ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

SCIENZE STATISTICHE

Ciclo 37

Settore Concorsuale: 13/D4 - METODI MATEMATICI DELL'ECONOMIA E DELLE SCIENZE ATTUARIALI E FINANZIARIE

Settore Scientifico Disciplinare: SECS-S/06 - METODI MATEMATICI DELL'ECONOMIA E DELLE SCIENZE ATTUARIALI E FINANZIARIE

Essays on Latent Factor Models in Financial Markets: A Focus on Commodity Prices and Volatility Dynamics

Presentata da: Christian Tezza

Coordinatore Dottorato
Angela Montanari

Supervisore
Luca Vincenzo Ballestra

Co-Supervisore
Enzo D'Innocenzo

Esame finale anno 2025

Abstract

Alma Mater Studiorum - Università di Bologna

Department of Statistical Sciences "Paolo Fortunati"

Doctor of Philosophy in Statistical Sciences

Essays on Latent Factor Models in Financial Markets:
A Focus on Commodity Prices and Volatility Dynamics

by Christian TEZZA

This PhD dissertation investigates financial applications of factor models, or latent variable models, with a focus on model specifications that link observable processes to multiple underlying drivers in two distinct contexts: the modeling of volatility and commodity prices.

Empirical evidence indicates that the conditional volatility of financial assets is more accurately captured by models that incorporate multiple stochastic components. Numerous studies have demonstrated the superiority of two-factor models over single-factor models in effectively capturing the complex dynamics of volatility. The generalized autoregressive conditional heteroskedasticity (GARCH) model, originally proposed by Bollerslev (1986), is one of the most popular models for filtering volatility in discrete time. Despite recent advancements in the literature of two-factor GARCH models, theoretical properties, such as ergodicity and strict stationarity, and further rigorous empirical testing are still lacking. This dissertation contributes to the literature by proposing a novel two-factor GARCH model.

Regarding the modeling of commodity prices, it is well established that the uncertainty inherent in commodity price dynamics is a complex phenomenon, making it challenging to achieve accurate representations of reality and reliable future predictions. Consequently, there has been significant interest in models capable of capturing the stylized facts of different commodity prices. Among these, multi-factor affine models have been particularly valued for their analytical tractability and ease of estimation using the Kalman filter. This thesis demonstrates their practical relevance through a novel empirical application, highlighting their potential to address the complexities associated with commodity price dynamics.

Acknowledgements

I dedicate this thesis to my family.

Christian Tezza

Bologna, October 2024

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, there has been a significant increase in the use of latent factor modeling for financial purposes. This growth is driven by easier access to large amounts of data and advances in computing power, which enable the development of complex models to analyze extensive financial time series data. However, creating and fine-tuning latent variable models can be challenging, as their specification and estimation are often difficult to establish. Additionally, with so many models available for financial time series, it is crucial to compare the results from different models to identify the most effective specifications.

In this doctoral thesis, we focus on two financial topics: market volatility and commodity prices. The primary goal is to offer new approaches and methods in factor modeling that could be useful for analyzing complex financial data more effectively.

1.1 Overview

The modeling of equity returns volatility in this thesis builds on the successful Heston-Nandi model, developed in Heston and Nandi (2000), which provides a semi-analytical formula for computing European option prices and captures key empirical stylized features. This work contributes to the GARCH literature by proposing two distinct models. The first is a novel GARCH

model that integrates two sources of uncertainty to capture the complex, multi-component dynamics observed in financial asset volatility. The second is an improved GARCH model, building on the two-component approaches of Christoffersen et al. (2008) and Oh and Park (2023), ensuring the positivity of return volatility both theoretically and empirically. For both models, we provide a detailed statistical analysis, supported by theoretical and empirical results.

The dynamics of commodities are influenced by several latent factors, as discussed in, e.g., Schwartz (1997), Hughen (2010), and Schöne and Spinler (2017). State-space modeling is a widely used approach in this context, particularly suited for capturing these underlying factors. Within a linear framework that assumes Gaussian random errors, state-space models can be efficiently handled using the Kalman filter, originally introduced by Kalman (1960). By utilizing the Kalman filter, we model commodity prices based on multiple latent factors and conduct rigorous empirical tests to evaluate the model's effectiveness and performance.

1.2 Main Contributions of the Thesis

The first two chapters of this thesis are devoted to the analysis of financial volatility modeling. We propose two innovative nonlinear time series models of the GARCH type that are particularly useful for revealing the complex behavior of volatility. Our approach builds on the seminal work in the Heston-Nandi model class, which introduces asymmetric nonlinear GARCH models. As discussed later in this thesis, a distinctive feature of this class of models is their ability to provide a semi-closed form formula for pricing European options. This area has been explored extensively in recent years, as evidenced by works such as Christoffersen et al. (2008), Christoffersen et al. (2012), Ghanbari (2024), and Wang et al. (2024). We contribute to the existing literature by extending some of these models to a more general framework and providing both theoretical and empirical justifications. Our focus on GARCH component models is motivated by their ability to address the challenges of nonlinear time series modeling with a high degree of generality and clarity.

The last chapter of this thesis addresses the modeling of commodity prices. There has been

significant development in financial products within commodity markets and the mathematical modeling behind such transactions over the last 20 years. Oil futures, in particular, have become the most heavily traded commodity globally. There is a continuing interest in understanding how the prices of oil futures evolve over time and across maturities, and which factors influence their dynamics. In this work, we contribute to the existing literature, as highlighted by Schwartz (1997), Hughen (2010), and Schöne and Spinler (2017), by proposing a novel multi-factor model. The primary advantage of the proposed specification over previous models is the development of a joint framework that enables a more accurate estimation of the underlying factors influencing commodity prices. Consequently, this framework also facilitates the study of dependencies between futures and other asset classes, such as bond yields.

1.3 Structure of the Thesis

- *Chapter 2*: (based on a joint work with L.V. Ballestra and E. D'Innocenzo)
  We introduce a novel GARCH model that integrates two sources of uncertainty to better account for the rich and multi-component dynamics often observed in the volatility of financial assets. This model offers a quasi closed-form representation of the characteristic function for future log-returns, from which semi-analytical formulas for option pricing can be obtained. A theoretical analysis is conducted to establish sufficient conditions for strict stationarity and geometric ergodicity, and to derive the continuous-time diffusion limit. Moreover, we perform both in-sample and out-of-sample empirical analyses using the S&P500 time series, and we show that the proposed model outperforms popular benchmark models on both returns and option prices.

- *Chapter 3*: (based on a joint work with L.V. Ballestra and E. D'Innocenzo)
  Christoffersen et al. (2008) proposed an improved Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model for valuing European options, in which the volatility of returns comprises two different components. They provided evidence suggesting that their model outperforms the single-component GARCH model developed by Heston and

Nandi (2000) and fits options data better than an approach combining conditional heteroskedasticity and Poisson-normal jumps. However, as demonstrated in Bormetti et al. (2015) and Oh and Park (2023), the modeling specification proposed by Christoffersen et al. (2008) permits the variance process to become negative. Oh and Park (2023) partially addressed this issue by developing a closely related model to the one proposed in Christoffersen et al. (2008), but the positivity of the volatility components is still not guaranteed, either theoretically or experimentally. The present paper proposes a new model that improves the models proposed by Christoffersen et al. (2008) and by Oh and Park (2023) and guarantees the positivity of the volatility of log-returns. Compared to the approaches in Christoffersen et al. (2008) and Oh and Park (2023), the novel methodology demonstrates similar in-sample performance on returns data and superior performance on S&P500 options data.

- *Chapter 4*: (based on a joint work with L.V. Ballestra)

  We propose a novel multi-factor model for commodity pricing that accounts for four distinct risk factors, including stochastic volatility, convenience yield, and stochastic interest rates. The model accommodates key stylized facts of commodity prices, such as dynamic correlation structure and time-varying risk premiums. We employ a Kalman filter-based estimation framework, which enables simultaneous estimation of model parameters and filtering of state variables using a joint term structure of futures prices and bond yields. Through an empirical analysis of crude oil futures, we benchmark our model against popular models. Our findings demonstrate that the proposed four-factor model effectively captures the complexities of futures term structures and outperforms existing models. This advancement is expected to enhance risk management strategies and support informed decision-making in commodity markets.

Contribution statement

C. Tezza: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **L.V. Ballestra**: Conceptualization,

Methodology, Formal analysis, Investigation, Writing – review & editing, Supervision. **E. D'Innocenzo**: Conceptualization, Methodology, Writing – review & editing, Supervision, Data curation.

Chapter 2

A GARCH model with two volatility components and two driving factors

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name

Department of Statistical Sciences

"Paolo Fortunati"

Doctor of Philosophy in Statistical Sciences

Essays on Latent Factor Models in Financial Markets:

A Focus on Commodity Prices and Volatility Dynamics

by Christian TEZZA

We introduce a novel GARCH model that integrates two sources of uncertainty to better capture the rich, multi-component dynamics often observed in the volatility of financial assets. This model provides a quasi closed-form representation of the characteristic function for future log-returns, from which semi-analytical formulas for option pricing can be derived. A theoretical analysis is conducted to establish sufficient conditions for strict stationarity and geometric ergodicity, while also obtaining the continuous-time diffusion limit of the model. Empirical evaluations, conducted both in-sample and out-of-sample using S&P500 time series data, show that our model outperforms widely used single-factor models in predicting returns and option prices. The code for estimating the model, as well as for computing option prices, is made accessible in MATLAB language[1].

[1]The model code is available at the GitHub repository: github.com/tezzachris/GARCH

2.1 Introduction

Empirical evidence indicates that the conditional volatility of financial assets is influenced by several stochastic components. Researchers have demonstrated the superiority of employing two factors over a single factor in capturing and modeling the rich and multifaceted dynamics of volatility. For example, Alizadeh et al. (2002), Bollerslev and Zhou (2002), and de Magistris and Grassi (2015) support stochastic volatility specifications that incorporate two sources of uncertainty, highlighting the inadequacy of a simple one-factor volatility model to fully account for the dynamic dependencies observed in the daily volatility of exchange rates. Additionally, Gallant et al. (1999) and Chernov et al. (2003) find that two-component volatility specifications provide higher accuracy than single-factor models when fitting equity market data and conclude that at least two components are necessary to capture the dynamics of volatility.

In addition, Bates (1996), Taylor and Xu (1994), Fouque et al. (2004), Schwartz and Trolle (2009a), and Fouque and Lorig (2011) demonstrate that the introduction of a second volatility factor is crucial for capturing the behavior of the volatility surface implied by options. Moreover, Bates (2000) and Christoffersen et al. (2009) propose two-factor stochastic volatility models that significantly outperform closely related one-factor models both in-sample and out-of-sample when applied to equity options data. Furthermore, Christoffersen et al. (2012) highlight the potential of multi-factor models for capturing time and cross-sectional variations in the implied volatility structure.

The aforementioned research documents the significantly superior performance achieved by multi-factor volatility models in continuous time. However, theoretical and empirical studies have also explored the multi-factor nature of volatility in discrete time. The generalized autoregressive conditional heteroskedasticity (GARCH) model, originally proposed by Bollerslev (1986), is one of the most popular and effective single-component models for filtering volatility in discrete time, as it inherently allows for direct estimation of volatility from historical returns. Consequently, given the complex and multifaceted nature of volatility observed empirically in financial markets, researchers have generalized the original GARCH model to incorporate two

volatility components.

Ding and Granger (1996) propose a novel approach to modeling volatility persistence by employing two GARCH components, demonstrating a significant likelihood improvement over the standard GARCH model when applied to S&P500 daily returns data. Another two-component GARCH model is developed by Engle and Lee (1999), who provide evidence in support of volatility decomposition into two components by investigating the US and Japanese stock equity markets. Adrian and Rosenberg (2008) further advance this area by proposing a GARCH model with short-run and long-run components that significantly outperforms one-component GARCH-type models. Leveraging this approach and building on the work of Engle and Lee (1999), Christoffersen et al. (2008) introduce a Heston-Nandi GARCH model with two volatility components, demonstrating excellent option pricing performance compared to a one-component GARCH model. The model of Christoffersen et al. (2008) was generalized by Christoffersen et al. (2014) and Bormetti et al. (2015) by also including the realized variance. Additional evidence supporting the effectiveness of two-component models on equity returns data is presented in Christoffersen et al. (2010) and Conrad and Kleen (2020).

All the two-component models mentioned so far have considered a single innovation term in the return equation. However, there is no inherent rationale to assume that both volatility components are driven by the same stochastic factor. In contrast, Fouque et al. (2000) and Fouque et al. (2003) extensively document the presence of two distinct volatility scales in returns driven by different random processes. Employing two different innovation terms seems appropriate, as the volatility components may relate to diverse sources of uncertainty. For example, Adrian and Rosenberg (2008), Engle and Rangel (2008), and Engle et al. (2013), who propose two-component models with single innovation factors, indicate that one volatility component captures market skewness risk, which may be interpreted as a measure of the tightness of financial constraints, while the other volatility component models business cycle risk.

This underscores the advantage of employing a different stochastic factor for each volatility component. Notably, Wang et al. (2024) recently developed a non-affine model utilizing two variance components with two different stochastic factors by incorporating the realized variance.

However, their approach implicitly assumes that both variance components are predictable, as both the returns and the realized variance are observable. Moreover, Ghanbari (2024) proposed a two-factor GARCH model that aims to capture the volatility of returns through two independent stochastic components. Nevertheless, the equations describing the two volatility components are independent, so the model does not account for spillovers between the volatility components, which may limit its ability to fully capture the complexities of the interaction mechanisms driving volatility dynamics.

Therefore, in the present work, filling a gap in the GARCH literature, we propose a GARCH model that incorporates two distinct components, each driven by an independent and unobservable innovation factor. In particular, by including two different stochastic factors, consistent with Engle and Rangel (2008), Adrian and Rosenberg (2008), and Engle et al. (2013), we allow the model to effectively adapt to diverse market conditions and discern different market trends. For instance, as evidenced by the empirical analysis conducted in this paper, we can accurately identify two volatility components with different reactions to external shocks. Furthermore, our approach provides flexible modeling of volatility spillovers, incorporating the influence of each volatility component's lagged value on the other, thereby capturing dynamic interactions between the two volatility components.

From a mathematical standpoint, the model we propose is a bivariate affine specification that extends the one-factor GARCH model of Heston and Nandi (2000), allowing for a closed-form expression of the moment-generating function, which allows for quasi closed-form derivatives pricing. Notably, our proposed framework nests the recent model by Ghanbari (2024), as it allows for a richer interaction structure between the volatility components. Additionally, the continuous-time limit of our volatility process recovers the popular stochastic volatility specification by Christoffersen et al. (2009). Importantly, this study is the first in the literature to derive conditions for geometric ergodicity and strict stationarity in a GARCH model driven by two stochastic factors and two volatility components with coupled dynamics.

We test the empirical performance of the proposed two-factor GARCH model by comparing it against closely related models such as the one-factor model of Heston and Nandi (2000), the

component GARCH model of Christoffersen et al. (2008), and the two-factor GARCH model of Ghanbari (2024). We conduct extensive in-sample and out-of-sample exercises, focusing on the S&P500 total return time series. The results show that our model exhibits superior performance compared to the benchmark models in explaining the cross-section of equity returns. Specifically, the introduction of a second innovation term significantly enhances the empirical fit compared to single-innovation benchmarks. Furthermore, the advantage of incorporating volatility spillovers leads to a marked improvement in predictive accuracy, both in-sample and out-of-sample, when applied to returns data.

Finally, after deriving the risk-neutralized version of the proposed model, we utilize it for pricing options written on the S&P500 index. The empirical findings show that the inclusion of a second factor generally enhances the option pricing performance compared to single-innovation benchmark models across various moneyness and maturity levels.

The remainder of the paper is organized as follows. In Section 2.2, we introduce the GARCH model with two innovations, establish sufficient conditions ensuring ensuring strict stationarity and geometric ergodicity, and derive a continuous-time limit. Section 2.3 discusses the parameter estimation procedure in detail. In Section 2.4, we briefly review some popular (affine) GARCH models that we use as benchmarks. Section 2.5 presents the empirical performance of the proposed model and the benchmarks when applied to the S&P500 total return time series. In Section 2.6, we perform the risk-neutralization, derive the expression for the characteristic function, and in Section 2.7, we test the models in pricing options written on the S&P500. Finally, Section 2.8 concludes. All mathematical proofs are gathered in the appendix.

2.2 Modeling Returns with Two Factors

In this section, we introduce a novel bivariate GARCH model with two innovation factors to capture the different components of the volatility of log-returns. This model will be labeled GARCH with two factors (*GARCH-2F*).

Let S_t denote the price of a risky asset at time $t \in \mathbb{Z}$ and let us consider the total log-return

$R_t = \ln(\frac{S_t+D_t}{S_{t-1}})$, including the dividend D_t. We model the return process as follows:

$$R_t = \mu_t + \sqrt{v_{1,t}}Z_{1,t} + \sqrt{v_{2,t}}Z_{2,t}, \tag{2.1}$$

where $v_{1,t}$ and $v_{2,t}$ denote the time-varying variance components. Moreover, we specify the conditional expected return as $\mu_t = r + \lambda(v_{1,t} + v_{2,t})$, where r represents the (constant) risk free rate and λ is the (constant) risk premia parameter. Unlike traditional GARCH models and similarly to continuous time stochastic volatility models, the specification in (2.1) contains two sources of risk, $Z_{1,t}$ and $Z_{2,t}$, and we assume $Z_{1,t} \overset{\text{i.i.d.}}{\sim} N(0,1)$, $Z_{2,t} \overset{\text{i.i.d.}}{\sim} N(0,1)$ with $Z_{1,t}$ independent from $Z_{2,t}$ for all $t \in \mathbb{Z}$. Note that the (conditional) total variance of returns is equal to $v_{1,t} + v_{2,t}$.

We model the time-varying variance components via the following bivariate system of equations:

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} v_{1,t-1} \\ v_{2,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \left(Z_{1,t-1} - \gamma_1\sqrt{v_{1,t-1}}\right)^2 \\ \left(Z_{2,t-1} - \gamma_2\sqrt{v_{2,t-1}}\right)^2 \end{bmatrix}. \tag{2.2}$$

Equation (2.2) specifies two volatility processes governed by Heston-Nandi type GARCH(1,1) dynamics, see Heston and Nandi (2000), allowing spillovers between each variance component and innovation component. The specification includes only one lag in order to keep a parsimonious structure, but further lags can be added. Note that there is no spillover if $\beta_{12} = \beta_{21} = \alpha_{12} = \alpha_{21} = 0$. Moreover, if $\omega_2 = \beta_{12} = \beta_{22} = \beta_{21} = \alpha_{12} = \alpha_{22} = \alpha_{21} = 0$ we obtain a model with a single volatility component driven by the standard Heston-Nandi dynamic, see Heston and Nandi (2000). We note that γ_1 and γ_2 introduce asymmetry in how volatility responds to shocks, which reflect different sensitivities to negative news between the two volatility components.

Assumption 2.1. *To ensure the positivity of the variances we assume that the parameters of*

model equations (2.2) *satisfy the following conditions:*

$$\begin{aligned} &\omega_i > 0, \quad \beta_{ij} \geq 0, \quad \alpha_{ij} \geq 0, \quad i,j = 1,2, \\ &\alpha_{11} + \alpha_{12} > 0, \quad \alpha_{21} + \alpha_{22} > 0, \\ &\beta_{11} + \beta_{12} > 0, \quad \beta_{21} + \beta_{22} > 0. \end{aligned} \tag{2.3}$$

2.2.1 Stationarity and Ergodicity

In this section, we derive sufficient conditions under which the model given by (2.1) and (2.2) generates stationary and ergodic trajectories.

We prove the geometric ergodicity of $\{R_t, \boldsymbol{v}_t\}$, where $\boldsymbol{v}_t = (v_{1,t}, v_{2,t})^T$, based on the Markov chain stability theory of Nummelin (1984) and Meyn and Tweedie (1993). To this aim, let us define $\mathcal{D} = [\omega_1, \infty) \times [\omega_2, \infty)$ and let us consider the following (non-linear) state representation of model (2.1)-(2.2):

$$R_t = \mu_t + G(\boldsymbol{v_t}, \boldsymbol{Z}_t), \tag{2.4}$$

$$\boldsymbol{v}_t = \boldsymbol{F}(\boldsymbol{v}_{t-1}, \boldsymbol{W}_t), \tag{2.5}$$

where $\boldsymbol{W}_t = \boldsymbol{Z}_{t-1}$ and $G : \mathcal{D} \times \mathbb{R}^2 \to \mathbb{R}$, and $\boldsymbol{F} : \mathcal{D} \times \mathbb{R}^2 \to \mathcal{D}$, are defined as follows:

$$G(\boldsymbol{v}_t, \boldsymbol{Z}_t) = \begin{bmatrix} \sqrt{v_{1,t}} & \sqrt{v_{2,t}} \end{bmatrix} \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix},$$

$$\boldsymbol{F}(\boldsymbol{v}_{t-1}, \boldsymbol{W}_t) = \boldsymbol{\omega} + \boldsymbol{\beta}\boldsymbol{v}_{t-1} + \boldsymbol{\alpha} \begin{bmatrix} \left(W_{1,t} - \gamma_1 \sqrt{v_{1,t-1}}\right)^2 \\ \left(W_{2,t} - \gamma_2 \sqrt{v_{2,t-1}}\right)^2 \end{bmatrix},$$

where $\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & \omega_2 \end{bmatrix}^T$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$ and $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$.

The process for $\boldsymbol{v}_t$ is a homogeneous Markov chain with state space $(\mathcal{D}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{D}$. Following Chapter 7 of Meyn and Tweedie (1993), we define inductively a sequence of functions $\boldsymbol{F}_t$ by $\boldsymbol{F}_{t+1}(\boldsymbol{x}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_{t+1}) = \boldsymbol{F}(\boldsymbol{F}_t(\boldsymbol{x}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_t), \boldsymbol{W}_{t+1})$ for $t = 1, 2, 3, \ldots$, so that, for any initial condition $\boldsymbol{v}_0 = \boldsymbol{x}$ and $t \geq 1$, we can use equation (2.5) recursively to obtain $\boldsymbol{v}_t = \boldsymbol{F}_t(\boldsymbol{x}, \boldsymbol{W}_1, ..., \boldsymbol{W}_t)$. If we replace the random disturbances $\{\boldsymbol{W}_t\}$ with a deterministic control sequence, say $\{\boldsymbol{w}_t\}$, we obtain the so-called deterministic control model associated to the non-linear state space model (2.5), see Chapter 7 of Meyn and Tweedie (1993) for more details.

