}$ implied by SDF$^{(n)}$. Similarly, a rejection of $H_0^{n,2}$ can be interpreted as a joint statistical significance of RRA$^{(n)}$ and RRA$^{(n-1)}$ implied by SDF$^{(n)}$. The sample covers the period 1996-2017 and contains 1158 weeks.

Figure 2.4: Option RIVMSE as a function of the ERP parameter λ



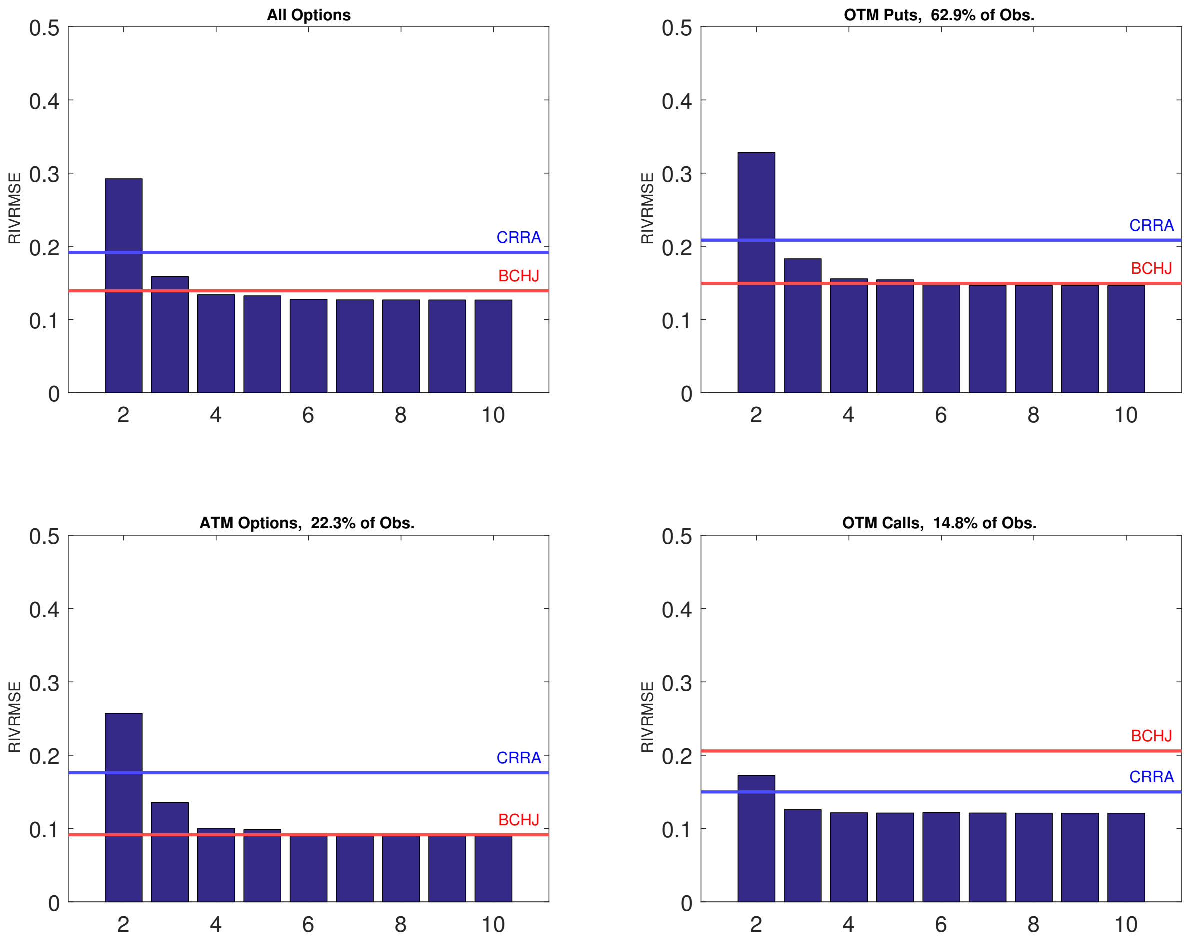
This figure graphs the options root mean square of the relative implied volatility (RIVRMSE) as a function of the ERP parameter λ for one typical week in our sample. For this illustration, we use the week for which the value of the volatility index VIX is the closest to its first quartile. In our sample, this week runs from September 20, 2004 to September 24, 2004. For the values of λ, we use 250 uniformly spaced points between -10 and 10. Options are priced assuming SDF4 is the "true" pricing kernel. For each value of λ, we maximize the options log-likelihood over the SDF4 parameters. The dotted vertical black line shows the λ value that minimizes the RIVRMSE and the dotted horizontal black line shows the corresponding RIVRMSE. This optimal λ implied a conditional ERP of 7.54%.

Figure 2.5: In-sample Goodness of Fit by Maturity



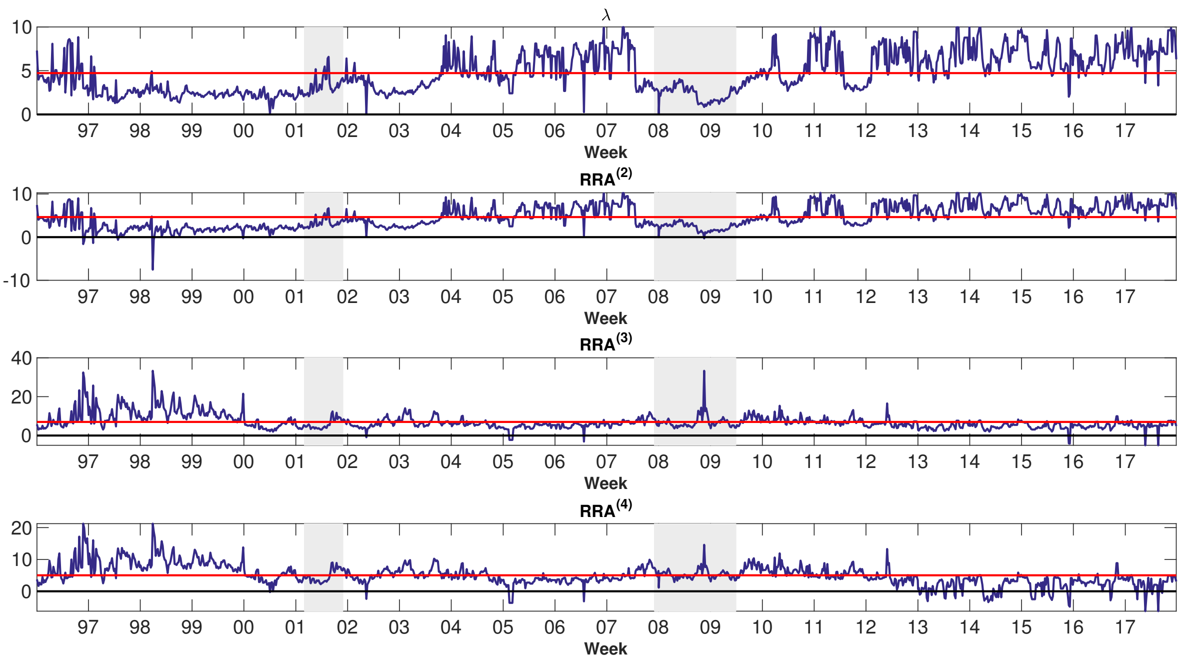
We report the in-sample RIVRMSE results by maturity buckets. For each plot, the n^{th} $(2 \leq n \leq 10)$ bar provides the RIVRMSE whenSDF $^{(n)}$ is used as pricing kernel. The top left figure gives the RIVRMSE for the full dataset of options. The top right figure provides the RIVRMSE for short-maturity options. The bottom left graph gives the RIVRMSE for medium-maturity options and the last graph provides the same information for long-maturity options. Short-term options have between 6 and 45 days to maturity (DTM), and long-term options have between 92 and 365 DTM. We estimate the RRA parameters using options data from 1996 to 2017 on a weekly basis.

Figure 2.6: In-sample Goodness of Fit by Moneyness



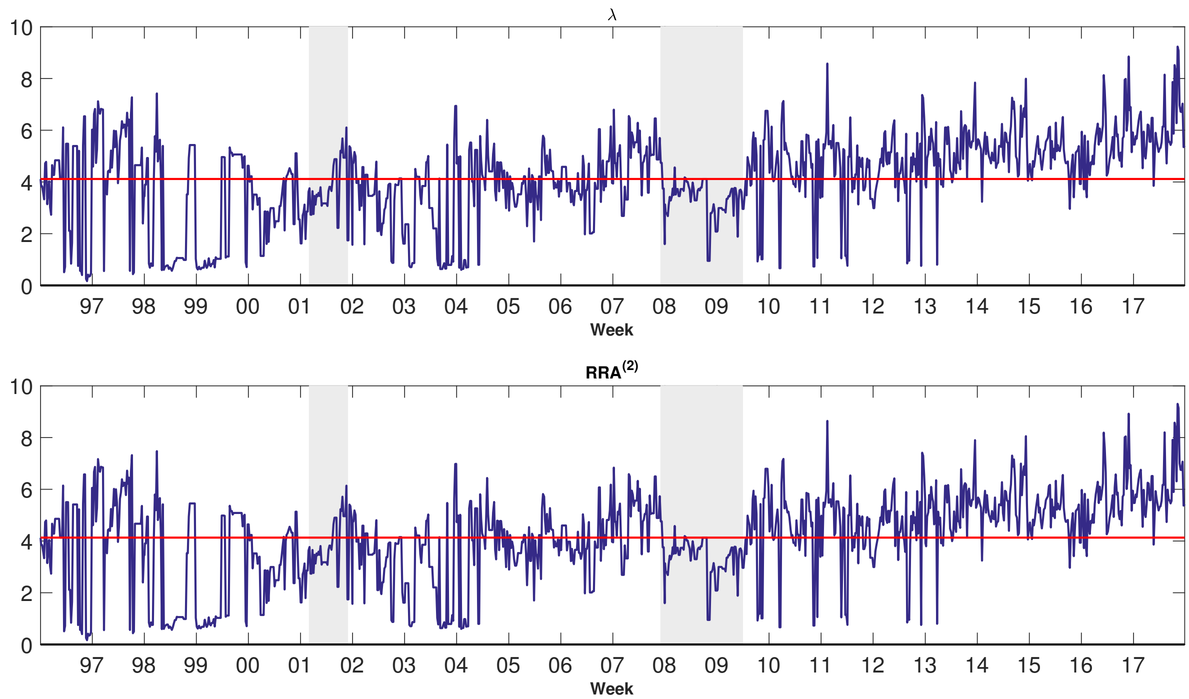
We report the in-sample RIVRMSE results by moneyness buckets. For each plot, the n^{th} $(2 \leq i \leq 10)$ bar provides the RIVRMSE when $SDF^{(n)}$ is used as pricing kernel. The top left figure gives the RIVRMSE for the full dataset of options. The top right figure provides the RIVRMSE for Out-of-money (OTM) puts. The bottom left graph gives the RIVRMSE for At-the-money (ATM) options and the last graph provides the same information for OTM calls. Following Israelov and Kelly (2017), we measure the moneyness using the proxy $\log(K/S) / \left(\text{VIX}\sqrt{DTM}\right)$, where VIX is the CBOE VIX index. OTM puts are options with moneyness less than -1, ATM options are those options having their moneyness between -1 and 0 and finally OTM calls have their moneyness greater than 0.

Figure 2.7: SDF$^{(4)}$ Risk Aversion Estimates



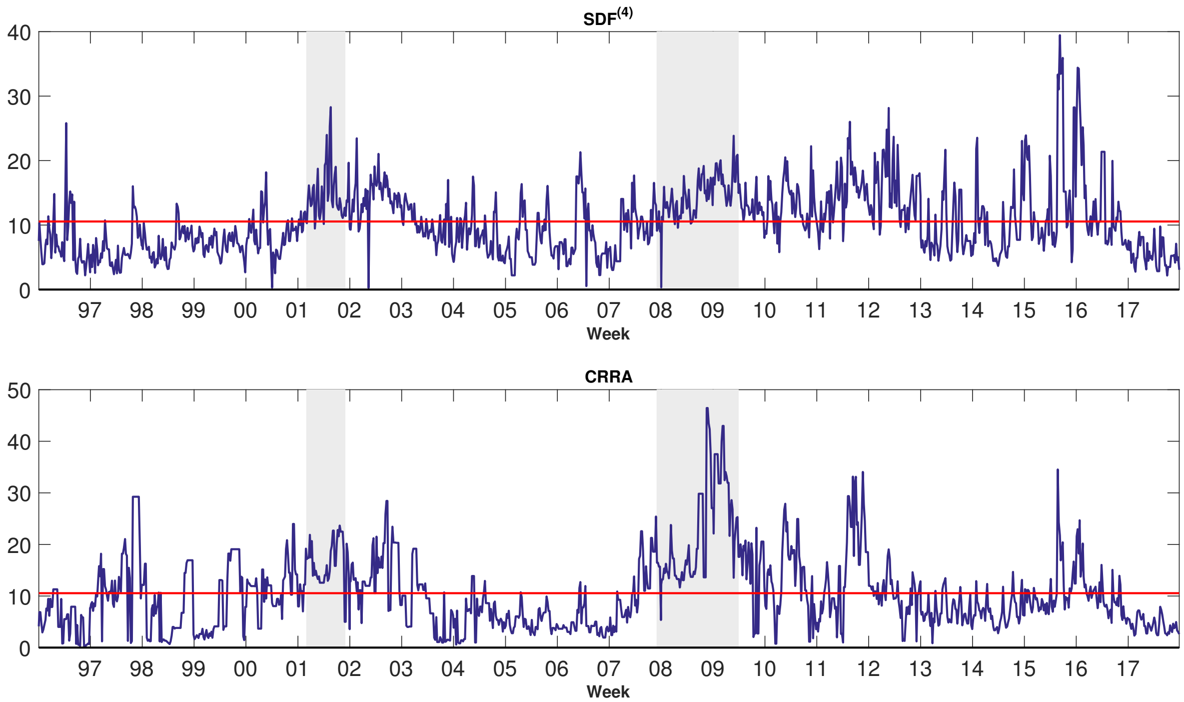
This figure depicts the ERP parameter λ, RRA$^{(2)}$, RRA$^{(3)}$ and RRA$^{(4)}$ estimated from SDF$^{(4)}$. The red (black) lines show the average of each time-series (zero- level threshold). Each $RRA^{(n)}$ parameter is a standardized version of $(-1)^n S_t^n U^{(n+1)}(S_t)/U'(S_t)$, where U is the utility function of the representative agent. For a CRRA utility function, the latter quantity is related to the RRA$^{(2)}$ by a well known parametric function. The standardized $RRA^{(n)}$ is obtained by applying the reciprocal of this parametric function to the estimated $(-1)^n S_t^n U^{(n+1)}(S_t)/U'(S_t)$. If the utility is CRRA, then we have RRA $^{(n)}$=RRA $^{(2)}$ for all n. The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis. The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis.

Figure 2.8: CRRA Risk Aversion Estimates



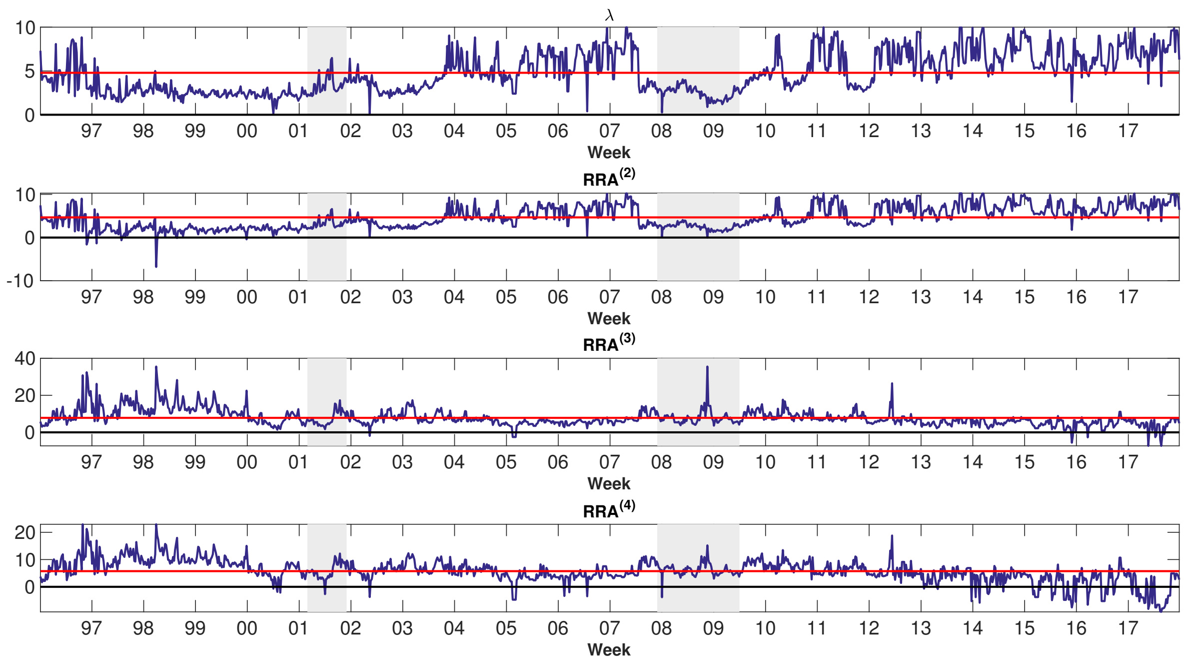
This figure depicts the ERP parameter λ and $\text{RRA}^{(2)}$ estimated from the CRRA pricing kernel. The red (black) lines show the average of each time-series (zero- level threshold). The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis.

Figure 2.9: SDF$^{(4)}$ and CRRA Expected Risk Premium



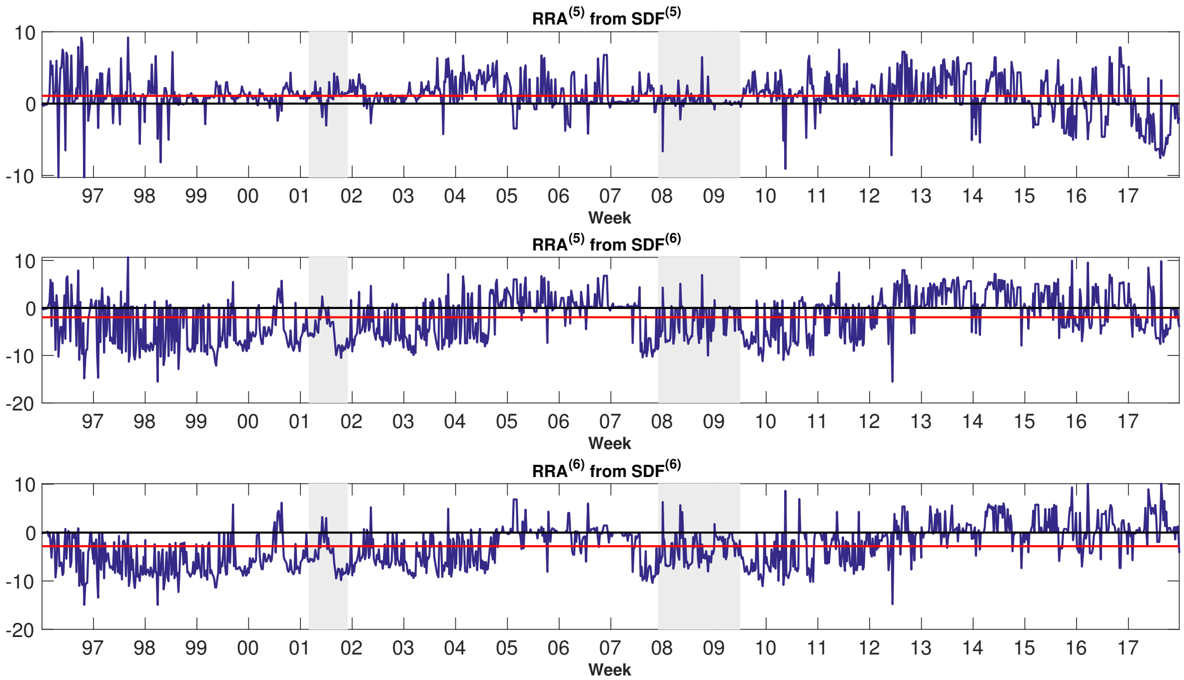
This figure shows the annualized equity risk premium estimated from the SDF$^{(4)}$ and the CRRA pricing kernels. The red (black) lines show the average of each time-series (zero- level threshold). The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis.

Figure 2.10: Relative Risk Aversion Parameters Implied by SDF $^{(6)}$



This figure presents the relative risk aversion parameters implied by SDF$^{(6)}$ together with the estimate of the physical equity risk premium parameter λ. Each $RRA^{(n)}$ parameter is a standardized version of $(-1)^n S_t^n U^{(n+1)}(S_t)/U'(S_t)$, where U is the utility function of the representative agent. For a CRRA utility function, the latter quantity is related to the $RRA^{(2)}$ by a well known parametric function. The standardized $RRA^{(n)}$ is obtained by applying the reciprocal of this parametric function to the estimated $(-1)^n S_t^n U^{(n+1)}(S_t)/U'(S_t)$. For a CRRA utility, we have $RRA^{(n)}{=}RRA^{(2)}$ for all n. The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis.

Figure 2.11: SDF$^{(6)}$ Higher-Order Risk Aversion Estimates



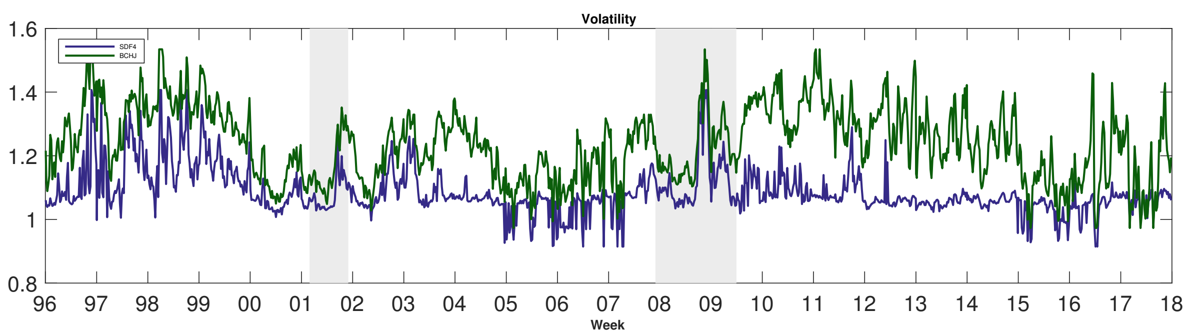
The top figure presents the time-series of RRA $^{(5)}$ estimated from SDF$^{(5)}$. The remaining two figures plots RRA$^{(5)}$ and RRA $^{(6)}$ estimated from SDF $^{(5)}$. Each RRA $^{(n)}$ parameter is a standardized version of $(-1)^n S_t^n U^{(n+1)}(S_t)/U'(S_t)$, where U is the utility function of the representative agent. The options data covers the period from 1996 to 2017 and the risk aversion parameters are estimated on a weekly basis.

Figure 2.12: Moments Risks Premia

This figure presents the time-series of the first three moments premium implied by our nonparametric $\text{SDF}^{(4)}$ pricing kernel (purple line) and the benchmark BCHJ pricing kernel (green line). For the first order moment, the risk premium is simply the annualized standard equity risk premium. For the higher-orders moments, we define the risk premium as the ratio of risk-neutral centred moment and its physical counterpart powered by the inverse of the moment order, that is, for $n > 1$:

$$\text{RP}^{(n)} = \left(\frac{\mathbb{E}^{\mathbb{Q}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{Q}}(R_{t\to T})\right)^n\right)}{\mathbb{E}^{\mathbb{P}}\left((R_{t\to T} - \mathbb{E}^{\mathbb{P}}(R_{t\to T}))^n\right)} \right)^{1/n}.$$

As standard in the literature, we winsorize the moments premium data at the 1% level.



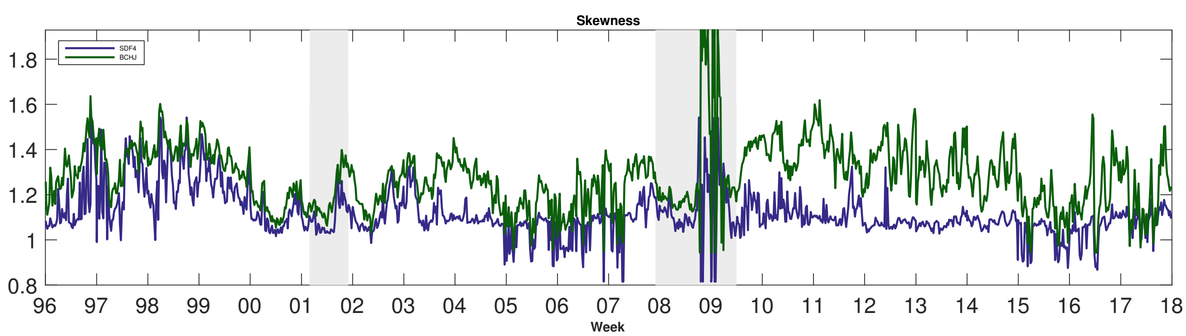


Figure 2.13: Moments Risks Premia (continued)

This figure presents the time-series of orders 4,5 and 6 moments premium implied by our nonparametric $\mathrm{SDF}^{(4)}$ pricing kernel (purple line) and the benchmark BCHJ pricing kernel (green line). For the first order moment, the risk premium is simply the annualized standard equity risk premium. For the higher-orders moments, we define the risk premium as the ratio of risk-neutral centered moment and its physical counterpart powered by the inverse of the moment order, that is, for $n > 1$:

$$\mathrm{RP}^{(n)} = \left(\frac{\mathbb{E}^{\mathbb{Q}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{Q}}(R_{t\to T})\right)^{n}\right)}{\mathbb{E}^{\mathbb{P}}\left((R_{t\to T} - \mathbb{E}^{\mathbb{P}}(R_{t\to T}))^{n}\right)} \right)^{1/n}.$$

As standard in the literature, we winsorize the moments premium data at the 1% level.

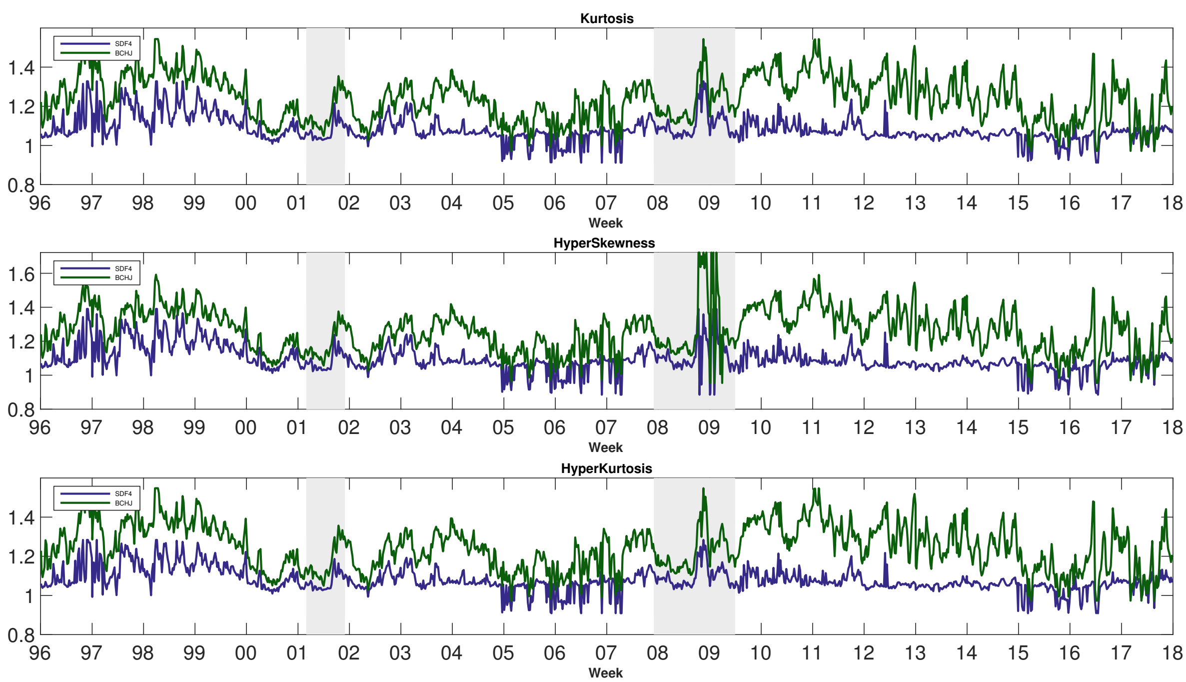
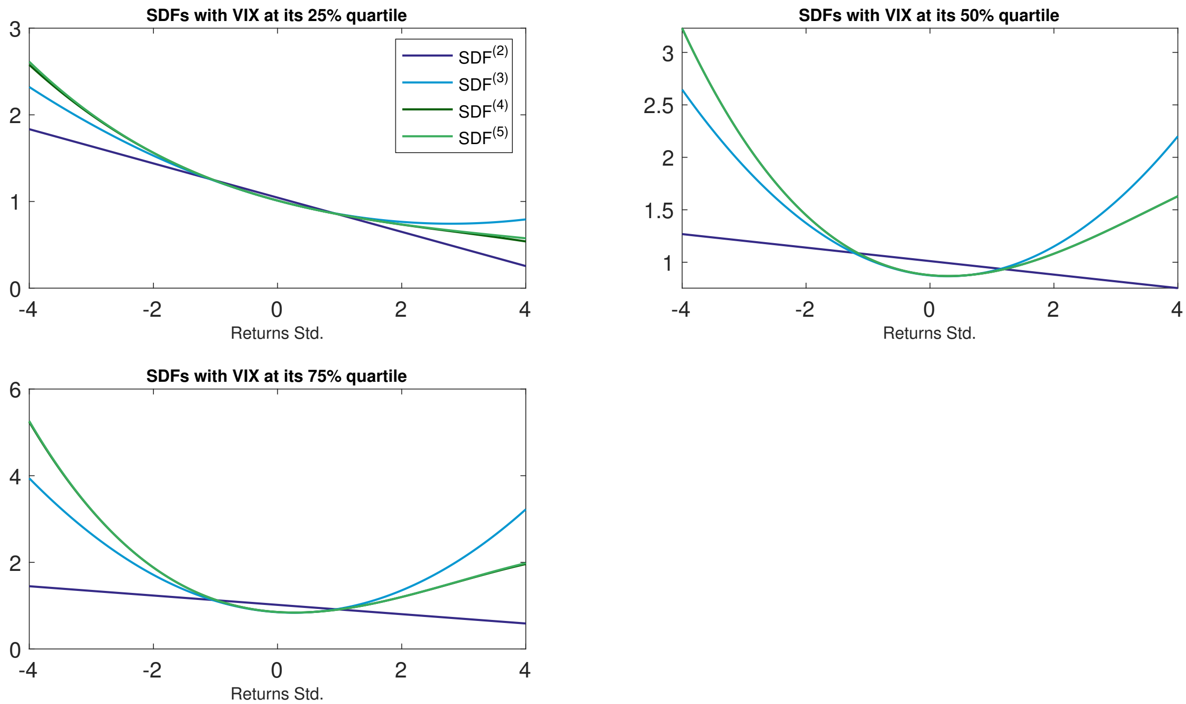


Figure 2.14: One-month Pricing Kernel



We plot the one-month ahead SDFs on three specific weeks. The top left (right) figure displays the shape of the SDFs for the week in which the VIX is at its 25% quartile (50%). The bottom left figure graphs the SDFs on the week in which the VIX is at its 75% quartile. On each graph, the purple line depicts SDF$^{(2)}$ as a function of future index price expressed in terms of returns standard deviations. The blue line provides the same information for SDF$^{(3)}$, the green line plots SDF$^{(4)}$ and the black line graphs SDF$^{(5)}$.

Table 2.1: Physical Parameters Estimates

σ^2	λ	η	a_h	b_h
1.02e-04	1.42e+00	-4.66e-04	2.83e-01	8.82e-01
		(1.15e-06)	(3.94e-03)	(5.93e-04)
c_h	a_q	b_q	c_q	LogLike
3.73e-03	4.26e-01	9.91e-01	4.50e-03	4.30e+04
(2.96e-05)	(4.44e-03)	(3.94e-06)	(1.55e-05)	

This table reports the MLE estimates of the parameters Component IG-GARCH(1,1) parameters under the physical measure. The estimation uses returns data from 19620705 to 20171229. Below each parameter, we report the parameter's standard error obtained using the outer product of the gradient. The equity risk premium parameter is calibrated using the sample average of the market returns and the long-run volatility parameter σ^2 is calibrated using the sample variance.

Table 2.2: Physical Moments Correlations

This table reports the correlations between the signed physical moments for different orders ranging from two to ten. As standard in the literature, the moments are computed at the one-month horizon. M_2 is simply the annualized square root of the variance. For $n \geq 3$, the standardized moment M_n is defined as:

$$M_n = \frac{\mathrm{E}^{\mathbb{P}}\left(\left(R_{t\to t+30} - \mathrm{E}^{\mathbb{P}}(R_{t\to t+30})\right)^n\right)}{\left(\mathrm{E}^{\mathbb{P}}\left(R_{t\to t+30} - \mathrm{E}^{\mathbb{P}}(R_{t\to t+30})\right)^2\right)^{n/2}}.$$

	M_2	$-M_3$	M_4	$-M_5$	M_6	$-M_7$	M_8	$-M_9$	M_{10}
M_2		-0.97	-0.85	-0.93	-0.82	-0.85	-0.76	-0.55	-0.13
$-M_3$			0.94	0.99	0.92	0.94	0.87	0.64	0.16
M_4				0.98	1.00	1.00	0.98	0.74	0.20
$-M_5$					0.97	0.98	0.93	0.69	0.18
M_6						1.00	0.99	0.75	0.21
$-M_7$							0.98	0.75	0.21
M_8								0.81	0.27
$-M_9$									0.76

Table 2.3: Options Data Descriptive Statistics

	Panel A: By maturity			
	Short Maturity	Medium Maturity	Long Maturity	All Options
% of Contracts	51.20	29.89	18.91	100.00
Average Implied Vol	0.23	0.24	0.27	0.24
Average Price	6.03	11.10	18.21	9.85
Average Bid-Ask Spread	0.52	0.94	1.27	0.79
Average Volume	869.95	651.33	558.50	745.71
	Panel B: By moneyness			
	OTM Puts	ATM Options	OTM Calls	All Options
% of Contracts	62.91	22.27	14.82	100.00
Average Implied Vol	0.27	0.20	0.16	0.24
Average Price	5.06	23.53	9.63	9.85
Average Bid-Ask Spread	0.59	1.33	0.84	0.79
Average Volume	611.95	1240.03	570.67	745.71

This table reports some descriptive statistics for the options data along the maturity dimension (Panel A) and moneyness dimension (Panel B). For each maturity/moneyness bucket, we report the percentage of contracts, the average of the option price, the average of the option implied volatility, the average of the bid-ask spread, and finally the average volume. Short-term options have between 6 and 45 days to maturity (DTM), and long-term options have between 91 and 365 DTM. Following Israelovr and Kelly (2017), we define the moneyness as $\log\left(K/S\right)/\left(\mathrm{VIX}\sqrt{DTM}\right)$, where VIX is the CBOE VIX index. OTM puts are options with moneyness less than -1, ATM options are those options having their moneyness between -1 and 0 and finally OTM calls have their moneyness greater than 0. The options sample covers the period 1996-2017.

Table 2.4: Out-Of-Sample Goodness-Of-Fit Assessment

	CRRA	BCHJ	SDF$^{(2)}$	SDF$^{(3)}$	SDF$^{(4)}$	SDF$^{(5)}$	SDF$^{(6)}$	SDF$^{(7)}$	SDF$^{(8)}$	SDF$^{(9)}$	SDF$^{(10)}$
Ins	1.43***	1.04*	2.18***	1.18***	1.00	0.99***	0.95***	0.95***	0.95***	0.95***	0.95***
	(12.85)	(1.71)	(9.23)	(14.28)		(-10.74)	(-13.21)	(-13.77)	(-13.91)	(-13.92)	(-13.99)
Oos_1	1.42***	1.01	2.21***	1.15***	1.00	0.99***	0.97***	0.96***	0.96***	0.96***	0.96***
	(8.59)	(0.34)	(7.12)	(11.03)		(-8.50)	(-7.82)	(-8.91)	(-9.12)	(-8.08)	(-7.14)
Oos_2	1.39***	0.99	2.11***	1.12***	1.00	0.99***	0.98***	0.97***	0.97***	0.97***	0.97***
	(10.65)	(-0.48)	(10.32)	(11.63)		(-3.31)	(-4.52)	(-4.73)	(-4.97)	(-4.95)	(-3.70)
Oos_3	1.40***	0.96	2.03***	1.08***	1.00	1.01	1.00	0.99	0.99	0.99	0.98**
	(8.25)	(-1.21)	(9.87)	(6.03)		(0.64)	(-0.26)	(-0.34)	(-0.43)	(-0.43)	(-2.37)
Oos_4	1.39***	0.92**	1.98***	1.04*	1.00	1.01	1.00	0.99	0.99	0.99	0.97**
	(5.94)	(-2.15)	(6.98)	(1.85)		(0.74)	(-0.46)	(-0.81)	(-0.92)	(-0.88)	(-2.12)

This table reports, for different specifications of the SDF, ratios of average root mean square of the relative errors in the implied volatility domain (RIVRMSE). Following the forecasting literature, we report the average RIVRMSE of each specification as a ratio with the average RIVRMSE of the baseline specification, SDF$^{(4)}$ in our study. In parentheses below each ratio, we report Diebold-Mariano (DM) statistics, based on time-series of RIVRMSEs, comparing the forecast accuracy of the different specification with that of SDF$^{(4)}$. This statistic accounts for autocorrelation in errors and is normally distributed asymptotically. It is positive (negative) whenever the SDF$^{(4)}$ outperforms (underperforms, which never happens) the alternate specification. The n weeks-horizon out-of-sample results are obtained by pricing week t options using the relative risk aversion (RRA) parameters estimated for week $t-n$.

Table 2.5: Descriptive Statistics for the RRA parameters and the ERP parameter λ

Panel A: $SDF^{(4)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$
% of RRA > 0	96.55	95.85	98.10	92.23
Min	-8.62	-8.49	-5.04	-6.24
25% Percentile	2.57	2.36	4.75	3.35
Average	4.44	4.37	6.98	4.94
75% Percentile	6.48	6.63	8.16	6.69
Max	12.17	12.61	33.31	21.24
Coefficient Of Variation	0.68	0.74	0.57	0.69

Panel B: $SDF^{(5)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$
% of RRA > 0	96.55	95.85	97.93	91.62	80.83
Min	-8.62	-8.49	-6.19	-8.15	-10.27
25% Percentile	2.58	2.36	4.99	3.80	0.00
Average	4.46	4.37	7.16	5.19	1.20
75% Percentile	6.52	6.61	8.35	7.17	2.23
Max	12.17	12.61	32.47	22.82	9.20
Coefficient Of Variation	0.68	0.74	0.57	0.75	2.05

Panel C: $SDF^{(6)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	$RRA^{(6)}$
% of RRA > 0	96.55	95.77	97.75	91.11	44.39	26.08
Min	-8.63	-8.49	-7.31	-9.18	-15.52	-14.93
25% Percentile	2.67	2.36	5.32	3.94	-6.05	-6.51
Average	4.49	4.38	7.80	5.63	-1.81	-2.92
75% Percentile	6.52	6.61	9.60	8.17	0.86	0.00
Max	12.17	12.61	35.54	22.94	10.66	10.11
Coefficient Of Variation	0.67	0.73	0.59	0.77	2.64	1.43

This table presents some descriptive statistics for the RRA parameters and the ERP parameter λ . The statistics include the percentage of positive values, the minimum value, the first quartile, the mean, the third quartile, the maximum value, and finally the coefficient of variation defined as the standard deviation divided by the absolute value of the mean. The RRA parameters are first suitable normalized following the procedure outlined in the main text. Panel A (Panel B, Panel C) reports the results for $SDF^{(4)}$ ($SDF^{(5)}$ and $SDF^{(6)}$)

Table 2.6: Correlations between RRA parameters and volatility factors

Panel A: $SDF^{(4)}$ Parameters

	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	λ	$\sqrt{252 * h_t}$	$\sqrt{252 * q_t}$
$RRA^{(3)}$	-0.19					
$RRA^{(4)}$	-0.14	0.88				
λ	0.99	-0.08	-0.04			
$\sqrt{252 * h_t}$	-0.44	0.18	0.26	-0.45		
$\sqrt{252 * q_t}$	-0.45	0.17	0.26	-0.46	0.97	

Panel B: $SDF^{(5)}$ Parameters

	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	λ	$\sqrt{252 * h_t}$	$\sqrt{252 * q_t}$
$RRA^{(3)}$	-0.18						
$RRA^{(4)}$	-0.08	0.84					
$RRA^{(5)}$	-0.05	-0.03	0.23				
λ	0.99	-0.07	0.02	-0.07			
$\sqrt{252 * h_t}$	-0.44	0.16	0.21	-0.09	-0.45		
$\sqrt{252 * q_t}$	-0.45	0.15	0.21	-0.10	-0.46	0.97	

Panel C: $SDF^{(6)}$ Parameters

	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	$RRA^{(6)}$	λ	$\sqrt{252 * h_t}$	$\sqrt{252 * q_t}$
$RRA^{(3)}$	-0.23							
$RRA^{(4)}$	-0.11	0.86						
$RRA^{(5)}$	0.19	-0.47	-0.44					
$RRA^{(6)}$	0.39	-0.61	-0.62	0.78				
λ	0.99	-0.11	-0.00	0.15	0.34			
$\sqrt{252 * h_t}$	-0.43	0.25	0.28	-0.33	-0.35	-0.44		
$\sqrt{252 * q_t}$	-0.44	0.24	0.28	-0.34	-0.35	-0.45	0.97	

This table presents the correlations between the RRA parameters implied by $SDF^{(4)}$ (Panel A), $SDF^{(5)}$ (Panel B), and $SDF^{(6)}$ (Panel C). The correlations between these parameters and the two volatility factors are also provided. The RRA parameters are first suitable normalized following the procedure outlined in the main text.

Table 2.7: Determinants of RRA parameters

Panel A: $SDF^{(4)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$
Intercept	4.85***	4.93***	0.95***	0.16
	(7.95)	(8.31)	(3.40)	(0.71)
AR(1)	0.50***	0.52***	0.70***	0.75***
	(9.93)	(10.98)	(18.93)	(28.79)
$\sqrt{252 * q_t}$	-17.37***	-18.55***	8.09***	7.41***
	(-6.92)	(-7.27)	(3.47)	(4.70)
$\sqrt{252 * h_t} - \sqrt{252 * q_t}$	-10.33**	-9.92*	8.76	4.30
	(-2.01)	(-1.85)	(1.32)	(0.96)
ADS	-0.43***	-0.53***	0.61***	0.45***
	(-5.14)	(-5.60)	(4.30)	(4.75)
Adjusted R^2	0.44	0.45	0.56	0.64
Adjusted R^2 with AR(1) only	0.39	0.40	0.55	0.63

Panel B: $SDF^{(5)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$
Intercept	4.82***	4.93***	1.06***	0.17	1.44***
	(8.10)	(8.41)	(3.35)	(0.51)	(3.35)
AR(1)	0.51***	0.52***	0.69***	0.66***	0.28***
	(10.61)	(11.40)	(18.34)	(21.33)	(6.26)
$\sqrt{252 * q_t}$	-17.26***	-18.54***	8.17***	10.76***	-3.52
	(-6.99)	(-7.31)	(3.38)	(5.13)	(-1.57)
$\sqrt{252 * h_t} - \sqrt{252 * q_t}$	-10.04*	-9.52*	7.53	1.51	3.69
	(-1.95)	(-1.78)	(1.10)	(0.27)	(0.76)
ADS	-0.43***	-0.53***	0.65***	0.66***	0.05
	(-5.23)	(-5.61)	(4.37)	(5.57)	(0.52)
Adjusted R^2	0.44	0.45	0.55	0.52	0.09
Adjusted R^2 without controls	0.39	0.40	0.54	0.51	0.08

Panel C: $SDF^{(6)}$ Parameters

	λ	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	$RRA^{(6)}$
Intercept	4.76***	4.87***	0.77**	-0.18	3.57***	2.20***
	(8.08)	(8.39)	(2.26)	(-0.46)	(5.94)	(4.80)
AR(1)	0.51***	0.52***	0.68***	0.66***	0.36***	0.44***
	(10.61)	(11.47)	(18.70)	(21.92)	(10.09)	(12.71)
$\sqrt{252 * q_t}$	-16.73***	-18.18***	12.25***	14.29***	-31.16***	-25.68***
	(-6.89)	(-7.24)	(4.20)	(5.75)	(-7.93)	(-8.29)
$\sqrt{252 * h_t} - \sqrt{252 * q_t}$	-9.90*	-9.35*	14.19*	4.18	-5.82	-12.47*
	(-1.92)	(-1.74)	(1.91)	(0.69)	(-0.62)	(-1.76)
ADS	-0.41***	-0.53***	0.81***	0.78***	-1.11***	-1.14***
	(-5.10)	(-5.58)	(4.95)	(5.50)	(-5.10)	(-6.60)
Adjusted R^2	0.43	0.45	0.56	0.54	0.28	0.37
Adjusted R^2 without with AR(1)	0.38	0.40	0.54	0.52	0.22	0.32

We regress each RRA parameter on its lagged value, the annualized long-run volatility $\sqrt{252 * q_t}$, the difference between the short-run and the long-run volatilities, $\sqrt{252 * h_t}$-$\sqrt{252 * q_t}$, and finally the ADS business cycle indicator. Panel A (Panel B, Panel C) shows the outputs of this regression for $SDF^{(4)}$ ($SDF^{(5)}$, SDF^{6}) parameters. HAC corrected t-stats are presented in parenthesis below each estimate. The RRA parameters are first suitable normalized following the procedure outlined in the main text.

Table 2.8: Average moments risk premium

This table reports the sample averages of the one-month ahead moments premium for different pricing kernels. $\mathrm{RP}^{(1)}$ is the standard equity risk premium, that is the difference between the expected returns and the risk free rate. For $n > 1$, we define a multiplicative risk premium as the ratio of risk-neutral centred moment and its physical counterpart powered by the inverse of the moment order, that is, for $n > 1$:

$$\mathrm{RP}^{(n)} = \left(\frac{\mathbb{E}^{\mathbb{Q}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{Q}}(R_{t\to T}) \right)^n \right)}{\mathbb{E}^{\mathbb{P}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{P}}(R_{t\to T}) \right)^n \right)} \right)^{1/n}.$$

In parentheses below each average premium, we report the Newey-West (1987) t-stat for testing the hypothesis H_0: "$\mathrm{E}(\mathrm{RP}^{(n)})=0$" for the first order moment and H_0: "$\mathrm{E}(\mathrm{RP}^{(n)})=1$" for higher-order moments. We winsorize the moments premium data at the 1% level.

	CRRA	BCHJ	SDF$^{(2)}$	SDF$^{(3)}$	SDF$^{(4)}$	SDF$^{(5)}$	SDF$^{(6)}$
Equity Risk	10.53***	10.12***	6.23***	7.00***	10.12***	10.19***	10.38***
	(15.17)	(20.10)	(12.14)	(15.98)	(21.12)	(21.08)	(20.48)
Variance Risk	1.04***	1.24***	1.01***	1.05***	1.09***	1.09***	1.10***
	(23.66)	(23.66)	(10.09)	(13.69)	(14.39)	(14.11)	(13.82)
Skewness Risk	1.06***	1.28***	1.02***	1.06***	1.12***	1.12***	1.13***
	(22.59)	(23.85)	(8.16)	(13.35)	(14.05)	(13.81)	(13.86)
Kurtosis Risk	1.04***	1.24***	1.01***	1.04***	1.08***	1.08***	1.09***
	(22.77)	(23.75)	(9.26)	(14.18)	(15.08)	(14.75)	(14.90)
HyperSkewness Risk	1.05***	1.26***	1.01***	1.05***	1.09***	1.09***	1.10***
	(23.47)	(23.89)	(8.00)	(14.30)	(14.99)	(14.57)	(15.31)
HyperKurtosis Risk	1.00	1.24***	1.01***	1.04***	1.08***	1.08***	1.01
	(0.08)	(23.94)	(8.53)	(14.29)	(15.59)	(14.64)	(0.72)

Table 2.9: Moments risk premium correlations

This table reports the correlations between the one-month ahead moments premium for different pricing kernels. For the higher-orders moments, we define a multiplicative risk premium as the ratio of risk-neutral centred moment and its physical counterpart powered by the inverse of the moment order, that is, for $n > 1$:

$$\mathrm{RP}^{(n)} = \left(\frac{\mathbb{E}^{\mathbb{Q}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{Q}}(R_{t\to T})\right)^n\right)}{\mathbb{E}^{\mathbb{P}}\left(\left(R_{t\to T} - \mathbb{E}^{\mathbb{P}}(R_{t\to T})\right)^n\right)} \right)^{1/n}.$$

Panel A: $\mathrm{SDF}^{(4)}$ implied RPs

	$\mathrm{RP}^{(1)}$	$\mathrm{RP}^{(2)}$	$\mathrm{RP}^{(3)}$	$\mathrm{RP}^{(4)}$	$\mathrm{RP}^{(5)}$	$\mathrm{RP}^{(6)}$
$\mathrm{RP}^{(1)}$						
$\mathrm{RP}^{(2)}$	0.11					
$\mathrm{RP}^{(3)}$	0.03	0.93				
$\mathrm{RP}^{(4)}$	0.10	0.99	0.94			
$\mathrm{RP}^{(5)}$	0.06	0.94	0.97	0.96		
$\mathrm{RP}^{(6)}$	0.08	0.97	0.94	0.99	0.96	
$\sqrt{252 * h_t}$	0.47	0.40	0.28	0.35	0.29	0.30

Panel B: BCHJ implied RPs

	$\mathrm{RP}^{(1)}$	$\mathrm{RP}^{(2)}$	$\mathrm{RP}^{(3)}$	$\mathrm{RP}^{(4)}$	$\mathrm{RP}^{(5)}$	$\mathrm{RP}^{(6)}$
$\mathrm{RP}^{(1)}$						
$\mathrm{RP}^{(2)}$	0.23					
$\mathrm{RP}^{(3)}$	0.28	0.90				
$\mathrm{RP}^{(4)}$	0.22	1.00	0.90			
$\mathrm{RP}^{(5)}$	0.26	0.95	0.98	0.95		
$\mathrm{RP}^{(6)}$	0.22	1.00	0.90	1.00	0.95	
$\sqrt{252 * h_t}$	0.48	0.06	0.12	0.04	0.10	0.02

Table 2.10: RRA parameters correlations across the SDF orders

		Panel A: $\text{RRA}^{(2)}$ correlations		
		$\text{RRA}^{(2)}$ from $\text{SDF}^{(4)}$	$\text{RRA}^{(2)}$ from $\text{SDF}^{(5)}$	$\text{RRA}^{(2)}$ from $\text{SDF}^{(6)}$
$\text{RRA}^{(2)}$	0.35	0.30	0.30	0.30
		Panel B: $\text{RRA}^{(3)}$ correlations		
		$\text{RRA}^{(3)}$ from $\text{SDF}^{(4)}$	$\text{RRA}^{(3)}$ from $\text{SDF}^{(5)}$	$\text{RRA}^{(3)}$ from $\text{SDF}^{(6)}$
$\text{RRA}^{(3)}$		0.70	0.68	0.65
		Panel C: $\text{RRA}^{(4)}$ correlations		
			$\text{RRA}^{(4)}$ from $\text{SDF}^{(5)}$	$\text{RRA}^{(4)}$ from $\text{SDF}^{(6)}$
$\text{RRA}^{(4)}$			0.87	0.86
		Panel D: $\text{RRA}^{(5)}$ correlations		
				$\text{RRA}^{(5)}$ from $\text{SDF}^{(6)}$
$\text{RRA}^{(5)}$				0.25

This table reports the correlations between the RRA parameters estimated using different SDFs. The observation (n, m) of this table gives the correlation between $\text{RRA}^{(n)}$ estimated using $\text{SDF}^{(n)}$ and the same $\text{RRA}^{(n)}$ estimated using $\text{SDF}^{(m)}$.