We use $\mu_{\text{Leb}}(\cdot)$ to denote the Lebesgue measure and $P^n(\boldsymbol{x}, A) = P(\boldsymbol{v}_n \in A | \boldsymbol{v}_0 = \boldsymbol{x})$, for $\boldsymbol{x} \in \mathcal{D}$ and $A \in \mathcal{B}$, to denote the n–step transition probability measure of the Markov chain $\boldsymbol{v}_t$. For $n = 0$ we have $P^0(\boldsymbol{x}, A) = \mathbb{1}_A(\boldsymbol{x})$, that is the indicator of the set A: $\mathbb{1}_A(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in A$, $\mathbb{1}_A(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \notin A$. For $n = 1$ the notation $P(\boldsymbol{x}, A)$ is used. We denote with $\|\cdot\|$ the L_2-norm of any vector and matrix and by $\rho(\boldsymbol{A})$ the spectral radius of any square matrix $\boldsymbol{A}$, *i.e.*, $\rho(\boldsymbol{A}) = \max\{|\varphi_i| : \varphi_i \text{ is an eigenvalue of } \boldsymbol{A}\}$.

For later purposes, we introduce the autoregressive matrix of the model (2.2):

$$\boldsymbol{B} = \begin{bmatrix} \beta_{11} + \alpha_{11}\gamma_1^2 & \beta_{12} + \alpha_{12}\gamma_2^2 \\ \beta_{21} + \alpha_{21}\gamma_1^2 & \beta_{22} + \alpha_{22}\gamma_2^2 \end{bmatrix}. \tag{2.6}$$

We state the definitions of the main concepts related to stationarity and ergodicity.

Definition 2.1 (Meyn and Tweedie (1993))**.** *A Markov process $\{\boldsymbol{v}_t\}$ is said to be geometrically ergodic if there exist a probability measure π on $(\mathcal{D}, \mathcal{B})$, a constant $0 < c < 1$, and a π-integrable nonnegative measurable function Q such that:*

$$\|P^n(\boldsymbol{x}, \cdot) - \pi(\cdot)\|_{TV} \leq c^n Q(\boldsymbol{x}), \quad n \geq 0, \quad \forall \boldsymbol{x} \in \mathcal{D},$$

where $\|\cdot\|_{TV}$ *denotes the total variation norm, that is* $\|\mu\|_{TV} = \sup_{|g|\leq 1} |\mu(g)|$ *for any signed measure* μ *on* $\mathcal{B}$*.*

Definition 2.2 (Meyn and Tweedie (1993))**.** *The Markov chain* $\{\boldsymbol{v}_t\}$ *is* ψ*-irreducible if there exists a measure* ψ *on* $\mathcal{B}$ *such that, for all* $\boldsymbol{x} \in \mathcal{D}$*, whenever* $\psi(A) > 0$*, there exists some* $k > 0$*, possibly depending on both* A *and* $\boldsymbol{x}$*, such that* $P^k(\boldsymbol{x}, A) > 0$*.*

If $\{\boldsymbol{v}_t\}$ is ψ-irreducible there exists a maximal irreducibility measure M on $(\mathcal{D}, \mathcal{B})$, *i.e.*, an irreducibility measure such that all other irreducibility measures are absolutely continuous with respect to M (Meyn and Tweedie (1993)). Moreover, we set $\mathcal{B}^+ = \{A \in \mathcal{B} : M(A) > 0\}$.

Definition 2.3 (Nummelin (1984))**.** *A sequence* $(E_0, \ldots, E_{d-1})$ *of* d *non-empty disjoint sets in* $\mathcal{B}$ *is called a* d*-cycle if for all* $i = 0, \ldots, d-1$ *and all* $\boldsymbol{x} \in E_i$*:*

$$P(\boldsymbol{x}, E_j^c) = 0, \quad j = i + 1 \mod (d).$$

If $d = 1$*, the chain is called aperiodic, otherwise is periodic.*

Definition 2.4 (Nummelin (1984))**.** *A set* $C \in \mathcal{B}^+$ *is called small if there are* $n \geq 1$*,* $k > 0$ *and a positive measure* $\nu(\cdot)$ *such that*

$$P^n(\boldsymbol{x}, \cdot) \geq k\nu(\cdot), \quad \forall \boldsymbol{x} \in C.$$

The following lemma appropriately defines the measure ψ so that we achieve ψ-irreducibility and aperiodicity under mild-assumptions.

Lemma 2.1. *If Assumption 2.1 holds,* $\rho(\boldsymbol{B}) < 1$*,* $\rho(\boldsymbol{\beta}) < 1$ *and* $\det(\boldsymbol{\alpha}) \neq 0$*, the Markov chain* $\{\boldsymbol{v}_t\}$ *is* ψ*-irreducible, where* $\psi(\cdot) = \mu_{Leb}(\cdot \cap \Omega)$ *denotes the Lebesgue measure restricted to the set* Ω *defined in equation* (A.2) *below. Moreover,* $\{\boldsymbol{v}_t\}$ *is aperiodic.*

Proof. See Appendix A.1.1. □

We note that in the proof of Lemma 2.1 we also derive an explicit form for the density of the transition probability associated to (2.2), see equation (A.4).

Studying the small sets of an irreducible chain allows us to derive its long-run probabilistic behavior. Specifically, to prove geometric ergodicity, we first establish the following proposition and preliminary lemma.

Proposition 2.1. *If the assumptions of Lemma 2.1 hold and using the irreducibility measure ψ, the set $C = \{\boldsymbol{x} \in \mathcal{D} : \|\boldsymbol{x}\| \leq r\}$, for any $r > \bar{r}$ where $\bar{r}$ is specified in equation* (A.5) *of the appendix, is a small set.*

Proof. See Appendix A.1.2. □

Then, given that $\{\boldsymbol{v}_t\}$ is irreducible and aperiodic, an appropriate small set exists and we can provide sufficient conditions for the geometric ergodicity of the process $\{R_t, \boldsymbol{v}_t\}$.

Theorem 2.1. *If the assumptions of Lemma 2.1 hold, then $\{R_t, \boldsymbol{v}_t\}$ is geometrically ergodic.*

Proof. See Appendix A.1.3. □

2.2.2 Continuous Time Limit

In this section, we derive the continuous time limit of the model in equations (2.1)-(2.2). We adopt the convergence scheme of Nelson (1990) and also use the same notation.

Proposition 2.2. *If Assumption 2.1 holds, as the time interval shrinks, the sequence $\{{}_hR_t, {}_h\boldsymbol{v}_t\}$ (defined in the Appendix A.2.1) converges weakly (in distribution) to the solution $\{R_t, \boldsymbol{v}_t\}$ of the following stochastic differential equations:*

$$
\begin{aligned}
dR_t &= (r + \lambda(v_{1,t} + v_{2,t}))dt + \sqrt{v_{1,t}}dW_{1,t} + \sqrt{v_{2,t}}dW_{2,t}, \\
dv_{1,t} &= (\omega_1 + \alpha_{11} + \alpha_{12} - \theta_{11}v_{1,t} - \theta_{12}v_{2,t})dt + 2\alpha_{11}\gamma_1\sqrt{v_{1,t}}dW_{3,t} + 2\alpha_{12}\gamma_2\sqrt{v_{2,t}}dW_{4,t}, \\
dv_{2,t} &= (\omega_2 + \alpha_{21} + \alpha_{22} - \theta_{21}v_{1,t} - \theta_{22}v_{2,t})dt + 2\alpha_{21}\gamma_1\sqrt{v_{1,t}}dW_{3,t} + 2\alpha_{22}\gamma_2\sqrt{v_{2,t}}dW_{4,t},
\end{aligned}
\tag{2.7}
$$

where the only non-zero correlations between Brownian motions are given by

$$\mathbb{E}\left[dW_{1,t}dW_{3,t}\right] = -sign(\gamma_1)dt, \quad \mathbb{E}\left[dW_{2,t}dW_{4,t}\right] = -sign(\gamma_2)dt.$$

Proof. See Appendix A.2.1. □

We note that the set of equations (2.7) is analogous to the bivariate square-root stochastic volatility specification considered, for example, in Christoffersen et al. (2009).

2.2.3 Three Nested Models

In this section, we present three parsimonious GARCH specifications nested within the *GARCH-2F* model. This enables a comparison between the full *GARCH-2F* model and its nested versions, where spillovers are either restricted or entirely absent.

The *GARCH-2Fβ* Model

In this model we partially remove the spillover between the two volatilities in Equation (2.2), by imposing:

$$\beta_{12} = \beta_{21} = 0.$$

The *GARCH-2Fα* Model

In this model we remove the spillover between each volatility and the innovation driving the other. This amounts to imposing:

$$\alpha_{12} = \alpha_{21} = 0.$$

The *GARCH-2F$\alpha\beta$* Model

In this model we remove all the spillover components given by the α and β coefficients. This amounts to setting:

$$\beta_{12} = \beta_{21} = \alpha_{12} = \alpha_{21} = 0. \tag{2.8}$$

By implementing these restrictions, we can completely eliminate the spillovers between the two volatility components, resulting in a more parsimonious variance process. Notably, under the specification in (2.8), we recover the interesting two-factor GARCH model introduced by Ghanbari (2024).

2.3 Parameter Estimation and Filtering

In this section, we illustrate a feasible and computationally efficient approach to estimate model (2.1)-(2.2). Equation (2.1) contains two sources of unobservable uncertainty, $Z_{1,t}$ and $Z_{2,t}$, which implies that we cannot estimate the model as done in traditional GARCH models, rather we need a filtering method. Following Christoffersen et al. (2012), we note that a key property of a filter is that the filtered states are equal to their expected values conditional on the relevant information set. Hence, we define the filtered estimates of $Z_{1,t}$ and $Z_{2,t}$ conditional to the filtration $\mathcal{F}_t = \sigma(\{R_t, R_{t-1}, \ldots\})$ at time t as

$$\widetilde{Z}_{i,t} = \mathbb{E}[Z_{i,t} \mid \mathcal{F}_t], \quad i = 1, 2. \tag{2.9}$$

Since $Z_{i,t}$ is normally distributed, then also R_t is (conditionally) normally distributed so that the conditional expectation in equation (2.9) can be computed using a well-known result on Normal distributions conditioning, see Chapter 2 of Carlin et al. (2016), that is:

$$\widetilde{Z}_{i,t} = \frac{1}{\sqrt{\widetilde{v}_{i,t}}} \left(\frac{R_t - \widetilde{\mu}_t}{\widetilde{v}_{1,t} + \widetilde{v}_{2,t}} \right), \quad i = 1, 2, \tag{2.10}$$

where $\widetilde{\mu}_t = r + \lambda(\widetilde{v}_{1,t} + \widetilde{v}_{2,t})$ and $\widetilde{v}_{1,t}$ and $\widetilde{v}_{2,t}$ denoting the filtered estimates of the conditional variances $v_{1,t}$ and $v_{2,t}$ respectively, which can be computed using equation (2.2)

$$\begin{bmatrix} \widetilde{v}_{1,t} \\ \widetilde{v}_{2,t} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \widetilde{v}_{1,t-1} \\ \widetilde{v}_{2,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \left(\widetilde{Z}_{1,t-1} - \gamma_1 \sqrt{\widetilde{v}_{1,t-1}} \right)^2 \\ \left(\widetilde{Z}_{2,t-1} - \gamma_2 \sqrt{\widetilde{v}_{2,t-1}} \right)^2 \end{bmatrix}. \tag{2.11}$$

The log-likelihood function can now be constructed as the product of the conditional distributions across the sample. Specifically, conditional on $\widetilde{v}_{1,0}$ and $\widetilde{v}_{2,0}$[2], the log-likelihood function of returns is given by

$$\ell_{\text{returns}}(\boldsymbol{\theta}) = -\frac{T}{2} \log \left(2\pi(\widetilde{v}_{1,t} + \widetilde{v}_{2,t}) \right) - \frac{1}{2} \sum_{t=1}^{T} \frac{(R_t - \widetilde{\mu}_t)^2}{\widetilde{v}_{1,t} + \widetilde{v}_{2,t}}, \tag{2.12}$$

where T denote the length of the daily log-returns series.

2.4 Benchmarks Models

Given the affine and two-component structure of model (2.1)-(2.2), we will compare it with the following models: the GARCH model developed by Heston and Nandi (2000), hereafter *GARCH-HN*, the two-component GARCH model introduced by Christoffersen et al. (2008), hereafter *GARCH-CJOW* model, and the *GARCH-2F*$\alpha\beta$ of Ghanbari (2024). Both models employ a single innovation factor, and, for the reader's convenience, they are briefly recalled below.

[2]We note that in the estimation the initial values $\widetilde{v}_{1,0}$ and $\widetilde{v}_{2,0}$ have been set equal to their long-run mean values.

2.4.1 The *GARCH-HN* Model

The return and the volatility processes are defined as follows

$$R_t = r + \lambda v_{1,t} + \sqrt{v_{1,t}} Z_t,$$
$$v_{1,t+1} = \omega_1 + \beta_{11} v_{1,t} + \alpha_{11} (Z_t - \gamma_1 \sqrt{v_{1,t}})^2,$$

where $Z_t \overset{\text{iid}}{\sim} N(0,1)$.

2.4.2 The *GARCH-CJOW* Model

The component GARCH model proposed by Christoffersen et al. (2008) comprises three equations, one for the return process, one for the long-term variance component q_t, and one for the short-term variance component s_t:

$$R_t = r + \lambda(q_t + s_t) + \sqrt{q_t + s_t} Z_t,$$
$$q_{t+1} = \omega_1 + \beta_{11} q_t + \alpha_{11} (Z_t^2 - 1 - 2\gamma_1 Z_t \sqrt{q_t + s_t}),$$
$$s_{t+1} = (\beta_{22} + \alpha_{22}\gamma_2^2) s_t + \alpha_{22} (Z_t^2 - 1 - 2\gamma_2 Z_t \sqrt{q_t + s_t}),$$

where $Z_t \overset{\text{iid}}{\sim} N(0,1)$.

2.5 Empirical Results for Returns

In this section, we present empirical results for the total S&P500 daily adjusted log-returns time series. Our data span from January 5, 1988 to December 29, 2023 (9069 daily observations),

as shown in Figure 2.1.

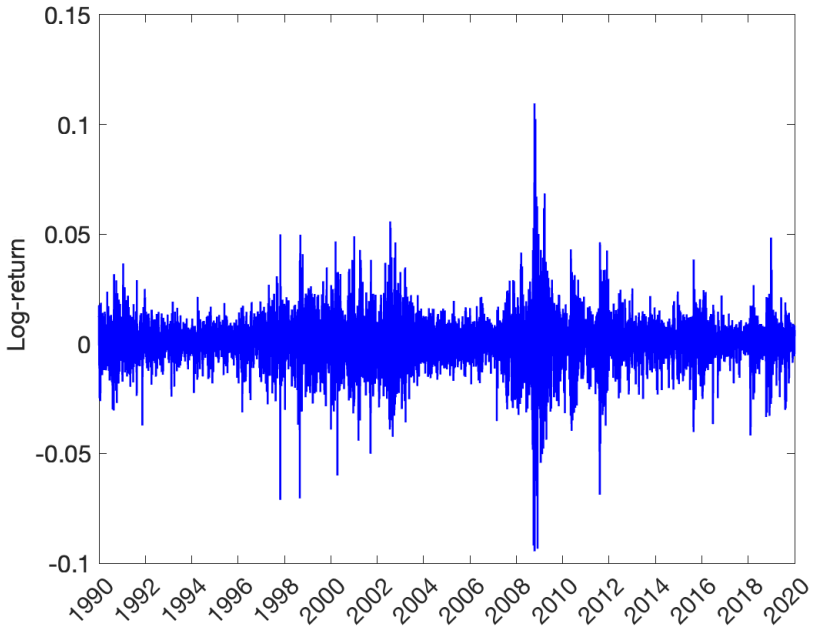


Figure 2.1: Daily time series of total S&P500 log-returns from January 5, 1988 to December 29, 2023.

2.5.1 In-Sample Analysis

The estimation procedure described in Section 2.3 requires, besides the time series of log-returns, also the risk-free interest rate, which we proxy using the 3-month US Treasury Bill rate. The data for the daily levels of the Adjusted Close Price for the S&P500 total Returns series is retrieved from Refinitiv Datastream, whereas the data for the 3-month Treasury Bill rate are gathered from the Federal Funds Effective Rate (FRED) dataset.

The in-sample results, based on the maximum likelihood estimation described in Section 2.3, are illustrated in Table 4.2. To assess the model performance we consider the Akaike (AIC) and the Bayesian information criteria (BIC). To further assess the goodness-of-fit of our model specification we performed a likelihood ratio (LR) test for the *GARCH-2F* model and its three nested models versus the one factor *GARCH-HN* model. The LR statistics are reported at the bottom of Table 4.2.

As we may see, the *GARCH-2F* model and its nested versions provide superior goodness-of-

fit to the returns data compared to the one-factor benchmark models. The LR statistic is significantly higher, and the AIC and BIC values are lower by at least 60 points. Moreover, the most parsimonious model, the *GARCH-2Fαβ* proposed by Ghanbari (2024), achieves 30 likelihood extra points compared to *GARCH-CJOW* and 130 points compared to *GARCH-HN*. However, in comparison to *GARCH-2Fαβ*, results also highlight the significance of including volatility spillovers, either through α for the *GARCH-2Fβ* model or the β coefficient for the *GARCH-2Fα* model specification.

We further assessed the volatility persistence for all the considered models. The persistences of the two variance components $v_{1,t}$ and $v_{2,t}$ in model (2.1)-(2.2) are defined as the two eigenvalues of the autoregressive matrix $\boldsymbol{B}$ in (2.6). Whereas, the persistences, of the *GARCH-CJOW* model, as discussed in Christoffersen et al. (2008), are equal to $\beta_{11}+\alpha_{11}\gamma_1^2$ for the one volatility component and $\beta_{22}+\alpha_{22}\gamma_2^2$ for the other volatility component. At the bottom of Table 4.2 we present the persistences of the volatilities for all the models. A feature that characterize all the component models is that the component displaying the highest β coefficient is also the one possessing the lowest reaction to innovation shocks, as indicated by the α coefficient. However, the persistence values do not differ significantly among the nested model specifications and are aligned to the peristences of the *GARCH-CJOW* model.

When analyzing the model parameters, we observe that for all considered models, the first volatility component returns a β_{11} estimate around 0.9, which is commonly encountered in the GARCH literature (see Heston and Nandi (2000), Christoffersen et al. (2008)). A similar pattern holds for α_{11} and γ_1. The parameter ω_1 is largely non-significant across most models, except for the *GARCH-CJOW*. In the component models, the second volatility component exhibits higher sensitivity to innovations, as reflected by α_{22} and γ_2. The autoregressive parameter β_{22} for the second component is generally lower than β_{11} and is not always significant in the *GARCH-2F*, *GARCH-2Fβ* and *GARCH-2Fα* models.

If we consider the spillover effects between the two volatility components in the *GARCH-2F* and the *GARCH-2Fβ* models, the coefficients α_{12} and α_{21} are highly significant. In the *GARCH-2Fα* model, the spillovers through the β_{12} and β_{21} coefficients are also significant. Consequently, the

more parsimonious *GARCH-2Fαβ* specification proposed by Ghanbari (2024), which eliminates spillovers between the two volatility components, is not fully capable of capturing the dynamics of return data.

Table 2.1: Maximum likelihood estimation results.

	GARCH-HN	*GARCH-CJOW*	*GARCH-2F*	*GARCH-2Fβ*	*GARCH-2Fα*	*GARCH-2Fαβ*
ω_1	0	1.115e-06***	0	0	0	0
	(8.849e-09)	(2.193e-07)	(6.925e-08)	(8.911e-09)	(1.695e-08)	(2.792e-10)
ω_2			0	0	0	0
			(5.268e-08)	(3.351e-09)	(2.644e-08)	(6.743e-10)
α_{11}	5.055e-06***	1.764e-06***	2.967e-07**	2.966e-07***	4.676e-07***	1.503e-07***
	(9.089e-07)	(3.426e-07)	(1.327e-07)	(1.069e-08)	(1.108e-07)	(5.654e-10)
γ_1	169.418***	115.328***	107.640***	107.638***	145.708**	358.980***
	(17.864)	(26.286)	(32.944)	(25.913)	(88.501)	(6.042)
β_{11}	0.812***	0.991***	0.986***	0.986***	0.975***	0.978***
	(0.016)	(0.001)	(0.004)	(0.004)	(0.014)	(0.001)
α_{22}		2.441e-06***	5.334e-06**	5.333e-06***	4.195e-06***	1.039e-05***
		(7.818e-07)	(8.491e-07)	(8.110e-07)	(1.556e-06)	(2.069e-07)
γ_2		369.261***	419.987***	420.007***	469.691***	154.782***
		(115.072)	(37.253)	(33.263)	(85.395)	(2.691)
β_{22}		0.539***	2.599e-07	1.607e-06	0.011	0.684***
		(0.127)	(8.989e-07)	(1.903e-06)	(0.019)	(0.013)
α_{12}			2.921e-08***	2.921e-08***		
			(1.155e-08)	(1.185e-08)		
α_{21}			1.716e-06***	1.716e-06***		
			(6.431e-07)	(6.072e-07)		
β_{12}			0		0.005***	
			(2.107e-08)		(0.001)	
β_{21}			0		0.058**	
			(1.373e-08)		(0.030)	
λ	1.101	1.115***	1.630	1.632	2.170	0.903
	(1.073)	(0.349)	(4.790)	(1.575)	(1.830)	(0.953)
Log-likelihood	29,667	29,761	29,831	29,831	29,812	29,797
AIC	-59,324	-59,506	-59,636	-59,640	-59,602	-59,576
BIC	-59,288	-59,449	-59,543	-59,561	-59,523	-59,511
LR	0		328	328	290	260
Persistence	0.96	0.99 | 0.87	0.99 | 0.94	0.99 | 0.94	0.99 | 0.93	0.99 | 0.93

The estimation period spans from January 5, 1988 to December 29, 2023 (9069 observations). The standard errors, reported in parenthesis, are computed by inverting the negative Hessian matrix evaluated at the optimum parameter values. [*],[**],[***] denote statistical significance at the 0.1, 0.05, 0.01 levels, respectively.

In Figure 2.2, we plot the filtered conditional volatility of the *GARCH-HN* model and the two volatility components of the *GARCH-CJOW* model and *GARCH-2F* model. One of the volatility components of the *GARCH-2F* model mimics the dynamics of the one-factor *GARCH-HN*, and the other volatility component is a slower moving process with much lower variation over the considered time span.