Table 2.11: Determinants of $\text{RRA}^{(2)}$ implied by the CRRA utility

Intercept	1.69***	3.21***	3.02***	3.36***
	(15.47)	(13.44)	(12.14)	(11.75)
AR(1)	0.54***	0.47***	0.47***	0.46***
	(20.28)	(15.48)	(15.36)	(15.44)
$\sqrt{252 * q_t}$		-7.91***	-6.89***	-9.07***
		(-8.44)	(-6.90)	(-6.80)
$\sqrt{252 * h_t} - \sqrt{252 * q_t}$			-7.90**	-7.57**
			(-2.45)	(-2.32)
ADS				-0.21***
				(-2.90)
Adjusted R^2	0.29	0.33	0.33	0.33

We regress each the CRRA-implied $\text{RRA}^{(2)}$ on its lagged value, the annualized long-run volatility $\sqrt{252 * q_t}$, the difference between the short-run and the long-run volatilities, $\sqrt{252 * h_t}$-$\sqrt{252 * q_t}$, and finally the ADS business cycle indicator. HAC corrected t-stats are presented in parenthesis below each estimate.

Table 2.12: Monte Carlo Results Summary

Panel A: Average Relative Absolute Error

	Avg. RIVMSE	$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	$RRA^{(6)}$
$CRRA^{(4)}$	0.03329	0.00301	0.10066	0.33488		
$CRRA^{(5)}$	0.00328	0.00376	0.07465	0.07920	0.08735	
$CRRA^{(6)}$	0.00130	0.00096	0.02251	0.03650	0.04699	0.09508
$SDF^{(4)}$	0.00013	0.00000	0.00001	0.00003		
$SDF^{(5)}$	0.00426	0.00696	0.02206	0.03128	0.24021	
$SDF^{(6)}$	0.00316	0.00806	0.02062	0.03851	0.25929	0.10093

Panel B: Median Relative Absolute Error

		$RRA^{(2)}$	$RRA^{(3)}$	$RRA^{(4)}$	$RRA^{(5)}$	$RRA^{(6)}$
$CRRA^{(4)}$	0.03329	0.00000	0.00000	0.00000		
$CRRA^{(5)}$	0.00328	0.00000	0.00000	0.00000	0.00000	
$CRRA^{(6)}$	0.00130	0.00050	0.01434	0.02337	0.01702	0.04421
$SDF^{(4)}$	0.00013	0.00000	0.00000	0.00000		
$SDF^{(5)}$	0.00426	0.00000	0.00001	0.00001	0.00032	
$SDF^{(6)}$	0.00316	0.00003	0.00009	0.00014	0.00824	0.00375

To assess the robustness of our estimation procedure, we implement a Monte Carlo exercice whose results are reported in this table. For one hundred randomly chosen weeks, we generate synthetic option prices using some known pricing kernels reported in the table's rows. $CRRA^{(n)}$ is the CRRA pricing kernel truncated at order n using the standard Taylor expansion. Recall that the RRAs of the CRRA pricing kernel are functions of the $RRA^{(2)}$ which we calibrate using the data-based weekly estimates of the CRRA risk aversion parameter $RRA^{(2)}$. $SDF^{(n)}$ is the estimated nonparametric pricing kernel of order n whose parameters are calibrated using the weekly estimates of the RRAs. Once the synthetic options prices are generated, we use our estimation methodology to recover the true parameters. The first column of the table provides the average pricing errors measured by the RIVRMSE. The remaining columns report the average of the absolute relative errors on the RRAs (Panel A) and the median of the absolute relative errors on the RRAs (Panel B).

Appendix for Chapter 2 (B)

B APPENDIX

B.1 Option Valuation using $\mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\; X_T\right]$

We depart from the standard risk-neutral valuation approach, which determines an option's price based on the $\mathbb{Q}$-distribution of its payoff. Here, we untangle the two ingredients of this risk-neutral distribution. The first is the SDF and the second is the physical distribution of the payoff. First, note that equation (2.6) is equivalent to

$$\mathrm{SDF}_{t,T}^{(k)} = \sum_{j=1}^{k} \kappa_{j,t,T}^{(k)} \frac{\mathrm{RRA}_t^{(j)}}{S_t^{j-1}} (S_T - S_t)^{j-1},$$

assuming, for convenience, that $\mathrm{RRA}^{(1)} = 1$.

Consider a call option with a strike price K that matures at time T. Using equation (2.1), the price of the call option can be expressed as

$$\mathrm{Call}_t\big(T, K\big) = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}(S_T - K)^+\right] \tag{A.1}$$

$$= \sum_{j=1}^{k} \kappa_{j,t,T}^{(k)} \frac{\mathrm{RRA}_t^{(j)}}{S_t^{j-1}}\, \mathrm{E}_t^{\mathbb{P}}\left[(S_T - S_t)^{j-1}(S_T - K)^+\right]. \tag{A.2}$$

Furthermore, we have

$$\mathrm{E}_t^{\mathbb{P}}\left[(S_T - S_t)^{j-1}\big(S_T - K)^+\big)\right] = \sum_{q=0}^{j-1}(-1)^{j-1-q} C_q^{j-1} (S_t)^{j-1-q}\, \mathbb{E}\Big(S_T^q (S_T - K) 1_{\{S_T \geq K\}}\Big),$$

where $1_{\{A\}}$ is an indicator function that equals 1 when the condition A is met and 0 elsewhere. The latter expectation is simply

$$\mathbb{E}\Big(S_T^q (S_T - K) 1_{\{S_T \geq K\}}\Big) = \int_{\ln(K)}^{\infty} e^{qx}(e^x - K)p(x)dx.$$

where $p(x)$ is the density of $\ln(S_T)$. Hence, it follows that the call price can be computed in semi-closed form using the following identity

$$\int_{\ln(K)}^{\infty} e^{qx} p(x)dx = f(q)\left[\frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \frac{1}{f(q)} Re\left[\frac{e^{-i\phi\ln(K)}f(i\phi + q)}{i\phi}\right]d\phi\right], \tag{A.3}$$

as long as the log-price's moment generating function f is known in closed-form.

B.2 Rewriting the Maximization Problem Constraints

Consider the k^{th}-order approximation to the SDF. Equation (2.3) becomes

$$1 = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\left(1 + R_{t,T}\right)\right] \tag{A.4}$$

$$= \frac{1}{1 + R_{f,t}} + \sum_{j=1}^{k} \kappa_{j,t,T}^{(k)}\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^j\right]. \tag{A.5}$$

where $R_{f,t} = (1 - B_t)/B_t$ is the return on the risk-free bond B, $\mathrm{RRA}_t^{(1)} = 1$ for convenience, and $\kappa_{j,t,T}^{(k)}$ is given by

$$\kappa_{j,t,T}^{(k)} = \frac{w_j}{(1 + R_{f,t})\Big(\sum_{l=1}^{k} w_l\, \mathrm{RRA}_t^{(l)}\, \mathrm{E}_t\left[R_{t,T}^{l-1}\right]\Big)} \qquad \text{with} \qquad w_j = \frac{(-1)^{j-1}}{(j-1)!}, \qquad j = 1, \ldots, k. \tag{A.6}$$

The Euler equation can then be rewritten as:

$$R_{f,t} = \frac{\sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^j\right]}{\sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^{j-1}\right]} \tag{A.7}$$

$$\Leftrightarrow \qquad R_{f,t}\left(\sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^{j-1}\right]\right) = \sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^j\right] \tag{A.8}$$

$$\Leftrightarrow \qquad 0 = R_{f,t}\left(\sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^{j-1}\right]\right) - \sum_{j=1}^{k} w_j\, \mathrm{RRA}_t^{(j)}\, \mathrm{E}_t\left[R_{t,T}^j\right] \tag{A.9}$$

$$\Leftrightarrow \quad 0 = \sum_{j=1}^{k-1} w_j\, \mathrm{RRA}_t^{(j)}\, \left(R_{f,t}\, \mathrm{E}_t\left[R_{t,T}^{j-1}\right] - \mathrm{E}_t\left[R_{t,T}^j\right]\right)$$

$$+ w_k \, \mathrm{RRA}_t^{(k)} \left(R_{f,t} \, \mathrm{E}_t \left[R_{t,T}^{k-1} \right] - \mathrm{E}_t \left[R_{t,T}^{k} \right] \right) \tag{A.10}$$

or equivalently

$$\mathrm{RRA}_t^{(k)} = - \frac{\sum_{j=1}^{k-1} w_j \, \mathrm{RRA}_t^{(j)} \left(R_{f,t} \, \mathrm{E}_t \left[R_{t,T}^{j-1} \right] - \mathrm{E}_t \left[R_{t,T}^{j} \right] \right)}{w_k \left(R_{f,t} \, \mathrm{E}_t \left[R_{t,T}^{k-1} \right] - \mathrm{E}_t \left[R_{t,T}^{k} \right] \right)} . \tag{A.11}$$

Hence, the maximization problem becomes:

$$\max_{\lambda, \, \mathrm{RRA}_t^{(2)}, \, \ldots, \, \mathrm{RRA}_t^{(k)}} L_{\text{Options}} \left(\lambda, \, \mathrm{RRA}_t^{(2)}, \, \ldots, \, \mathrm{RRA}_t^{(k)} \right) \quad \text{s.t.} \quad \text{Equations (A.6) and (A.11).} \tag{A.12}$$

B.3 The Returns' Model in BCHJ's Notation

Babaoglu, Christoffersen, Heston, and Jacobs (2018, hereafter BCHJ) specify their model as follows:

$$r_t = r_f + \mu h_t + \eta y_t, \tag{A.13}$$

$$h_t = w + b_1 h_{t-1} + b_2 h_{t-2} + c_1 y_{t-1} + c_2 y_{t-2} + a_1 h_{t-1}^2 / y_{t-1} + a_2 h_{t-2}^2 / y_{t-2}. \tag{A.14}$$

We here demonstrate that this formulation is observationally equivalent to the one in equations (2.8), (2.9) and (2.10). First, remark that the η parameter is the same in both formulation and is identified by the conditional skewness of the innovations (in particular, the negative skewness observed in the data requires a negative η). Second, the convexity correction is:

$$\xi_t = \left(\frac{1}{\eta^2} + \frac{\sqrt{1 - 2\eta}}{\eta^2} - \frac{1}{\eta} \right) h_t. \tag{A.15}$$

Thus, equating equation (2.8), at t, with (A.13), we obtain

$$\mu = \lambda - \left(\frac{1}{\eta^2} + \frac{\sqrt{1 - 2\eta}}{\eta^2} - \frac{1}{\eta} \right) - \frac{1}{\eta}, \tag{A.16}$$

$$= \lambda - \frac{1}{\eta^2} - \frac{\sqrt{1 - 2\eta}}{\eta^2}. \tag{A.17}$$

Given that η is identified by the conditional skewness, BCHJ's μ is identified by the market price of equity risk, which is simply λ in our formulation.

Turning to the variance process, we first rewrite equation (2.10) as

$$q_t = (1 - \rho_q) \sigma^2 + \rho_q q_{t-1} + \nu_{q,t-1}, \tag{A.18}$$

Then, we insert q_t into h_t (equation 2.9)

$$
\begin{aligned}
h_t &= (1-\rho_q)\sigma^2 + \rho_q q_{t-1} + \nu_{q,t-1} + \rho_h(h_{t-1} - \underbrace{q_{t-1}}_{(\text{A.18})}) + \nu_{h,t-1} \\
&= (1-\rho_q)\sigma^2 + \rho_q q_{t-1} + \nu_{q,t-1} + \rho_h h_{t-1} - \rho_h \left(\overbrace{(1-\rho_q)\sigma^2 + \rho_q q_{t-2} + \nu_{q,t-2}}\right) + \nu_{h,t-1} \\
&= \tilde{w} + \rho_h h_{t-1} + \rho_q \underbrace{(q_{t-1} - \rho_h q_{t-2})}_{\text{rewrite (2.9)}} + \nu_{h,t-1} + \nu_{q,t-1} - \rho_h \nu_{q,t-2}, \qquad \text{(A.19)}
\end{aligned}
$$

where $\tilde{w} = (1-\rho_h)(1-\rho_q)\sigma^2$ highlights that BCHJ's w parameter (which we will obtain shortly) is identified by the unconditional variance. Now, we rewrite equation (2.9) as

$$
q_{t-1} - \rho_h q_{t-2} = h_{t-1} - \rho_h h_{t-2} - \nu_{h,t-2} \qquad \text{(A.20)}
$$

and make use of this result in equation (A.19)

$$
\begin{aligned}
h_t &= \tilde{w} + \rho_h h_{t-1} + \rho_q(h_{t-1} - \rho_h h_{t-2} - \nu_{h,t-2}) + \nu_{h,t-1} + \nu_{q,t-1} - \rho_h \nu_{q,t-2} \\
&= \tilde{w} + \tilde{b}_1 h_{t-1} + \tilde{b}_2 h_{t-2} + \nu_{h,t-1} + \nu_{q,t-1} - \rho_q \nu_{h,t-2} - \rho_h \nu_{q,t-2}, \qquad \text{(A.21)}
\end{aligned}
$$

where the autoregression coefficients $\tilde{b}_1 = (\rho_h + \rho_q)$ and $\tilde{b}_2 = -\rho_h\rho_q$ are identified by short- and long-term persistence.

Focusing on the innovations, we have

$$
\begin{aligned}
&\nu_{h,t-1} + \nu_{q,t-1} - \rho_q \nu_{h,t-2} - \rho_h \nu_{q,t-2} \\
&\quad= (a_h + a_q) h_{t-1}\nu_{t-1} + (c_h + c_q)\nu_{r,t-1} - (a_h\rho_q + a_q\rho_h) h_{t-2}\nu_{t-2} - (c_h\rho_q + c_q\rho_h)\nu_{r,t-2} \\
&\quad= a_1 \left(\frac{h_{t-1}^2}{y_{t-1}} - \eta^2 h_{t-1} - \eta^4\right) + \frac{c_1}{\eta}\left(\eta y_{t-1} - \frac{h_{t-1}}{\eta}\right) \\
&\qquad + a_2 \left(\frac{h_{t-2}^2}{y_{t-2}} - \eta^2 h_{t-2} - \eta^4\right) + \frac{c_2}{\eta}\left(\eta y_{t-2} - \frac{h_{t-2}}{\eta}\right) \\
&\quad= a_1 \frac{h_{t-1}^2}{y_{t-1}} + c_1 y_{t-1} + a_2 \frac{h_{t-2}^2}{y_{t-2}} + c_2 y_{t-2} \\
&\qquad - a_1\eta^2 h_{t-1} - a_1\eta^4 - \frac{c_1 h_{t-1}}{\eta^2} - a_2\eta^2 h_{t-2} - a_2\eta^4 - \frac{c_2 h_{t-2}}{\eta^2}, \qquad \text{(A.22)}
\end{aligned}
$$

where $a_1 = (a_h + a_q)\sigma^{-4}$, $a_2 = -(a_h\rho_q + a_q\rho_h)\sigma^{-4}$, $c_1 = (c_h + c_q)\eta$, and $c_2 = -(c_h\rho_q + c_q\rho_h)\eta$.

We conclude by using equation (A.22) in equation (A.21), which yields

$$
h_t = \left[\tilde{w} - (a_1 + a_2)\eta^4\right]
$$

$$
\begin{aligned}
&+\left[\tilde{b}_1-\left(a_1\eta^2+\frac{c_1}{\eta^2}\right)\right]h_{t-1}+\left[\tilde{b}_2-\left(a_2\eta^2+\frac{c_2}{\eta^2}\right)\right]h_{t-2}\\
&+a_1\frac{h_{t-1}^2}{y_{t-1}}+c_1y_{t-1}+a_2\frac{h_{t-2}^2}{y_{t-2}}+c_2y_{t-2}\\
=\,&w+b_1h_{t-1}+b_2h_{t-2}+a_1\frac{h_{t-1}^2}{y_{t-1}}+c_1y_{t-1}+a_2\frac{h_{t-2}^2}{y_{t-2}}+c_2y_{t-2},
\end{aligned}
\tag{A.23}
$$

where

$$
w=\tilde{w}-(a_1+a_2)\eta^4,\qquad b_1=\tilde{b}_1-\left(a_1\eta^2+\frac{c_1}{\eta^2}\right),\qquad b_2=\tilde{b}_2-\left(a_2\eta^2+\frac{c_2}{\eta^2}\right).
$$

B.4 Computing the standardized physical moments

The returns physical distribution can be analyzed by looking at the returns standardized moments for various orders. For $n\geq 1$, let's define the returns centred moments as

$$
M_n=\mathrm{E}_t^{\mathbb{P}}\left[\left(R_{t,T}-\mathrm{E}_t^{\mathbb{P}}(R_{t,T}\right)^n\right].
$$

Next, recall that

$$
R_{t,T}=\frac{S_T-S_t}{S_t},
$$

which implies that,

$$
R_{t,T}-\mathrm{E}_t^{\mathbb{P}}(R_{t,T})=\frac{S_T-\mathrm{E}_t^{\mathbb{P}}(S_t)}{S_t}.
$$

Using the latter equation, we can express the centred returns moments in terms of the underlying conditional moments,

$$
\begin{aligned}
M_n&=\frac{1}{S_t^n}\,\mathrm{E}_t^{\mathbb{P}}\left[\left(S_T-\mathrm{E}_t^{\mathbb{P}}(S_T\right)^n\right]\\
&=\frac{1}{S_t^n}\Big(\sum_{q=0}^{n}C_n^q\mathrm{E}_t^{\mathbb{P}}\big(S_T^q\big)\big(-\mathrm{E}_t^{\mathbb{P}}(S_T)\big)^{n-q}\Big).
\end{aligned}
$$

If we denote by f, the moment generating function of S_T, then for every k, $\mathrm{E}_t^{\mathbb{P}}\big(S_T^k\big)$ is simply $f(k)$. For $n\geq 3$, the normalized moment of order n is

$$
\frac{M_n}{M_2^{n/2}}=\frac{\mathrm{E}_t^{\mathbb{P}}\left[\left(S_T-\mathrm{E}_t^{\mathbb{P}}(S_T)\right)^n\right]}{\mathrm{E}_t^{\mathbb{P}}\Big(\left(S_T-\mathrm{E}_t^{\mathbb{P}}(S_T\right)^2\Big)^{n/2}}.
$$

When $n = 2$, the normalized moment is simply $\sqrt{M_2}$.

B.5 Computing the risk-neutral moments

Since the risk-neutral distribution implied by $SDF^{(k)}$ is not available in closed form, one elegant way to access this distribution is through its moments. Fortunately the Q-moments can be obtained straightforwardly by combining the the P-moments and the RRA parameters.
To see this, first recall that

$$\mathrm{SDF}_{t,T}^{(k)} = \sum_{j=1}^{k} \kappa_{j,t,T}^{(k)} \frac{\mathrm{RRA}_t^{(j)}}{S_t^{j-1}} (S_T - S_t)^{j-1}.$$

For $n \geq 0$, we have

$$\mathrm{E}_t^{\mathbb{Q}}(\mathrm{S}_T^n) = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{SDF}_{t,T}^{(k)}\, S_T^n\right] = \sum_{j=1}^{k} \kappa_{j,t,T}^{(k)} \frac{\mathrm{RRA}_t^{(j)}}{S_t^{j-1}} \mathrm{E}_t^{\mathbb{P}}\left[(S_T - S_t)^{j-1} S_T^n\right].$$

Since,

$$(S_T - S_t)^{j-1} = \sum_{q=0}^{j-1} C_q^{j-1} S_T^q (-S_t)^{j-1-q},$$

we have

$$\mathrm{E}_t^{\mathbb{P}}\left[S_T^n (S_T - S_t)^{j-1}\right] = \sum_{q=0}^{j-1} C_q^{j-1} (-S_t)^{j-1-q}\, \mathrm{E}_t^{\mathbb{P}}\left[S_T^{n+j}\right].$$

B.6 More on the Model

Moment Generating Function

The moment generating function of the returns under the IG-GARCH(C) model is given by

$$\begin{aligned}
f(\phi) &= \mathbf{E}_t[S(T)^{\phi}] = S(t)^{\phi} \exp(A(t) + B(t)h(t+\Delta) + C(t)q(t+\Delta)),\\
A(t) &= A(t+\Delta) + r\phi\Delta - (w_q - a_h\eta^4 - a_q\eta^4)B(t+\Delta) + (w_q - a_q\eta^4)C(t+\Delta)\\
&\quad - 0.5\ln(1 - 2(a_h + a_q)\eta^4 B(t+\Delta) - 2a_q\eta^4 C(t+\Delta))\\
B(t) &= \varphi\mu + (\rho_h - (c_h + c_q)\eta^{-2} - (a_h + a_q)\eta^2)B(t+\Delta) - (c_q\eta^{-2} + a_q\eta^2)C(t+\Delta) + \eta^{-2} -\\
&\quad \frac{\sqrt{(1 - 2(a_q + a_h)\eta^4 B(t + \Delta - 2a_q\eta^4 C(t+\Delta)))(1 - 2\eta\varphi - 2(c_q + c_h)B(t+\Delta) - 2c_q C(t+\Delta))}}{\eta^2}\\
C(t) &= (\rho_q - \rho_h)B(t+\Delta) + \rho_q C(t+\Delta)
\end{aligned}$$

with

$$A(T) = 0, B(T) = 0, C(T) = 0.$$

Setting $T = t + 1$, the last line becomes $A(t + 1) = B(t + 1) = C(t + 1) = 0$. Next,

$$\begin{aligned}\mathbb{E}_t\big(\exp(r_{t+1})\big) &= \frac{\mathbb{E}_t(S_{t+1})}{S_t} \\ &= \exp(A_t + B_t h_{t+1} + C_t q_{t+1}) = \exp(r + \lambda h_{t+1}),\end{aligned}$$

with $C_t = 0$, $A_t = r$, and $B_t = \mu + \frac{1}{\eta^2} + \frac{\sqrt{1-2\eta}}{\eta^2}$

$$\mathbb{E}_t\big(\exp(r_{t+1})\big) = S(t)\exp\Big(r + \big(\mu + \frac{1}{\eta^2} + \frac{\sqrt{1-2\eta}}{\eta^2}\big)h_{t+1}\Big),$$