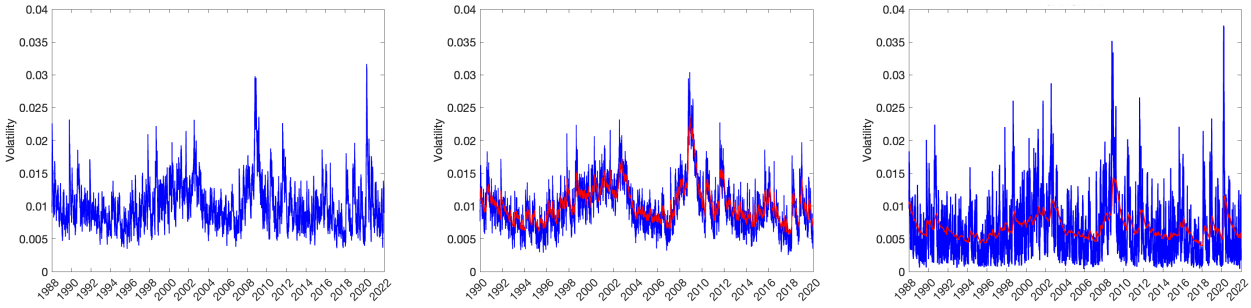


Figure 2.2: Filtered conditional volatility for the *GARCH-HN* (left); *GARCH-CJOW* total volatility $\sqrt{q_t + s_t}$ (center blue) and long-term volatility component $\sqrt{q_t}$ (center red); *GARCH-2F* first volatility component $\sqrt{v_{1,t}}$ (right red) and second volatility component $\sqrt{v_{2,t}}$ (right blue).

In Figure 2.3, we compare the conditional variances of the one-factor *GARCH-HN* model and of the two-factor *GARCH-2F* model, that is $v_{1,t} + v_{2,t}$. The variance of both models displays a similar pattern, however during highly volatile periods, as in 2008 and 2020, the higher spikes of the *GARCH-2F* might yield a better descprition of the high volatility of such periods.

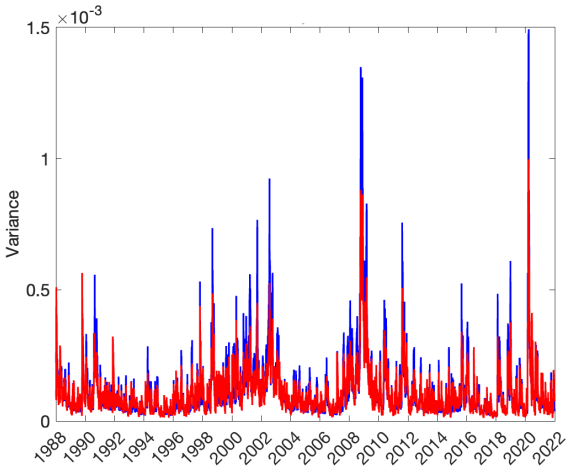


Figure 2.3: *GARCH-HN* conditional variance (red) and *GARCH-2F* total conditional variance $v_{1,t} + v_{2,t}$ (blue).

Following Durham et al. (2015) we performed an experiment to assess the bias in state estimates associated with using the filter in propose in equations (2.9)-(2.11). In particular, we investigate the extent to which the filtered states $\tilde{v}_{1,t}$ and $\tilde{v}_{2,t}$ are informative about the true states $v_{1,t}$ and $v_{2,t}$. We use simulated data generated from the *GARCH-2F* model with the parameter vector reported in Table 4.2. We simulate one million observations for one trajectory with a burn-in period of one thousand observations to minimize the effect of initial conditions. We then apply the filter (2.9)-(2.11) to extract the filtered states $\tilde{v}_{1,t}$ and $\tilde{v}_{2,t}$ and innovations $\tilde{Z}_{1,t}$ and $\tilde{Z}_{2,t}$.

We report in Table 2.2 the summary statistics for the true and filtered states.

		True	Filtered
v_1	Mean $\times 1000$	0.1543	0.1040
	SD $\times 1000$	0.1698	0.1374
	Skewness	2.1929	2.7557
	Kurtosis	10.0504	14.1985
v_2	Mean $\times 1000$	0.1062	0.0790
	SD $\times 1000$	0.0378	0.0287
	Skewness	1.4715	1.7746
	Kurtosis	6.7062	8.3708
Z_1	Mean $\times 1000$	1.2473	7.5746
	SD $\times 1000$	999.8772	778.0238
	Skewness	0.0033	0.0118
	Kurtosis	3.0033	4.1288
Z_2	Mean $\times 1000$	2.4500	11.2246
	SD $\times 1000$	1000.3257	984.2194
	Skewness	0.0028	0.0268
	Kurtosis	3.0044	4.2878

Table 2.2: Summary Statistics.
This table reports summary statistics for the true and filtered states and innovations using data simulated from the *GARCH-2F* model with the parameter vector reported in Table 4.2. SD = standard deviation.

2.5.2 Out-of-Sample Analysis

To test the forecasting performances of the rival models, we forecast the Value-at-Risk (VaR) at the 1% and 5% significance levels from one to five days ahead. We test the models during different periods of crisis by splitting the dataset into two subperiods (4534 observations each) and running separate out-of-sample analyses: from January 5, 1988 to December 21, 2005 (to include the September 11, 2001 crisis) and from December 22, 2005 to December 29, 2023 (to include the 2008 financial crisis and the COVID-19 period). For each of the considered subsamples, we estimate the model parameters using a rolling window equal to 60% of the days in the period. That is, we start by using the first 60% of the daily S&P500 total returns to calibrate the models, leaving the remaining data for the out-of-sample analysis. Then, for each day in the out-of-sample, we re-estimate the model parameters using the past 60% daily observations and compute forecasts from one to five days ahead.

In particular, for all the competing models, at each forecasting time, we simulate $100,000$ sample paths of $R_{t+\ell}$, with $\ell = 1, 2, 3, 4, 5$, according to the return equation (2.1). This is done by randomly simulating the $Z_{1,t}$ and $Z_{2,t}$ standard normal random variables. Then, we estimate the VaR at the 1% and 5% significance levels by computing the empirical quantiles of the simulated sample paths of $R_{t+\ell}$.

We then measure the statistical accuracy of the VaR forecasts by performing the proportion of failures test of Kupiec (1995) and the correct conditional coverage test of Christoffersen (1998). The results obtained for the proposed *GARCH-2F*, *GARCH-2Fα* and *GARCH-2Fβ*, together with the competing *GARCH-2Fαβ*, *GARCH-HN* and *GARCH-CJOW* models, are reported in Tables 2.3 and 2.4.

We note that, overall, the novel two-factor specifications result in a significantly lower number of failures compared to the one-factor models. For the period from January 1988 to December 2005, both the Kupiec and Christoffersen tests confirm the superior performance of the *GARCH-2F* models, particularly the *GARCH-2Fβ*, which consistently performs well at both the 95% and 99% confidence levels, outperforming the *GARCH-2Fαβ* model with no spillovers between the volatility components. For the period from December 2005 to December 2023, the *GARCH-2F* and *GARCH-2Fβ* models exhibit similar performances.

In addition, we further compare the accuracy of the VaR forecast using the Model Confidence Set (MCS) introduced by Hansen et al. (2011). Following González-Rivera et al. (2004), we measure performance by means of the quantile loss function as in Bassett and Koenker (1978), defined as

$$\mathrm{QL}_{t+\ell}(\vartheta) = (\vartheta - \mathbb{1}_{\{R_{t+\ell} < \mathrm{VaR}_{t+\ell}(\vartheta)\}})(R_{t+\ell} - \mathrm{VaR}_{t+\ell}(\vartheta)),$$

where $\mathrm{VaR}_{t+\ell}(\vartheta)$ denotes the VaR on day $t+\ell$ at the significance level ϑ. Moreover, to check if the forecasts of the models *GARCH-CJOW*, *GARCH-2F*, *GARCH-2Fα*, *GARCH-2Fβ* and *GARCH-2Fαβ* significantly differ from the ones of the *GARCH-HN* benchmark we performed the test of Diebold and Mariano (2002). The quantile losses and the p-values are reported in

Table 2.5.

Overall, the *GARCH-2F* model, its three nested specifications, and the *GARCH-CJOW* model return significantly different forecasts compared to the ones of the *GARCH-HN* as indicated by the p-values for both periods and significance levels. For the period from January 5, 1988 to December 21, 2005, all the four *GARCH-2F* models return lower quantile lossess, thus indicating a superior VaR forecasting accuracy and are included in the MCS, contrary to the *GARCH-CJOW* model. In the period from December 22, 2005 to December 29, 2023, only the *GARCH-2Fβ* is always included in the MCS and outperform the benchmark models by returning lower quantile losses.

2.6 Option Pricing

In this section, we derive results that allow us to value option contracts using the four proposed two-factor models. First, we perform the risk-neutralization of these models. Second, we estimate the *GARCH-2F* model (and its nested models) by jointly using returns and options data.

2.6.1 The *GARCH-2F* Risk Neutral Dynamics

Equations (2.1)-(2.2) define the return and the volatility dynamics under the physical measure. In this section, we risk-neutralize the model, so that we can compute option prices. Following Christoffersen et al. (2012), we consider a pricing kernel with affine dynamic

$$M_t = \frac{e^{-r-\xi(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}}{\mathbb{E}[e^{-\xi(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})} \mid \mathcal{F}_{t-1}]}, \tag{2.13}$$

where $\mathbb{E}$ denotes the expectation operator under the physical measure. Accordingly, we guess the conditional Radon–Nikodym derivative in the following form

$$\frac{dQ_t}{dP_t} = \frac{e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}}{\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}, \tag{2.14}$$

and we impose

$$\frac{dQ_t}{dP_t} = M_t e^r, \tag{2.15}$$

so that using equations (2.13)-(2.15) we obtain $\Lambda = -\xi$. Therefore, we have the following result.

Proposition 2.3. *Let us consider the return process specified by equation* (2.1)*, then the equivalent martingale measure* $\mathbb{Q}$ *associated to the Radon–Nikodym derivative in* (2.14) *exists if and only if the coefficient* Λ *satisfy*

$$\Lambda = -\lambda - \frac{1}{2}. \tag{2.16}$$

Proof. See Appendix A.3.1. □

Finally, the following proposition yields the dynamic of the risk-neutralized processes $\{R_t\}$, $\{v_{1,t}\}$ and $\{v_{2,t}\}$.

Proposition 2.4. *Under the risk-neutral measure* $\mathbb{Q}$*, the dynamics of the return and of the two variance components are as follows:*

$$R_t = r - \frac{1}{2}\left(v_{1,t} + v_{2,t}\right) + \sqrt{v_{1,t}}Z^*_{1,t} + \sqrt{v_{2,t}}Z^*_{2,t}, \tag{2.17}$$

$$\begin{bmatrix} v_{1,t+1} \\ v_{2,t+1} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \left(Z^*_{1,t} - \gamma^*_1\sqrt{v_{1,t}}\right)^2 \\ \left(Z^*_{2,t} - \gamma^*_2\sqrt{v_{2,t}}\right)^2 \end{bmatrix}, \tag{2.18}$$

where $Z^*_{1,t} \sim N(0,1)$ *and* $Z^*_{2,t} \sim N(0,1)$ *denote the risk-neutral shocks and* $\gamma^*_1 = \gamma_1 + \lambda + \frac{1}{2}$ *and* $\gamma^*_2 = \gamma_2 + \lambda + \frac{1}{2}$ *are risk-neutralized parameters.*

Proof. See Appendix A.3.2. □

2.6.2 Moment Generating Function

Equations (2.1)-(2.2) allow us to derive the moment generating function using a standard procedure as in Heston and Nandi (2000).

Proposition 2.5. *The conditional moment generating function for the logarithm of the terminal stock price* $\ln S_T$*, that is*

$$f_t(\phi) = \mathbb{E}\left[e^{\phi \ln S_T} \middle| \mathcal{F}_t\right], \quad t < T, \tag{2.19}$$

can be computed as

$$f_t(\phi) = e^{\phi \ln S_t + A_t(\phi) + B_{1,t}(\phi) v_{1,t+1} + B_{2,t}(\phi) v_{2,t+1}}, \tag{2.20}$$

where the expressions for $A_t(\phi)$*,* $B_{1,t}(\phi)$ *and* $B_{2,t}(\phi)$ *can be computed using the recursion equations* (A.42) *in the Appendix.*

Proof. See Appendix A.4.1. □

2.6.3 Valuation of Options

Let us consider a European Call option on the underlying asset S_t with strike price K and maturity T. Using the inversion theorem in Gil-Pelaez (1951), the option price at a generic time $t < T$, which we denote by C_t, is computed as follows:

$$\begin{aligned} C_t = \frac{1}{2} S_t + \frac{e^{-r(T-t)}}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi} f_t^*(i\phi + 1)}{i\phi}\right] d\phi \\ - K e^{-r(T-t)} \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi} f_t^*(i\phi)}{i\phi}\right] d\phi\right), \end{aligned} \tag{2.21}$$

where, as in Heston and Nandi (2000), f_t^* denotes the (conditional) moment generating function under the risk-neutral measure, that is function (3.32) evaluated using risk-neutral parameters.

Put option values can be calculated using the well-known put-call parity as in Heston and Nandi (2000). The valuation of the integrals in (2.21) can be performed using Fourier pricing as illustrated in Papapantoleon et al. (2010).

2.7 Option Valuation Empirics

We consider European options, both Put and Call, written on the S&P500 index, with data retrieved from Thomson Reuters Eikon Datastream. Specifically, we consider options with maturities ranging from 2021 to 2023 and the time series of their daily prices from February 10, 2021 to December 29, 2023.

As a common practice, see for example Christoffersen et al. (2012), Ballestra et al. (2023) and Ballestra et al. (2024), we apply several exclusion filters to obtain the final panel of option contracts. We keep only the options with time-to-maturity between 14 and 365 days and we select only out-of-the-money Put and Call options (we compute the moneyness as K/S_t, where K is the strike price and S_t is the underlying index level), and we filter out illiquid quotes by selecting only the six most liquid strikes at each maturity, and we consider option quotes only on Wednesday. Finally, we remove price quotes lower than 3.8$.

In Table 2.6, we report the resulting number of option prices on the S&P500 index, sorted by moneyness and days to maturity, for a total of $N = 9372$ options prices.

2.7.1 Parameter Estimation using both Returns and Options

A common practice for estimating parameters using daily time series of log-returns and a panel of option contracts is using a suitably weighted log-likelihood function, see Christoffersen et al. (2012) and Ballestra et al. (2023). The log-likelihood for the returns series, $\ell_{\text{returns}}(\boldsymbol{\theta})$, is already available in equation (2.12), whereas, for the N option prices, we proceed as follows. First, we define the error on the volatility implied by the i-th option price as

$$e_i = \frac{IV_i^{MKT} - IV_i^{MOD}}{IV_i^{MKT}}, \tag{2.22}$$

where IV_i^{MKT} and IV_i^{MOD} denote the market and the model volatilities implied by the i-th option price, which we compute via the Black-Scholes model. Then, by assuming a Gaussian distribution for e_i, we can derive the log-likelihood associated to the option pricing errors as

$$\ell_{\text{options}}(\boldsymbol{\theta}) = -\frac{N}{2}\log\left(2\pi\sigma_e^2\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{e_i^2}{\sigma_e^2}, \tag{2.23}$$

where σ_e^2 represents the variance of the option pricing errors.

The number N of data points available from the option panel could be significantly higher than the number T of daily returns, see Section 4. Therefore, following Christoffersen et al. (2012), we assign an equal weight to returns and option prices by considering the following weighted joint (total) log-likelihood:

$$\ell_{\text{joint}}(\boldsymbol{\theta}) = \frac{N+T}{2}\frac{\ell_{\text{options}}(\boldsymbol{\theta})}{N} + \frac{N+T}{2}\frac{\ell_{\text{returns}}(\boldsymbol{\theta})}{T}. \tag{2.24}$$

The joint estimation results are presented in Table 2.7. We calculated standard information metrics AIC and BIC, and we also report the persistence values, as we did in Table 4.2. We observe different persistence values around 0.98 and 0.97 for the two volatility components, compared to the returns data-only estimation, which were 0.99 and 0.92. Additionally, we report the variance of the option pricing errors. The proposed *GARCH-2Fβ* model return the highest joint likelihood values and lower AIC and BIC compared to the benchmark models, indicating improved performance in fitting both option and returns data. The *GARCH-2F* models allowing for spillovers have highly significant parameters and provide a superior fit compared to the *GARCH-2F$\alpha\beta$* model.

2.7.2 Empirical Results for Options

To assess the performance of the six competing models in option pricing, we follow Christoffersen et al. (2012) and employ the (percentage) implied volatility root mean square error:

$$IVRMSE(\%) = 100 \times \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(IV_i^{MKT} - IV_i^{MOD}\right)^2}. \tag{2.25}$$

The results obtained, shown in Table 2.8, reveal that the novel two-factor specifications yield better performance compared to the benchmark approaches when pricing options. In particular, the *GARCH-2F* model and its nested models return lower (from 4% to 6%) overall implied volatility pricing errors compared to the one factor *GARCH-HN* benchmark model. By sorting the options for moneyness and days to maturity, the *GARCH-2F* and its nested models provide consistent results, that is lower *IVRMSE* than the one-factor benchmark models, and reveal more robust performance in pricing out-of-the money options ($0.8 \leq K/S_t \leq 0.9$). Notably, the parsimonious *GARCH-2F*$\alpha\beta$ model provides consistent performance across both moneyness and days to maturity which might indicate that the volatilities of the two series are es- sentially independent or that any interaction between them is negligible.

2.8 Conclusions

Accurately capturing the the complex and multi-faceted nature of volatility is crucial for effective financial analysis and risk management. In this paper, we introduce a novel GARCH specification with two volatility components each driven by an independent stochastic factor, an approach commonly used in the continuous time (stochastic) volatility models, see, e.g., Christoffersen et al. (2008) and Fouque and Lorig (2011), yet still overlooked in discrete-time GARCH models.

We conducted a theoretical investigation that establishes, for the first time, sufficient conditions for strict stationarity and geometric ergodicity in a two-component GARCH model driven by independent stochastic factors. Additionally, we derived a continuous-time limit for the proposed *GARCH-2F* model, which recovers a well-known bivariate square-root stochastic volatility specification. Finally, we considered three more parsimonious nested models: the *GARCH-2F*α, *GARCH-2F*β, and *GARCH-2F*$\alpha\beta$, the latter introduced by Ghanbari (2024).

We empirically tested the two-factor models on S&P500 total log-returns data and compare their in-sample and out-sample performance against one-factor affine benchmark models. Overall, the two-factor models demonstrate superior in-sample and out-sample VaR prediction accuracy. Then, we also considered the pricing of options on the S&P500 index. The two-factors model provide lower implied volatility pricing errors, particularly for out-of-the money options, when estimated jointly using returns and options data.

The approach proposed in this article can be extended in many directions. Potential extensions include include incorporating fat-tailed distributions for the innovations while preserving model affinity, exploring different specifications for the drift of the return process (e.g., an autoregressive model), and considering additional lags.

Table 2.3: Likelihood ratio test statistic of Kupiec (1995), p-values in parentheses.

	January 1988 to December 2005				
	$\vartheta = 5\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	52.206(0.000)	52.206(0.000)	54.805(0.000)	53.499(0.000)	54.805(0.000)
CJOW	36.641(0.000)	40.024(0.000)	37.755(0.000)	34.457(0.000)	35.542(0.000)
2F	7.113(0.007)	6.496(0.011)	7.761(0.005)	7.113(0.007)	7.113(0.007)
2Fβ	2.275(0.132)	1.644(0.200)	1.644(0.200)	1.644(0.200)	1.946(0.163)
2Fα	7.415(0.007)	7.965(0.005)	9.120(0.003)	9.724(0.002)	13.008(0.000)
2F$\alpha\beta$	0.984(0.321)	1.440(0.230)	1.202(0.273)	1.202(0.273)	1.202(0.273)
	December 2005 to December 2023				
	$\vartheta = 5\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	9.688(0.001)	10.309(0.001)	9.688(0.001)	10.309(0.001)	9.688(0.001)
CJOW	2.264(0.132)	2.923(0.087)	2.923(0.087)	2.264(0.132)	2.583(0.108)
2F	5.379(0.020)	5.379(0.020)	5.379(0.020)	4.852(0.021)	4.354(0.037)
2Fβ	1.961(0.161)	2.648(0.104)	2.648(0.104)	2.291(0.130)	2.648(0.104)
2Fα	2.648(0.104)	2.648(0.104)	2.648(0.104)	1.961(0.161)	1.961(0.161)
2F$\alpha\beta$	0.034(0.854)	0.006(0.940)	0.006(0.940)	0.006(0.940)	0.006(0.940)
	January 1988 to December 2005				
	$\vartheta = 1\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	81.442(0.000)	73.481(0.000)	76.104(0.000)	73.481(0.000)	70.891(0.000)
CJOW	115.931(0.000)	115.931(0.000)	98.179(0.000)	95.317(0.000)	89.679(0.000)
2F	0.781(0.377)	0.437(0.509)	0.001(0.976)	0.001(0.976)	0.042(0.838)
2Fβ	2.352(0.125)	2.352(0.125)	2.352(0.125)	2.352(0.125)	1.743(0.187)
2Fα	0.001(0.976)	0.073(0.788)	0.263(0.608)	0.073(0.788)	0.073(0.788)
2F$\alpha\beta$	7.611(0.006)	4.654(0.031)	3.042(0.081)	3.042(0.081)	3.042(0.081)
	December 2005 to December 2023				
	$\vartheta = 1\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	49.055(0.000)	40.323(0.000)	38.239(0.000)	38.239(0.000)	38.239(0.000)
CJOW	26.628(0.000)	21.439(0.000)	19.808(0.000)	18.228(0.000)	15.226(0.000)
2F	0.266(0.606)	0.583(0.445)	0.583(0.445)	1.632(0.201)	1.632(0.201)
2Fβ	5.558(0.018)	3.801(0.051)	3.033(0.082)	3.033(0.082)	2.344(0.126)
2Fα	15.226(0.000)	12.445(0.000)	11.141(0.001)	9.897(0.002)	8.715(0.003)
2F$\alpha\beta$	16.700(0.000)	12.445(0.000)	8.715(0.003)	6.543(0.011)	5.558(0.018)

Table 2.4: Likelihood ratio test statistic of Christoffersen (1998), p-values in parentheses.

	January 1988 to December 2005				
	$\vartheta = 5\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	53.399(0.000)	53.399(0.000)	55.774(0.000)	54.577(0.000)	56.362(0.000)
CJOW	37.093(0.000)	41.201(0.000)	39.923(0.000)	37.138(0.000)	40.089(0.000)
F2	9.294(0.009)	8.515(0.014)	10.111(0.006)	9.294(0.009)	13.271(0.001)
2Fβ	3.143(0.208)	2.326(0.313)	2.326(0.313)	2.326(0.313)	2.718(0.257)
2Fα	7.462(0.024)	8.211(0.017)	9.688(0.008)	11.661(0.003)	18.748(0.000)
2Fαβ	1.030(0.598)	2.341(0.310)	3.154(0.207)	3.154(0.207)	3.154(0.207)
	December 2005 to December 2023				
	$\vartheta = 5\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	11.631(0.003)	12.103(0.002)	11.631(0.003)	12.103(0.002)	11.631(0.003)
CJOW	3.662(0.160)	5.028(0.081)	5.028(0.081)	4.698(0.096)	4.850(0.089)
2F	8.689(0.013)	8.689(0.013)	8.689(0.013)	7.957(0.019)	7.262(0.026)
2Fβ	3.839(0.147)	4.841(0.089)	4.841(0.089)	5.878(0.053)	6.456(0.040)
2Fα	8.416(0.015)	8.416(0.015)	8.416(0.015)	7.176(0.028)	7.176(0.028)
2Fαβ	2.774(0.250)	2.566(0.277)	2.566(0.277)	2.566(0.277)	2.566(0.277)
	January 1988 to December 2005				
	$\vartheta = 1\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	82.208(0.000)	74.554(0.000)	77.067(0.000)	74.554(0.000)	72.078(0.000)
CJOW	115.996(0.000)	115.996(0.000)	98.494(0.000)	95.692(0.000)	91.254(0.000)
2F	1.322(0.516)	0.929(0.628)	0.362(0.834)	0.362(0.834)	0.444(0.801)
2Fβ	3.206(0.201)	3.206(0.201)	3.051(0.218)	3.051(0.218)	2.387(0.303)
2Fα	0.362(0.834)	0.395(0.821)	0.548(0.760)	0.395(0.821)	0.395(0.821)
2Fαβ	10.126(0.006)	5.219(0.074)	3.791(0.150)	3.791(0.150)	3.791(0.150)
	December 2005 to December 2023				
	$\vartheta = 1\%$				
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$	$\ell = 5$
HN	49.385(0.000)	40.478(0.000)	38.360(0.000)	38.360(0.000)	38.360(0.000)
CJOW	26.633(0.000)	21.445(0.000)	19.824(0.000)	18.258(0.000)	16.768(0.000)
2F	2.558(0.278)	3.111(0.211)	3.111(0.211)	1.820(0.403)	1.820(0.403)
2Fβ	6.043(0.049)	4.454(0.108)	3.783(0.151)	3.783(0.151)	3.199(0.202)
2Fα	19.300(0.000)	17.065(0.000)	16.054(0.000)	15.116(0.001)	14.254(0.001)
2Fαβ	16.750(0.000)	12.586(0.002)	8.999(0.012)	6.954(0.031)	6.043(0.049)

Table 2.5: $\mathrm{QL}_{t+\ell}(\vartheta)$ ratios with respect to *GARCH-HN* and p-values of the test of Diebold and Mariano (2002) are reported in parentheses (left column); Models included in the MCS of Hansen et al. (2011) according to $\mathrm{QL}_{t+\ell}(\vartheta)$ function (right column).

	January 1988 to December 2005									
	$\vartheta = 5\%$									
	$\ell = 1$		$\ell = 2$		$\ell = 3$		$\ell = 4$		$\ell = 5$	
	QL	MCS	QL	MCS	QL	MCS	QL	MCS	QL	MCS
HN	1		1		1		1		1	
CJOW	1.021(0.000)		1.023(0.000)		1.025(0.000)		1.025(0.000)		1.025(0.000)	
2F	1.073(0.000)	✓	1.074(0.000)	✓	1.075(0.000)	✓	1.077(0.000)	✓	1.078(0.000)	✓
2Fβ	1.059(0.000)	✓	1.060(0.000)	✓	1.060(0.000)	✓	1.061(0.000)	✓	1.062(0.000)	✓
2Fα	1.050(0.000)	✓	1.050(0.000)	✓	1.050(0.000)	✓	1.049(0.000)	✓	1.048(0.000)	✓
2Fαβ	1.056(0.000)	✓	1.056(0.000)	✓	1.057(0.000)	✓	1.057(0.000)	✓	1.058(0.000)	✓
	December 2005 to December 2023									
	$\vartheta = 5\%$									
	$\ell = 1$		$\ell = 2$		$\ell = 3$		$\ell = 4$		$\ell = 5$	
	QL	MCS	QL	MCS	QL	MCS	QL	MCS	QL	MCS
HN	1	✓	1		1	✓	1	✓	1	✓
CJOW	1.015(0.001)	✓	1.017(0.000)		1.018(0.000)	✓	1.018(0.000)	✓	1.018(0.000)	✓
2F	0.983(0.000)		0.986(0.000)		0.987(0.000)		0.988(0.000)	✓	0.988(0.000)	✓
2Fβ	1.054(0.000)	✓	1.054(0.000)	✓	1.053(0.000)	✓	1.050(0.000)	✓	1.049(0.000)	✓
2Fα	0.985(0.000)		0.985(0.000)		0.985(0.000)		0.985(0.000)		0.985(0.000)	
2Fαβ	1.027(0.000)	✓	1.028(0.000)		1.028(0.000)	✓	1.027(0.000)	✓	1.026(0.000)	✓
	January 1988 to December 2005									
	$\vartheta = 1\%$									
	$\ell = 1$		$\ell = 2$		$\ell = 3$		$\ell = 4$		$\ell = 5$	
	QL	MCS	QL	MCS	QL	MCS	QL	MCS	QL	MCS
HN	1		1		1		1		1	
CJOW	0.919(0.000)		0.927(0.000)		0.932(0.000)		0.934(0.000)		0.936(0.000)	
2F	1.257(0.000)	✓	1.236(0.000)	✓	1.231(0.000)	✓	1.221(0.000)	✓	1.224(0.000)	✓
2Fβ	1.244(0.000)	✓	1.234(0.000)	✓	1.233(0.000)	✓	1.223(0.000)	✓	1.221(0.000)	✓
2Fα	1.240(0.000)	✓	1.229(0.000)	✓	1.221(0.000)	✓	1.210(0.000)	✓	1.201(0.000)	✓
2Fαβ	1.237(0.000)	✓	1.234(0.000)	✓	1.234(0.000)	✓	1.230(0.000)	✓	1.226(0.000)	✓
	December 2005 to December 2023									
	$\vartheta = 1\%$									
	$\ell = 1$		$\ell = 2$		$\ell = 3$		$\ell = 4$		$\ell = 5$	
	QL	MCS	QL	MCS	QL	MCS	QL	MCS	QL	MCS
HN	1		1		1		1		1	
CJOW	1.097(0.000)	✓	1.104(0.000)	✓	1.111(0.000)	✓	1.114(0.000)	✓	1.119(0.000)	✓
2F	1.138(0.001)	✓	1.134(0.001)	✓	1.129(0.001)	✓	1.121(0.001)	✓	1.120(0.001)	✓
2Fβ	1.127(0.000)	✓	1.131(0.000)	✓	1.129(0.000)	✓	1.132(0.000)	✓	1.134(0.000)	✓
2Fα	1.101(0.002)	✓	1.099(0.001)	✓	1.096(0.001)	✓	1.095(0.001)	✓	1.094(0.002)	✓
2Fαβ	1.144(0.007)	✓	1.151(0.003)	✓	1.151(0.002)	✓	1.154(0.001)	✓	1.157(0.001)	✓

Values greater than one indicate better out-sample performance.

Table 2.6: Number of S&P500 option prices considered, sorted by moneyness and day-to-maturity (DTM).

	Number of option prices			
	Maturity			
Moneyness	$14 \leq DTM \leq 50$	$50 < DTM \leq 150$	$150 < DTM \leq 365$	All
$0.8 \leq K/S_t \leq 0.9$	156	573	1133	1862
$0.9 < K/S_t \leq 1.02$	817	1997	2527	5341
$1.02 < K/S_t \leq 1.2$	245	807	1117	2169
All	1218	3377	4777	9372

Table 2.7: Joint maximum likelihood estimation results.

	GARCH-HN	*GARCH-CJOW*	*GARCH-2F*	*GARCH-2Fβ*	*GARCH-2Fα*	*GARCH-2F$\alpha\beta$*
ω_1	6.529e-07***	8.109e-06***	1.207e-12	1.207e-12	1.207e-12	1.207e-12
	(2.828e-11)	(3.053e-08)	(3.195e-12)	(1.624e-12)	(1.804e-11)	(1.898e-12)
ω_2			1.207e-12	1.207e-12	4.928e-08***	4.928e-08***
			(2.887e-12)	(4.240e-12)	(4.081e-11)	(6.855e-12)
α_{11}	1.738e-06***	4.972e-06***	3.545e-11	3.545e-11***	3.545e-11	3.545e-11
	(2.279e-10)	(7.820e-09)	(3.411e-12)	(3.532e-12)	(5.671e-12)	(2.660e-12)
γ_1	335.931***	291.608***	132.891***	132.892***	132.891***	132.891***
	(0.016)	(0.443)	(0.043)	(0.091)	(0.028)	(0.014)
β_{11}	0.772***	0.907***	0.985***	0.985***	0.984***	0.985***
	(1.337e-04)	(0.002)	(1.107e-04)	(5.428e-04)	(1.617e-04)	(2.719e-04)
α_{22}		2.525e-06***	2.175e-06***	2.175e-06***	2.175e-06***	2.168e-06***
		(4.856e-07)	(2.618e-10)	(4.533e-10)	(1.704e-09)	(6.660e-10)
γ_2		149.761***	274.319***	274.323***	274.319***	274.318***
		(0.017)	(0.087)	(0.434)	(0.064)	(0.072)
β_{22}		0.741***	0.803***	0.802***	0.803***	0.804***
		(5.498e-04)	(1.007e-04)	(3.695e-04)	(8.350e-05)	(2.134e-04)
α_{12}			1.135e-11***	1.135e-11***		
			(1.475e-12)	(1.255e-12)		
α_{21}			1.230e-13	1.231e-13		
			(1.965e-12)	(3.158e-12)		
β_{12}			0.000		1.622e-05***	
			(4.466e-08)		(1.350e-08)	
β_{21}			3.220e-05***		1.622e-05***	
			(4.902e-09)		(1.414e-08)	
λ	0.158***	-6.816***	3.179***	3.177***	3.179***	3.178***
	(1.583e-05)	(0.007)	(5.890e-04)	(0.002)	(0.004)	(3.443e-04)
Log-likelihood	28,039	27,251	28,135	28,135	28,126	28,118
AIC	-56,068	-54,486	-56,244	-56,248	-56,230	-56,218
BIC	-56,029	-54,443	-56,142	-56,162	-56,144	-56,148
σ_e^2	0.084	0.076	0.085	0.083	0.085	0.086
Persistence	0.97	0.96 | 0.81	0.98 | 0.97	0.98 | 0.97	0.98 | 0.97	0.98 | 0.97

The estimation period spans from January 5, 1988 to December 29, 2023 (9069 observations) for the returns, and spans from February 10, 2021 to November 29, 2023 (9372 observations) for the option prices. The standard errors, reported in parenthesis, are computed by inverting the negative Hessian matrix evaluated at the optimum parameter values. The variance σ_e^2 is calculated as the empirical variance estimator of the N option implied volatility errors. [*],[**],[***] denote statistical significance at the 0.1, 0.05, 0.01 levels, respectively.

Table 2.8: Implied volatility root mean square error, in percentage points.

Panel A: *IVRMSE(%)* overall						
	HN	*CJOW*	*F2*	*F2β*	*F2α*	*F2αβ*
	5.383	5.573	5.273	5.273	5.200	5.161
Normalized	1.000	1.035	0.980	0.980	0.966	0.959
Panel B: *IVRMSE(%)* by moneyness						
Moneyness	*HN*	*CJOW*	*F2*	*F2β*	*F2α*	*F2αβ*
$0.8 \leq K/S_t \leq 0.9$	7.212	7.418	6.936	6.935	6.831	6.765
Normalized	1.000	1.029	0.962	0.962	0.947	0.938
$0.9 < K/S_t \leq 1.02$	4.918	5.091	4.835	4.834	4.769	4.735
Normalized	1.000	1.035	0.983	0.983	0.970	0.963
$1.02 < K/S_t \leq 1.2$	4.874	5.103	4.873	4.873	4.810	4.784
Normalized	1.000	1.047	1.000	1.000	0.987	0.982
Panel C: *IVRMSE(%)* by day-to-maturity (DTM)						
Maturity	*HN*	*CJOW*	*F2*	*F2β*	*F2α*	*F2αβ*
$14 \leq DTM \leq 50$	5.902	5.658	5.819	5.818	5.746	5.712
Normalized	1.000	0.959	0.986	0.986	0.974	0.968
$50 < DTM \leq 150$	5.050	5.200	4.949	4.948	4.872	4.832
Normalized	1.000	1.030	0.980	0.980	0.965	0.957
$150 < DTM \leq 365$	5.478	5.809	5.355	5.355	5.284	5.243
Normalized	1.000	1.061	0.978	0.978	0.965	0.957

Chapter 3

GARCH option valuation with multiple volatility components: Ensuring variance positivity via telescopic autoregression

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name

Department of Statistical Sciences

"Paolo Fortunati"

Doctor of Philosophy in Statistical Sciences

Essays on Latent Factor Models in Financial Markets:

A Focus on Commodity Prices and Volatility Dynamics

by Christian TEZZA

Christoffersen, Jacobs, Ornthanalai, and Wang (2008) (CJOW) introduced an enhanced Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model for valuing European options, incorporating two distinct volatility components. Empirical studies show that CJOW's model outperforms widely used single-component GARCH models. However, a key drawback is its allowance for negative variance values. Oh and Park (2023) (OP) partially addressed this issue with a related model, but the positivity of its volatility components remains neither theoretically nor empirically guaranteed. Bormetti et al. (2015) proposed a general framework that includes the CJOW and OP models, deriving specific conditions to ensure variance positivity. In this paper, we introduce a novel GARCH model within this broader framework and establish more general conditions for maintaining positive variance, extending beyond those proposed by Bormetti et al. (2015). Our model incorporates a telescopic autoregressive structure, enabling the construction of multi-component telescopic GARCH models. This, we further extend this approach to a three-component model, enhancing flexibility in capturing complex volatility dynamics. Compared to previous GARCH models, our methodology delivers comparable in-sample performance on returns data while achieving superior results in pricing S&P500 options.

3.1 Introduction

Building on the pioneering work of Engle and Lee (1999), Christoffersen, Jacobs, Ornthanalai, and Wang (2008) (CJOW) proposed an interesting GARCH model with two volatility components, hereafter referred to as the *GARCH-CJOW* model. One volatility component is a long-run component that can be modeled as a mean-reverting or as a fully persistent process, while the other is a short-run mean-reverting component with zero mean. As shown by the authors, the model allows for an efficient valuation of European options and its most distinctive feature is its ability to accurately describe the implied volatility term structure. In particular, the empirical analysis that CJOW conducted on both long-maturity and short-maturity options demonstrates superior pricing performance compared to the popular GARCH model with a single volatility component developed by Heston and Nandi (2000) and a GARCH(1,1) model augmented with Poisson-normal jumps. Owing to its well-documented empirical performance, the *GARCH-CJOW* model has gained popularity in the GARCH literature and has been used for pricing S&P500 options in Corsi et al. (2013) and VIX futures in Cheng et al. (2023).

Despite its success, Bormetti et al. (2015) and Oh and Park (2023) provide evidence that the *GARCH-CJOW* model fails to guarantee positive volatilities for parameter sets commonly encountered in financial practice. This raises significant modeling concerns for out-of-sample (volatility) analysis and for option pricing, especially for medium to long-maturing options. Even though Bormetti et al. (2015) state that the likelihood of obtaining negative volatilities is extremely low, Oh and Park (2023) further analyze the performances of the *GARCH-CJOW* model and find that it can generate a consistent number of negative volatility trajectories. Then, to address the issue, they propose an enhancement of the *GARCH-CJOW* model, hereafter referred to as *GARCH-OP*, which allows for the pricing of SPX options and VIX derivatives. However, this model also does not ensure positivite volatility.

In this paper, we investigate further the issue of negative volatility of the *GARCH-CJOW* model and its implications for option pricing. As previously established in Oh and Park (2023), we confirm that, for parameter sets commonly used in the empirical literature, the *GARCH-*

CJOW model consistently generates a significant number of negative volatility trajectories. Additionally, the *GARCH-OP* model also produces negative volatilities, a concern that becomes more pronounced for longer time horizons. We also find that this issue affects the computation of option prices, as it is based on an inversion formula where the integrands for longer maturities may diverge.

To fix these issues, we propose an improvement to the model developed by CJOW that ensures the positivity of the volatility process by properly specifying the impact of the innovation on volatility. This new model, which allows for a well-posed option semi-closed form valuation, will be referred to as corrected positive component GARCH model, or, in short, *GARCH-CPC*. In contrast to the *GARCH-CJOW* model, the semi-closed formula of the *GARCH-CPC* model remains reliable in option pricing scenarios involving large option maturities.

Moreover, we compare the performance of the *GARCH-CJOW*, *GARCH-CPC* and *GARCH-OP* models in the valuation of a large panel of options written on the S&P500 index. We find that the novel *GARCH-CPC* model offers a feasible and efficient solution for computing option prices using affine formulas with greater accuracy than both the *GARCH-CJOW* and the *GARCH-OP* models. The proposed model falls within the broader modeling framework introduced by Bormetti et al. (2015) and further generalized in **?**. Nevertheless, it incorporates a less restrictive positivity condition, allowing the difference between volatility components to change sign over time, which enhances the model's flexibility. In addition, the *GARCH-CPC* model follows a telescopic structure. Building on this, we explore a potential extension to a three-component version with the same telescopic structure, which provides a comparable goodness-of-fit to the two-component model on both returns and options data.

The remainder of this paper is organized as follows. Section 3.2 briefly revisits the *GARCH-CJOW* and *GARCH-OP* models, highlighting the negativity issues inherent in these approaches and their implications for option pricing. Section 3.3 introduces the *GARCH-CPC* model. We derive its risk-neutral dynamics and provide the corresponding option valuation formulas. Section 3.4 presents an empirical analysis using return and option data. Section 3.5 explores a telescopic extension to three components. Finally, Section 3.6 offers concluding remarks.

3.2 Return dynamics with volatility components

The volatility component model proposed by CJOW for the return process $R_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, where S_t denotes the spot price, is given by the following equations (under the physical measure):

$$R_{t+1} = r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}, \tag{3.1}$$

$$h_{t+1} = q_{t+1} + \tilde{\beta}(h_t - q_t) + \alpha\left(Z_t^2 - 1 - 2\gamma_1\sqrt{h_t} Z_t\right), \tag{3.2}$$

$$q_{t+1} = \omega + \rho q_t + \varphi\left(Z_t^2 - 1 - 2\gamma_2\sqrt{h_t} Z_t\right), \tag{3.3}$$

where r denotes the risk-free rate, $Z_t \overset{iid}{\sim} N(0,1)$. CJOW refer to q_t and $h_t - q_t$ as the long-run and the short-run variance component, respectively. The parameter ω sets the long-term means of both volatility components, while $\tilde{\beta}$ and ρ represent the persistence of the short-term and long-term components of volatility. The parameters α and φ capture the effects of mean-zero stochastic terms, and γ_1 and γ_2 model the leverage effect. CJOW impose the parameter constraints $\omega \geq 0$, $\alpha > 0$, $\varphi > 0$, $\tilde{\beta} < 1$ and $\rho < 1$. In the long-run we have that $\mathbb{E}[h_t] = \mathbb{E}[q_t] = \omega/(1-\rho)$.

As pointed out in Bormetti et al. (2015) and Oh and Park (2023), the model (3.1)-(3.3) does not guarantee that the total variance h_t remains positive. Therefore, in Oh and Park (2023), the following model, hereafter referred to as *GARCH-OP*, has been proposed as a more robust alternative to address the negative volatility issue that affects the *GARCH-CJOW* model:

$$R_{t+1} = r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}, \tag{3.4}$$

$$h_{t+1} = q_{t+1} + \tilde{\beta}(h_t - q_t) + \alpha\left(Z_t - \gamma_1\sqrt{h_t}\right)^2 - \omega - \alpha\gamma_1^2 h_t, \tag{3.5}$$

$$q_{t+1} = \omega + \rho q_t + \varphi\left(Z_t - \gamma_2\sqrt{h_t}\right)^2, \tag{3.6}$$

where the long-term means are given by $\begin{pmatrix} \mathbb{E}[h_t] & \mathbb{E}[q_t] \end{pmatrix}^\top = (I_2 - P)^{-1}R$, where I_2 denote a 2×2 identity matrix, $P = \begin{pmatrix} \tilde{\beta} + \varphi\gamma_2^2 & \rho - \tilde{\beta} \\ \varphi\gamma_2^2 & \rho \end{pmatrix}$ and $R = \begin{pmatrix} \alpha + \varphi \\ \omega + \varphi \end{pmatrix}$.

We note that both the *GARCH-CJOW* and *GARCH-OP* models can be derived from the framework proposed by Bormetti et al. (2015). Specifically, by setting

$$f_t^1 = h_t - q_t, \quad f_t^2 = q_t,$$

we introduce the linear operator $\boldsymbol{\mathcal{L}} : \mathbb{R}^2 \to \mathbb{R}_+$, which acts on $\boldsymbol{f}_t = (f_t^1, f_t^2)^\top$ and provide the log-return variance as

$$\mathcal{L}(\boldsymbol{f}_t) = f_t^1 + f_t^2.$$

Then, we can rewrite the *GARCH-CJOW* and *GARCH-OP* model as follows:

$$R_{t+1} = r + \lambda\mathcal{L}(\boldsymbol{f}_{t+1}) + \sqrt{\mathcal{L}(\boldsymbol{f}_{t+1})}Z_{t+1}, \tag{3.7}$$

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{M}_1\boldsymbol{f}_t + \boldsymbol{N}_1 \begin{pmatrix} \left(Z_t - \gamma_1\sqrt{\mathcal{L}(\boldsymbol{f}_t)}\right)^2 \\ \left(Z_t - \gamma_2\sqrt{\mathcal{L}(\boldsymbol{f}_t)}\right)^2 \end{pmatrix}, \tag{3.8}$$

where for the *GARCH-CJOW* model we have

$$\boldsymbol{\omega} = \begin{pmatrix} -\alpha \\ \omega - \varphi \end{pmatrix}, \quad \boldsymbol{M}_1 = \begin{pmatrix} \tilde{\beta} - \alpha\gamma_1^2 & -\alpha\gamma_1^2 \\ -\varphi\gamma_2^2 & \rho - \varphi\gamma_2^2 \end{pmatrix}, \quad \boldsymbol{N}_1 = \begin{pmatrix} \alpha & 0 \\ 0 & \varphi \end{pmatrix},$$

and for the *GARCH-OP* model we have

$$\boldsymbol{\omega} = \begin{pmatrix} -\omega \\ \omega \end{pmatrix}, \quad \boldsymbol{M}_1 = \begin{pmatrix} \tilde{\beta} - \alpha\gamma_1^2 & -\alpha\gamma_1^2 \\ 0 & \rho \end{pmatrix}, \quad \boldsymbol{N}_1 = \begin{pmatrix} \alpha & 0 \\ 0 & \varphi \end{pmatrix}.$$

3.2.1 Model issues

In this section, we will document the issues of model (3.1)-(3.3) and model (3.4)-(3.6) related to the negative values that the variance can take through a few test cases. Using parameters established in the empirical literature, we will show via statistical simulation how many variance trajectories h_t can turn negative, and we will examine the implications for option pricing.