It follows that:

$$\begin{aligned}\lambda &= \mu + \frac{1}{\eta^2} + \frac{\sqrt{1-2\eta}}{\eta^2} \\ \lambda - \xi = \mu + \frac{1}{\eta} &= \mu + \frac{1}{\eta^2} + \frac{\sqrt{1-2\eta}}{\eta^2} - \xi,\end{aligned}$$

which imply that

$$\xi = \frac{1}{\eta^2} + \frac{\sqrt{1-2\eta}}{\eta^2} - \frac{1}{\eta}.$$

B.7 More on benchmark parametric SDFs

Deriving the RRA parameters for a CRRA utility function

A CRRA utility function is defined: as

$$\mathrm{U}(S_t) = \frac{S_t^{1-\mathrm{RRA}^{(2)}}}{1-\mathrm{RRA}^{(2)}},$$

from which it follows that

$$\begin{aligned}\mathrm{U}'(S_t) &= S_T^{-\mathrm{RRA}^{(2)}} \\ \mathrm{U}''(S_t) &= -\mathrm{RRA}^{(2)}\, S_t^{-(\mathrm{RRA}^{(2)}+1)} \\ \mathrm{U}^{(3)}(S_t) &= \mathrm{RRA}^{(2)}(\mathrm{RRA}^{(2)}+1)S_t^{-(\mathrm{RRA}^{(2)}+2)} \\ \mathrm{U}^{(4)}(S_t) &= -\mathrm{RRA}^{(2)}(\mathrm{RRA}^{(2)}+1)(\mathrm{RRA}^{(2)}+2)S_t^{-(\mathrm{RRA}^{(2)}+3)}\end{aligned}$$

$$\mathrm{U}^{(5)}(S_t) = \mathrm{RRA}^{(2)}(\mathrm{RRA}^{(2)}+1)(\mathrm{RRA}^{(2)}+2)(\mathrm{RRA}^{(2)}+3)S_t^{-(\mathrm{RRA}^{(2)}+4)},$$

which we generalize as

$$\mathrm{U}^{(k)}(S_t) = (-1)^{k-1}\Big(\prod_{j=0}^{k-2}(\mathrm{RRA}^{(2)}+j)\Big)S_t^{-(\mathrm{RRA}^{(2)}+k-1)}, \qquad k \geq 0.$$

Now recalling the definition of the RRA parameters:

$$\mathrm{RRA}_t^{(k)} = (-1)^{k-1}\frac{S^{k-1}\,\mathrm{U}^{(k)}(S_t)}{\mathrm{U}'(S_t)}, \qquad k \geq 2,$$

we deduce that

$$\mathrm{RRA}^{(k)} = \prod_{j=0}^{k-2}(\mathrm{RRA}^{(2)}+j).$$

B.8 Alternative representation of the physical model

Babaoglu, Christoffersen, Heston, and Jacobs (2018) additionally show that the initial Component IGGARCH model can be rewritten as an IGGARCH(2,2) model

$$\begin{aligned}
&\log(S_(t+\Delta) = \log(S_(t) + r\Delta + \mu h(t+\Delta) + \eta y(t+\Delta),\\
&h(t+\Delta) = w + b_1 h(t) + b_2 h(t-\Delta) + c_1 y(t) + c_2 y(t-\Delta),\\
&+ a_1\frac{h(t)^2}{y(t)} + a_2\frac{h(t-\Delta)^2}{y(t-\Delta)}.
\end{aligned}$$

Seting $\tilde{b}_i = b_i + \frac{c_i}{\eta^2} + a_i\eta^2$, for i=1,2, then we have

$$\begin{aligned}
&a_1 = a_h + a_q, \quad a_2 = -\rho 2 a_h - \rho_1 a_q,\\
&\tilde{b}_1 = \rho_1 + \rho_2, \quad \tilde{b}_2 = -\rho_1\rho_2\\
&c_1 = c_h + c_q, \quad c_2 = -\rho_2 c_h - \rho_1 c_q.
\end{aligned}$$

The parameters of the Component model are also functions of the parameters in the IGGARCH(2,2) model as follows

$$\begin{aligned}
&\tilde{\mu} = \mu + \eta^{-1},\\
&\tilde{w} = w + a_1\eta^4 + a_2\eta^4,\\
&w_q = \frac{\tilde{w}}{1-\rho_1},
\end{aligned}$$

$$
\begin{aligned}
\sigma^2 &= \frac{\tilde{w}}{(1-\rho_1)(1-\rho_2)}, \\
a_h &= \frac{-\rho_1}{\rho_2-\rho_1}a_1 - \frac{1}{\rho_2-\rho_1}a_1, \\
a_q &= \frac{-\rho_2}{\rho_2-\rho_1}a_1 + \frac{1}{\rho_2-\rho_1}a_2, \\
c_h &= \frac{-\rho_1}{\rho_2-\rho_1}c_1 - \frac{1}{\rho_2-\rho_1}c_1, \\
c_q &= \frac{-\rho_2}{\rho_2-\rho_1}c_1 + \frac{1}{\rho_2-\rho_1}c_2.
\end{aligned}
$$

B.9 Change of measure under the volatility-dependent pricing kernel

Assuming the volatility-dependent pricing kernel, the risk-neutral dynamics is given by:

$$
\begin{aligned}
&\log(S_(t+\Delta) = \log(S_(t) + r\Delta + \mu^\star h^\star(t+\Delta) + \eta^\star y^\star(t+\Delta), \\
&h^\star(t+\Delta) = w^\star + b_1^\star h(t) + b_2^\star h(t-\Delta) + c_1^\star y(t) + c_2^\star y(t-\Delta), \\
&+ a_1^\star \frac{h^\star(t)^2}{y^\star(t)} + a_2^\star \frac{h^\star(t-\Delta)^2}{y^\star(t-\Delta)},
\end{aligned}
$$

where

$$
\begin{aligned}
&h^\star(t) = s_h h(t), \quad y^\star(t) = s_y y(t), \\
&\mu^\star = \mu/s_h, \quad \eta^\star = \eta/s_h,, \quad w^\star = s_h w, \\
&a_i^\star = s_h a_i/s_y, \quad c_i^\star = s_h c_i/s_y, \quad \text{for } i = 1, 2.
\end{aligned}
$$

with

$$
\begin{aligned}
s_y &= \frac{\Big(0.5\mu^2\eta^4 + \big(1 - 2a_1\Phi\eta^4\big)\eta\Big)^2}{\big(1 - 2a_1\Phi\eta^4\big)\mu^2\eta^4}, \\
s_h &= \frac{\mu\eta^2}{s_y^2\big(\sqrt{1 - 2 * \eta/s_y} - 1\big)}.
\end{aligned}
$$

B.10 Component model representation of the purely risk-neutral dynamics

$$
\begin{aligned}
&\log(S_(t+\Delta) = \log(S_(t) + r\Delta + \mu^\star h^\star(t+\Delta) + \eta^\star y(t+\Delta), \\
&h^\star(t+\Delta) = q^\star(t+\Delta) + \rho_1(h^\star(t) - q^\star(t)) + \nu_h^\star(t), \\
&q^\star(t+\Delta) = \sigma^{\star 2} + \rho_2(q^\star(t) - \sigma^{\star 2} + \nu_q^\star(t),
\end{aligned}
$$

where the innovations $\nu_h(t)$ and $\nu_q(t)$ are defined as

$$\nu_h^\star(t) = c_h^\star y^\star(t) + a_h^\star \frac{h^\star(t)^2}{y^\star(t)} - c_h^\star \frac{h^\star(t)}{\eta^{\star 2}} - a_h \eta^{\star 2} h^\star(t),$$
$$\nu_q^\star(t) = c_q^\star y^\star(t) + a_q^\star \frac{h^\star(t)^2}{y^\star(t)} - c_q^\star \frac{h^\star(t)}{\eta^{\star 2}} - a_q \eta^{\star 2} h^\star(t).$$

Chapter 3

Multi-Horizon Pricing of Co-Skewness Risk: Evidence from Equity Returns and Options*

3.1 Introduction

The risk-return relationship has received considerable attention in the asset pricing literature. Research in this area has mainly focused on finding risk factors that have explanatory power in the cross-section of returns. Co-skewness risk is an important risk factor that has been introduced to capture nonlinear dependence between stock returns and market returns.[1] This risk represents the component of the stock returns asymmetry that arises due to market returns asymmetry. Co-skewness risk naturally complements the single factor CAPM model of Sharpe (1964b) and Lintner (1965a). As shown by Harvey and Siddique (2000a), this risk is relevant as it has the potential to explain some anomalies that have been observed in the finance literature.[2]

Economic theory is unambiguous regarding the sign of the risk premia of co-skewness

*I benefited from discussions with Ilze Kalnina, Christian Dorion, Marine Carrasco, Benoit Perron, Réne Garcia, Bruno Feunou, Caio Almeida, Elise Gourier and Erik Vogt.

[1]Kraus and Litzenberger (1976) are the first to highlight the importance of this risk on a theoretical level.
[2]Harvey and Siddique (2000a) find that co-skewness partially explains the momentum anomaly.

risk.[3] However, little is yet known about the evolution of the co-skewness risk premia across investment horizons from both empirical and theoretical viewpoints. Term structure information is relevant for investors who are willing to earn such risk premia as it helps them determine the timing of their investment. In addition, studying empirically the term structure of the risk premia associated with co-skewness risk could help discriminate between competing structural asset pricing models.

My contribution is two-fold. First, I propose a hybrid methodology that uses options and returns data to estimate option-implied co-skewness at any desirable maturity. Options data is known to contain forward-looking information on investors' anticipation of future risk, making it conducive in estimating expected risk and risk premia.[4]

Second, I use the proposed methodology to perform a multi-horizon analysis of the pricing of co-skewness risk. A multi-horizon analysis aims at studying the risk-return relationship across investment horizons. It has the advantage of providing a model-free insight into the term structure of risk premia. A multi-horizon analysis needs to account for the possibility that investors' anticipation of future risk could change across risk horizons. Theoretically, options traded at a given horizon should be the most informative for capturing the risk premia at that horizon. However, in practice, this prediction is fallible. It is always possible that option-implied risk for a given maturity could be more informative for estimating risk premia for an investment horizon that does not coincide with this maturity. To address this issue, for a given investment horizon, I look for the option horizon that best captures the price of the risk[5] and use the estimate obtained for this option horizon in my multi-horizon analysis.

My approach to estimating option-implied co-skewness relies on a proper modelling of the risk-neutral joint density of the market and stock returns across maturities. I treat the risk-neutral joint density of market and stock returns conditional on a maturity as a smooth trivariate function. Next, I approximate this trivariate function using Hermite polynomials across its three dimensions. It is important to note that the maturity dimension of this density function is flexibly modeled using Hermite polynomials. The approximation involves

[3] As noted by Dittmar (2002b), co-skewness is priced negatively when investors have decreasing absolute risk aversion.

[4] This observation can be traced back to papers such as Rubinstein (1994) and Jackwerth and Rubinstein (1996).

[5] Since co-skewness risk is priced negatively, I select the option horizon that minimizes the t-statistics of the estimated average price of risk

some coefficients that are to be estimated each day. Option-implied risk-neutral joint density cannot be obtained in a model-free way due to the absence of option contracts whose payoffs depend on the market and stock returns. Given this issue, additional restrictions are needed.[6] To solve this problem, I use the "zero correlation risk premia" assumption of Aït-Sahalia and Brandt (2007) and French, Groth, and Kolari (1983), which states that the risk-neutral correlation should not differ considerably from the historical correlation. This restriction is mild and is supported by the Girsanov's theorem, as explained by Aït-Sahalia and Brandt (2007). The approximate density function is required to correctly price options on the market and options on the stock across strikes and maturities. It should also fulfill the correlation restriction across risk horizons.[7] Provided that such restrictions are satisfied, the polynomial coefficients are obtained by minimizing the discrepancy between the risk-neutral joint density and a preliminary estimate of the physical joint density across horizons. The flexibility of the methodology is exploited when varying the orders of the Hermite polynomials involved in the approximation. This offers the possibility of checking "to some extent" the robustness of the term structure conclusions to model misspecification. The latter property is clearly an advantage when compared with adopting a fully parametric approach.

I estimate the model using options and returns data on 96 stocks of the S&P500 covering the period of 2001-2014.[8] For each month and investment horizon, the price of risk is estimated using Fama and MacBeth (1973) regressions for option horizon ranging from one to twelve months. Following the literature, I proxy the average risk premia by the average price of risk multiplied by the standard deviation of the option-implied risk measure.

First, the results show that, on average, the magnitude of the option-implied co-skewness risk decreases with the option horizon. This finding can be interpreted as a signal of investors expecting this risk to matter less over a long-horizon.[9]

Second, my results demonstrate that co-skewness risk premia are negative and exhibit a decreasing term structure (in absolute value). In particular, I find that the annual risk

[6] Cross-options whose payoffs depend on the market and stock returns are required to estimate non-parametrically the risk-neutral joint density.

[7] My paper appears to be the first to use the "zero correlation risk premia" assumption for the estimation of option-implied co-skewness risk.

[8] The sample starts from 2001 to maximize the liquidity of the stock options. The selected stocks are the ones that have at least 15 traded options at the end of each month between 2001 to 2014.

[9] Following the literature, I use a co-skewness measure that is appropriately standardized so that the measure is comparable across risk horizons.

premia decrease from 3.83% at a monthly horizon to 1.26% at an annual horizon in absolute value. These results are economically significant and robust to the inclusion of Carhart (1997) factors and past returns. The decreasing risk premia are also robust across different model specifications.

These results have direct implication for investors who are willing to earn the co-skewness risk premia. They suggest that such investors are better off keeping their position (in portfolios that load on co-skewness risk) on short time periods such as one-month (and possibly rebalance it), rather than holding it over longer period such as one year.

To the best of my knowledge, my paper is the first to study the multi-horizon pricing of co-skewness risk accounting for forward-looking information from option data. To that end, a brief overview of the related literature follows.

There is a growing literature in finance that studies the term structure of risk premia. Using data on dividend strips, van Binsbergen, Brandt, and Koijen (2012) show that the equity risk premium has a decreasing term structure. Andries, Eisenbach, Thomas, Schmalz, Schmalz, and Wang (2015) and Dew-Becker, Giglio, Le, and Rodriguez (2015) find that the variance risk premium also has a decreasing term structure (in absolute value). Although I adopt a methodology that is different from the ones used in these papers,[10] my results regarding the term structure of the co-skewness risk premia are not implied by such existing term structure results. In fact, as shown by Chabi-Yo (2002), in a dynamic two-period model, the price of co-skewness is distinct from that of the volatility risk. Recently, Christoffersen, Fournier, Jacobs, and Karoui (2016) argue that both prices should be the same in a one-period model. However, their theory has no direct implications for the multi-horizon pricing of co-skewness risk. Tédongap (2014) studies the multi-horizon pricing of consumption volatility in the cross-section of returns using a long-run risk model. My multi-horizon analysis is model-free and uses options data across different maturities. Moreover, my results could be used to discriminate between competing asset pricing models.

Only a few papers have estimated conditional co-skewness using options data. Conrad, Dittmar, and Ghysels (2013) estimate conditional co-skewness for several option maturities but only examine the risk-return relationship at the one-month horizon. Bali, Hu, and Murray

[10] Most of the existing papers use information on the profitability of investment strategies in options with different maturities to learn about the term structure of risk premia. Here, investment strategies are realized using stocks; options are only used to learn about investors' expectations of future risk.

(2016) use monthly options to estimate ex-ante co-skewness and further test its ability to explain variation in a yearly measure of expected returns based on the analyst price target. Both papers use the one-factor CAPM model together with estimates of total moments obtained using the methodology of Bakshi, Kapadia, and Madan (2003). The latter methodology only uses options from a single maturity. When the number of such options is scarce (as is the case at the long-horizon), the results can be quite sensitive to the interpolation and extrapolation techniques used. In such case, it is better to add more structure and use all the options data (covering all strikes and maturities) when estimating the moments for a given maturity, which is what my paper does. This illiquidity issue has also been tackled by Aliouchkin (2015), who relies on a fully parametric two-factor model to estimate ex-ante instantaneous co-skewness and study the risk pricing at the one-month horizon. Using a particular parametric model does not leave room for addressing issues related to the robustness of the results with respect to distributional assumptions, which my paper tackles.[11] My paper is also the first to make use of the entire term structure of option-implied co-skewness to study the term structure of the risk premia. Additionally, I formally show that the option-implied co-skewness delivered by my methodology more accurately predicts future realized co-skewness risk than the past historical measure used in Harvey and Siddique (2000a).

Vogt (2015) is the only other paper that uses Hermite polynomials to model the maturity dimension of the risk-neutral density. Vogt (2015) estimates marginal risk-neutral moments of the market, but does not consider any joint moments. My paper furthers the study in two ways. First, the problem I study is more complex due to market incompleteness arising from the absence of option-linking market and stock returns. Second, the Vogt (2015) methodology requires a large number of daily options, which is restrictive when looking at individual stock options, for which less traded options are available. By using a joint model, my approach allows for the use of additional mild restrictions between the market and the individual stock, which arguably is more helpful in identifying stock risk-neutral dynamics while still using only one day of option data.

Furthermore, my methodology can be used to investigate the multi-horizon pricing of co-variance and co-kurtosis, which will be discussed in the latter sections of this paper. It can also be used to investigate the multi-horizon pricing idiosyncratic risks (variance, skewness,

[11] Affine models are often used in the option pricing literature. To obtain option prices in semi closed form, only a few number of distributions can be used, which is a huge constraint for robustness analysis.

kurtosis). This analysis can be useful because there is no consensus on the signs of the prices of idiosyncratic risks.[12]

The rest of the paper is organized as follows. Section 3.2 describes the term structure model proposed for the option-implied joint distribution of market and stock returns, which is required to compute option-implied co-skewness. Section 3.5 presents the data used for the estimation. Section 3.6 presents the empirical results on the multi-horizon pricing of co-skewness risk. Section 3.7 concludes.

3.2 The Model

In this section, I propose a flexible term structure model for the option-implied joint distribution (or joint risk-neutral density) that can be used to compute option-implied co-skewness at any risk horizon. I start by presenting the model and its key properties and subsequently show how to estimate it.

Co-skewness risk captures the portion of the stock returns asymmetry that arises as a result of market returns asymmetry. As shown by Kraus and Litzenberger (1976) and Harvey and Siddique (2000a), it can be well proxied by the covariance between the stock returns and the squared market returns.[13] Option-implied co-skewness for a given stock is intended to capture stock co-skewness risk as reflected in traded options on the market and options on the stock.

For a given maturity, the option-implied co-skewness can be straightforwardly computed after estimating the joint risk-neutral density of the market and the stock returns at the horizon defined by this maturity.[14] My target is to come up with an estimate of the term structure of this joint density for each day. For the estimation, I rely on the options traded on this day as well as historical observed prices. The most recent options data available is likely to be more informative about the investors' perception on future risk than options traded on previous days. This is especially important in periods of turmoil.

[12]For example, Ang, Hodrick, Xing, and Zhang (2006) find that idiosyncratic volatility is priced negatively while Fu (2009) finds that it is priced positively. Multi-horizon analysis of idiosyncratic risk may become challenging when all option maturities do not yield the same sign for the estimated price of risk.

[13]This is the most popular proxy of co-skewness risk used in the literature due to its extreme simplicity. Alternative proxies can be found in Harvey and Siddique (2000a).

[14]This computation involves a numerical integration that is simple to implement and reliable if the density is available in closed form.

Let I_t denote the value of the market index, and S_t the value of the individual stock at time t. For any maturity τ, the aim is to estimate the joint risk-neutral density of the pair $(I_{t+\tau}, S_{t+\tau})$ without specifying the full dynamics of the two processes I and S, as is the case for parametric models.[15] I achieve this goal by relying on the sieve literature pioneered by Gallant and Nychka (1987) and Fenton and Gallant (1996). The key idea of this literature is that we can always reasonably approximate any unknown density by a combination of known functions called basis functions (which are usually of polynomial form).

A direct application of Gallant and Nychka (1987) in our context requires inferring a combination of the basis functions for each maturity using only options traded for this particular maturity (see for example Ángel Léon, Mencía, and Sentana (2009)). It is more challenging to estimate the long-horizon risk using such procedure due to the illiquidity of long-maturity options, which are less traded. To better model long-horizon risk, it is desirable to incorporate an additional structure that will allow for the use of the full panel of options data to recover in a single step the option-implied joint density for any maturity. The option panel is indexed by two dimensions: the maturity dimension and the moneyness dimension (see e.g. Andersen, Fusari, and Todorov (2015)). To achieve this desirable property, one needs to specify the form of the density of the triplet $(I_{t+\tau}, S_{t+\tau}, \tau)$. Vogt (2015) uses a similar approach in the special case with no stocks.[16]

I start by performing a change of variables that allows switching from the pair (I, S) to some intermediate variables that are more suited to follow the approximate multivariate sieve density of Gallant and Nychka (1987). To be precise, I model the τ-periods returns on the market $(r_{I,\tau})$ and τ-periods returns on the stock $(r_{S,\tau})$ as follows,

$$r_{I,\tau} = \log\left(\frac{I_{t+\tau}}{I_t}\right) = \mu_I(\tau) + V_I(\tau)X_\tau, \tag{3.1}$$

and

$$r_{S,\tau} = \log\left(\frac{S_{t+\tau}}{S_t}\right) = \mu_S(\tau) + V_S(\tau)Y_\tau, \tag{3.2}$$

[15] Parametric models rely on tight restrictions and are subject to misspecification errors.

[16] The Vogt (2015) approach is designed to infer the joint risk-neutral distribution of $(I_{t+\tau}, \tau)$. Extending his main idea to the case of the trivariate process $(I_{t+\tau}, S_{t+\tau}, \tau)$ requires additional technology, which I develop in this paper, see section 3.4 for a more comprehensive comparison.

with

$$V_I(\tau) = \sigma_I\sqrt{\tau},\ V_S(\tau) = \sigma_S\sqrt{\tau},\ \mu_I(\tau) = (r_f - \sigma_I^2/2)\tau,\ \mu_S(\tau) = (r_f - \sigma_S^2/2)\tau, \tag{3.3}$$

where r_f is the risk free rate, σ_I (resp σ_S) is the market (resp stock) volatility at time t. The change of variables in (3.1) and (3.2) allows for movements from strictly positive variables (I and S) that have their own ranges of values to new variables X_τ and Y_τ that are expected to be stationary with ranges approximately centered around zero.

Note that this is exactly the same change of variables that is performed is the classical Black and Scholes (1973a) model.[17] However, in the Black and Scholes (1973a) model, the univariate variables X_τ and Y_τ are normally distributed. More importantly, the Black and Scholes (1973a) model has no restrictions of the joint behaviour of X_τ and Y_τ. It is now widely recognized that the Black and Scholes (1973a) model does not explain well the implied volatility smile observed in option prices; this is mainly due to the fact that it does not allow for the conditional non-normality of X_τ and Y_τ (see, for example, Christoffersen, Heston, and Jacobs (2006b)). The sieve framework allows for the calibration of the degree of non-normality of X_τ and Y_τ that is required to suitably fit the options data. This paper is the first to use the sieve framework to uncover the joint behaviour of X_τ and Y_τ under the risk-neutral measure.

Specifically, I assume that the true unknown density $f^{(X_\tau,Y_\tau,\tau)}$ of (X_τ,Y_τ,τ) can be well approximated by the sieve multivariate density $f_K^{(X_\tau,Y_\tau,\tau)}$ defined as follows,

$$f_K^{(X_\tau,Y_\tau,\tau)}(x,y,\tau) = \left[\sum_{k=0}^{K_x}\left(\sum_{j=0}^{K_y}\Big(\sum_{l=0}^{K_\tau} B_{kjl}H_l(\tau)\Big)H_j(y)\right)H_k(x)\right]^2 e^{-\tau^2/2}e^{-x^2/2}e^{-y^2/2}, \tag{3.4}$$

subject to the constraint

$$\sum_{k=0}^{K_x}\sum_{j=0}^{K_y}\sum_{l=0}^{K_\tau} B_{kjl}^2 = 1, \tag{3.5}$$

where $H_j(x)'s$ are the Hermite polynomials, B is the three-dimensional matrix of coefficients,

[17] Specifically, $\mu_I(\tau)$ and $V_I(\tau)$ are the mean and volatility of the τ periods index returns. Jackwerth (2000) use a similar change of variables to estimate nonparametrically the univariate risk-neutral density for one single maturity.

and $K = (K_x, K_y, K_\tau)$ defines the orders of the Hermite polynomials.[18] By definition, the multivariate density in (3.4) is always positive and the squared term that appears in this expression allows for polynomials with order greater than K_x, K_y or K_τ. The constraint (3.5) ensures that the density integrates to one. It is worth stressing that setting $K_x = 0$, $K_y = 0$, and $K_\tau = 0$ returns to the Black and Scholes (1973a) model with X_τ and Y_τ being independent. Thus, our framework nests this popular model.