Negative volatility trajectories

For the model (3.1)-(3.3) we initially consider two sets of parameters available in the literature, namely, those in Christoffersen et al. (2008) and in Cheng et al. (2023), which we denote as *CJOW08* and *CCLT23*, respectively. These parameters are reported in Table 3.1. We note that equation (1) in Cheng et al. (2023) specifies $\lambda - \frac{1}{2}$ instead of just λ as in equation (3.1) in the present paper, so in Table 3.1 we adjusted the value of λ accordingly. Whereas, for model in (3.4)-(3.6) we consider the same parameters as in Oh and Park (2023), which we denote *OP23* and we report in Table 3.1.

Table 3.1: Model parameters.

Parameter	ω	α	γ_1	$\tilde{\beta}$	φ	γ_2	ρ	λ
CJOW08	8.208e-07	1.580e-06	415.100	0.6437	2.480e-06	63.240	0.9896	2.092
CCLT23	7.776e-07	1.380e-06	402.352	0.862	1.795e-06	73.205	0.991	1.357
OP23	-1.57e-06	0.190e-06	7050	0.922	2.62e-06	89	0.983	-7.88

We perform a Monte Carlo exercise where we simulate the trajectories of the variance in equations (3.2) and (3.5) to check if, and how many of them, turn negative. The simulation consists in generating 1,000,000 paths of innovations Z_t from $t = 0$ to a given time horizon $T > 0$. Using equations (3.2)-(3.3) and (3.5)-(3.6), for each path we compute the variance h_t and count the number of trajectories such that h_t becomes negative for some $t \leq T$. We set $S_0 = 100$, $r = 10^{-5}$ and, following Oh and Park (2023), we choose q_0 and h_0 such that the (annualized) initial volatility states $\sqrt{252q_0}$ and $\sqrt{252h_0}$ are equal to 5% and 10%, respectively. The results, illustrated in Table 3.2, show that as T increases, the number of negative trajectories also

increases for both the *GARCH-CJOW* and *GARCH-OP* models. However, as documented in Oh and Park (2023), the number of negative trajectories generated by the *GARCH-OP* model is significantly lower compared to the *GARCH-CJOW* model.

Table 3.2: Negative trajectories for the variance h_t out of 1,000,000 simulations for time horizon T.

	Initial volatility $\left(\sqrt{252q_0}, \sqrt{252h_0}\right)$					
	(5% , 5%)			(10% , 10%)		
	GARCH-CJOW		*GARCH-OP*	*GARCH-CJOW*		*GARCH-OP*
T	*CJOW08*	*CCLT23*	*OP23*	*CJOW08*	*CCLT23*	*OP23*
15	226,386	185,403	2,328	0	0	0
30	287,888	235,937	7,848	317	0	0
50	315,161	251,841	10,422	3,671	115	7
80	330,745	258,352	11,023	10,034	481	33
120	339,795	260,234	11,112	16,883	891	41
252	351,374	261,183	11,114	29,129	1,465	44

Option pricing

In this subsection, after briefly recalling formulas commonly used to compute option prices, we empirically analyze the consequences of generating negative volatility trajectories for option valuation in the *GARCH-CJOW* and *GARCH-OP* models.

As shown by Heston and Nandi (2000) and further utilized by CJOW, the inversion formula developed in Gil-Pelaez (1951) yields the following expression for a European call option on a non-dividend paying stock with spot price S_t, strike price K and expiration date T:

$$\begin{aligned} C_t = \frac{1}{2}S_t + \frac{e^{-r(T-t)}}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi}f^*(t,T;i\phi+1)}{i\phi}\right]d\phi \\ - Ke^{-r(T-t)}\left(\frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{K^{-i\phi}f^*(t,T;i\phi)}{i\phi}\right]d\phi\right), \end{aligned} \tag{3.9}$$

where $\mathrm{Re}[\cdot]$ denotes the real part of a complex number and $f^*(t,T;i\phi)$ represents the conditional characteristic function of the terminal log-stock price under the risk-neutral measure. The risk-

neutral conditional characteristic function of the *GARCH-CJOW* and *GARCH-OP* models can be computed as follows:

$$f^*(t,T;\phi) = S_t^{\phi} \exp\left\{A_t(\phi) + B_{1,t}(\phi)(h_{t+1} - q_{t+1}) + B_{2,t}(\phi)q_{t+1}\right\}, \quad t < T, \tag{3.10}$$

where the expressions of $A_t(\phi)$, $B_{1,t}(\phi)$ and $B_{2,t}(\phi)$ can be obtained using equations (25) in the paper by Christoffersen et al. (2008) for the *GARCH-CJOW* model and using equations (8) in the paper by Oh and Park (2023) for the *GARCH-OP* model.

Let us now focus on the integrand functions in (3.9):

$$I_1 = \text{Re}\left[\frac{K^{-i\phi} f^*(t,T;i\phi+1)}{i\phi}\right], \quad I_2 = \text{Re}\left[\frac{K^{-i\phi} f^*(t,T;i\phi)}{i\phi}\right]. \tag{3.11}$$

For both the *GARCH-CJOW* and *GARCH-OP* models, we check the integrability of I_1 and I_2, by analyzing the behavior of I_1 and I_2 as functions of ϕ and for different option maturities $T = 15, 30, 50, 80, 120, 252$. For comparison purposes, we also consider the values of I_1 and I_2 for the popular GARCH model developed in Heston and Nandi (2000), hereafter *GARCH-HN*:

$$R_{t+1} = r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}, \tag{3.12}$$

$$h_{t+1} = \omega + \tilde{\beta} h_t + \alpha \left(Z_t - \gamma_1 \sqrt{h_t}\right)^2. \tag{3.13}$$

Similarly to the *GARCH-CJOW* model, for the *GARCH-HN* model we use parameters obtained by Christoffersen et al. (2008) and Cheng et al. (2023), which are reported in Table 3.3. We note that, for the *GARCH-HN* model, option prices can be computed using equation (3.9) in conjunction with the risk-neutral conditional characteristic function obtained by Heston and Nandi (2000).

Table 3.3: Parameters of the *GARCH-HN* obtained by CJOW and Cheng et al. (2023).

Parameter	ω	α	γ_1	$\tilde{\beta}$	λ
CJOW08	2.101e-17	3.317e-06	1.276e+02	0.9552	2.231
CCLT23	1.744e-06	3.098e-06	120.967	0.935	1.395

Figures 3.1-3.5 show the behavior of $|I_1|$ and $|I_2|$ for the *GARCH-HN*, *GARCH-CJOW* and *GARCH-OP* models. In particular, Figures 3.1 and 3.2 display the behavior of $|I_1|$ and $|I_2|$, respectively, for the *GARCH-HN* model using the parameters *CJOW08* and *CCLT23* in Table 3.1. As the initial (annualized) volatility state we set $\sqrt{252h_0}$ equal to either 5% or 10% and we set $r = 10^{-5}$ and $S_0 = K = 100$. Instead, for the *GARCH-CJOW* model we show the behavior of $|I_1|$ and $|I_2|$ in Figures 3.3 and 3.4 and for the *GARCH-OP* model in Figure 3.5. As for the initial volatility states, again we set $\sqrt{252h_0}$ and $\sqrt{252q_0}$ equal to 5% and 10%, and we set $r = 10^{-5}$ and $S_0 = K = 100$.

As we may notice, for both the *CJOW08* and *CCLT23* parameters, the *GARCH-HN* model does not show any integrability issue for I_1 and I_2 in equation (3.11), whereas Figures 3.3, 3.4 and 3.5 clearly show that I_1 and I_2 explode for almost all the considered maturities. Only for $T = 15$ and for the 10% initial volatility state both the *GARCH-CJOW* and *GARCH-OP* models do not exhibit any issues.

Figure 3.1: Behavior of I_1 (left) and I_2 (right), as functions of ϕ for different values of T for the *GARCH-HN* model. We set $S_0 = K = 100$, $r = 10^{-5}$ and $\sqrt{252h_0} = 5\%$. We use the *CJOW08* parameters (top panels) and the *CCLT23* parameters as in Table 3.1 (bottom panels).

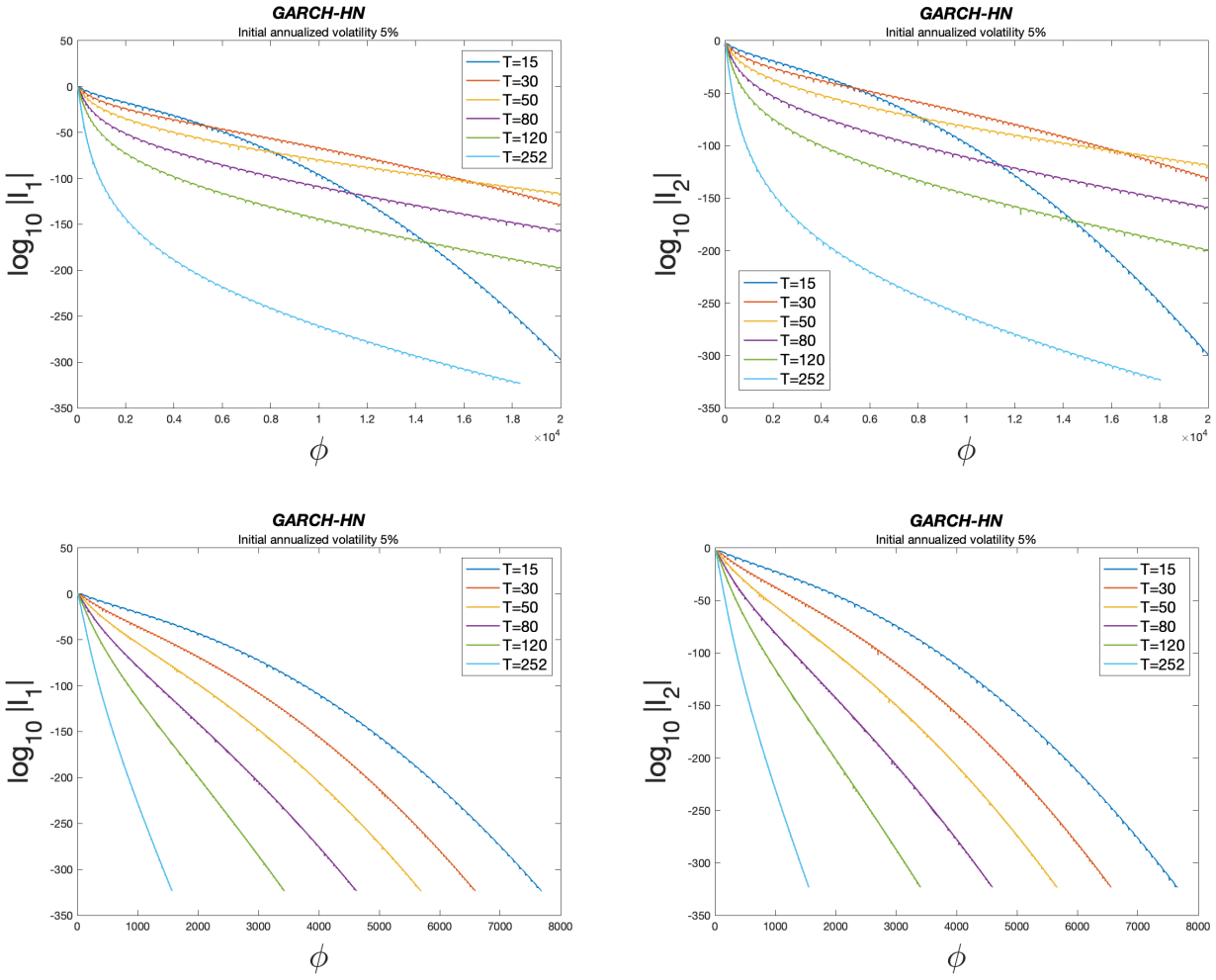


Figure 3.2: Behavior of I_1 (left) and I_2 (right), as functions of ϕ for different values of T for the *GARCH-HN* model. We set $S_0 = K = 100$, $r = 10^{-5}$ and $\sqrt{252h_0} = 10\%$. We use the *CJOW08* parameters (top panels) and the *CCLT23* parameters as in Table 3.1 (bottom panels).

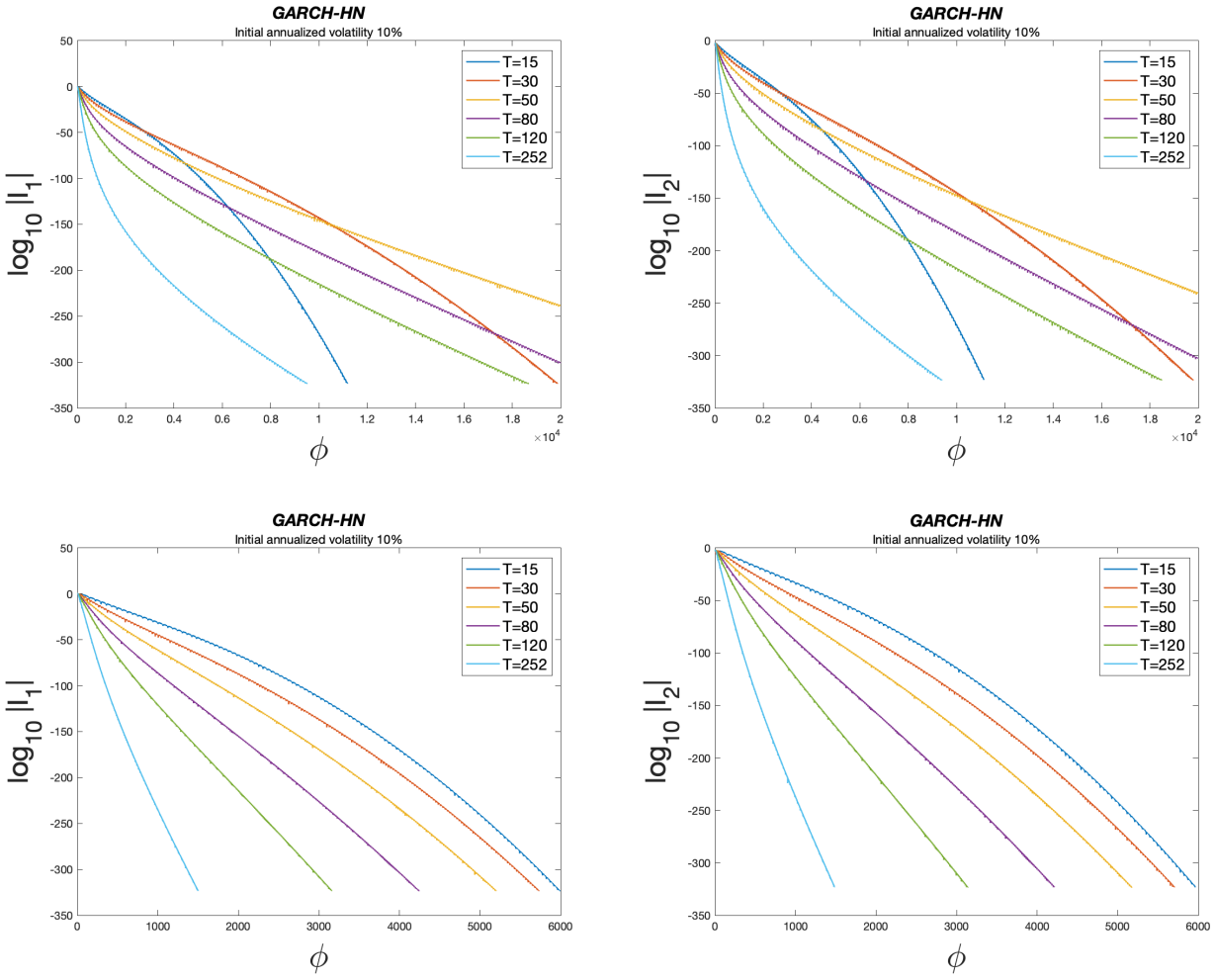


Figure 3.3: Behavior of I_1 (left) and I_2 (right), as functions of ϕ for different values of T for the *GARCH-CJOW* model. We used $S_0 = K = 100$, $r = 10^{-5}$ and the initial volatility states $\sqrt{252h_0} = 5\%$ and $\sqrt{252q_0} = 5\%$. We use the *CJOW08* parameters (top panels) and the *CCLT23* parameters as in Table 3.1 (bottom panels).

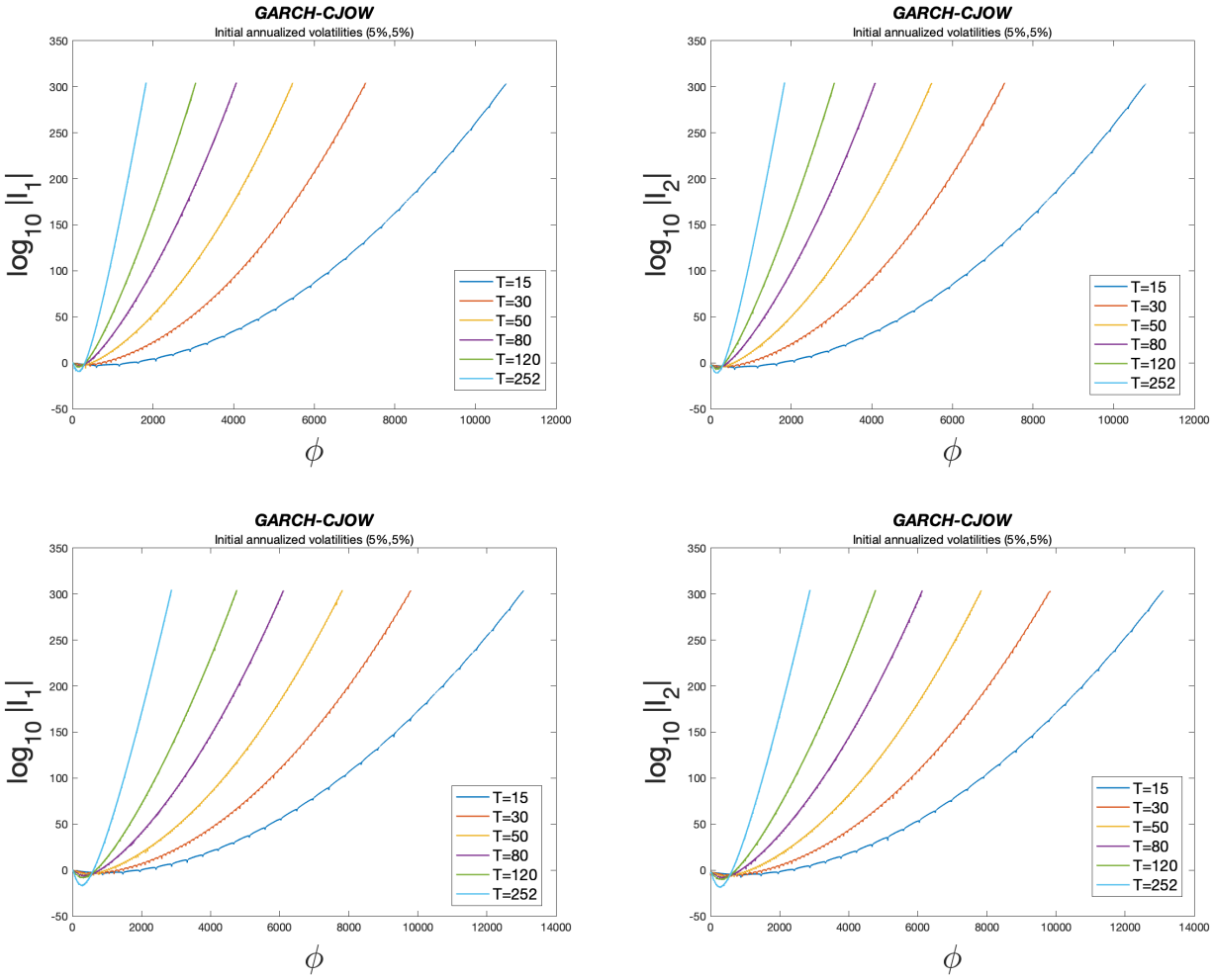


Figure 3.4: Behavior of I_1 (left) and I_2 (right), as functions of ϕ for different values of T for the *GARCH-CJOW* model. We used $S_0 = K = 100$, $r = 10^{-5}$ and the initial volatility states $\sqrt{252h_0} = 10\%$ and $\sqrt{252q_0} = 10\%$. We use the *CJOW08* parameters (top panels) and the *CCLT23* parameters as in Table 3.1 (bottom panels).