Most importantly, the density specification in (3.4) implies that for each maturity τ, the risk-neutral distribution of $(X_\tau, Y_\tau|\tau)$, which is required to compute the option-implied co-skewness, is given by

$$f_K^{(X_\tau,Y_\tau|\tau)}(x, y|\tau) = \frac{f_K^{X_\tau,Y_\tau,\tau}(x, y, \tau)}{\int_{\mathbb{R}}\int_{\mathbb{R}} f_K^{X_\tau,Y_\tau,\tau}(x, y, \tau)}. \tag{3.6}$$

The conditional density in (3.6) can be rewritten in the following compact form,

$$f_K^{(X_\tau,Y_\tau|\tau)}(x, y|\tau) = \left[\sum_{j=0}^{K_x}\sum_{k=0}^{K_y} \alpha_{jk}(B, \tau)H_j(x)H_k(y)\right]^2 \phi(x)\phi(y) \bigg/ \Big(\sum_{j=0}^{K_x}\sum_{k=0}^{K_y}[\alpha_{jk}(B, \tau)]^2\Big),$$

where

$$\alpha_{jk}(B, \tau) = \sum_{l=0}^{K_\tau} B_{jkl}H_l(\tau).$$

Given some mild assumptions on the space of the true density $f^{(X_\tau,Y_\tau,\tau)}$, Gallant and Nychka (1987) show that the approximated density $f_K^{(X_\tau,Y_\tau,\tau)}$ converges asymptotically to $f^{(X_\tau,Y_\tau,\tau)}$ as long as the total number of coefficients $(K_x + 1) \times (K_y + 1) \times (K_\tau + 1)$ of the B matrix in (3.4) tends to infinity.[19]

[18] The Hermite polynomials are orthogonalized polynomials defined as follows,

$$H_0(x) = 1, H_1(x) = x,$$
$$H_k(x) = \frac{xH_{k-1}(x) - \sqrt{k-1}H_{k-2}(x)}{\sqrt{k}}, k \geq 2$$

[19] While the conditional density in (3.6) is enough to compute derivative prices, we can always obtain the density of $(I_{t+\tau}, S_{t+\tau}|\tau)$ using the fact that the change of variables defined in (3.1) and (3.2) is a bijection.

3.3 Option Valuation

Given that my aim is to estimate the co-skewness risk using option prices, it is important to ensure that I am able to compute option prices swiftly and reliably. In this regard, an important feature from a practical point of view of my method is that it gives closed-form expressions for the univariate conditional densities and option prices. In particular, I show that the conditional marginal densities $f_K^X(x|\tau)$ and $f_K^Y(y|\tau)$ can be written as linear combinations of the Hermite polynomials. Indeed, using some key properties of the Hermite polynomials, I obtain

$$f_K^{X_\tau}(x|\tau) = \sum_{j=0}^{2K_x}\sum_{k=0}^{K_y} \alpha_{\bullet,k}(B,\tau)^\top A_j^{K_x} \alpha_{\bullet,k}(B,\tau) H_j(x)\phi(x) \Big/ \Big(\sum_{j=0}^{K_x}\sum_{k=0}^{K_y}[\alpha_{jk}(B,\tau)]^2\Big), \tag{3.7}$$

$$f_K^{Y_\tau}(y|\tau) = \sum_{j=0}^{2K_y}\sum_{k=0}^{K_x} \alpha_{k,\bullet}(B,\tau)^\top A_j^{K_y} \alpha_{k,\bullet}(B,\tau) H_j(y)\phi(y) \Big/ \Big(\sum_{j=0}^{K_x}\sum_{k=0}^{K_y}[\alpha_{jk}(B,\tau)]^2\Big), \tag{3.8}$$

where $\alpha_{\bullet,k} = (\alpha_{0,k},\ldots,\alpha_{K_x,k})^\top$, $\alpha_{k,\bullet} = (\alpha_{k,0},\ldots,\alpha_{k,K_y})^\top$, $\phi(x)$ is the pdf of the normal density and $A_j^{K_u}$ is a $(K_u+1)\times(K_u+1)$ symmetric matrix whose typical elements are

$$a_{uv,j} = \frac{(u!v!j!)}{(\frac{u+v-j}{2})!(\frac{u+j-v}{2})!(\frac{j+v-u}{2})!}$$

if $j \in \Gamma$ and zero otherwise with

$$\Gamma = \{j \in \mathbb{N} : |u-v| \le j \le u+v; \frac{u-v+j}{2} \in \mathbb{N}\}.$$

Note that (3.7) (respectively (3.8)) shows that the larger K_x (respectively K_y) becomes, the more complex and richer the conditional density $f_K^{X_\tau}$ (respectively $f_K^{Y_\tau}$) is. In practice, it is desirable to have both K_x and K_y large enough. Obtaining the density of the future assets returns (over different horizons) facilitates the computation of option prices.[20]

Using the classical risk-neutral valuation formula, the price of European put options on

[20]This is clearly an attractive feature since this density is not available in closed form in parametric affine option pricing models, only the characteristic function is available in closed form.

the market index I with maturity τ and strike price κ is given by

$$P^I(\kappa, \tau) = e^{-r_f \tau} \mathbb{E}^{\mathbb{Q}} \Big(\max\left(\kappa - I_{t+\tau}, 0\right) \Big), \tag{3.9}$$

where $\mathbb{Q}$ denotes the risk-neutral measure. The probability density of $I_{t+\tau}$ under $\mathbb{Q}$ is provided by (3.7). Proposition 1 below shows that the price $P^I(\kappa_I, \tau_I)$ can be computed in closed form.[21]

Proposition 1.

$$\begin{aligned} P^I(\kappa, \tau) = \kappa \exp^{-r_f \tau} \Big[& \Phi(d(\tau, \kappa)) - \sum_{k=1}^{2K_x} \frac{\gamma_k(B, \tau)}{\sqrt{k}} H_{k-1}(d(\tau, \kappa)) \Phi(d(\tau, \kappa)) \Big] \\ - I_t \exp^{r_f \tau + \mu_I(\tau)} \Big[& \exp\left(V_I(\tau)^2/2\right) \Phi(d(\tau, \kappa) - V_I(\tau)) + \sum_{k=1}^{2Kx} \gamma_k(B, \tau) J_k^*(d(\tau, K)) \Big], \end{aligned}$$

where

$$\begin{aligned} J_k^*(d(\tau, \kappa)) &= \frac{V_I(\tau)}{\sqrt{(k)}} J_{k-1}^*(d(\tau, \kappa)) - \frac{1}{\sqrt{k}} \exp\left(V_I(\tau) d(\tau, \kappa)\right) H_{k-1}(d(\tau, \kappa)) \Phi(d(\tau, \kappa)) \\ J_0^*(d(\tau, \kappa)) &= \exp\left(V_I(\tau)^2/2\right) \Phi(d(\tau, \kappa) - V_I(\tau)), \textit{ and} \\ d(\tau, \kappa) &= \frac{\log(\kappa/I) - \mu_I(\tau)}{V_I(\tau)} \end{aligned}$$

and $\Phi(.)$ *is the cumulative CDF of the normal distribution, and*
$\gamma_j(B, \tau) = \sum_{k=0}^{K_y} \alpha_{\bullet,k}(B, \tau)^\top A_j^{K_x} \alpha_{\bullet,k}(B, \tau) \Big/ \Big(\sum_{k=0}^{K_y} \alpha_{\bullet,k}(B, \tau)^\top \alpha_{\bullet,k}(B, \tau) \Big)$.

The price of a put option on the individual stock can be obtained in a similar manner and is provided in the Appendix C.2. Call options prices can be obtained in closed forms using the call-put parity.

3.4 Model Estimation

To obtain the option-implied co-skewness risk at any maturity, we need to estimate the risk-neutral joint density. The estimation of the risk-neutral joint density is more challenging

[21] The closed form expression is obtained by exploiting the closed-form expression of the univariate density in (3.7).

than that of the univariate density.[22] Option contracts whose payoffs depend on the future joint realizations of the market and stock prices should be informative for this estimation. Unfortunately, however, such options are not traded.

To overcome this limitation, I impose some additional restrictions on my multivariate sieve density defined in (3.4). In particular, I adopt the correlation assumption of Aït-Sahalia and Brandt (2007), which restricts the risk-neutral correlation to be equal to the conditional historical correlation. As explained by Aït-Sahalia and Brandt (2007), over short time intervals, this assumption can be motivated by the Girsanov's theorem according to which the second moments are unaffected by a change of measure in the continuous time limit. As previously discussed, I refer to this restriction as the "zero correlation risk premia" assumption. As also argued by Aït-Sahalia and Brandt (2007), we can expect this assumption to hold approximately over long-horizon.[23]

I propose to obtain the risk-neutral joint distribution by minimizing a discrepancy measure between the risk-neutral joint distribution and an estimate of the conditional physical joint distribution. This is subject to the restrictions that the options on both the market and the stock are well priced, the correlation restrictions are met, and some no-arbitrage conditions that guarantee both the stock and the market prices are martingales under the risk-neutral measure are satisfied. The discrepancy measure is aggregated over an appropriate set of maturities, which allows for the handling of all maturities in a single estimation stage. I use the quadratic criterion as the discrepancy measure.[24] To be more precise, the

[22] Theoretically, for a given maturity the univariate risk-neutral density is the second derivative with respect to the strike price of the call options, see Breeden and Litzenberger (1978).

[23] This assumption can be traced back to French, Groth, and Kolari (1983), who use it to estimate option-implied market betas for one single option maturity. Aït-Sahalia and Brandt (2007) use this assumption together with a parametric copula to obtain the risk-neutral joint distribution of the market returns and the bond returns for one single option maturity. Chabi-Yo and Song (2013) use the methodology of Aït-Sahalia and Brandt (2007) to derive the risk-neutral joint distribution of the market returns and the innovations of the VIX index.

[24] One alternative is to minimize a convex function of the stochastic discount factor defined as the ratio between the risk-neutral and the $\mathbb{P}$ densities. Since I do not have closed form solution for the optimization problem, dividing by probabilities that are close to zero causes some numerical instabilities in the optimization algorithm. These numerical difficulties are also highlighted by Jackwerth and Rubinstein (1996) in the univariate case for a single option maturity.

problem to be solved has the following form,

$$\min_{B} \quad \left[\frac{1}{n_\tau}\sum_{\tau}\frac{\int_{\mathbb{R}^2}\left[f_K^{\mathbb{Q}}(r_{I,\tau}, r_{S,\tau}|\tau) - \widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau)\right]^2 dr_{I,\tau}dr_{S,\tau}}{\int_{\mathbb{R}^2}\left[\widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau)\right]^2 dr_{I,\tau}dr_{S,\tau}}\right]$$

$$\text{subject to} \quad \begin{cases} P_i^I = P^I(\kappa_i^I, \tau_i^I), \quad \forall i \in \{1, \ldots, n_I\}, \\ P_j^S = P^S(\kappa_j^S, \tau_j^S), \quad \forall j \in \{1, \ldots, n_S\}, \\ \mathbb{E}^{\mathbb{Q}}\left(r_{I,\tau}\right) = r_f \times \tau, \quad \forall \tau \in \{\tau_1, \ldots, \tau_{n_\tau}\}, \\ \mathbb{E}^{\mathbb{Q}}\left(r_{S,\tau}\right) = r_f \times \tau, \quad \forall \tau \in \{\tau_1, \ldots, \tau_{n_\tau}\}, \\ \rho_\tau^{\mathbb{Q}} = \rho_\tau^{\mathbb{P}}, \quad \forall \tau \in \{\tau_1, \ldots, \tau_{n_\tau}\}, \end{cases}$$

where r_f is the risk-free rate and

$$f_K^{\mathbb{Q}}(r_{I,\tau}, r_{S,\tau}|\tau) = \frac{1}{V_I(\tau)V_S(\tau)} f_K^{(X_\tau, Y_\tau|\tau)}(x, y|\tau), \tag{3.10}$$

$f_K^{(X_\tau, Y_\tau|\tau)}(x, y|\tau)$ is defined in (3.6) and involves the B matrix, and $f_K^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau)$ is an estimate of the physical joint density, which will be defined below. Note that this problem is equivalent to finding a minimum variance pricing kernel in a multivariate setting with multiple maturities. Recently, Schneider and Trojani (2019) solve a similar problem in the univariate case with a single maturity.

Option-inferred risk is usually a biased predictor of the future realized risk (see e.g. Bliss and Panigirtzoglou (2004)). The correlation and density restrictions are intended to reduce this bias. It is also worth emphasizing that, with the exception of the correlation assumption, my estimation methodology can be regarded as a multivariate generalization of the procedure proposed by Jackwerth and Rubinstein (1996), where the estimate of the physical joint density serves as prior information on the risk-neutral joint density. Here, I am able to handle more than one option maturity at the same time, which is a desirable feature. Given my sieve framework, only some coefficients of the B matrix need to be estimated to obtain the risk-neutral joint density.

In practice, option prices are observed with errors, and that some imperfect measures are to be used to approximate the historical correlations, which implies that the theoretical constraints of the optimization problem cannot be fully satisfied. Due to this limitation, it is easier to solve the following problem,

$$
\begin{aligned}
\min_{B} \quad & \left[\frac{1}{n_\tau} \sum_{\tau} \frac{\int_{\mathbb{R}^2} \left[f_K^{\mathbb{Q}}(r_{I,\tau}, r_{S,\tau}|\tau) - \widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau) \right]^2 dr_{I,\tau} dr_{S,\tau}}{\int_{\mathbb{R}^2} \left[\widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau) \right]^2 dr_{I,\tau} dr_{S,\tau}} \right. \\
& + \omega^I \left(\frac{1}{n_I} \sum_{i=1}^{n_I} \frac{\left[P_i^I - P^I(\kappa_i^I, \tau_i^I) \right]^2}{\overline{(P^I)^2}} \right) + \omega^S \left(\frac{1}{n_S} \sum_{i=1}^{n_S} \frac{\left[P_i^S - P^S(\kappa_i^S, \tau_i^S) \right]^2}{\overline{(P^S)^2}} \right) \\
& \left. + \omega^\rho \left(\frac{1}{n_\tau} \sum_{\tau} \frac{[\rho_\tau^{\mathbb{Q}} - \rho_\tau^{\mathbb{P}}]^2}{\overline{(\rho^{\mathbb{P}})^2}} \right) \right]
\end{aligned}
$$

$$
\text{subject to} \quad \begin{cases} \mathbb{E}^{\mathbb{Q}}\left(r_{I,\tau}\right) = r_f \times \tau, & \forall \tau \in \{\tau_1, \ldots, \tau_{n_\tau}\} \\ \mathbb{E}^{\mathbb{Q}}\left(r_{S,\tau}\right) = r_f \times \tau, & \forall \tau \in \{\tau_1, \ldots, \tau_{n_\tau}\}. \end{cases}
$$

where $\overline{P^I}$ is the average of the index option prices across the maturity and moneyness dimensions, $\overline{P^S}$ is the average of the stock option prices, and $\overline{\rho^{\mathbb{P}}}$ is the average of this historical conditional correlations across the range of maturities.

Note that, now, the minimization criterion involves four discrepancy measures. The first two measures account for option prices restrictions. The third one incorporates the correlation restriction and the last one handles the density restriction. These four measures are suitably standardized to make them comparable to each other. They are further weighted by some scale factors ω^I, ω^S and ω^ρ.

The physical density $\widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau)$ should reflect the current state of the economy. It can be estimated using returns data over a short time period or estimated over longer time period conditional on some key variables. Since I am also interested in estimating risk at long-horizon, I opt for the second approach. The key conditioning variables are the market volatility and the stock volatility. For each horizon τ, $\widehat{f}^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau)$ is estimated using the Nadara-Watson kernel as follows,[25]

$$
f^{\mathbb{P}}(r_{I,\tau}, r_{S,\tau}|\tau) = \frac{1}{b_{r_I} b_{r_S} b_{\sigma_I} b_{\sigma_S}} \times
$$

[25] More complex estimation of the conditional physical distribution includes the local polynomial approach by Song and Xiu (2016).

$$\sum_{t'=0}^{t-\tau} \mathcal{K}\Big(\frac{r_{I,t',\tau} - r_{I,\tau}}{b_{r_I}}\Big)\mathcal{K}\Big(\frac{r_{S,t',\tau} - r_{S,\tau}}{b_{r_S}}\Big)\mathcal{K}\Big(\frac{\sigma_{I,t'} - \sigma_{I,t}}{b_{\sigma_I}}\Big)\mathcal{K}\Big(\frac{\sigma_{S,t'} - \sigma_{S,t}}{b_{\sigma_S}}\Big) \Big/ \sum_{t'=0}^{T-\tau} \mathcal{K}\Big(\frac{\sigma_{I,t'} - \sigma_{I,t}}{b_{\sigma_I}}\Big)\mathcal{K}\Big(\frac{\sigma_{S,t'} - \sigma_{S,t}}{b_{\sigma_S}}\Big),$$

where $\mathcal{K}$ is the Gaussian kernel, $r_{I,t',\tau} = \log(I_{t'+\tau}/I_{t'})$, and $\sigma_{I,t'}$ (resp $\sigma_{S,t'}$) is an estimate of the market (stock) volatility at time t'. Recall that the risk-neutral distribution is to be estimated at time t. The bandwiths ($b_{r_I}, b_{r_S}, b_{\sigma_I}$and b_{σ_S}) are chosen as in Jackwerth (2000).[26]

Empirically, the estimation is to be performed for a given set of days.[27] For each day in this set, I solve the optimization problem for each stock. This procedure allows for the use of the market options when estimating the univariate risk-neutral density of the stock returns. This feature is attractive and can be achieved in parametric models through the use of factor models. It is important to make sure that the univariate distribution of the market is not significantly affected by the stock data. This key property has recently been emphasized by Gouriéroux and Monfort (2017) and Boloorforoosh, Christoffersen, Fournier, and Gouriéroux (2017). Empirically, I achieve this property by setting ω^I sufficiently higher compared to ω^S and ω^ρ. Indeed, I set $\omega^I = 10^7$, $\omega^S = 10^5$, and $\omega^\rho = 10^5$. This ensures that the market option data is well fitted and that the fit is similar across the stocks. Papers relying on factor models (see e.g. Bégin, Dorion, and Gauthier (2017), Gourier (2016b), Aliouchkin (2015)) proceed by estimating first the dynamics of the market relying on market options, then estimating the dynamics of the stocks conditional on that of the market. A disadvantage of such methodology is that it requires the use of several days of options data, which makes the dynamic estimation of the model computationally fairly intensive. In contrast, my methodology requires only one day of options data.

Vogt (2015) is the only other paper that uses Hermite polynomials to model the maturity dimension of the risk-neutral density. Vogt (2015) estimates marginal risk-neutral moments of the market, but does not consider any joint moments. This paper furthers the analysis in two ways. First, the problem I study is more complex due to market incompleteness arising from the absence of option linking market and stock returns. Second, the Vogt (2015) methodology requires a large number of daily options, which is restrictive when looking at individual stock

[26] In particular, the bandwith is $1.8 \times \sigma \times n^{-1/5}$, where n is the number of observations and σ is the standard deviation of the sample returns.

[27] In the empirical section, the estimation is performed for 168 days.

options for which less traded options are available. By using a joint model, my approach allows for the use of additional mild restrictions between the market and the individual stock, which I argue more accurately identifies stock risk-neutral dynamics while still using only one day of option data.

For the implementation, I use several combinations to check the robustness of my results. The main results are reported for the combination $K_x = 4$, $K_y = 4$, and $K_\tau = 2$.[28] The range of maturities is set as $\{1, 2, \ldots, 12\}$ months. The market volatility ($\sigma_{I,t}$) at time t is estimated using the square root of the realized volatility computed using the market daily returns observed over the last 21 business days.[29] Similarly, the correlations are computed using historical returns data over the period $[t - \tau, t]$.[30]

3.5 Data

For the estimation, I rely on the daily stock data and monthly options data. The options data used come from OptionMetrics and include both options on the market (S&P 500) and options on individual stocks. The index options are of European style, while options on the individual stocks are of American style. Following the literature, I implement different filters to clean the data. In doing so, I removed options that violate the no-arbitrage rules, options with negative bid-ask bounds, and in-the money options. For the index options, I use the mid-quote as the observed price. For the individual stock options, I convert the implied volatility provided by OptionMetrics into option prices.[31] I use options covering the period of 2001-2014.[32] I select stocks that have traded options at the end of every month between 2001 and 2014. Subsequently, I keep stocks that have at least 15 options each day.[33] This leaves a total of 96 stocks. Call options are converted into put options using the call-put parity for

[28] I estimate the model for two other combinations $(K_x, K_y, K_\tau) = (4, 4, 3), (5, 5, 2)$.

[29] An appealing alternative approach is to estimate volatility using high-frequency data (see e.g. Andersen, Bollerslev, Diebold, and Labys (2003)). The stock volatility ($\sigma_{S,t}$) is estimated in a similarly manner. As the number of individual stocks covered in the study is relatively large, using daily data simplifies the computation.

[30] Buss and Vilkov (2012) use the same proxy for historical correlations.

[31] OptionMetrics uses binomial trees to extract European style implied volatility from the American style implied volatility.

[32] Since my methodology uses only daily options to estimate the risk measures for all risk horizons, it is desirable to have a large number of options per day. Starting the analysis from 1996 and using the same selection criterion considerably lowers the final number of stocks.

[33] The higher the number of daily options, the more precise the estimates.

estimation purposes. I use data on risk free rate and dividends from OptionMetrics.[34] I also use stock prices data from CRSP covering the period of 1980-2015. Additionally, I use data on Carhart (1997) factors from the Fama-French Library.

Table 3.1 reports the daily average number of available index options in my sample by maturity buckets. Table 3.2 provides the same information averaged across all stocks. It appears that the average number of options available each day has increased throughout the sample. The average daily number of options seems to be much higher for the index than for the individual stocks. In both cases, this number seems to be much smaller at long maturity.

3.6 Empirical Results

This section presents the results of the estimation and addresses the implications for the pricing of the co-skewness risks in the cross-section of returns.

3.6.1 Model Fit of the Option Data

Before studying the multi-horizon pricing of co-skewness risk, I analyze the fit of the model to the cross-section of options and dissect the key results across time and along the maturity dimension. My methodology can be thought of as being a semi-nonparametric method that involves the estimation of a relatively large number of coefficients in order to learn from the data as freely as possible. It is important to ensure that the method does not overfit the options data. To analyze the fit of the model to data, I use the relative root mean square error of the prices defined as follows

$$RRMSE = \sqrt{\sum_{i=1}^{n} \left(\hat{P}_i / P_i - 1\right)^2}, \tag{3.11}$$

where P_i is the observed option price, $\hat{P}_i$ is the model predicted option price, and n denotes the number of options.

In Section 3.4, I propose a larger weight on the constraint of the index option prices to ensure that the market marginal density is not significantly affected by stock data. Here, I proceed to assess the performance of my methodology. A large amount of options are traded

[34] My model does not account directly for dividends. Thus, I follow the literature and use ex-dividend stock prices for the option valuation.

on the index each day and, as shown by Vogt (2015), this information is sufficient to estimate market marginal density across maturities. This means that the marginal density should not differ, given that the fit of the market option data is comparable across the stocks. To analyze this feature, I look at the time series of the coefficient of variation of the index option RRMSE across the stocks. The coefficient of variation is the ratio of the standard deviation to the mean. Figure 3.1 displays this time series. Having the coefficient of variation identically equal to zero will mean that the RRMSE is the same across stocks. Here, I note that the coefficient of variation of the RRMSE is quite small in magnitude (below 1.2%). It rises during the financial crisis as a signal of how relatively complicated it was to ensure this consistency property during the turmoil period. Overall, however, the result is satisfactory.

Table 3.3 reports the RRMSE for the market and the stocks. RRMSE reported for the stock corresponds to a cross-sectional average. As observed in Table 3.3, the model fits quite well to both the market and the individual stock options. The overall RRMSE is 45% for the index and 33% for individual stock options. For both the market and individual stocks, the RRMSE has somewhat increased over time due to the increasing number of options. As shown in Table 3.4, the long-maturity options seem to be slightly better fitted than the short-maturity ones, which can be explained by the relative number of options in these option categories. There is no evidence of the model overfitting the data since the overall RRMSEs are not considerably small.

3.6.2 Behaviour of Co-skewness risk and its pricing

Descriptive Statistics for Co-skewness risk

In this section, I study the behaviour of the time series of option-implied co-skewness and contrast this behaviour with that of the historical co-skewness. I also look at the predictive content of the option-implied risk for future realized risk.[35]

Let r_m be the market excess returns and r_j be the excess returns on a given stock j, I denote by $\mathrm{Var}_{t,m}$ the variance of the market returns and by $\mathrm{Var}_{t,j}$ that of the stock returns,

$$\mathrm{Var}_{t,m} = \mathbb{E}_t\Big(\big(r_{t,m} - \mathbb{E}_t(r_{t,m})\big)^2\Big), \quad \mathrm{Var}_{t,j} = \mathbb{E}_t\Big(\big(r_{t,j} - \mathbb{E}_t(r_{t,j})\big)^2\Big). \tag{3.12}$$

[35] The historical co-skewness is calculated using returns data only, and does not incorporate option data. The future realized co-skewness is computed using daily returns data over the next year. It also represents a future value of the time series of the historical measure.

Following the literature, the co-skewness risk is proxied by

$$\text{Coskew}_{t,j} = \frac{\mathbb{E}_t\Big(\big(r_{t,j} - \mathbb{E}_t(r_{t,j})\big)\big(r_{t,m} - \mathbb{E}_t(r_{t,m})\big)^2\Big)}{\text{Var}_{t,m}\text{Var}_{t,j}^{1/2}}. \tag{3.13}$$

Note that, similar to the classical definition of skewness, the co-skewness measure in (3.13) is suitably standardized by some appropriate functions of the variances in (3.12), making it scale-invariant and comparable across risk horizons. By replacing the theoretical expectations in (3.13) with their empirical counterparts, one obtains the most popular historical co-skewness measure of Harvey and Siddique (2000a). To account for the conditional nature of co-skewness risk, the historical measure is computed using one year of past daily returns.