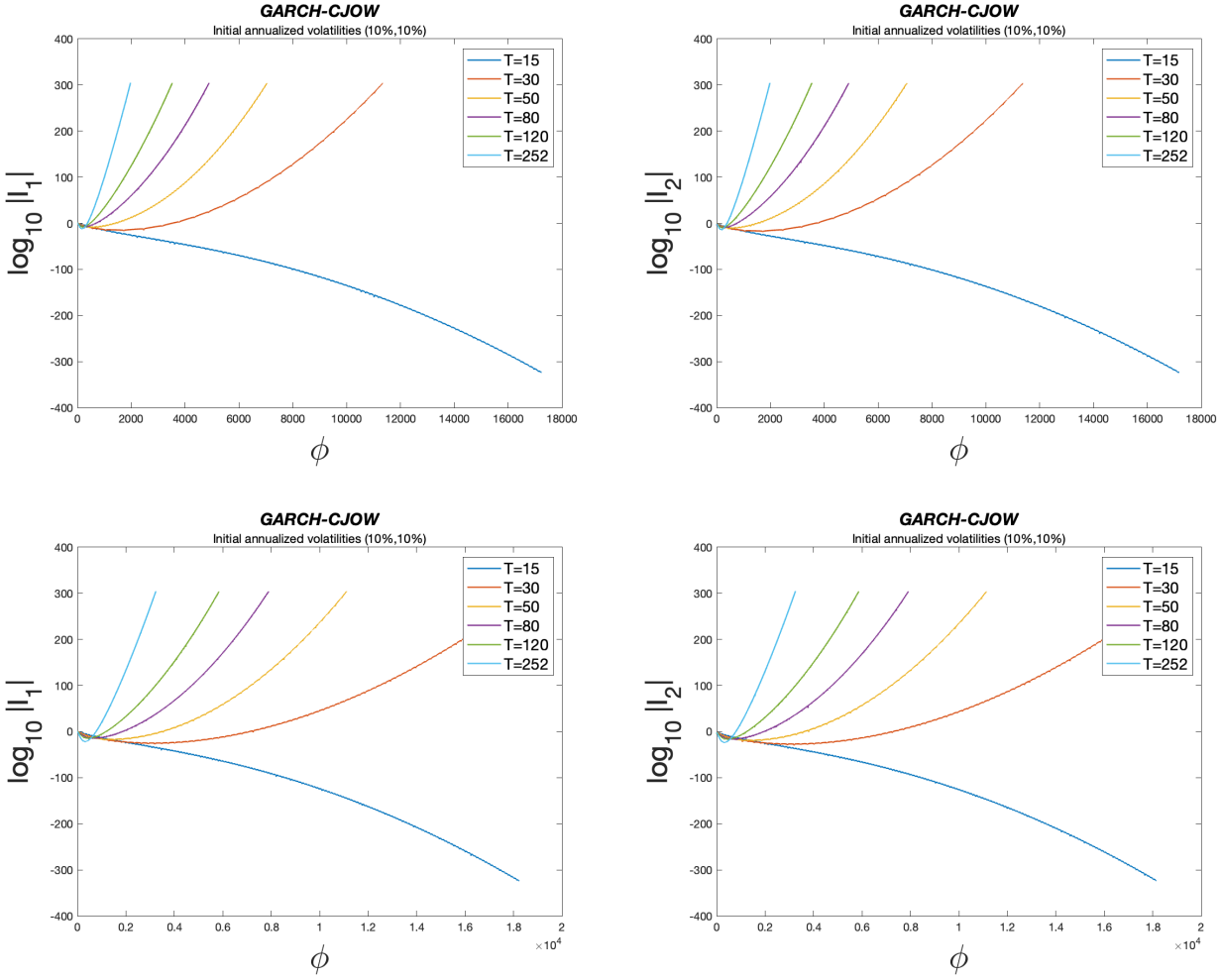
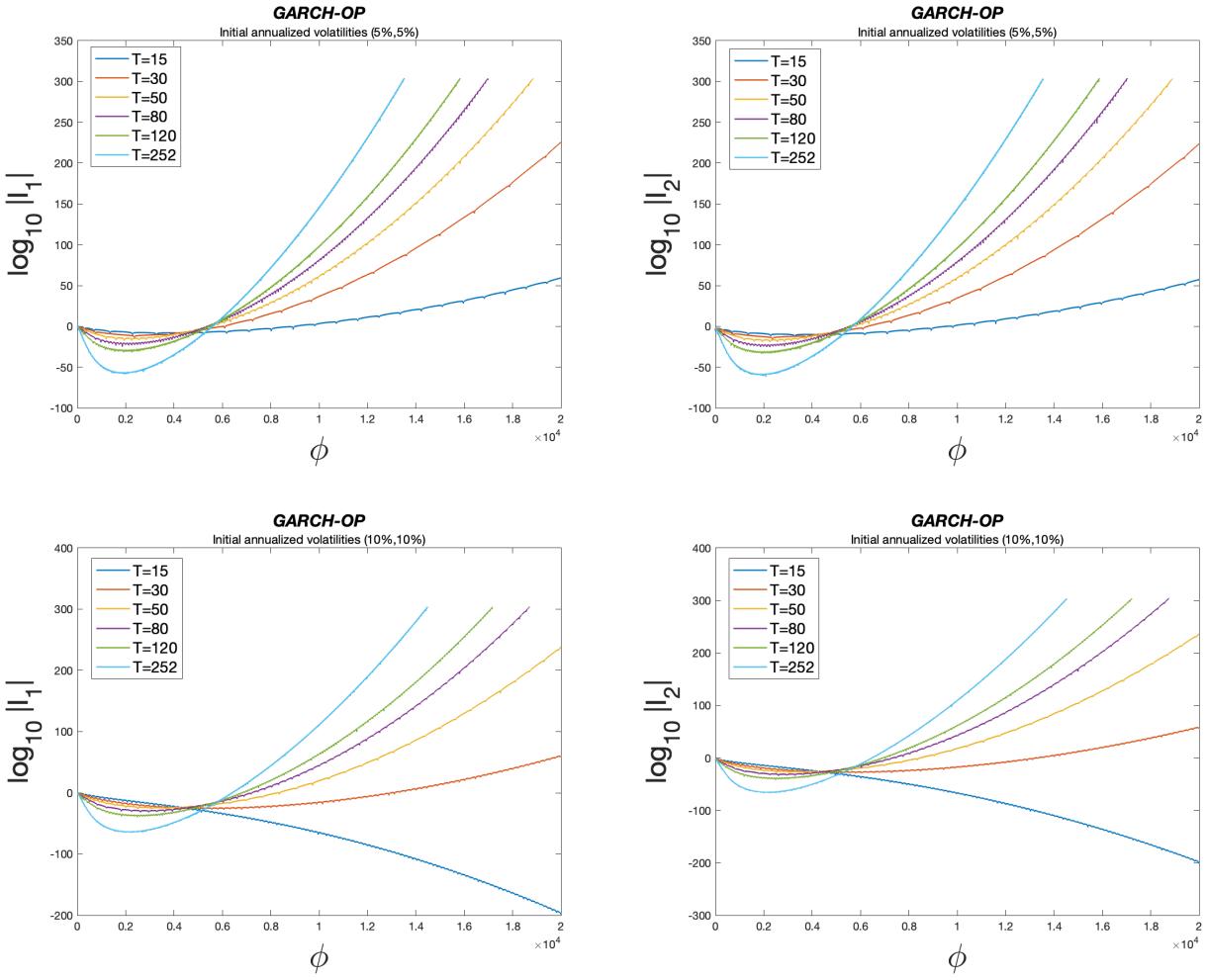


Figure 3.5: Behavior of I_1 (left) and I_2 (right), as functions of ϕ for different values of T for the *GARCH-OP* model with the parameters *OP23* of Table 3.1. We set $S_0 = K = 100$, $r = 10^{-5}$. We use $\sqrt{252h_0} = \sqrt{252q_0} = 5\%$ (top panels) and $\sqrt{252h_0} = \sqrt{252q_0} = 10\%$ (bottom panels).



3.3 An improved two-component GARCH model for option valuation

In this section, we propose a component volatility model that enhances the one developed by CJOW by ensuring that the volatility remains positive. We name this model *GARCH-CPC*, which stands for Corrected Positive Component GARCH. The model consists of the following equations:

$$R_{t+1} = r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}, \tag{3.14}$$

$$h_{t+1} = q_{t+1} + \tilde{\beta}(h_t - q_t) + \alpha \left\{ \left(Z_t - \gamma_1 \sqrt{h_t} \right)^2 - \gamma_1^2 q_t \right\}, \tag{3.15}$$

$$q_{t+1} = \omega + \rho q_t + \varphi \left(Z_t - \gamma_2 \sqrt{h_t} \right)^2, \tag{3.16}$$

where $Z_t \overset{iid}{\sim} N(0,1)$ and q_t represents a process related to the long-run volatility component as in the *GARCH-CJOW* model.

To ensure the positivity of h_t, we assume the following conditions for the parameters: $\omega \geq 0$, $\alpha > 0$, $\varphi > 0$ and

$$\tilde{\beta} + \alpha \gamma_1^2 < \rho, \quad \tilde{\beta} > 0, \quad \rho > 0. \tag{3.17}$$

Substituting equation (3.16) into (3.15) immediately shows that these parameter conditions are sufficient to guarantee the positivity of h_t for all t. We remark that the *GARCH-CPC* model is included in the modeling framework of Bormetti et al. (2015), using in equation (3.8) the following specification for the coefficients:

$$\boldsymbol{\omega} = \begin{pmatrix} 0 \\ \omega \end{pmatrix}, \quad \boldsymbol{M}_1 = \begin{pmatrix} \tilde{\beta} & -\alpha\gamma_1^2 \\ 0 & \rho \end{pmatrix}, \quad \boldsymbol{N}_1 = \begin{pmatrix} \alpha & 0 \\ 0 & \varphi \end{pmatrix}. \tag{3.18}$$

Moreover, condition (3.17) relaxes the positivity conditions previously imposed by Bormetti et al. (2015). In particular, Bormetti et al. (2015) require that $\boldsymbol{M}_1$ is component-wise positive, which specifically implies that $h_t \geq q_t$ for all t. This condition ensures that the total volatility is always greater than or equal to the long-term volatility, enforcing component-wise positivity to guarantee variance positivity. Conversely, in our approach, equation (3.17) does not impose any restrictions on the sign of the difference $h_t - q_t$, similar to what occurs in the CJOW model.

The long-run means are given by

$$\begin{pmatrix} \mathbb{E}[h_t] & \mathbb{E}[q_t] \end{pmatrix}^\top = (I_2 - P)^{-1} R,$$

where I_2 denote a 2×2 identity matrix, $P = \begin{pmatrix} \tilde{\beta} + \alpha\gamma_1^2 + \varphi\gamma_2^2 & \rho - \tilde{\beta} - \alpha\gamma_1^2 \\ \varphi\gamma_2^2 & \rho \end{pmatrix}$ and $R = \begin{pmatrix} \omega + \alpha + \varphi \\ \omega + \varphi \end{pmatrix}$.

To avoid the divergence of the recursive equations (3.15)-(3.16) we must ensure that the spectrum of the matrix P is bounded by one. It readily follows that eigenvalues of P are $\lambda_1 = \tilde{\beta} + \alpha\gamma_1^2$ and $\lambda_2 = \rho + \varphi\gamma_2^2$, so that in addition to condition (3.17) we need to assume

$$|\tilde{\beta} + \alpha\gamma_1^2| < 1, \quad |\rho + \varphi\gamma_2^2| < 1. \tag{3.19}$$

We note that, for the model in equations (3.2)-(3.3), $\mathbb{E}[h_t - q_t] = 0$. In contrast, for equations (3.5)-(3.6) it holds that $\mathbb{E}[h_t - q_t] = \alpha/(1-\tilde{\beta})$ and for equations (3.15)-(3.16) we have $\mathbb{E}[h_t - q_t] = \alpha/(1 - \tilde{\beta} - \alpha\gamma_1^2)$. Therefore, unlike in the *GARCH-CJOW* model, the long-run volatility component in the *GARCH-OP* and *GARCH-CPC* models is represented not by q_t, but rather by $q_t + \bar{h}$, where $\bar{h} = \frac{\alpha}{1-\tilde{\beta}}$ for the *GARCH-OP* model, and $\bar{h} = \frac{\alpha}{1-\tilde{\beta}-\alpha\gamma_1^2}$ for the *GARCH-CPC* model.

3.3.1 The risk-neutral *GARCH-CPC* dynamics

Following Christoffersen et al. (2008), we risk-neutralize the dynamics for the *GARCH-CPC* model as follows:

$$R_{t+1} = r - \frac{1}{2}h_{t+1} + \sqrt{h_{t+1}}Z^*_{t+1}, \tag{3.20}$$

$$h_{t+1} = q_{t+1} + \tilde{\beta}(h_t - q_t) + \alpha\left(Z^*_t - \gamma^*_1\sqrt{h_t}\right)^2 - \alpha(\gamma^*_1)^2 q_t, \tag{3.21}$$

$$q_{t+1} = \omega + \rho q_t + \varphi\left(Z^*_t - \gamma^*_2\sqrt{h_t}\right)^2, \tag{3.22}$$

with $\gamma^*_i = \gamma_i + \frac{1}{2} + \lambda$ for $i = 1, 2$ and $Z^*_t \sim N(0, 1)$.

3.3.2 Moment generating function

To perform option valuation, we derive the moment generating function associated to the terminal log-spot price of equation (3.20).

Proposition 3.1. *The conditional moment generating function for the logarithm of the terminal log-price* $\ln S_T$*, denoted by* $f(t, T; \phi)$*, can be computed as*

$$f(t, T; \phi) = e^{\phi \log S_t + A_t(\phi) + B_{1,t}(\phi)(h_{t+1} - q_{t+1}) + B_{2,t}(\phi)q_{t+1}}, \quad t < T, \tag{3.23}$$

where the expressions for $A_t(\phi)$*,* $B_{1,t}(\phi)$ *and* $B_{2,t}(\phi)$ *are derived in equation* (B.5) *in the Appendix B.1.1.*

Proof. See Appendix B.1.1. □

3.4 Empirical results for returns and options

In this section, we examine the goodness-of-fit of the *GARCH-CJOW*, *GARCH-OP* and the *GARCH-CPC* models on both returns and options data.

The data for the daily levels of Adjusted Close Price for the S&P500 returns series is retrieved from Refinitiv Datastream. As a proxy of the risk-free interest rate, we utilize the 3-month

Treasury Bill rate and we gathered its time series data from the Federal Funds Effective Rate (FRED) dataset. For the returns, we considered two different periods: one covering the same period considered by CJOW, from July 2, 1962, to December 31, 2001 (labeled as "Period 1") resulting in 9943 days, and another covering more recent data from January 2, 2002, to December 29, 2023 (labeled as "Period 2") for a total of 5537 days.

We consider both Put and Call European options written on the S&P500 index, with data retrieved from Thomson Reuters Eikon Datastream. The options we examined have maturities ranging from 2020 to 2023 and we consider the option daily prices from February 10, 2020, to December 29, 2023. As a common practice, following Christoffersen et al. (2012), Ballestra et al. (2023), and Ballestra et al. (2024), we apply several exclusion filters retaining a total of 14,247 options prices. In particular, we keep only the options with time-to-maturity between 14 and 365 days and we select only out-of-the-money Put and Call options (we compute the moneyness as K/S_t, where K is the strike price and S_t is the underlying index level), and we filter out illiquid quotes by selecting only the six most liquid strikes at each maturity, and we consider option quotes only on Wednesday. Finally, we remove price quotes lower than 3.8$.

Following CJOW, the estimation we conducted relies on maximum likelihood estimation using only the log-returns. The results for Period 1 and Period 2 are shown in Table 3.4. For Period 1, since we considered the same time frame as CJOW, we included the parameters they estimated in their article. To evaluate the goodness-of-fit on the returns data, we also include log-likelihood values alongside standard information criteria such as AIC and BIC. Finally, the fit to the option data is assessed using a standard metric, the implied volatility root mean square error:

$$\text{IVRMSE}(\%) = 100 \times \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(IV_i^{MKT} - IV_i^{MOD}\right)^2}, \tag{3.24}$$

where IV_i^{MKT} and IV_i^{MOD} denote the market and the model implied volatilities of the i-th option price, respectively.

Generally speaking, for both periods, the *GARCH-CPC* and *GARCH-OP* models exhibit

slightly worse fit on the returns data compared to the *GARCH-CJOW*, as indicated by AIC and BIC values. However, for the *GARCH-CJOW* model it was not possible to calculate all the option prices due to the issues of negative variance outlined previously in Section 3.2.1.

Thus, the fit to option data was assessable only for the *GARCH-CPC* and *GARCH-OP* models. In particular, the *GARCH-CPC* displayed IVRMSE values of 5.7% and 4.9% for Period 1 and Period 2, respectively. For the *GARCH-CJOW* model the "NaN" values, reported at the bottom of Table 3.4, indicate that it was not possible to compute prices for all the N options based on the numerical integration of (3.9). Whereas, for the parameters estimated in Table 3.4, the *GARCH-OP* did not encounter any issues in pricing the options we considered but returned higher pricing errors compared to the *GARCH-CPC* model.

Table 3.4: Maximum likelihood estimation results.

	Period 1			Period 2		
	GARCH-CPC	*GARCH-OP*	*GARCH-CJOW*	*GARCH-CPC*	*GARCH-OP*	*GARCH-CJOW*
ω	1.546e-16	8.678e-12	8.208e-07***	6.177e-14	6.689e-09	7.735e-07***
	(7.420e-09)	(1.911e-08)	(7.620e-08)	(3.860e-09)	(5.651e-09)	(1.957e-07)
α	2.923e-06***	1.337e-06***	1.580e-06***	1.003e-06**	3.004e-06***	3.520e-06***
	(1.392e-06)	(1.339e-07)	(2.430e-07)	(5.463e-07)	(6.094e-07)	(1.234e-06)
γ_1	140.269***	438.588***	4.151e+02**	343.652***	337.450***	227.209***
	(29.868)	(73.865)	(6.341e+01)	(123.759)	(59.820)	(82.479)
$\tilde{\beta}$	0.374***	0.776***	6.437e-01***	0.626***	0.887***	0.704***
	(0.151)	(0.180)	(2.759e-02)	(0.232)	(0.033)	(0.082)
φ	2.205e-06***	2.152e-06***	2.480e-06***	5.146e-06***	1.684e-06**	1.510e-06***
	(4.226e-07)	(1.035e-07)	(1.160e-07)	(1.016e-06)	(8.926e-07)	(3.910e-07)
γ_2	134.469***	58.924***	6.324e+01***	148.223***	120.697***	188.654***
	(16.244)	(15.322)	(5.300)	(21.761)	(40.545)	(65.588)
ρ	0.925***	0.960***	9.896e-01***	0.836***	0.949***	0.993***
	(0.009)	(0.033)	(9.630e-01)	(0.033)	(0.016)	(0.002)
λ	0.472	0.843	2.092***	-2.957	-3.957^{***}	-3.412^{***}
	(0.813)	(0.852)	(7.729e-01)	(1.415)	(1.405)	(1.264)
Log-lik.	33,978	33,979	34,102	17,993	17,997	18,065
AIC	-67,940	-67,942	-68,188	-35,970	-35,978	-36,114
BIC	-67,882	-67,884	-68,130	-35,917	-35,925	-36,061
IVRMSE(%)	5.787	6.237	NaN	4.965	5.163	NaN

Note: The standard errors, reported in parenthesis, are computed by inverting the negative Hessian matrix evaluated at the optimum parameter values. [*],[**],[***] denote statistical significance at the 0.1, 0.05, 0.01 levels, respectively. As initial volatility states we use the long-term means.

In Figure 3.6, we display the S&P500 daily log-return data for both Period 1 and Period 2 along with the filtered variance h_t and the long-run component $q_t + \bar{h}$ of the *GARCH-CPC* model. Overall, the filtered long-run component is centered around the return variance h_t and

the model seems to correctly capture the spikes corresponding to higher volatility periods.

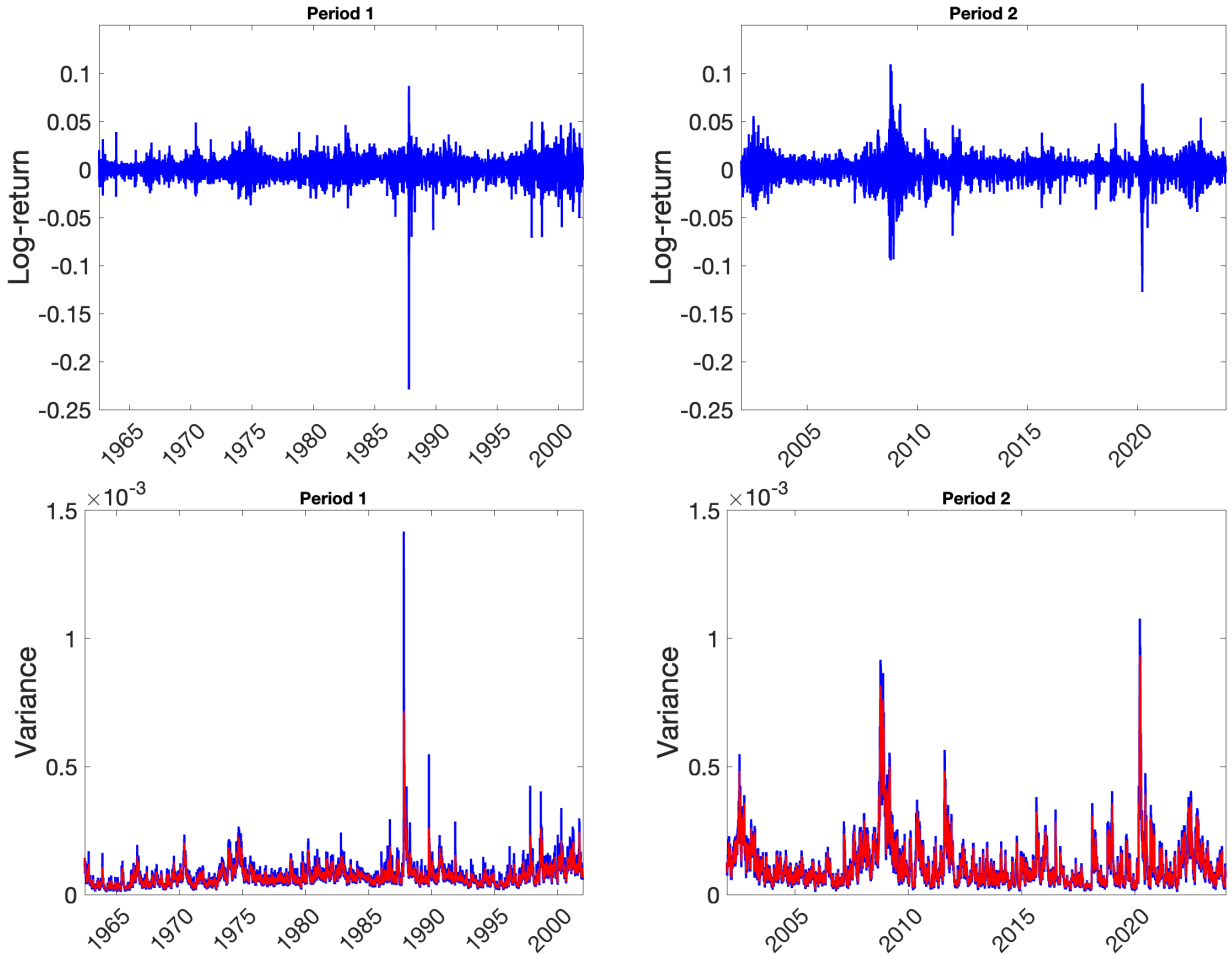


Figure 3.6: Log-returns S&P 500 series (top panels) and *GARCH-CPC* model (conditional) variance (bottom panels): h_t (blue) and $q_t + \bar{h}$ (red) where $\bar{h} = \alpha/(1 - \tilde{\beta} - \alpha\gamma_1^2)$.

As a further comparison, in Figure 3.7, we display the filtered conditional variance h_t, the long-term component q_t, and the short-term component $h_t - q_t$, as considered by CJOW, for the *GARCH-CJOW*, *GARCH-OP*, and *GARCH-CPC* models. The filtered variances are computed using the parameters from Table 3.4 for both Period 1 and Period 2. Overall, the filtered variance components of the three models exhibit similar patterns. Moreover, the short-term component $h_t - q_t$ for the *GARCH-OP* and *GARCH-CPC* models is not mean-zero, as explained in Section 3.3.