Panel B of Table 3.5 presents the descriptive statistics for co-skewness risk. Following the literature, these descriptive statistics are obtained after pooling the measure across its cross-sectional and time series dimensions. As can be seen in Panel B of Table 3.5, the option-implied co-skewness is negative on average, which implies that the stock returns load negatively on the squared market factor. In other words, on average, the stocks in my sample perform worse when the market is volatile.

Table 3.5 shows that the average option-implied co-skewness has a decreasing term structure (in absolute value) from -0.34 at the monthly horizon to -0.10 at the yearly horizon. This result implies that market participants expect the magnitude of stock sensitivities to this risk factor to decrease with horizon. However, note that this decreasing term structure of average risk has no direct implication for the term structure of the price of risk, as the latter can be cyclical or countercyclical. The option-implied co-skewness measure appears to be smoother at the long-horizon than at the short-horizon. Its standard deviation decreases from 0.25 at the one-month horizon to 0.19 at the one-year horizon. However, note that this standard deviation remains constant when moving from the mid-term horizon (six months) to the long-term (one year). The one year option-implied co-skewness appears to be smoother than the historical measure, which has a standard deviation of 0.22, as presented in the last row of Panel B of Table 3.5. The average of the one year option-implied co-skewness (-0.10) is close to the average of the historical measure (-0.09). This means that this measure is almost mean unbiased.[36] The option-implied co-skewness measure displays a non-trivial cor-

[36] This property could be useful in predictive regression.

relation with its historical counterpart, as can be seen in Table 3.6. The correlation between the option-implied co-skewness and the historical co-skewness decreases from 0.533 to 0.410 as the horizon rises. It is worth stressing that my analysis uses the option-implied measure obtained for one single day. This measure can be smoothed further by averaging it quarterly or monthly. This was omitted as it would require estimating the model for every single day in my sample (instead of every end of month as discussed).[37] In contrast, I show that the single day estimate allows for the recovery of a risk-return trade-off that is in line with the economic theory (see Section 3.6.2).

The intuition for using options in the estimation of risk measures is that we expect to better capture investors' anticipation of the future risk. I now proceed to formally test this intuition for co-skewness risk. More precisely, I regress the twelve months ahead average historical risk (which is my proxy for future risk) on the average option-implied measure (for different option horizons). For comparison purposes, I repeat the same analysis using the past average historical measure. The results, which are reported in Panel A of Table 3.7, show that the option implied co-skewness measures contain predictive information for the future risk. The adjusted R^2 of the regression lies between 3.6% and 6.7%, depending on the option horizon. This is clearly far better compared to an adjusted R^2 of -0.5% obtained using the past average co-skewness as a predictor. The results show that, while the option-implied measures are non-trivially correlated with the past historical measure, only the former seems to have an important predictability power for capturing future risk. The results also show that the option-implied co-skewness is a biased predictor of the future risk, since the intercept of the regression is statistically different from zero and the slope different from one.[38] The slope is positive and lies between 0.25 and 0.36.

Multi-horizon pricing of co-skewness risk

I now look at the multi-horizon pricing of co-skewness risk in the cross-section of returns. In Section 3.6.2, I demonstrate that the option-implied measure seems to be to perform well in predicting future co-skewness risk. I now test whether this predictability gain translates

[37]Note that while averaging the risk measure over a time period allows us to obtain less noisy estimate, the procedure has the disadvantage of yielding a measure that reflects less of the conditional ex-ante information.

[38]To account for the overlapping nature of the regression, the t-statistics are calculated using the HAC procedure.

into a better correlation between conditional skewness and future stock returns across risk horizons. To achieve this goal, I use the Fama and MacBeth (1973) procedure.

Following the literature, at the end of each month, I run a regression of stocks excess returns on some risk measures. There are 168 months in the sample. The prices of risk, which are the coefficients from this regression, are then averaged over the sample period, which is 2001-2014. When looking at the k month investment horizon with k greater that one, the estimated prices of risk are obtained using overlapping returns. To adjust for this effect, I use the Newey-West t-statistics with $k-1$ lags. To ensure that the results are not driven by other factors, I control for some standard systematic factors. Specifically, I use the Carhart (1997) factors as well as two momentum factors (past monthly returns and past yearly returns). The Carhart (1997) factors include the market, the size, the value, and the momentum factors. Following the literature, stock betas with respect to these factors are computed using OLS regression with one year of past data.[39]

The Fama and MacBeth (1973) regression takes the following form

$$r_{t,t+k} = \alpha + \lambda_{t,k} F_{t,t+h} + \lambda_Z Z_t + \epsilon_{t,t+h}, \tag{3.14}$$

where $r_{t,t+k}$ is a cumulative stock excess returns over k months, $F_{t,t+h}$ is an option-implied measure corresponding to horizon h, and Z is a matrix of control variables. Varying k from one to twelve allows for the uncovering of the term structure of the price of risk, which is important in learning the optimal timing of an investment strategy that is aimed to earn this price.

Table 3.8 provides the t-statistics for the average price of risk of the co-skewness risk in the univariate regression (F equals the co-skewness measure in (3.14) and the Z factor is empty). With the exception of the one-month option-implied measure, all the coefficients presented in this table are negative, which means that this risk is negatively priced in the cross-section of returns. This result is consistent with the theoretical predictions of Kraus and Litzenberger (1976) and Harvey and Siddique (2000a). Looking at the magnitude of the t-statistics, it appears that for most of the investment horizons, the risk-return relation looks stronger when the long-horizon implied co-skewness is used instead of the short-horizon

[39] The market beta is computed in a simple univariate regression. To estimate the beta of the size, value, and momentum factors, I use a multivariate regression that includes the market factor as an additional regressor. This allows for the reduction of possible correlation between these anomalies betas and the market beta.

one. This result is perhaps not surprising given that the long-horizon measure was less noisy than that of the short horizon, as documented in Section 3.6.2. However, when the yearly option-implied co-skewness is used (see the last row of Table 3.8), the average price of risk is statistically significant at the 5% level only for the two and three months horizons. To test the robustness of this conclusion, I add the control variables discussed above (namely, the Carhart (1997) factors, the past monthly returns, and the past yearly returns) in the regression. The t-statistics are reported in Table 3.9. The results are very similar to the ones obtained in the univariate regression. These results are also robust to the inclusion of the historical measure of Harvey and Siddique (2000a), as shown in Table 3.10.

Theoretically, options traded at a given horizon should be the most informative for capturing the risk premia at that horizon. That is, short-term options information should better capture risk premia at the short-horizon and long-term options information should be useful in capturing risk premia at the long-horizon. My results show that this simple prediction is not always satisfied. Since the end goal is to investigate risk premia behaviour across horizons, I adopt the following strategy. For a given investment horizon, I look for the option horizon that best captures the price of the risk (that is, the one that minimizes the t-statistics of the estimated average price of risk) and use the estimate obtained for this option horizon to learn about the risk premia. I proxy the risk premia by the average price of risk scaled by the standard deviation of the option-implied moment that best captures the pricing relationship. This strategy fully exploits the information content of the term structure of option-implied risk for every single investment horizon.

I adopt the above strategy to robustly estimate co-skewness risk premia across investment horizons. Figure 3.3 depicts the term structure of risk premia both in the univariate and multivariate regressions. In both cases, the risk premia have a decreasing term structure (in absolute value). Accounting for control variables, the co-skewness risk premium (in annual term) decreases from 3.83% at the monthly horizon to 1.26% in absolute value. The term structure becomes almost flat after the six months horizon. The short-term premium is close to the 3.6% found by Harvey and Siddique (2000a) and a bit smaller than the 5.64% reported by Aliouchkin (2015). However, none of these two papers estimated the multi-periods risk premia that I obtain by exploiting the information content of the entire term-structure of option-implied co-skewness.

To further check the robustness of my results for the co-skewness risk, I examine an alternative measure of this risk. As argued by Harvey and Siddique (2000a), this risk can also be proxied by the beta from a multivariate regression of stock returns on market returns and the squared market returns. I evaluate the option-implied version of this beta. My paper appears to be the first to investigate how the option-implied version of this proxy affects future returns.[40] Table 3.11 reports the t-statistics from running the Fama and MacBeth (1973) regression with this sole measure as the regressor. As can be seen from the table, this measure carries a negative average price of risk, which is consistent with the theoretical prediction of Harvey and Siddique (2000a). For most horizons, the estimates are statistically significant at the 5% level. Table 3.12 shows that after controlling for Carhart (1997) and past returns, the estimates are still negative, while the t-statistics are reduced to some extent. This means that part of the variation previously attributed to the co-skewness measure is due to the systematic factors. Figure 3.3 depicts the term structure of risk premia both in the univariate and multivariate regression. The risk premium is estimated as before. From the univariate regression, we see that the co-skewness risk premia have a decreasing term structure in absolute value, which is mildly reversed in the mid-term when we control for other systematic factors. In annual terms, the risk premium is 2.3% at the monthly horizon and 0.83% at the yearly horizon. These premia are less economically significant than the ones documented using the first proxy. However, the decreasing magnitude of the premia seems to be a robust conclusion.

Overall, the decreasing term structure of co-skewness risk premia suggests that market participants who are willing to earn the co-skewness risk premia should keep their investment position over shorter time period rather than holding it over longer time period in order to maximize their earnings. There is a growing literature in finance that studies the term structure of risk premia. Using data on dividend strips, van Binsbergen, Brandt, and Koijen (2012) show that the equity risk premia have a decreasing term structure. Andries, Eisenbach, Thomas, Schmalz, Schmalz, and Wang (2015) and Dew-Becker, Giglio, Le, and Rodriguez (2015) find that the variance risk premia also have a decreasing term structure (in absolute value). Here, I show that the co-skewness risk premia also have a decreasing term structure (in absolute value) and I argue that this result is not implied by the existing result

[40] Papers that use the historical version of this measure include Bali, Hu, and Murray (2016) and Christoffersen, Fournier, Jacobs, and Karoui (2016).

on the term structure of variance risk premia. In fact, as shown by Chabi-Yo (2002) in a dynamic two-period model, the price of co-skewness is distinct from that of the volatility risk.

3.6.3 Behaviour of co-kurtosis risk and its pricing

Although my paper aims at studying co-skewness risk, I briefly describe some results for the co-kurtosis risk, which captures the sensitivity of stock returns to extreme market returns.

Descriptive Statistics for Co-Kurtosis risk

In this section, I study the behaviour of the time series of option-implied co-kurtosis risk and contrast this behaviour with that of the historical co-kurtosis. I also look at the predictive content of the option-implied risk for future realized risk.
Following the literature, the co-kurtosis risk is proxied using,

$$\text{Cokurt}_{t,j} = \frac{\mathbb{E}_t\Big(\big(r_{t,j} - \mathbb{E}_t(r_{t,j})\big)\big(r_{t,m} - \mathbb{E}_t(r_{t,m})\big)^3\Big)}{\text{Var}_{t,m}^{3/2}\text{Var}_{t,j}^{1/2}}. \tag{3.15}$$

Panel C of Table 3.5 provides the descriptive statistics on the co-kurtosis risk. It shows that the option-implied co-kurtosis is positive on average and exhibits a term structure that is decreasing from an average value of 1.84 at the monthly horizon to 1.49 at the yearly horizon. This means that on average, the stocks load positively on the markets returns when the latter experiences very extreme variations and that investors expect this sensibility to be less important over long-horizon.[41] Similar to the option-implied co-skewness measure, the option-implied co-kurtosis measure appears to be smoother at the longer horizon as opposed to the shorter horizon (the standard deviation decreases from 0.75 to 0.63 from the monthly horizon to the yearly horizon, but still a little bit noisier than the historical measure, which has a standard deviation of 0.52). The average yearly implied co-kurtosis (1.49) is smaller than the average of the historical measure (1.81), which means that on average the measure is biased.[42] This relatively poor behaviour of the co-kurtosis measure compared to the co-skewness measure can potentially be explained by the fact that co-kurtosis risk is much harder

[41] In general, the probability of having extreme market returns over longer maturity is small. Crashes or jumps are more likely to occur over the short-horizon.

[42] Aliouchkin (2015) also reports some average bias in his option-implied co-kurtosis measure.

to capture.

The option-implied co-kurtosis risk has a non-trivial correlation with its historical counterpart, as reported in Table 3.6. This correlation increases from 0.31 at the monthly horizon to 0.41 at the yearly horizon. Table 3.7 reports the results of the future risk predictability analysis. Compared to the results reported for co-skewness risk in Section 3.6.2, the option-implied co-kurtosis measure delivered by my methodology does not seem to significantly predict the future risk. Its performance is worse than that of the historical measure. This result can perhaps be imputable to the measure being noisier than its historical counterpart.

Multi-horizon pricing of co-kurtosis risk

I now turn to the analysis of the pricing of co-kurtosis in the cross-section of returns. This risk is proxied by the measure defined in (3.15). The t-statistics for the estimated prices of risk in the univariate regression are reported in Table 3.14. As observable from this table, the price of co-kurtosis is positive across all risk horizons when using the option-implied horizon specific expectations. This result is consistent with the theoretical prediction of Dittmar (2002b) and implies that on average, stocks that load strongly on the co-kurtosis factor tend to outperform stocks that load weakly or negatively on this factor. However, the estimates are rarely significant, even though the t-statistics increase when using long-horizon expectation. After controlling for other systematic factors, the signs of the estimates obtained using short-horizon expectation are reverted while the ones obtained using long-horizon information remain positive (see Table 3.15). As depicted in Figure 3.5, contrary to the results obtained for co-skewness risk, there is no clear evidence of a decreasing term structure of the risk premia for co-kurtosis risk.

3.7 Conclusion

This paper studies the multi-horizon pricing of co-skewness risk in the cross-section of stock returns and aims to provide a model-free insight on the term structure of co-skewness risk premia. To achieve this goal, I introduce a flexible term structure model that allows for the estimation of option-implied co-skewness risk at any maturity. The entire term structure of option-implied co-skewness is used in combination with Fama and MacBeth (1973) regressions to study the term structure of co-skewness risk premia across investment horizons.

Using both returns and options data on 96 stocks of the S&P 500 index, the analysis reveals that investors expect co-skewness risk to matter less over long-horizon and that the risk premia they demand for this risk is decreasing in (absolute value) with the investment horizon. Overall, my results suggest that market participants who are willing to earn the co-skewness risk premia maximize their earnings by keeping their investment position over a shorter time period rather than holding it over a longer time period.

This paper is the first to study multi-horizon pricing of co-skewness risk using options data. It contributes to the term structure literature by showing that co-skewness risk premia have a decreasing term structure (in absolute value), a result that is not implied by existing evidence on the term structure of equity risk premia (see, for example, van Binsbergen, Brandt, and Koijen (2012)) or variance risk premia (see for example Andries, Eisenbach, Thomas, Schmalz, Schmalz, and Wang (2015) and Dew-Becker, Giglio, Le, and Rodriguez (2015)).

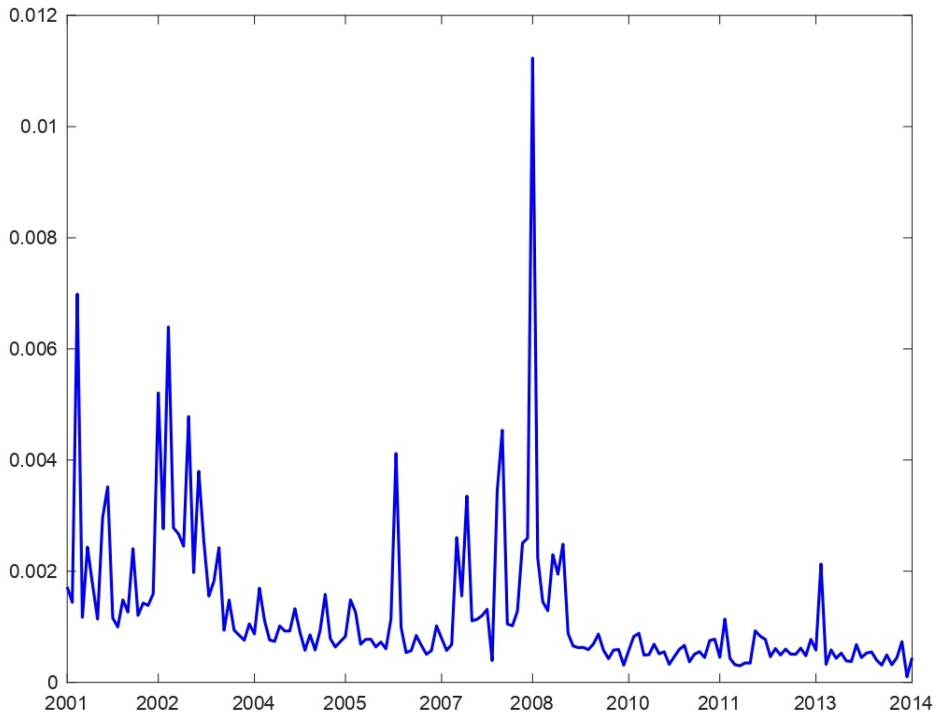
Figures and Tables of Chapter 3

Table 3.1: Daily Average Number of Index Options by Year and Maturity Buckets

Year	DTM $\leq$ 90	90 $<$ DTM $\leq$ 180	180 $<$ DTM $\leq$ 270	270 $<$ DTM $\leq$ 365	Average
2001	87.833	30.833	29.083	26.750	174.500
2002	85.083	27.333	29.000	27.750	169.167
2003	87.667	27.500	27.667	26.667	169.500
2004	94.583	33.583	31.500	29.750	189.417
2005	107.667	29.083	29.583	26.250	192.583
2006	141.917	31.167	26.500	23.500	223.083
2007	191.000	50.250	43.083	36.000	320.333
2008	250.333	75.167	53.833	53.750	433.083
2009	315.500	99.417	65.417	62.083	542.417
2010	335.083	103.917	76.000	77.583	592.583
2011	378.833	116.583	75.583	70.583	641.583
2012	486.583	141.083	90.667	92.333	810.667
2013	659.833	167.917	114.583	115.333	1057.667
2014	1259.833	220.917	122.083	124.750	1727.583
Average	320.125	82.482	58.185	56.649	517.440

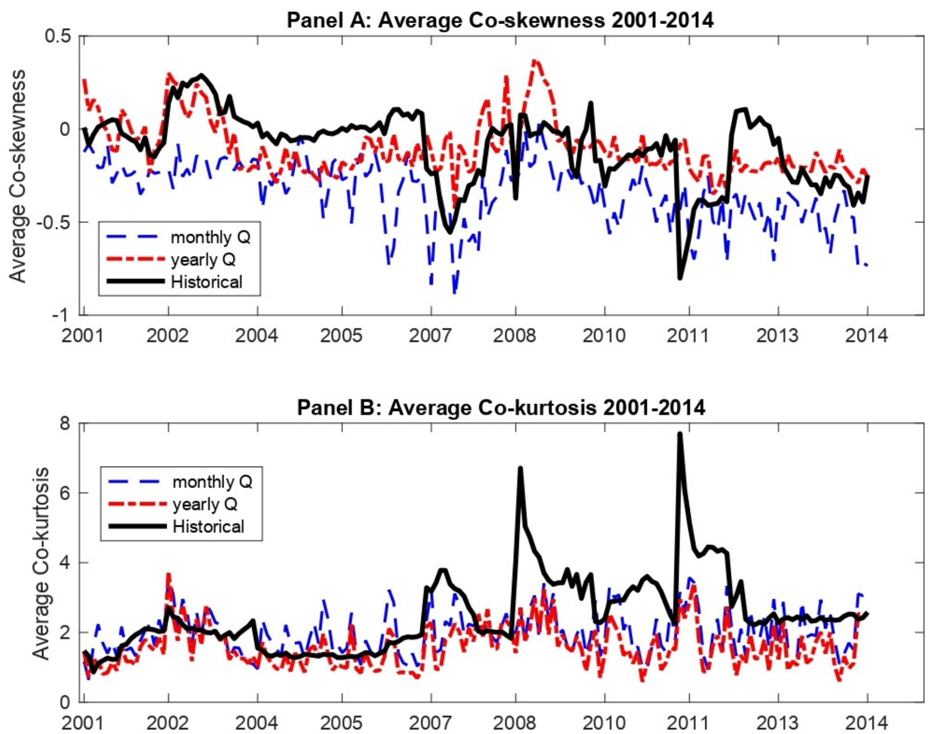
I categorize the index option data into four buckets according to the value on the days to maturity (DTM). For each class, I report the average number of option contracts.

Figure 3.1: Time-Series of the Coefficient of Variation of the Index Option RRMSE across stocks



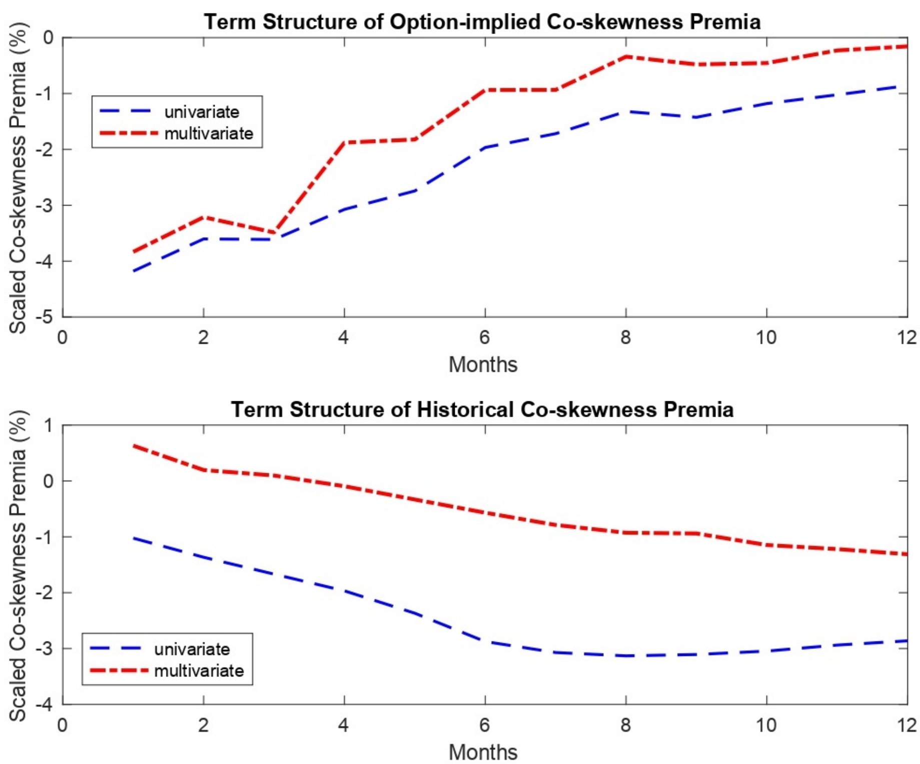
For each month, I compute the coefficient of variation of RRMSE of the model predicted market option across stocks.

Figure 3.2: Time-Series of Average Co-Skewness and Co-Kurtosis



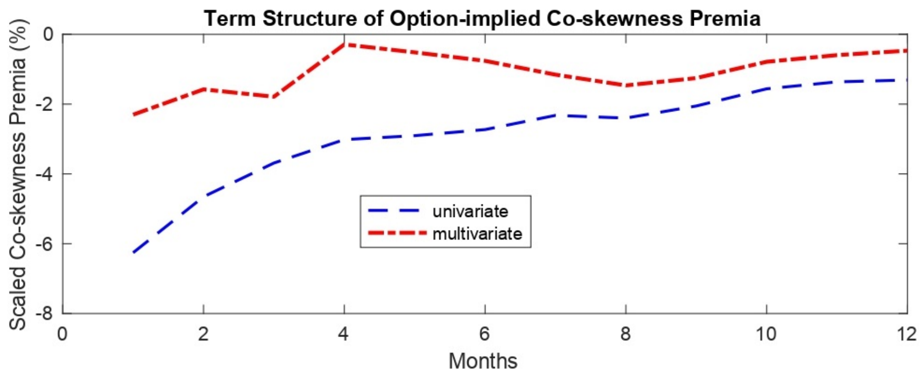
The dashed line represents the one-month option-implied measure (co-skewness or co-kurtosis), the dash-dotted line plots the one year measure, and lastly, the solid line depicts the time series of the historical measure.

Figure 3.3: Term Structure of Co-Skewness Risk



In the top box, each line plots (across months) the annualized product of the average coefficient of an option-implied co-skewness measure in a Fama and MacBeth (1973) regression and the standard deviation of the measure. The dashed line is obtained from the univariate regression. The dashed-dotted line is obtained after controlling for Carhart (1997) factors, the past monthly returns and the past yearly returns.

Figure 3.4: Term Structure of Co-Skewness Risk, robustness analysis



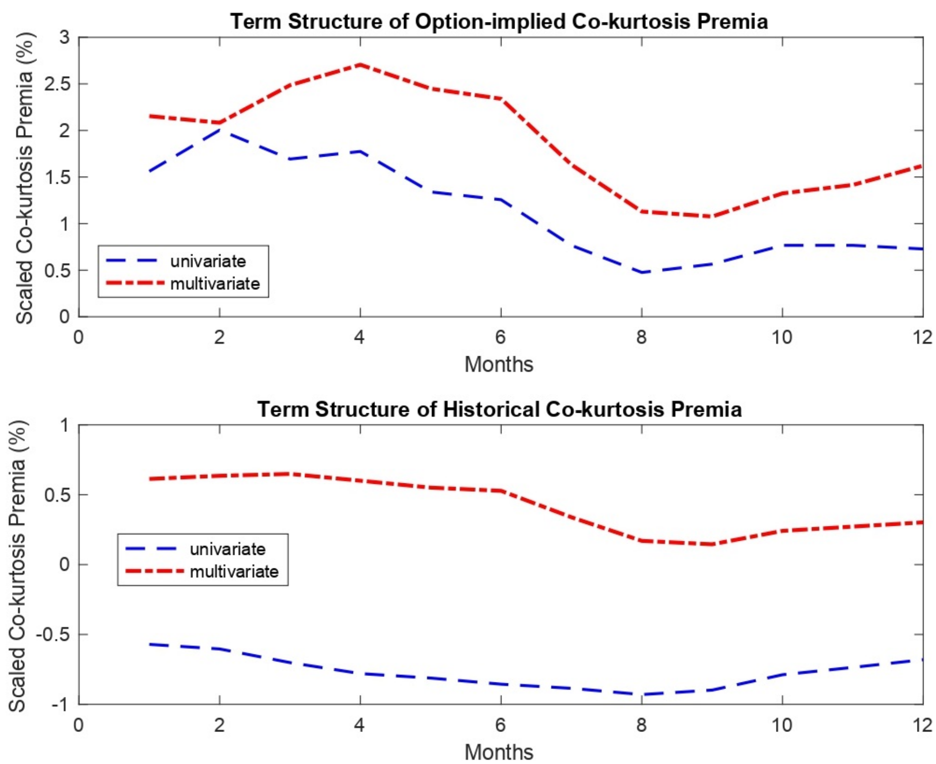
In the top box, each line plots (across months) the annualized product of the average coefficient of an option-implied co-skewness measure in a Fama and MacBeth (1973) regression and the standard deviation of the measure. The dashed line is obtained from the univariate regression. The dashed-dotted line is obtained after controlling for Carhart (1997) factors, the past monthly returns and the past yearly returns.

Table 3.2: Daily Average Number of Stock Options by Year and Maturity Buckets

Year	DTM $\leq$ 90	90 $<$ DTM $\leq$ 180	180 $<$ DTM $\leq$ 270	270 $<$ DTM $\leq$ 365	Average
2001	12.822	8.688	5.774	1.738	29.023
2002	12.178	8.912	5.909	1.641	28.640
2003	10.213	8.057	5.740	1.646	25.656
2004	9.911	8.252	6.383	1.914	26.459
2005	9.786	8.241	6.671	2.016	26.714
2006	10.982	9.239	7.257	1.979	29.457
2007	13.146	10.938	8.243	2.512	34.839
2008	17.212	13.549	9.403	2.882	43.046
2009	19.774	16.351	11.058	3.607	50.790
2010	22.105	18.181	12.224	3.331	55.840
2011	26.555	22.316	15.411	3.919	68.201
2012	25.656	24.560	19.012	4.447	73.675
2013	34.971	23.168	18.447	4.290	80.876
2014	65.375	20.955	16.988	5.181	108.499
Average	20.763	14.386	10.609	2.936	48.694

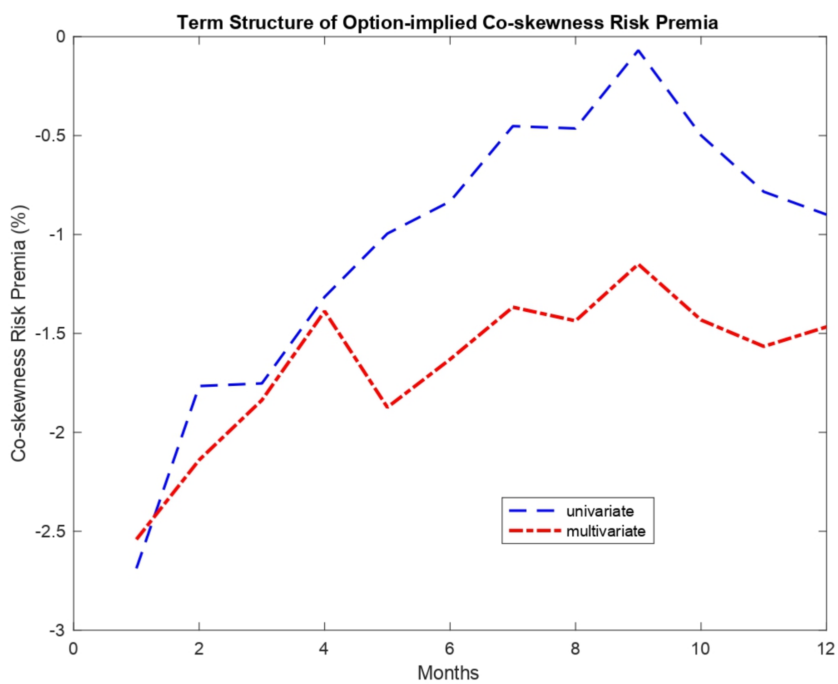
I categorize the stock option data into four buckets according to the value on the days to maturity (DTM). For each class, I report the average number of option contracts.

Figure 3.5: Term Structure of Co-Kurtosis Risk



In the top box, each line plots (across months) the annualized product of the average coefficient of an option-implied co-kurtosis measure in a Fama and MacBeth (1973) regression and the standard deviation of the measure. The dashed line is obtained from the univariate regression. The dashed-dotted line is obtained after controlling for Carhart (1997) factors, the past monthly returns and the past yearly returns.

Figure 3.6: Term Structure of Co-Skewness Risk



In the top box, each line plots (across months) the annualized product of the average coefficient of an option-implied co-skewness measure in a Fama and MacBeth (1973) regression and the standard deviation of the measure. The dashed line is obtained from the univariate regression. The dashed-dotted line is obtained after controlling for Carhart (1997) factors, the past monthly returns and the past yearly returns. Here the model is estimated using the combination (k_x, k_y, k_τ)=(5,5,2).

Table 3.3: Option Relative Root Mean Square Error By Year

Year	Index	Stock
2001	0.284	0.234
2002	0.234	0.231
2003	0.362	0.254
2004	0.462	0.277
2005	0.476	0.319
2006	0.503	0.316
2007	0.369	0.322
2008	0.307	0.336
2009	0.464	0.338
2010	0.552	0.394
2011	0.542	0.418
2012	0.589	0.470
2013	0.602	0.460
2014	0.673	0.441
All	0.458	0.343

The table reports the relative root mean square error (RRMSE) of the model predicted option price by year for both the index and the stocks (see definition in 3.11). The last row provides the RRMSE for the entire dataset. The value reported for the stocks is a cross-sectional average.

Table 3.4: Option Relative Root Mean Square Error By Maturity

Year	Index	Stock
DTM $\leq$ 90	0.592	0.466
90 $<$ DTM $\leq$ 180	0.530	0.384
180 $<$ DTM $\leq$ 270	0.500	0.340
270 $<$ DTM $\leq$ 365	0.465	0.370
All	0.458	0.343

The table reports the relative root mean square error (RRMSE) of the model predicted option price by maturity (DTM) buckets for both the index and the stocks.

Table 3.5: Descriptive Statistics for Beta, Co-Skewness and Co-Kurtosis

	Panel A: Market Beta				Panel B: Co-Skewness				Panel C: Co-Kurtosis			
h	P_{25}	Mean	P_{75}	Std	P_{25}	Mean	P_{75}	Std	P_{25}	Mean	P_{75}	Std
1	0.670	0.970	1.209	0.473	-0.490	-0.339	-0.176	0.246	1.311	1.835	2.600	0.755
2	0.656	0.928	1.145	0.434	-0.427	-0.288	-0.146	0.225	1.273	1.758	2.430	0.675
3	0.648	0.908	1.116	0.417	-0.378	-0.246	-0.119	0.210	1.238	1.724	2.420	0.659
4	0.643	0.897	1.101	0.409	-0.340	-0.214	-0.097	0.201	1.206	1.696	2.415	0.653
5	0.641	0.892	1.091	0.404	-0.314	-0.189	-0.079	0.196	1.176	1.664	2.389	0.644
6	0.639	0.889	1.087	0.401	-0.293	-0.168	-0.062	0.193	1.152	1.636	2.355	0.637
7	0.638	0.887	1.085	0.401	-0.277	-0.151	-0.046	0.193	1.130	1.614	2.307	0.636
8	0.635	0.886	1.086	0.401	-0.264	-0.137	-0.034	0.193	1.110	1.590	2.262	0.635
9	0.633	0.885	1.084	0.400	-0.253	-0.126	-0.025	0.193	1.087	1.564	2.214	0.631
10	0.628	0.883	1.083	0.400	-0.245	-0.116	-0.015	0.193	1.063	1.534	2.174	0.621
11	0.626	0.883	1.083	0.404	-0.237	-0.108	-0.009	0.193	1.040	1.518	2.126	0.628
12	0.622	0.881	1.085	0.406	-0.231	-0.101	-0.002	0.193	1.014	1.494	2.084	0.626
Hist	0.749	0.998	1.225	0.382	-0.233	-0.094	0.055	0.224	1.647	1.809	2.120	0.522

I report the average (Mean), the standard deviation (Std), the first quartile (P_{25}), and the third quartile (P_{75}) of the market beta. The co-skewness and the co-kurtosis measures for horizons h range from one to twelve months. The measures are estimated using the methodology outlined in Section 3.2 at the end of each month from 2001 to 2014. The statistics are computed across both the time series and cross-sectional dimensions. The last row of the Table presents the same statistics for the historical measures.

Table 3.6: Correlation between Option-Implied and historical measures

h	1	2	3	4	5	6	7	8	9	10	11	12
Corr(Coskew(h), HistCoskew)	0.533	0.546	0.543	0.532	0.515	0.495	0.477	0.460	0.446	0.433	0.421	0.410
Corr(Cokurt(h), HistCokurt)	0.307	0.325	0.344	0.356	0.364	0.369	0.374	0.379	0.386	0.392	0.399	0.406

I report the correlation between the option-implied high-order co-moments measures and their historical counterpart. The option-implied measures are computed for horizon h ranging from one to twelve months.

Table 3.7: Predicting Average Realized Co-Skewness and Co-Kurtosis

	Panel A: Co-Skewness					Panel B: Co-Kurtosis				
h	α	tstat (α)	β	tstat	AdjR^2	α	tstat (α)	β	tstat(β)	AdjR^2
1	-0.020	-0.582	0.251	2.681	0.038	2.401	8.503	0.094	0.694	-0.003
2	-0.014	-0.452	0.319	3.301	0.060	2.340	8.651	0.128	0.967	-0.000
3	-0.019	-0.665	0.353	3.507	0.068	2.302	8.730	0.152	1.150	0.002
4	-0.028	-1.096	0.360	3.506	0.067	2.286	8.828	0.164	1.237	0.003
5	-0.038	-1.577	0.356	3.446	0.065	2.283	8.950	0.169	1.269	0.004
6	-0.046	-2.057	0.348	3.366	0.062	2.287	9.068	0.170	1.272	0.004
7	-0.054	-2.532	0.335	3.260	0.058	2.294	9.187	0.169	1.256	0.004
8	-0.060	-2.981	0.319	3.114	0.053	2.301	9.309	0.168	1.238	0.003
9	-0.066	-3.379	0.303	2.974	0.048	2.309	9.421	0.167	1.218	0.003
10	-0.070	-3.718	0.288	2.835	0.043	2.318	9.532	0.164	1.188	0.003
11	-0.074	-3.998	0.275	2.712	0.039	2.327	9.629	0.162	1.158	0.002
12	-0.077	-4.222	0.264	2.605	0.036	2.332	9.709	0.161	1.143	0.002
Hist	-0.105	-6.118	-0.046	-0.548	-0.005	1.795	8.963	0.316	4.347	0.050

I regress the 12 months ahead (average) historical co-skewness (co-kurtosis) on the average of the horizon h option-implied historical co-skewness (co-kurtosis). The table reports the intercept (α), the slope (β), its t-statistics (tstat), and the adjusted R^2 of this regression. For comparative purposes, in the last row, I report the results obtained using the historical measures instead of the option-implied ones. The estimation is performed using the HAC procedure.

Table 3.8: T-statistics for average co-skewness price in univariate regression

(h, k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.621	-0.130	-0.209	0.010	0.240	0.254	0.263	0.287	0.363	0.242	0.254	0.042
2	-1.353	-0.772	-0.599	-0.306	-0.338	-0.306	-0.235	-0.197	-0.081	-0.263	-0.312	-0.495
3	-1.302	-0.951	-0.801	-0.505	-0.671	-0.551	-0.483	-0.417	-0.261	-0.555	-0.616	-0.786
4	-1.061	-0.961	-0.940	-0.676	-0.829	-0.625	-0.577	-0.454	-0.313	-0.630	-0.659	-0.801
5	-1.033	-1.006	-1.083	-0.792	-0.949	-0.673	-0.605	-0.433	-0.294	-0.587	-0.600	-0.722
6	-1.247	-1.279	-1.262	-1.000	-1.138	-0.833	-0.733	-0.551	-0.455	-0.655	-0.685	-0.790
7	-1.583	-1.686	-1.612	-1.315	-1.352	-1.023	-0.873	-0.709	-0.704	-0.808	-0.841	-0.922
8	-1.593	-1.819	-1.724	-1.440	-1.502	-1.149	-0.992	-0.850	-0.913	-0.912	-0.923	-0.965
9	-1.639	-1.976	-1.929	-1.654	-1.684	-1.361	-1.171	-1.038	-1.130	-1.072	-1.051	-1.049
10	-1.416	-1.947	-2.003	-1.670	-1.675	-1.399	-1.237	-1.108	-1.181	-1.095	-1.054	-1.004
11	-1.374	-1.924	-2.026	-1.701	-1.694	-1.433	-1.289	-1.096	-1.212	-1.071	-1.014	-0.913
12	-1.520	-2.045	-2.030	-1.692	-1.689	-1.419	-1.309	-1.045	-1.150	-0.972	-0.901	-0.731

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-skewness at horizon h inferred from the option prices (using the methodology outlined in Section 3.2). The co-skewness measure is a risk-neutral version of the one defined in (3.13). This table reports the Newey-West t-statistics (obtained using $k - 1$ lags) for the average co-skewness price.

Table 3.9: T-statistics for average co-skewness price with control variables

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.117	-0.506	-0.548	-0.375	-0.525	-0.482	-0.577	-0.389	-0.430	-0.552	-0.479	-0.673
2	-1.762	-1.461	-1.236	-0.805	-1.189	-1.155	-1.256	-1.029	-0.973	-1.126	-1.029	-1.193
3	-1.681	-1.630	-1.347	-0.855	-1.390	-1.259	-1.391	-1.183	-1.091	-1.331	-1.224	-1.359
4	-1.310	-1.439	-1.269	-0.799	-1.310	-1.080	-1.254	-1.036	-1.000	-1.255	-1.143	-1.265
5	-0.933	-1.239	-1.219	-0.649	-1.152	-0.808	-0.959	-0.764	-0.761	-1.020	-0.909	-1.035
6	-0.926	-1.327	-1.286	-0.652	-1.219	-0.843	-0.943	-0.772	-0.777	-0.962	-0.852	-0.968
7	-1.005	-1.428	-1.512	-0.786	-1.263	-0.906	-0.949	-0.825	-0.852	-0.966	-0.886	-0.995
8	-1.058	-1.604	-1.676	-0.958	-1.479	-0.952	-0.910	-0.765	-0.864	-0.932	-0.834	-0.961
9	-1.247	-1.803	-1.875	-1.205	-1.649	-1.152	-0.963	-0.777	-0.917	-0.951	-0.828	-0.939
10	-1.290	-1.865	-1.955	-1.123	-1.463	-1.064	-0.884	-0.599	-0.820	-0.825	-0.665	-0.764
11	-1.552	-2.041	-2.189	-1.200	-1.472	-1.090	-0.888	-0.452	-0.753	-0.727	-0.497	-0.560
12	-1.896	-2.294	-2.395	-1.218	-1.460	-1.132	-0.922	-0.343	-0.603	-0.551	-0.265	-0.238

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-skewness at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns. The co-skewness measure is a risk-neutral version of the one defined in (3.13). This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average co-skewness price.

Table 3.10: T-statistics for average co-skewness price with control variables and historical co-skewness

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.866	-0.109	-0.119	-0.007	-0.186	-0.169	-0.294	-0.191	-0.300	-0.402	-0.304	-0.477
2	-1.493	-1.035	-0.759	-0.395	-0.844	-0.868	-0.998	-0.857	-0.865	-1.004	-0.877	-1.022
3	-1.471	-1.255	-0.937	-0.486	-1.041	-0.955	-1.121	-1.019	-0.989	-1.213	-1.075	-1.196
4	-1.100	-1.133	-0.918	-0.444	-0.953	-0.769	-0.980	-0.881	-0.909	-1.141	-0.999	-1.114
5	-0.728	-0.990	-0.906	-0.313	-0.801	-0.521	-0.717	-0.634	-0.700	-0.928	-0.786	-0.914
6	-0.677	-1.068	-0.967	-0.304	-0.858	-0.568	-0.735	-0.656	-0.729	-0.886	-0.752	-0.877
7	-0.694	-1.142	-1.172	-0.417	-0.888	-0.649	-0.772	-0.724	-0.815	-0.906	-0.802	-0.924
8	-0.728	-1.326	-1.364	-0.599	-1.107	-0.721	-0.768	-0.677	-0.840	-0.885	-0.768	-0.909
9	-0.913	-1.522	-1.571	-0.842	-1.300	-0.942	-0.837	-0.685	-0.884	-0.896	-0.756	-0.886
10	-1.028	-1.592	-1.680	-0.792	-1.171	-0.881	-0.766	-0.506	-0.778	-0.760	-0.586	-0.703
11	-1.408	-1.774	-1.906	-0.873	-1.204	-0.915	-0.766	-0.364	-0.690	-0.642	-0.412	-0.490
12	-1.834	-2.060	-2.129	-0.918	-1.217	-0.972	-0.806	-0.280	-0.536	-0.461	-0.196	-0.191

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as left hand side and a right hand side populated by the forward-looking co-skewness at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns and the historical co-skewness. The co-skewness measure is a risk-neutral version of the one defined in (3.13). This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average price of the option-implied co-skewness.

Table 3.11: T-statistics for average co-skewness price in univariate regression, robustness analysis

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.493	-0.965	-1.444	-1.230	-1.450	-1.464	-1.328	-1.277	-1.392	-1.193	-1.168	-1.153
2	-0.883	-1.591	-1.532	-1.330	-1.652	-1.724	-1.655	-1.677	-1.821	-1.524	-1.498	-1.469
3	-1.164	-1.929	-1.641	-1.417	-1.719	-1.909	-2.046	-2.069	-2.139	-1.864	-1.929	-1.815
4	-1.240	-1.998	-1.684	-1.495	-1.812	-2.000	-2.173	-2.155	-2.150	-1.912	-1.973	-1.855
5	-1.290	-2.005	-1.684	-1.520	-1.904	-2.112	-2.257	-2.183	-2.113	-1.835	-1.905	-1.798
6	-1.471	-2.100	-1.684	-1.392	-1.861	-2.120	-2.236	-2.192	-2.094	-1.799	-1.906	-1.792
7	-1.628	-2.236	-1.774	-1.480	-1.954	-2.207	-2.245	-2.203	-2.163	-1.841	-1.901	-1.802
8	-1.684	-2.308	-1.844	-1.486	-1.943	-2.223	-2.250	-2.290	-2.282	-1.972	-2.005	-1.899
9	-1.853	-2.395	-2.005	-1.599	-2.023	-2.326	-2.282	-2.314	-2.329	-2.054	-2.045	-1.952
10	-1.793	-2.339	-1.935	-1.576	-2.014	-2.323	-2.266	-2.302	-2.279	-2.008	-1.955	-1.850
11	-1.870	-2.280	-1.938	-1.605	-2.013	-2.372	-2.339	-2.381	-2.363	-2.056	-1.975	-1.876
12	-2.056	-2.357	-2.110	-1.839	-2.120	-2.403	-2.415	-2.333	-2.348	-2.026	-1.919	-1.845

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-skewness at horizon h inferred from option prices (using the methodology outlined in Section 3.2). The co-skewness is proxied by the beta obtained in a regression of excess stock returns on the excess market returns and its square. This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average co-skewness price.

Table 3.12: T-statistics for average co-skewness price with control variables, robustness analysis

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	0.104	0.241	-0.354	-0.096	-0.532	-0.796	-1.385	-1.266	-1.327	-1.128	-0.800	-0.646
2	-0.204	-0.399	-0.841	-0.583	-1.182	-1.310	-2.042	-2.070	-2.143	-1.948	-1.611	-1.354
3	-0.165	-0.648	-1.006	-0.691	-1.266	-1.346	-2.094	-2.209	-2.229	-2.112	-1.999	-1.651
4	-0.119	-0.741	-1.174	-0.678	-1.274	-1.209	-1.956	-2.047	-2.073	-1.891	-1.807	-1.442
5	-0.346	-0.902	-1.269	-0.670	-1.320	-1.285	-1.799	-1.846	-1.798	-1.544	-1.428	-1.111
6	-0.448	-0.998	-1.200	-0.417	-1.141	-1.234	-1.668	-1.740	-1.700	-1.386	-1.266	-0.991
7	-0.512	-1.031	-1.157	-0.332	-1.027	-1.230	-1.547	-1.596	-1.646	-1.261	-1.117	-0.884
8	-0.602	-1.196	-1.228	-0.276	-0.956	-1.185	-1.536	-1.700	-1.787	-1.400	-1.242	-0.999
9	-0.815	-1.289	-1.370	-0.306	-0.973	-1.306	-1.520	-1.693	-1.848	-1.474	-1.289	-1.068
10	-0.654	-1.229	-1.292	-0.199	-0.821	-1.228	-1.424	-1.673	-1.788	-1.345	-1.140	-0.936
11	-0.578	-1.196	-1.348	-0.258	-0.805	-1.296	-1.539	-1.822	-1.976	-1.488	-1.245	-1.053
12	-1.092	-1.378	-1.500	-0.216	-0.662	-1.222	-1.698	-1.948	-2.050	-1.458	-1.162	-0.993

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-skewness at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns. The co-skewness is proxied by the beta obtained in a regression of excess stock returns on the excess market returns and its square. This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average co-skewness price.

Table 3.13: T-statistics for average co-skewness price with control variables and historical co-skewness, robustness analysis

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	0.323	0.268	-0.277	-0.123	-0.612	-1.060	-1.595	-1.616	-1.596	-1.380	-1.133	-0.941
2	0.008	-0.456	-0.876	-0.676	-1.309	-1.551	-2.246	-2.524	-2.567	-2.441	-2.118	-1.842
3	0.114	-0.691	-1.043	-0.743	-1.381	-1.498	-2.238	-2.674	-2.698	-2.646	-2.567	-2.143
4	0.171	-0.796	-1.244	-0.723	-1.396	-1.316	-2.099	-2.501	-2.501	-2.399	-2.362	-1.906
5	-0.035	-0.906	-1.299	-0.684	-1.475	-1.448	-1.960	-2.288	-2.141	-1.938	-1.849	-1.475
6	-0.121	-0.971	-1.194	-0.343	-1.245	-1.365	-1.829	-2.184	-2.052	-1.779	-1.682	-1.358
7	-0.153	-0.949	-1.100	-0.190	-1.079	-1.319	-1.665	-1.999	-1.959	-1.608	-1.471	-1.191
8	-0.270	-1.139	-1.198	-0.153	-0.992	-1.266	-1.656	-2.111	-2.102	-1.735	-1.566	-1.277
9	-0.449	-1.215	-1.308	-0.127	-0.961	-1.331	-1.626	-2.084	-2.128	-1.766	-1.553	-1.324
10	-0.342	-1.171	-1.258	-0.028	-0.789	-1.256	-1.507	-2.060	-2.044	-1.604	-1.352	-1.125
11	-0.276	-1.142	-1.326	-0.087	-0.757	-1.325	-1.640	-2.235	-2.235	-1.749	-1.448	-1.234
12	-0.874	-1.334	-1.502	-0.043	-0.580	-1.201	-1.919	-2.433	-2.294	-1.673	-1.305	-1.123

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-skewness at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns and the historical co-skewness. The co-skewness is proxied by the beta obtained in a regression of excess stock returns on the excess market returns and its square. This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average price of the option-implied co-skewness.