3.5 Telescopic extension to three components

The model introduced in equations (3.15)-(3.16) exhibits an autoregressive telescopic structure, as can be seen from matrix $\boldsymbol{M}_1$ in (3.18). Therefore, in this section we propose a telescopic extenstion to include three volatility components given by:

$$R_{t+1} = r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}, \tag{3.25}$$

$$h_{t+1} = p_{t+1} + \rho_1(h_t - p_t) + \alpha_1 \left(Z_t - \gamma_1 \sqrt{h_t}\right)^2 - \alpha_1 \gamma_1^2 p_t, \tag{3.26}$$

$$p_{t+1} = q_{t+1} + \rho_2(p_t - q_t) + \alpha_2 \left(Z_t - \gamma_2 \sqrt{h_t}\right)^2 - \alpha_2 \gamma_2^2 q_t, \tag{3.27}$$

$$q_{t+1} = \omega + \rho_3 q_t + \alpha_3 \left(Z_t - \gamma_3 \sqrt{h_t}\right)^2. \tag{3.28}$$

To retrieve the positivity conditions, we substitute equations (3.28)-(3.27) into (3.26), obtaining:

$$h_{t+1} = \omega + (\rho_3 - \rho_2 - \alpha_2 \gamma_2^2) q_t + (\rho_2 - \rho_1 - \alpha_1 \gamma_1^2) p_t + \rho_1 h_t + \sum_{i=1}^{3} \alpha_i \left(Z_t - \gamma_i \sqrt{h_t}\right)^2, \tag{3.29}$$

from which we derive

$$\rho_2 + \alpha_2 \gamma_2^2 < \rho_3, \quad \rho_1 + \alpha_1 \gamma_1^2 < \rho_2, \quad \rho_i > 0, \quad i = 1, 2, 3. \tag{3.30}$$

The long-run means are given by:

$$\begin{pmatrix} \mathbb{E}[h_t] & \mathbb{E}[p_t] & \mathbb{E}[q_t] \end{pmatrix}^\top = (I_3 - P)^{-1} R,$$

where I_3 denote a 3×3 identity matrix, $R = \begin{pmatrix} \omega + \alpha_3 + \alpha_2 + \alpha_1 \\ \omega + \alpha_3 + \alpha_2 \\ \omega + \alpha_3 \end{pmatrix}$ and

$$P = \begin{pmatrix} \rho_1 + \alpha_1\gamma_1^2 + \alpha_2\gamma_2^2 + \alpha_3\gamma_3^2 & \rho_2 - \rho_1 - \alpha_1\gamma_1^2 & \rho_3 - \rho_2 - \alpha_2\gamma_2^2 \\ \alpha_2\gamma_2^2 + \alpha_3\gamma_3^2 & \rho_2 & \rho_3 - \rho_2 - \alpha_2\gamma_2^2 \\ \alpha_3\gamma_3^2 & 0 & \rho_3 \end{pmatrix}.$$

It can be easily checked that the eigenvalues of P are $\lambda_1 = \alpha_1\gamma_1^2 + \rho_1$, $\lambda_2 = \alpha_2\gamma_2^2 + \rho_2$ and $\lambda_3 = \alpha_3\gamma_3^2 + \rho_3$, so that once again in addition to condition (3.30) we need to assume

$$|\alpha_1\gamma_1^2 + \rho_1| < 1, \quad |\alpha_2\gamma_2^2 + \rho_2| < 1, \quad |\alpha_3\gamma_3^2 + \rho_3| < 1.$$

Model (3.25)-(3.28) can still be reformulated within the framework established by Bormetti et al. (2015). By setting

$$f_t^1 = h_t - p_t, \quad f_t^2 = p_t - q_t, \quad f_t^3 = q_t,$$

we consider the linear operator $\mathcal{L} : \mathbb{R}^3 \to \mathbb{R}_+$ acting on $\boldsymbol{f}_t = (f_t^1, f_t^2, f_t^3)^\top$ and providing the log-return variance $\mathcal{L}(\boldsymbol{f}_t) = f_t^1 + f_t^2 + f_t^3$. Then, we obtain

$$\boldsymbol{f}_{t+1} = \boldsymbol{\omega} + \boldsymbol{M}_1\boldsymbol{f}_t + \boldsymbol{N}_1 \begin{pmatrix} \left(Z_t - \gamma_1\sqrt{\mathcal{L}(\boldsymbol{f}_t)}\right)^2 \\ \left(Z_t - \gamma_2\sqrt{\mathcal{L}(\boldsymbol{f}_t)}\right)^2 \\ \left(Z_t - \gamma_3\sqrt{\mathcal{L}(\boldsymbol{f}_t)}\right)^2 \end{pmatrix}, \tag{3.31}$$

where

$$\boldsymbol{\omega} = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix}, \quad \boldsymbol{M}_1 = \begin{pmatrix} \rho_1 & -\alpha_1\gamma_1^2 & -\alpha_1\gamma_1^2 \\ 0 & \rho_2 & -\alpha_2\gamma_2^2 \\ 0 & 0 & \rho_3 \end{pmatrix}, \quad \boldsymbol{N}_1 = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}.$$

To perform option valuation, we derive the moment generating function associated to the terminal log-spot price of equation (3.25).

Proposition 3.2. *The conditional moment generating function for the logarithm of the terminal log-price* $\ln S_T$*, denoted by* $f(t,T;\phi)$*, can be computed as*

$$f(t,T;\phi) = e^{\phi \log S_t + A_t(\phi) + B_{1,t}(\phi)(h_{t+1} - p_{t+1}) + B_{2,t}(\phi)(p_{t+1} - q_{t+1}) + B_{3,t}(\phi) q_{t+1}}, \quad t < T, \tag{3.32}$$

where the expressions for $A_t(\phi)$*,* $B_{1,t}(\phi)$*,* $B_{2,t}(\phi)$ *and* $B_{3,t}(\phi)$ *are derived in equation* (B.8) *in the Appendix B.1.2.*

Proof. See Appendix B.1.2. □

In Table 3.5, we report the estimation results of the three-component model extension on the same data considered in Table 3.1. In Period 1, the majority of the model coefficients are not significant at standard levels, whereas in Period 2, 9 out of 11 parameters are significant. The option pricing errors are slightly higher to those of the *GARCH-OP* model reported in Table 3.1 for Period 1 and lower to all other models for Period 2.

In Figure 3.8, we present the volatility dynamics of the three-component model for both periods. Similar to our observations for the two-component model in Figure 3.7, we find that the three components exhibit closely related dynamics and occasionally intersect with one another.

Table 3.5: Maximum likelihood estimation results.

	Period 1	Period 2
ω	3.162e-15	3.338e-11
	(5.807e-09)	(1.911e-08)
α_1	1.376e-06	2.321e-06**
	(1.340e-06)	(1.245e-06)
γ_1	299.745	108.112**
	(297.445)	(51.464)
ρ_1	0.321	0.386
	(0.448)	(0.420)
α_2	6.523e-07	2.556e-06***
	(9.865e-07)	(3.373e-07)
γ_2	355.052*	333.484***
	(266.136)	(25.002)
ρ_2	0.696	0.638***
	(0.616)	(0.079)
α_3	1.460e-06**	8.308e-07**
	(7.104e-07)	(3.581e-07)
γ_3	119.496**	114.837**
	(69.469)	(49.793)
ρ_3	0.955***	0.967***
	(0.031)	(0.010)
λ	0.202*	-3.613**
	(0.142)	(1.588)
Log-lik.	33,983	17,999
AIC	-67,946	-35,976
BIC	-67,867	-35,903
IVRMSE (%)	6.394	4.931

Note: The standard errors, reported in parentheses, are computed by inverting the negative Hessian matrix evaluated at the optimum parameter values. [*], [**], [***] denote statistical significance at the 0.1, 0.05, and 0.01 levels, respectively. Long-term means are used as the initial volatility states.

3.6 Conclusions

In the influential article by Christoffersen et al. (2008), it is demonstrated that the *GARCH-CJOW* component model significantly outperforms single component models in capturing stock return volatility. Moreover, the *GARCH-CJOW* model also shows excellent performance in explaining option prices.

However, a key limitation of the *GARCH-CJOW* model is its failure to not guarantee positive volatilities, leading to substantial practical challenges. Specifically, even with realistic parameter sets, it is unfeasible to compute option prices for several maturities using the popular semi-closed formula, thereby vanishing one of the main advantages of affine models.

These challenges were partially addressed by the model proposed in Oh and Park (2023). However, as we demonstrate in this paper, it still experiences negative volatilities. Therefore, we propose an enhancement of the *GARCH-CJOW* and *GARCH-OP* models, termed *GARCH-CPC* model, which guarantees the positivity of the variance by leveraging from the telescopic autoregressive structure. Moreover, we illustrate a possible extension to three volatility components. Our approach, which fits within the general framework established by Bormetti et al. (2015), demonstrates comparable performance on returns data and superior performance in pricing options compared to both the *GARCH-CJOW* and *GARCH-OP* models.

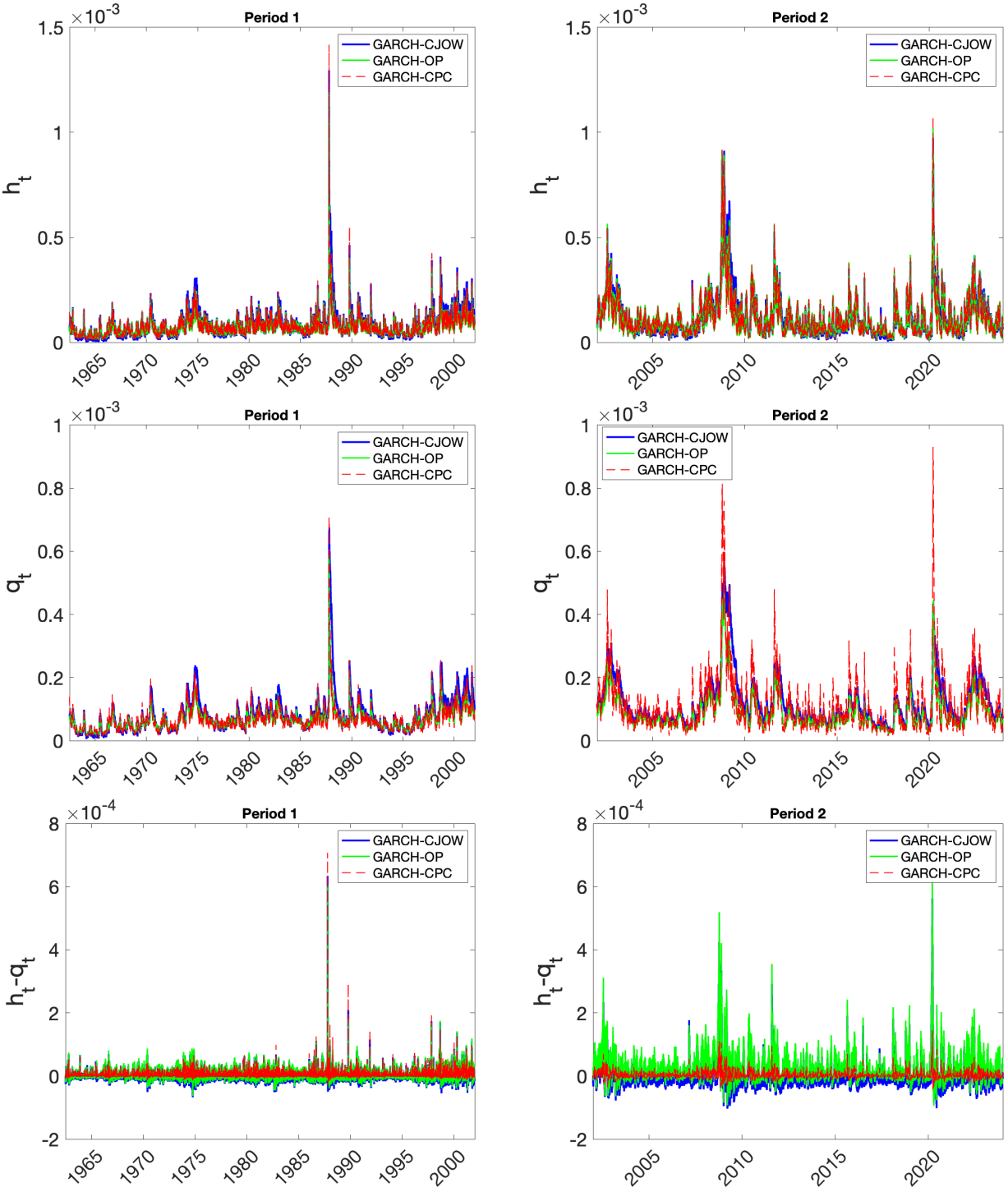


Figure 3.7: Filtered (conditional) variance.

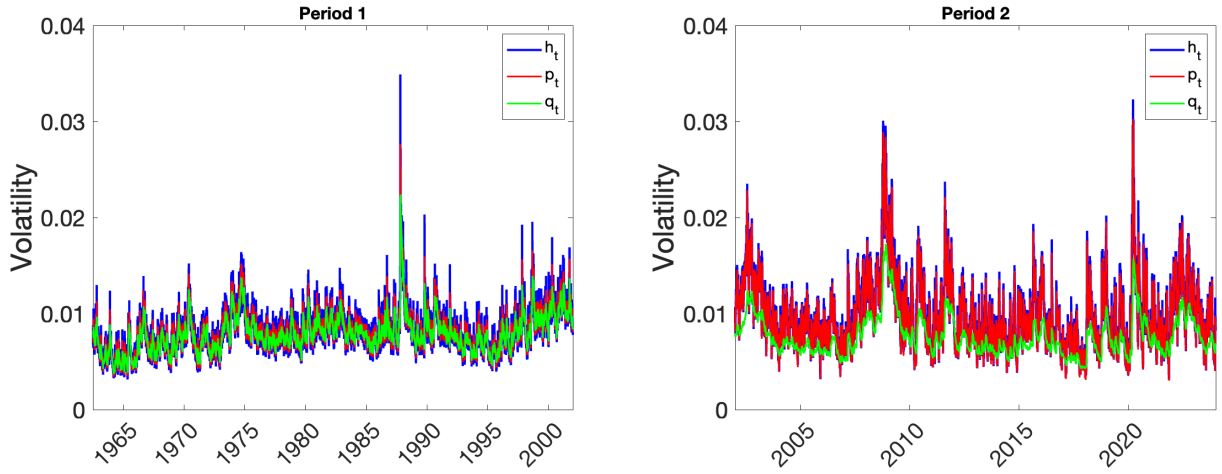


Figure 3.8: Filtered (conditional) volatility.

Chapter 4

A multi-factor model for improved commodity pricing: Calibration and an application to the oil market

ALMA MATER STUDIORUM - UNIVERSITÁ DI BOLOGNA

Abstract

Faculty Name

Department of Statistical Sciences

"Paolo Fortunati"

Doctor of Philosophy in Statistical Sciences

Essays on Latent Factor Models in Financial Markets:

A Focus on Commodity Prices and Volatility Dynamics

by Christian TEZZA

We present a new model for commodity pricing that enhances accuracy by integrating four distinct risk factors: spot price, stochastic volatility, convenience yield, and stochastic interest rates. While the influence of these four variables on commodity futures prices is well recognized, their combined effect has not been addressed in the existing literature. We fill this gap by proposing a model that effectively captures key stylized facts including a dynamic correlation structure and time-varying risk premiums. Using a Kalman filter-based framework, we achieve simultaneous estimation of parameters while filtering state variables through the joint term structure of futures prices and bond yields. We perform an empirical analysis focusing on crude oil futures, where we benchmark our model against established approaches. The results demonstrate that the proposed four-factor model effectively captures the complexities of futures term structures and outperforms existing models.

4.1 Introduction

In recent years, the crude oil market emerged as the largest and most influential commodity market globally. Accurate modeling of the key drivers of commodity price dynamics is essential for effective pricing, hedging, and risk management of commodity derivatives.

Empirically, as documented in Litzenberger and Rabinowitz (1995), Duffie et al. (1999), and Eydeland and Wolyniec (2003), many commodity prices are mean-reverting, strongly heteroscedastic, and influenced by the Samuelson (1965) effect, whereby volatility tends to increase as futures contracts approach maturity. Furthermore, commodity futures prices are often "backwardated," meaning they decline as the delivery date approaches. Notably, the analysis of crude oil prices in Routledge et al. (2000) shows that the degree of backwardation is positively related to volatility, implying that volatility encompasses a component spanned by futures contracts.

To address there stylized facts, as discussed in Duffie et al. (2003), affine factor models gained traction among practitioners due to their desirable analytical properties and because they allow for straightforward pricing of futures and bonds. However, a significant challenge remains in determining the optimal number and type of factors driving the spot price dynamics. In the seminal work of Schwartz (1997), it is assumed that all the uncertainty is summarized by one factor, namely, the spot price of the commodity. However, one-factor models imply that the returns of all futures in the term structure are perfectly correlated and that the degree of backwardation is time-invariant, which are overly restrictive assumptions inconsistent with empirical data.

To address the drawbacks of using a single factor, Gibson and Schwartz (1990) and Schwartz (1997) considered two stochastic factors, the spot price and the convenience yield. Under these models, futures prices are no longer perfectly correlated, allowing for richer term structures. However, Routledge et al. (2000) argued that the correlation between the spot price and the convenience yield needs to be time-varying and that two factors do not adequately capture price volatility dynamics. Consequently, several articles have proposed three-factor models, typically including either stochastic volatility (see Deng (2000), Geman and Nguyen (2005),

Hikspoors and Jaimungal (2008), Lutz (2010), and Hughen (2010)) or stochastic interest rate (e.g., Cortazar and Schwartz (1994), Schwartz (1997), Cortazar and Schwartz (2003), Casassus and Collin-Dufresne (2005), Tang (2012) and Mellios et al. (2016)).

While the above-mentioned articles suggest that effective pricing of commodity futures require at least three factors, the filtered estimates from these models are not always satisfactory. Hughen (2010) found that the pricing performance of three-factor models is often comparable to that of two-factor models, indicating the potential need for a fourth factor. Cortazar (2006) support this finding showing that a fourth factor is crucial for fitting the volatility term structure of crude oil futures. Four-factor model specifications proposed by Yan (2002), Schwartz and Trolle (2009b), Schwartz and Trolle (2009c), and Schöne and Spinler (2017) aim to enhance commodity futures pricing by incorporating factors like mean reversion, convenience yield, stochastic volatility, and jump clustering, albeit with different factor combinations.

To shed light on the appropriate number and types of factors needed to model commodity prices, we propose a novel four-factor model that incorporates the spot price, the convenience yield, the interest rate, and the volatility of the spot price, treating the variables in the drift term of the log spot price as stochastic. This combination of factors has not been considered in the literature, yet it allows for a more accurate evaluation of the impact of interest rates on the risk-neutral drift of log spot prices. In line with Hughen (2010) and Schöne and Spinler (2017), we utilize affine specifications for the drift and covariance terms, while maintaining a parsimonious structure, and we account for a time-varying correlation matrix and time-varying risk premia since, following Duffee (2002), Casassus and Collin-Dufresne (2005), and Cheridito et al. (2007), we model risk premia as affine functions of the state factors.

We develop an innovative estimation framework that incorporates bond yields, marking the first effort to jointly model the four factors that we model using panel data on futures prices and bond yields. Moreover, we adapt the discretization derived by Kelly and Lord (2023) for square-root processes to our state-space estimation framework to ensure the positivity of the state volatility process. We conduct rigorous empirical testing, both in-sample and out-of-sample, using a comprehensive panel dataset of crude oil futures prices and bond data. The

results show that our approach outperforms popular models like Schwartz (1997), Yan (2002), Hughen (2010), and Schöne and Spinler (2017), particularly when bond estimation is included, which enhances short-term futures accuracy.

The contribution of this paper is twofold. First, our four-factor model improves upon existing benchmarks for pricing commodity futures while accounting for the interaction between futures and bond yields. Second, our analysis highlights the crucial role volatility and interest rates in the drift and diffusion terms of log spot prices. Notably, the incorporation of bond yields not only enhances the accuracy of interest rate estimation, but also provides a more effective representation of commodity futures prices dynamics. While the empirical analysis in this paper focuses on the crude oil market, our model and estimation framework are applicable to a broad range of commodities.

The remainder of this article is structured as follows: In Section 4.2, we present our model and derive formulas for the valuation of commodity futures and bond yields. Section 4.3 briefly reviews the benchmark models. Section 4.4 details the data sources and the model estimation framework. Empirical results, including both in-sample and out-of-sample analyses, are presented in Section 4.5. Finally, Section 4.6 concludes the paper.

4.2 The four-factor model

In this section, we present our proposed model for commodity prices, which incorporates Stochastic interest Rate and Volatility, comprising a total of four factors. Thus, we refer to this model as *SRV-4f*.

We assume that the commodity spot price S is driven by four independent sources of uncertainty and is influenced on the convenience yield δ, the (instantaneous) interest rate r, and the variance of the log spot price v. We define the state vector $X(t) = \begin{pmatrix} \ln S(t) & \delta(t) & r(t) & v(t) \end{pmatrix}^\top$, which we assume satisfies the following stochastic differential equation, under the risk-neutral probability measure:

$$dX(t) = [A + BX(t)]dt + \Sigma^{1/2}(t, X(t))dZ(t), \tag{4.1}$$

$$A = \begin{pmatrix} 0 \\ k_2\mu_2 \\ k_3\mu_3 \\ k_4\mu_4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -1 & +1 & -\frac{1}{2} \\ 0 & -k_2 & 0 & 0 \\ 0 & 0 & -k_3 & 0 \\ 0 & 0 & 0 & -k_4 \end{pmatrix},$$

where $dZ(t)$ is a vector of four uncorrelated standard Brownian motion increments, $\Sigma^{1/2}(t, X(t))$ denotes the Cholesky decomposition of the instantaneous variance and covariance matrix given by

$$\Sigma(t, X(t)) = \Omega_0 + \Omega_1 v(t), \tag{4.2}$$

where

$$\Omega_0 = \begin{pmatrix} 0 & s_{12} & s_{13} & 0 \\ s_{12} & s_{22} & s_{23} & 0 \\ s_{13} & s_{23} & s_{33} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} 1 & \rho_{12}\sigma_{22} & \rho_{13}\sigma_{33} & \rho_{14}\sigma_{44} \\ \rho_{12}\sigma_{22} & \sigma_{22}^2 & \rho_{23}\sigma_{22}\sigma_{33} & \rho_{24}\sigma_{22}\sigma_{44} \\ \rho_{13}\sigma_{33} & \rho_{23}\sigma_{22}\sigma_{33} & \sigma_{33}^2 & \rho_{34}\sigma_{33}\sigma_{44} \\ \rho_{14}\sigma_{44} & \rho_{24}\sigma_{22}\sigma_{44} & \rho_{34}\sigma_{33}\sigma_{44} & \sigma_{44}^2 \end{pmatrix}.$$

As in Duffie and Kan (1996), both the drift term and $\Sigma(t, X(t))$ are affine functions of the state variables. We note that $\Omega_0(1,1) = 0$ and $\Omega_1(1,1) = 1$, so that $v(t)$ is actually the variance of the first state variable $\log S(t)$, and we require $\Omega_0 + \Omega_1 v(t)$ to be positive definite.