Table 3.14: T-statistics for average co-kurtosis price in univariate regression

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.247	0.095	0.175	0.195	0.390	0.486	0.392	0.042	-0.044	0.071	0.094	0.121
2	-0.001	0.428	0.493	0.501	0.684	0.694	0.569	0.279	0.236	0.376	0.403	0.388
3	0.098	0.570	0.625	0.647	0.821	0.741	0.552	0.335	0.311	0.471	0.462	0.403
4	0.189	0.637	0.682	0.704	0.788	0.692	0.472	0.303	0.306	0.472	0.431	0.377
5	0.240	0.666	0.726	0.752	0.793	0.684	0.445	0.296	0.297	0.464	0.433	0.383
6	0.355	0.778	0.742	0.792	0.802	0.698	0.455	0.306	0.284	0.473	0.444	0.394
7	0.475	0.903	0.857	0.893	0.866	0.749	0.474	0.300	0.305	0.506	0.479	0.449
8	0.436	0.901	0.820	0.904	0.859	0.750	0.462	0.287	0.293	0.480	0.468	0.443
9	0.542	1.003	0.888	0.946	0.877	0.764	0.491	0.300	0.308	0.485	0.483	0.461
10	0.539	1.067	0.928	0.974	0.865	0.759	0.469	0.283	0.287	0.465	0.479	0.452
11	0.586	1.163	0.982	1.016	0.904	0.809	0.508	0.298	0.315	0.497	0.508	0.474
12	0.692	1.320	1.077	1.075	0.934	0.844	0.527	0.340	0.379	0.541	0.535	0.489

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-kurtosis at horizon h inferred from option prices (using the methodology outlined in Section 3.2). The co-kurtosis measure is a risk-neutral version of the one defined in (3.15). This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average co-kurtosis price.

Table 3.15: T-statistics for average co-kurtosis price with control variables

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.171	-1.400	-0.924	-1.025	-0.585	-0.200	0.020	-0.539	-0.311	0.263	0.429	0.574
2	-1.038	-1.016	-0.540	-0.502	0.027	0.333	0.418	0.104	0.404	0.872	1.013	1.091
3	-0.933	-0.758	-0.346	-0.123	0.456	0.626	0.584	0.462	0.748	1.203	1.270	1.306
4	-0.935	-0.682	-0.227	0.136	0.557	0.726	0.617	0.528	0.842	1.307	1.353	1.395
5	-0.980	-0.648	-0.012	0.441	0.819	0.918	0.719	0.657	0.945	1.403	1.457	1.533
6	-0.684	-0.297	0.375	0.866	1.205	1.195	1.042	0.952	1.147	1.570	1.612	1.742
7	-0.263	0.102	0.844	1.317	1.520	1.443	1.241	1.071	1.174	1.577	1.617	1.882
8	-0.153	0.307	1.025	1.560	1.725	1.625	1.445	1.146	1.163	1.520	1.597	1.929
9	0.220	0.635	1.354	1.844	1.988	1.869	1.689	1.299	1.274	1.628	1.720	2.095
10	0.446	0.817	1.557	1.982	2.123	1.986	1.762	1.319	1.283	1.635	1.771	2.154
11	0.663	1.025	1.731	2.111	2.229	2.058	1.816	1.329	1.312	1.676	1.841	2.245
12	0.938	1.332	1.947	2.268	2.371	2.212	1.955	1.475	1.483	1.842	1.993	2.411

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-kurtosis at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns. The co-kurtosis measure is a risk-neutral version of the one defined in (3.15). This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average co-kurtosis price.

Table 3.16: T-statistics for average co-kurtosis price with control variables and historical co-kurtosis

(h,k)	1	2	3	4	5	6	7	8	9	10	11	12
1	-1.171	-1.400	-0.772	-0.803	-0.432	-0.145	0.014	-0.316	-0.171	0.138	0.205	0.272
2	-1.038	-1.016	-0.447	-0.380	0.019	0.216	0.254	0.056	0.206	0.426	0.461	0.486
3	-0.933	-0.758	-0.284	-0.091	0.310	0.382	0.331	0.236	0.364	0.568	0.564	0.568
4	-0.935	-0.682	-0.185	0.100	0.373	0.427	0.333	0.263	0.401	0.608	0.594	0.598
5	-0.980	-0.648	-0.010	0.321	0.544	0.535	0.380	0.324	0.444	0.647	0.637	0.652
6	-0.684	-0.297	0.306	0.627	0.802	0.700	0.543	0.470	0.532	0.718	0.703	0.725
7	-0.263	0.102	0.690	0.951	1.019	0.854	0.650	0.538	0.548	0.724	0.709	0.769
8	-0.153	0.307	0.837	1.127	1.162	0.973	0.765	0.586	0.550	0.703	0.705	0.785
9	0.220	0.635	1.108	1.333	1.349	1.142	0.910	0.672	0.606	0.758	0.763	0.855
10	0.446	0.817	1.275	1.431	1.443	1.221	0.954	0.686	0.614	0.765	0.790	0.891
11	0.663	1.025	1.422	1.525	1.514	1.272	0.990	0.693	0.629	0.787	0.824	0.938
12	0.938	1.332	1.605	1.643	1.619	1.382	1.077	0.771	0.718	0.871	0.902	1.025

At the end of each month between 2001 and 2014, a Fama and MacBeth (1973) regression is performed with the k months future excess returns as the left hand side and a right hand side populated by the forward-looking co-kurtosis at horizon h, the Carhart (1997) factors, the past monthly excess returns and the past yearly excess returns and the historical co-kurtosis. The co-kurtosis measure is a risk-neutral version of the one defined in (3.15). This table reports the Newey-West t-statistics (obtained using $k-1$ lags) for the average price of the option-implied co-kurtosis.

Appendix for Chapter 3

C Appendix

C.1 Individual Stock Option Valuation

From the classical risk-neutral valuation formula, we know that the price of European put options on the stock S with maturity τ and strike price κ is

$$P^S(\kappa,\tau) = e^{-r_f\tau}\mathbb{E}^{\mathbb{Q}}\Big((\kappa_{t+\tau} - S_{t+\tau})^+\Big), \tag{C.1}$$

where $\mathbb{Q}$ denotes the risk-neutral measure. The probability density of $S_{t+\tau}$ under $\mathbb{Q}$ is provided by (3.8).

The price of the put option can be obtained in closed form as follows,

$$\begin{aligned} P^S(\kappa,\tau) &= \kappa \exp^{-r_f\tau}\left[\Phi\big(d(\tau,\kappa)\big) - \sum_{k=1}^{2K_y}\frac{\delta_k(B,\tau)}{\sqrt{k}}H_{k-1}\big(d(\tau,\kappa)\big)\Phi\big(d(\tau,\kappa)\big)\right] \\ &- S_t \exp^{r_f\tau+\mu_S(\tau)}\left[\exp\big(V_I(\tau)^2/2\big)\Phi\big(d(\tau,\kappa)-V_S(\tau)\big) + \sum_{k=1}^{2Ky}\delta_k(B,\tau)J_k^*\big(d(\tau,\kappa)\big)\right], \end{aligned}$$

where

$$\begin{aligned} J_k^*\big(d(\tau,\kappa)\big) &= \frac{V_I(\tau)}{\sqrt{(k)}}J_{k-1}^*\big(d(\tau,\kappa)\big) - \frac{1}{\sqrt{k}}\exp\big(V_S(\tau)d(\tau,\kappa)\big)H_{k-1}\big(d(\tau,\kappa)\big)\Phi\big(d(\tau,\kappa)\big) \\ J_0^*\big(d(\tau,\kappa)\big) &= \exp\big(V_S(\tau)^2/2\big)\Phi\big(d(\tau,\kappa)-V_S(\tau)\big), \text{and} \\ d(\tau,\kappa) &= \frac{\log(\kappa/I)-\mu_S(\tau)}{V_S(\tau)} \end{aligned}$$

and $\Phi(.)$ is the cumulative CDF of the normal distribution, and $\delta_j(B,\tau) = \sum_{k=0}^{K_x}\alpha_{k,\bullet}(B,\tau)^\top A_j^{K_x}\alpha_{k,\bullet}(B,\tau)\Big/\Big(\sum_{k=0}^{K_y}\alpha_{k,\bullet}(B,\tau)^\top\alpha_{k,\bullet}(B,\tau)\Big)$.

C.2 Robustness Analysis

In the main text, the model is estimated using the following combination of orders for the sieve polynomials $(k_x, k_y, k_\tau) = (4, 4, 2)$. I now estimate the model using the combination $(k_x, k_y, k_\tau) = (5, 5, 2)$. Figure 3.6 displays the term structure of co-skewness risk premia. As evident from this graph, the term structure is decreasing in absolute value which corroborates the results obtained in the main text.

Bibliography

Aït-Sahalia, Y., and M. W. Brandt (2007): "Consumption and Portfolio Choice with Option-Implied State Prices," Discussion paper.

Aït-Sahalia, Y., J. Fan, R. Laeven, D. C. Wang, and X. Yang (2017): "The Estimation of Continuous and Discontinuous Leverage Effects," *Journal of the American Statistical Association*, 112, 1744–1758.

Aït-Sahalia, Y., J. Fan, and Y. Li (2013): "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency," *Journal of Financial Economics*, 109, 224–249.

Aït-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., I. Kalnina, and D. Xiu (2019): "High-Frequency Factor Models and Regressions," Discussion paper, The University of Chicago.

Aït-Sahalia, Y., and A. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

Ait-Sahalia, Y., and A. W. Lo (1998): "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

Ait-Sahalia, Y., and D. Xiu (2017): "Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-Frequency Data," *Journal of Econometrics*, 201, 384–399.

——— (2019): "Principal Component Analysis of High Frequency Data," *Journal of the American Statistical Association*, 114, 287–303.

Aldous, D. G., and G. K. Eagleson (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331.

Aliouchkin, R. L. (2015): "Option-Implied Idiosyncratic and Systematic Risk in the Cross-Section of Expected Stock Returns," Discussion paper.

Alizadeh, S., M. Brandt, and F. Diebold (2002): "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047–1091.

Andersen, T., N. Fusari, and V. Todorov (2015): "Parametric Inference and Dynamic State Recovery from Option Panels," *Econometrica*, 83, 1081–1145.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and J. Wu (2006): "Realized Beta: Persistence and Predictability," in *Econometric Analysis of Economic and Financial Time Series*, ed. by T. Fomby, and D. Terrell, vol. 20 of *Advances in Econometrics*, pp. 1–40. Elsevier.

Andries, M., Eisenbach, Thomas, Schmalz, M. C. Schmalz, and Y. Wang (2015): "The term structure of the price of variance risk," Discussion paper, Federal Reserve Bank of New York.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The cross-section of volatility and expected returns," *Journal of Finance*, 61(1), 259–299.

Ángel Léon, J. Mencía, and E. Sentana (2009): "Parametric properties of semi-nonparametric distributions, with applications to option valuation," *Journal of Business and Economics Statistics*, 27, 176–192.

Babaoglu, K., P. Christoffersen, S. Heston, and K. Jacobs (2018): "Option Valuation with Volatility Components, Fat Tails, and Non-Monotonic Pricing Kernels," *Review of Asset Pricing Studies*, 8.

Bakshi, G., N. Kapadia, and D. Madan (2003): “Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options,” *Review of Financial Studies*, 16, 101–143.

Bakshi, G., and D. Madan (2006): “A Theory of Volatility Spreads,” *Journal of Management Science*, 52, No. 12, 1945–1956.

Bali, T. G., J. Hu, and S. Murray (2016): “Option Implied Volatility, Skewness, and Kurtosis and the Cross-Section of Expected Returns,” Discussion paper.

Bandi, F., and R. Renò (2012): “Time-Varying Leverage Effects,” *Journal of Econometrics*, 12, 94–113.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): “Designing Realized Kernels to Measure Ex-post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E., and N. Shephard (2004): “Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics,” *Econometrica*, 72, 885–925.

Bates, D. (2000): “Post-87 Crash Fears in S&P500 Futures Options,” *Journal of Econometrics*, 94, 181–238.

Bégin, J. F., C. Dorion, and G. Gauthier (2017): “Idiosyncratic Jump Risk Matters: Evidence from Equity Returns and Options,” Discussion paper.

Black, F. (1976): “Studies of Stock Price Volatility Changes,” in *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, pp. 177–181.

Black, F., and M. Scholes (1973a): “Parametric properties of semi-nonparametric distributions, with applications to option valuation,” *The Pricing of Options and Corporate Liabilities*, 81, 637–654.

Black, F., and M. Scholes (1973b): “The Pricing of Options and Corporate Liabilities,” *Journal of Political Economy*, 81(3), 637–654.

Bliss, R., and N. Panigirtzoglou (2004): "Option-Implied Risk Aversion Estimates," *Journal of Finance*, 59, 407–46.

Bollerslev, T., and V. Todorov (2010): "Jumps and Betas: A New Framework for Disentangling and Estimating Systematic Risks," *Journal of Econometrics*, 157, 157, 220235.

Boloorforoosh, A., P. Christoffersen, M. Fournier, and C. Gouriéroux (2017): "Beta Risk in Cross-Section of Equities," Discussion paper.

Breeden, D., and R. Litzenberger (1978): "Prices of state-contingent claims implicit in option prices," *Journal of Finance*, 51, 621–651.

Broadie, M., M. Chernov, and M. Johannes (2007): "Model Specification and Risk Premia: Evidence from Futures Options," *Journal of Finance*, 62(3), 1453–1490.

Buss, A., and G. Vilkov (2012): "Measuring Equity Risk with Option Implied- Correlations," *Review of Financial Studies*, 25, 3113–3140.

Campbell, J. Y., M. Lettau, B. G. Malkiel, and Y. Xu (2001): "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56(1), 1–43.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Chabi-Yo, F. (2002): "Nonlinear pricing kernels," *Management Science*, 58, 624–640.

Chabi-Yo, F., and Z. Song (2013): "Recovering the Probability Weights of Tail Events with Volatility Risk from Option Prices," Discussion paper.

Chen, Z., and R. Petkova (2012): "Does idiosyncratic volatility proxy for risk exposure?," *Reviews of Financial Studies*, 25(9), 2745–2787.

Chernov, M., R. Gallant, E. Ghysels, and G. Tauchen (2003): "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225–257.

Christoffersen, P., M. Fournier, and K. Jacobs (2018): "The Factor Structure in Equity Options," *Review of Financial Studies*, 31, 595–637.

Christoffersen, P., M. Fournier, K. Jacobs, and M. Karoui (2016): “Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk,” Discussion paper.

Christoffersen, P., M. Fournier, K. Jacobs, and M. Karoui (2017): “Option-Based Estimation of the Price of Co-Skewness and Co-Kurtosis Risk,” *Working paper.*

Christoffersen, P., S. Heston, and K. Jacobs (2006a): “Option Valuation with Conditional Skewness,” *Journal of Econometrics*, 131, 253–284.

——— (2006b): “Option Valuation with Conditional Skewness,” *Journal of Econometrics*, 131, 253–284.

——— (2013): “Capturing Option Anomalies with a Variance-Dependent Pricing Kernel,” *Review of Financial Studies*, 26, 1963–2006.

Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang (2008): “Option valuation with long-run and short-run volatility components,” *Journal of Financial Economics*, 90, 272–297.

Christoffersen, P. F., C. Dorion, K. Jacobs, and Y. Wang (2010): “Volatility Components, Affine Restrictions and Non-Normal Innovations,” *Journal of Business and Economic Statistics*, 28, 483–502.

Chung, P., H. Johnson, and M. J. Schill (2006): “Asset Pricing When Returns Are Nonnormal: Fama-French Factors versus Higher-Order Systematic Comoments,” *The Journal of Business*, 79, No. 2, 923–940.

Connor, G., R. A. Korajczyk, and O. Linton (2006): “The common and specific components of dynamic volatility,” *Journal of Econometrics*, 132(1), 231–255.

Conrad, J., R. F. Dittmar, and E. Ghysels (2013): “Ex Ante Skewness and Expected Stock Returns,” *Journal of Finance*, 68, 85–124.

Corradi, V., and W. Distaso (2006): “Semi-Parametric Comparison of Stochastic Volatility Models using Realized Measures,” *Review of Economic Studies*, 73, 635–667.

Deck, C., and H. Schlesinger (2014): “Consistency of Higher Order Risk Preferences,” *Econometrica*, 82, No. 5, 1913–1943.

Denuit, M., and B. Rey (2010): "Prudence, temperance, edginess, and risk apportionment as decreasing sensitivity to detrimental changes," *Mathematical Social Sciences*, 60, 137–143.

Dew-Becker, I., S. Giglio, A. Le, and M. Rodriguez (2015): "The price of variance risk," Discussion paper.

Diebold, F., and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Diebold, F. X., and K. Yilmaz (2014): "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, 182, 119–132.

Dittmar, R. F. (2002a): "Nonlinear pricing kernels," *Journal of Finance*, 57, 369–403.

——— (2002b): "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns," *Journal of Finance*, 57, 369–403.

Dorion, C. (2016): "Option Valuation with Macro-Finance Variables," *Journal of Financial and Quantitative Analysis*, 51, 1359–1389.

Duan, J.-C. (1999): "Conditionally Fat-Tailed Distributions and the Volatility Smile in Options," *manuscript, Hong Kong University of Science and Technology.*

Duarte, J., A. Kamara, S. Siegel, and C. Sun (2014): "The Systematic Risk of Idiosyncratic Volatility," Discussion paper.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68(6), 1343–1376.

Eeckhoudt, and Schlesinger (2006): "Putting Risk in tis proper place," *American Economic Review*, 96 No 1, 280–289.

Ekern, S. (1980): "Increasing Nth degree risk," *Economics Letters*, 6, 329–333.

Engle, R. F., and G. Lee (1999): "A Permanent and Transitory Component Model of Stock Return Volatility," in *Cointegration, Causality, and Forecasting: a Festschrift in Honor of*

Clive W.J. Granger, ed. by R. Engle, and H. White, pp. 475–497. Oxford University Press, New York, NY.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and J. D. MacBeth (1973): "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Fan, J., A. Furger, and D. Xiu (2016): "Incorporating Global Industrial Classification Standard into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator with High Frequency Data," *Journal of Business and Economics*, 34, 489–503.

Fenton, V. M., and A. R. Gallant (1996): "Qualitative and asymptotic performance of SNP density estimators," *Journal of Econometrics*, 74, 77–118.

French, D. W., J. C. Groth, and J. W. Kolari (1983): "Journal of Portfolio Management," *Current Investor Expectations and Better Betas*, 10, 12–18.

Fu, F. (2009): "Idiosyncratic risk and the cross-section of expected stock returns," *Journal of Financial Economics*, 91(1), 24–37.

Gallant, A. R., and D. W. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Gormsen, N. J., and C. S. Jensen (2018): "Higher-Moment Risk," Working paper.

Gourier, E. (2016a): "Pricing of Idiosyncratic Equity and Variance Risks," Discussion paper, Queen Mary University.

——— (2016b): "Pricing of Idiosyncratic Equity and Variance Risks," Discussion paper.

Gouriéroux, C., and A. Monfort (2017): "Pairwise analysis of high dimensional stochastic volatility with market benchmark," Discussion paper.

Hansen, L. P., and R. Jagannathan (1991): "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99(2), 225–262.

Hansen, L. P., and K. J. Singleton (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational expectation models," *Econometrica*, 50 No. 5, 1269–1285.

Hansen, P. R., A. Lunde, and V. Voev (2014): "Realized Beta GARCH: a Multivariate GARCH Model with Realized Measures of Volatility," *Journal of Applied Econometrics*, 29, 774–799.

Harvey, C. R., and A. Siddique (2000a): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55(3), 1263–1295.

——— (2000b): "Conditional Skewness in Asset Pricing Tests," *Journal of Finance*, 55, 1263–1295.

Herskovic, B., B. Kelly, H. Lustig, and S. V. Nieuwerburgh (2016): "The Common Factor in Idiosyncratic Volatility: Quantitative Asset Pricing Implications," *Journal of Financial Economics*, 119, 249–283.

Heston, S., and S. Nandi (2000): "A Closed-Form GARCH Option Valuation Model," *The Review of Financial Studies*, 13, 585–626.

Hsieh, K. C., and P. Ritchken (2005): "An Empirical Comparison of GARCH Option Pricing Models," *Review of Derivatives Research*, 8(3), 129–150.

Israelovr, R., and B. T. Kelly (2017): "Forecasting the Distribution of Option Returns," Discussion paper.

Jackwerth, J. C. (2000): "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13, 433–451.

Jackwerth, J. C., and M. Rubinstein (1996): "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51, 1611–1631.

Jacod, J., and P. Protter (2012): *Discretization of Processes*. Springer.

Jacod, J., and M. Rosenbaum (2013): "Quarticity and Other Functionals of Volatility: Efficient Estimation," *Annals of Statistics*, 41, 1462–1484.

Jacod, J., and M. Rosenbaum (2015): "Estimation of Volatility Functionals: the Case of a $\sqrt{n}$ Window," in *Large Deviations and Asymptotic Methods in Finance*, ed. by P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, vol. 110, pp. 559–590.

Kalnina, I., and D. Xiu (2017): "Nonparametric Estimation of the Leverage Effect: A Trade-off between Robustness and Efficiency," *Journal of the American Statistical Association*, 112, 384–396.

Kanaya, S., and D. Kristensen (2016): "Estimation of Stochastic Volatility Models by Nonparametric Filtering," *Econometric Theory*, 32, 861–916.

Kimball, M. (1993): "Standard Risk Aversion," *Econometrica*, 61, No 3, 589–611.

Kraus, A., and R. H. Litzenberger (1976): "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, 31, 1085–1100.

Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-Based Approach," *Econometric Theory*, 26.

Lajeri, F., and Lajeri (2004): "Proper prudence, standard prudence and precautionary vulnerability," *Economic Letters*, 82, 29–34.

Li, J., Y. Liu, and D. Xiu (2019): "Efficient Estimation of Integrated Volatility Functionals via Multiscale Jackknife," *Annals of Statistics*, 47, 156–176.

Li, J., and A. J. Patton (2018): "Asymptotic Inference about Predictive Accuracy Using High Frequency Data," *Journal of Econometrics*, 232, 223–240.

Li, J., V. Todorov, and G. Tauchen (2013): "Inference Theory on Volatility Functional Dependencies," Discussion paper, Duke University and Northwestern University.

——— (2016): "Inference Theory on Volatility Functional Dependencies," *Journal of Econometrics*, 193, 17–34.

——— (2017a): "Adaptive Estimation of Continuous-Time Regression Models using High-Frequency Data," *Journal of Econometrics*, 200, 36–47.

——— (2017b): "Jump Regressions," *Econometrica*, 85, 173–195.

Lintner, J. (1965a): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

Lintner, J. (1965b): "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47(1), 13–37.

Liu, L., and J. Meyer (2013): "Substituting one risk increase for another: A method for measuring risk aversion," *Journal of Economic Theory*, 148, 2706–2718.

Liu, L., A. Patton, and K. Sheppard (2015): "Does Anything Beat 5-Minute RV? A Comparison of Realized Measures Across Multiple Asset Classes," *Journal of Econometrics*, 187, 293–311.

Mehra, R., and E. C. Prescott (1985): "The Equity Premium, a Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Mykland, P. A., and L. Zhang (2006): "ANOVA for diffusions and Itô processes," *Annals of Statistics*, 34(4), 1931–1963.

Pelger, M. (2019a): "Large-Dimensional Factor Modeling Based on High-Frequency Observations," *Journal of Econometics*, 208, 23–42.

——— (2019b): "Understanding Systematic Risk - A High-Frequency Approach," Discussion paper, Stanford University.

Renault, E. (1997): "Econometric Models of Option Pricing Errors," in *Advances in Economics and Econometrics*, ed. by D. Kreps, and K. Wallis, pp. 223–278. Econometric Society Monographs, Cambridge University Press.

Rosenberg, J. V., and R. F. Engle (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64, 341–372.

Ross, S. A. (1976): "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–376.

Rubinstein, M. (1994): "Implied Binomial Trees," *Journal of Finance*, 49, 771–818.

Schneider, P., and F. Trojani (2019): "(Almost) Model-Free Recovery," *Journal of Finance*, 74-1, 323–370.

Scott, R., and P. A. Horvath (1980a): "On the Direction of Preference for Moments of Higher Order Than the Variance," *Journal of Finance*, 35, 915–919.

Scott, R. C., and P. A. Horvath (1980b): "On the direction of the preference for moments of higher order than the variance," *Journal of Finance*, 35, No. 4, 1913–1943.

Sharpe, W. F. (1964a): "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19(3), 425–442.

Sharpe, W. F. (1964b): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–42.

Song, Z., and D. Xiu (2016): "A tale of two option markets: Pricing kernels and volatility risk," *Journal of Econometrics*, 190, 176–196.

Tédongap, R. (2014): "Consumption Volatility and the Cross-Section of Stock Returns," *Review of Finance*, 58, 1–36.

van Binsbergen, J., M. Brandt, and R. Koijen (2012): "On the Timing and Pricing of Dividends," *American Economic Review*, 102(4), 1596–1618.

Vetter, M. (2015): "Estimation of Integrated Volatility of Volatility with Applications to Goodness-of-fit Testing," *Bernoulli*, 21, 2393–2418.

Vogt, E. (2015): "Option-Implied Term Structures," Discussion paper, Federal Reserve Bank of New York Staff Reports.

Wang, D. C., and P. A. Mykland (2014): "The Estimation of Leverage Effect with High Frequency Data," *Journal of the American Statistical Association*, 109, 197–215.