We define the vector of risk-premia $\Lambda(t, X(t)) = \begin{pmatrix} \lambda_x(t, S(t)) & \lambda_\delta(t, \delta(t)) & \lambda_r(t, r(t)) & \lambda_v(t, v(t)) \end{pmatrix}^\top$ as an affine function of $X(t)$, that is

$$\Lambda(t, X(t)) = \Sigma(t, X(t))^{-1} \left(\widehat{A} - A + [\widehat{B} - B]X(t) \right), \tag{4.3}$$

where the physical drift parameters $\widehat{A}$ and $\widehat{B}$ are given by

$$\widehat{A} = \begin{pmatrix} \widehat{\mu}_1 \\ \widehat{k}_2\widehat{\mu}_2 \\ \widehat{k}_3\widehat{\mu}_3 \\ \widehat{k}_4\widehat{\mu}_4 \end{pmatrix}, \quad \widehat{B} = \begin{pmatrix} 0 & -1 & 0 & -\frac{1}{2} \\ 0 & -\widehat{k}_2 & 0 & 0 \\ 0 & 0 & -\widehat{k}_3 & 0 \\ 0 & 0 & 0 & -\widehat{k}_4 \end{pmatrix}. \tag{4.4}$$

From the Girsanov Theorem, see Theorem 11.3 in Björk (2009), we know that the process $Z(t)^{\mathbb{P}}$, defined by

$$Z(t)^{\mathbb{P}} = Z(t) - \int_0^t \Lambda(s, X(s))ds, \tag{4.5}$$

is a standard $\mathbb{P}$-Wiener process, where $\mathbb{P}$ denotes the physical probability measure.

Substituting (4.5) into (4.1) yields the continuous time model dynamics under $\mathbb{P}$:

$$dX(t) = \left[\widehat{A} + \widehat{B}X(t)\right] dt + \Sigma^{1/2}(t, X(t))dZ^{\mathbb{P}}(t). \tag{4.6}$$

From a mathematical standpoint, the diffusion term in (4.2) follows an affine specification similar to the model proposed by Schöne and Spinler (2017). However, we model different risk factors and we allow for a more parsimonious parametrization of the expressions for B and Ω_0 (see Section 4.3.6).

4.2.1 Futures prices

The futures price at time t with time to maturity $\tau = T - t$ can be computed as the risk-neutral expectation of the futures price:

$$F(t, \tau, X(t)) = \mathbb{E}\left[e^{X(t+\tau)} \middle| \mathcal{F}(t)\right],$$

where $\mathcal{F}(t)$ denotes the information available at time t. The futures price is then the solution to the following partial differential equation (PDE) (see Proposition 5.5 in Björk (2009)):

$$\frac{\partial F(t,\tau,X(t))}{\partial \tau} = \left(\nabla_x F(t,\tau,X(t))\right)^\top [A + BX(t)] + \frac{1}{2}\mathrm{Tr}\left\{\left(H_x F(t,\tau,X(t))\right)\Sigma(t,X(t))\right\}, \quad (4.7)$$

with the terminal condition $F(t,0,X(t)) = e^{X(t)}$, where $\nabla_x F(t,\tau,X(t))$ and $H_x F(t,\tau,X(t))$ are the gradient of $F(t,\tau,X(t))$ with respect to $X(t)$ and the Hessian matrix of $F(t,\tau,X(t))$ with respect to $X(t)$, respectively, and Tr is the trace operator.

Given that the drift and diffusion term in (4.1) are affine functions of $X(t)$, the solution to equation (4.7) can be guessed as follows:

$$\ln F(t,\tau,X(t)) = \alpha(\tau) + \beta(\tau)X(t), \quad (4.8)$$

where $\beta(\tau) = \begin{pmatrix}\beta_1(\tau) & \beta_2(\tau) & \beta_3(\tau) & \beta_4(\tau)\end{pmatrix}$. By substituting equation (4.8) into (4.7) we obtain

$$\frac{d\alpha(\tau)}{d\tau} + \frac{d\beta(\tau)}{d\tau}X(t) = \beta(\tau)[A + BX(t)] + \frac{1}{2}\left[\beta(\tau)\Omega_0\beta(\tau)^\top + \beta(\tau)\Omega_1\beta(\tau)^\top v(t))\right].$$

Separation of variables leads to the following system of ordinary differential equations (ODEs):

$$\frac{d\alpha(\tau)}{d\tau} = \beta(\tau)A + \frac{1}{2}\beta(\tau)\Omega_0\beta(\tau)^\top, \ \frac{d\beta_1(\tau)}{d\tau} = \beta(\tau)B_{(1)}, \ \frac{d\beta_2(\tau)}{d\tau} = \beta(\tau)B_{(2)}, \ \frac{d\beta_3(\tau)}{d\tau} = \beta(\tau)B_{(3)}, \quad (4.9)$$

$$\frac{d\beta_4(\tau)}{d\tau} = \beta(\tau)B_{(4)} + \frac{1}{2}\beta(\tau)\Omega_1\beta(\tau)^\top, \quad (4.10)$$

where $B_{(i)}$ denotes the i-th column of B, $i = 1,2,3,4$. Equations (4.9)-(4.10) must be solved

with terminal conditions $\alpha(0) = 0, \beta_1(0) = 1, \beta_2(0) = 0, \beta_3(0) = 0$ and $\beta_4(0) = 0$.

4.2.2 Bond yields

The present value at time t of a unit discount bond $P(t, \tau, X(t))$ with time to maturity τ can be computed as:

$$P(t, \tau, X(t)) = \mathbb{E}\left[e^{-\int_t^{t+\tau} r(s)ds} \middle| \mathcal{F}(t)\right]. \tag{4.11}$$

The conditional expectation (4.11) is the solution of the following PDE:

$$\begin{aligned} P(t, \tau, X(t))r(t) = -\frac{\partial P(t, \tau, X)}{\partial \tau} &+ \left(\nabla_x P(t, \tau, X(t))\right)^\top [A + BX(t)] \\ &+ \frac{1}{2}\mathrm{Tr}\left\{\left(H_x P(t, \tau, X(t))\right)\Sigma(t, X(t))\right\}, \end{aligned} \tag{4.12}$$

with terminal condition $P(t, 0, X(t)) = 1$. The PDE (4.12) can be solved as

$$\ln P(t, \tau, X(t)) = \gamma(\tau) + \zeta(\tau)X(t), \tag{4.13}$$

where $\zeta(\tau) = \begin{pmatrix} \zeta_1(\tau) & \zeta_2(\tau) & \zeta_3(\tau) & \zeta_4(\tau) \end{pmatrix}$. By substituting equation (4.13) into (4.12) we obtain

$$\frac{d\gamma(\tau)}{d\tau} + \frac{d\zeta(\tau)}{d\tau}X(t) = \zeta(\tau)[A + BX(t)] + \frac{1}{2}\left[\zeta(\tau)\Omega_0\zeta(\tau)^\top + \zeta(\tau)\Omega_1\zeta(\tau)^\top v(t)\right] - r(t),$$

which leads to the following system of ODEs:

$$\frac{d\gamma(\tau)}{d\tau} = \zeta(\tau)A + \frac{1}{2}\zeta(\tau)\Omega_0\zeta(\tau)^\top, \; \frac{d\zeta_1(\tau)}{d\tau} = \zeta(\tau)B_{(1)}, \; \frac{d\zeta_2(\tau)}{d\tau} = \zeta(\tau)B_{(2)}, \; \frac{d\zeta_3(\tau)}{d\tau} = \zeta(\tau)B_{(3)} - 1,$$
$$\frac{d\zeta_4(\tau)}{d\tau} = \zeta(\tau)B_{(4)} + \frac{1}{2}\zeta(\tau)\Omega_1\zeta(\tau)^\top,$$

and $\gamma(0) = \zeta(0) = 0$ so that $P(t, 0, X(t)) = 1$.

4.3 Benchmark models

In this section, for the readers convenience, we briefly recall well-known factor models for commodity prices which we will utilize as benchmarks.

4.3.1 Schwartz's (1997) one-factor model

The one-factor model of Schwartz (1997), hereafter *SCH-1f*, assumes that $\ln S(t)$ follows an Ornstein-Uhlenbeck stochastic process, under the risk-neutral measure:

$$d\ln S(t) = \kappa\left(\alpha - \lambda - \ln S(t)\right)dt + \sigma dZ(t),$$

where λ is the (constant) market price of risk and $dZ(t)$ denotes the increment to a standard Brownian motion.

According to equation (8) of Schwartz (1997), we can compute the logarithm of the futures price with time to maturity τ as:

$$\ln F(t, \tau, S(t)) = e^{-\kappa\tau}\ln S(t) + (1 - e^{-\kappa\tau})(\alpha - \lambda) + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa\tau}).$$

4.3.2 Schwartz's (1997) two-factor model

The two-factor model of Schwartz (1997), denoted *SCH-2f*, models the logarithm of the spot price and convenience yield δ via the following equations, under the risk-neutral measure:

$$
\begin{aligned}
d\ln S(t) &= \left(r - \delta(t) - \frac{1}{2}\sigma^2\right)dt + \sigma_1 dZ_1(t), \\
d\delta(t) &= \kappa[(\alpha - \delta(t)) - \lambda]dt + \sigma_2 dZ_2(t),
\end{aligned}
$$

where the increments of the standard Brownian motions are correlated with $\mathbb{E}[dZ_1(t)dZ_2(t)] = \rho dt$. The logarithm of the futures price with time to maturity τ is given as in equation (19) of Schwartz (1997):

$$
\ln F(t,\tau,S(t),\delta(t)) = \ln S(t) - \delta(t)\frac{1-e^{-\kappa\tau}}{\kappa} + A^*(\tau),
$$

where $A^*(\tau)$ is specified in equation (20) of Schwartz (1997).

4.3.3 Schwartz's (1997) stochastic interest rate model

The three-factor model of Schwartz (1997), which we denote by *SCH-3f*, comprises the following stochastic differential equations under the risk-neutral measure:

$$
\begin{aligned}
d\ln S(t) &= \left(r(t) - \delta(t) - \frac{1}{2}\sigma^2\right)dt + \sigma_1 dZ_1(t), \\
d\delta(t) &= \kappa\left(\alpha - \delta(t)\right)dt + \sigma_2 dZ_2(t), \\
dr(t) &= a\left(\mu - r(t)\right)dt + \sigma_3 dZ_3(t),
\end{aligned}
$$

where the Brownian motion increments are correlated via $\mathbb{E}[dZ_1(t)dZ_2(t)] = \rho_1 dt$, $\mathbb{E}[dZ_2(t)dZ_3(t)] = \rho_2 dt$ and $\mathbb{E}[dZ_1(t)dZ_3(t)] = \rho_3 dt$. The logarithm of the futures price with time to maturity τ is given by equation (27) of Schwartz (1997):

$$\ln F(t,\tau,S(t),\delta(t),r(t)) = \ln S(t) - \frac{\delta(t)(1-e^{-\kappa\tau})}{\kappa} + \frac{r(t)(1-e^{-a\tau})}{a} + C^*(\tau),$$

where $C^*(\tau)$ is specified in equation (28) of Schwartz (1997).

4.3.4 Hughen's (2010) stochastic volatility model

The model of Hughen (2010), hereafter *HU-3f*, considers $X(t) = \begin{pmatrix} \ln S(t) & \delta(t) & v(t) \end{pmatrix}^\top$ as state vector, which, under the risk-neutral measure, satisfies the following SDE:

$$\begin{aligned} dX(t) &= (A + BX(t))\,dt + \Sigma^{1/2}(t, X(t))dZ(t), \\ \Sigma(t, X(t)) &= \Omega_0 + \Omega_1 v(t), \end{aligned}$$

where $Z(t)$ is a vector of three uncorrelated Brownian motions, and

$$A = \begin{pmatrix} 0 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & -1/2 \\ \kappa_{12} & \kappa_{22} & \kappa_{23} \\ 0 & 0 & \kappa_{33} \end{pmatrix},$$

$$\Omega_0 = \begin{pmatrix} 0 & s_{12} & \sigma_{13}\vartheta \\ s_{21} & s_{22} & \sigma_{23}\vartheta \\ \sigma_{31}\vartheta & \sigma_{32}\vartheta & \sigma_{33}\vartheta \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix},$$

where Ω_1 is positive definite and $\Omega_0 - \Omega_1\vartheta$ is positive semi-definite with $\vartheta \leq 0$. The logarithm of the futures price with time to maturity τ is given by equation (27) of Hughen (2010), that

is:

$$\ln F(t,\tau,X(t)) = \alpha(\tau) + \beta(\tau)X(t),$$

where $\alpha(\tau)$ and $\beta(\tau) = \begin{pmatrix} \beta_1(\tau) & \beta_2(\tau) & \beta_3(\tau) \end{pmatrix}$ can be obtained via numerical integration of the following system of ODEs:

$$\frac{d\alpha(\tau)}{d\tau} = \beta(\tau)A + \frac{1}{2}\beta(\tau)\Omega_0\beta(\tau)^\top,\ \frac{d\beta_1(\tau)}{d\tau} = \beta(\tau)B_{(1)},\ \frac{d\beta_2(\tau)}{d\tau} = \beta(\tau)B_{(2)},$$
$$\frac{d\beta_3(\tau)}{d\tau} = \beta(\tau)B_{(3)} + \frac{1}{2}\beta(\tau)\Omega_1\beta(\tau)^\top,$$

where $B_{(i)}$ denotes the i-th column of B, $i = 1,2,3$, and the terminal conditions are $\alpha(0) = 0, \beta_1(0) = 1, \beta_2(0) = 0, \beta_3(0) = 0$.

4.3.5 Yan's (2002) stochactic volatility with jumps model

The four-factor model of Yan (2002), which we denote by *YAN-4f*, comprises the following set of equations under the risk-neutral measure:

$$\begin{aligned}
d\ln S(t) &= \left(r(t) - \delta(t) - \nu\mu_J - \frac{\sigma_x^2}{2} - \frac{1}{2}v(t)\right)dt + \sigma_x dZ_x(t) + \sqrt{v(t)}dZ_v(t) + dL(t),\\
d\delta(t) &= (\mu_\delta - \kappa_\delta\delta(t))\,dt + \sigma_\delta dZ_\delta(t),\\
dr(t) &= (\mu_r - \kappa_r r(t))\,dt + \sigma_r\sqrt{r(t)}dZ_r(t),\\
dv(t) &= (\mu_v - \kappa_v v(t))\,dt + \sigma_v\sqrt{v(t)}dZ_v(t) + dL_v(t),
\end{aligned}$$

where $(L(t), L_v(t))$ denote a two-dimensional compound Poisson process defined as

$$L(t) = \sum_{i=1}^{q(t)} J_i, \quad L_v(t) = \sum_{i=1}^{q(t)} J_{v,i}, \tag{4.14}$$

and $q(t)$ is a Poisson process with constant intensity ν. The jump sizes of the log spot price and volatility are modeled by J and J_v, respectively. Here, $J_v \sim \text{Exp}(\theta)$ with $\theta > 0$ and $\ln(1+J) \sim N\left(\ln(1+\mu_J) - \frac{\sigma_J^2}{2}, \sigma_J^2\right)$. The Brownian motion increments are correlated as $\mathbb{E}[dZ_x(t)dZ_\delta(t)] = \rho_{x\delta}dt$ and $\mathbb{E}[dZ_x(t)dZ_v(t)] = \rho_{xv}dt$. The logarithm of the futures price with time to maturity τ is given by:

$$\ln F(t,\tau,S(t),\delta(t),r(t)) = \ln S(t) + \beta_0(\tau) + \beta_\delta(\tau)\delta(t) + \beta_r(\tau)r(t),$$

where the expressions for $\beta_0(\tau)$, $\beta_\delta(\tau)$ and $\beta_r(\tau)$ are given in equation (13) of Yan (2002). We note that the futures price does not directly depend on the spot volatility. Moreover, only the convenience yield and the spot volatility are correlated with the return process, whereas all the other correlations are set to zero.

4.3.6 Schöne and Spinler (2017) four-factor model

The four-factor model of Schöne and Spinler (2017), denoted *SS-4f*, uses as state vector $X(t) = \begin{pmatrix} \ln S(t) & \theta(t) & q(t) & v(t) \end{pmatrix}^\top$, where $\theta(t)$ represent the long-term level of $\ln S(t)$ and $q(t)$ is a factor related to the cost of carry. Schöne and Spinler (2017) formulate their model, under the risk-neutral measure, as follows:

$$\begin{aligned} dX(t) &= [A + BX(t)]\,dt + \Sigma^{1/2}(t, X(t))dZ(t), \\ \Sigma(t, X(t)) &= \Omega_0 + \Omega_1 v(t), \end{aligned}$$

where $Z(t)$ is a vector of four uncorrelated Brownian motions and

$$A = \begin{pmatrix} 0 \\ \mu_2 \\ \kappa_{33}\mu_3 \\ \kappa_{44}\mu_4 \end{pmatrix}, \quad B = \begin{pmatrix} -\kappa_{11} & \kappa_{11} & -1 & \kappa_{14} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\kappa_{33} & 0 \\ 0 & 0 & 0 & -\kappa_{44} \end{pmatrix},$$

$$\Omega_0 = \begin{pmatrix} 0 & s_{12} & s_{13} & \vartheta\sigma_{14} \\ s_{21} & s_{22} & s_{23} & \vartheta\sigma_{24} \\ s_{31} & s_{32} & s_{33} & \vartheta\sigma_{34} \\ \vartheta\sigma_{41} & \vartheta\sigma_{42} & \vartheta\sigma_{43} & \vartheta\sigma_{44} \end{pmatrix}, \quad \Omega_1 = \begin{pmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix},$$

where Ω_1 and $\Omega_0 - \Omega_1\vartheta$ are positive semi-definite with $\vartheta \leq 0$. The logarithm of the futures price with time to maturity τ is given by equation (11) of Schöne and Spinler (2017), that is

$$\ln F(t, \tau, X(t)) = \alpha(\tau) + \beta(\tau)X(t),$$

where $\alpha(\tau)$ and $\beta(\tau) = \begin{pmatrix} \beta_1(\tau) & \beta_2(\tau) & \beta_3(\tau) & \beta_4(\tau) \end{pmatrix}$ can be obtained via numerical integration of the following system of differential equations:

$$\frac{d\alpha(\tau)}{d\tau} = \beta(\tau)A + \frac{1}{2}\beta(\tau)\Omega_0\beta(\tau)^\top, \; \frac{d\beta_1(\tau)}{d\tau} = \beta(\tau)B_{(1)}, \; \frac{d\beta_2(\tau)}{d\tau} = \beta(\tau)B_{(2)}, \; \frac{d\beta_3(\tau)}{d\tau} = \beta(\tau)B_{(3)},$$
$$\frac{d\beta_4(\tau)}{d\tau} = \beta(\tau)B_{(4)} + \frac{1}{2}\beta(\tau)\Omega_1\beta(\tau)^\top,$$

where $B_{(i)}$ denotes the i-th column of B, $i = 1, 2, 3, 4$, and the terminal solutions are $\alpha(0) = \beta_2(0) = \beta_3(0) = \beta_4(0) = 0$ and $\beta_1(0) = 1$.

4.4 Estimation methodology

Since the variables included in the state vector are typically unobserved, we utilize the estimation framework developed in Duan and Simonato (1999), which is based on the Kalman filter,

as it allows us to extract latent variables from a cross-section of observables and to estimate the parameters of the commodity price models (see Schwartz (1997) and Schöne and Spinler (2017)).

In subsection 4.4.1, we illustrate the traditional Kalman filter equations for estimations that rely solely on observed futures prices. In subsection 4.4.2, we derive a state-space system and the related filtering equations, which allow for the joint estimation of futures and bond prices. Lastly, subsection 4.4.3 derives a discretized transition equation for the volatility process, ensuring its positivity.

4.4.1 Futures prices

At each time t we observe the prices of H futures with different maturities τ_i, for $i = 1, \dots, H$. To proceed, we stack the corresponding pricing equations, derived from (4.8) by adding a measurement error term, resulting in the following representation:

$$y(t) = \begin{pmatrix} \ln F^{(1)}(t, \tau_1, X(t)) \\ \vdots \\ \ln F^{(H)}(t, \tau_H, X(t)) \end{pmatrix} = \alpha(\tau) + \beta(\tau) X(t) + \begin{pmatrix} \varepsilon^{(1)}(t) \\ \vdots \\ \varepsilon^{(H)}(t) \end{pmatrix}, \tag{4.15}$$

where $\alpha(\tau) = \begin{pmatrix} \alpha(\tau_1) & \cdots & \alpha(\tau_H) \end{pmatrix}^\top$, $\beta(\tau) = \begin{pmatrix} \beta_1(\tau_1) & \beta_2(\tau_1) & \beta_3(\tau_1) & \beta_4(\tau_1) \\ \vdots & \vdots & \vdots & \vdots \\ \beta_1(\tau_H) & \beta_2(\tau_H) & \beta_3(\tau_H) & \beta_4(\tau_H) \end{pmatrix}$ and $\varepsilon^{(i)}(t)$ represents the measurement error related to logarithm of futures price i with zero mean and standard deviation σ_{ε_i}, $i = 1, \dots, H$. This accounts for the possibility of bid-ask spreads, nonsimultaneity of observations and errors in the data. In a correctly specified model, the errors $\varepsilon(t)$ should be serially and cross-sectionally uncorrelated with zero mean.

To obtain the transition equation for the state-space system we need to derive the expressions for the conditional mean and variance of the (unobserved) state variables over a discrete time

interval of length h, which we set equal to $1/252$ to represent daily data. We discretize equation (4.6) over h using Euler discretization and we obtain the following transition equation:

$$\tilde{X}(t+h) = \widehat{A}h + (I_4 + \widehat{B}h)\tilde{X}(t) + \sqrt{h}\Sigma^{1/2}(t, \tilde{X}(t))\eta(t+h), \tag{4.16}$$

where $\tilde{X}(t)$ represents the discretized approximation of the process at time step t, $\eta(t+h)$ is a normally distributed (4×1) error vector of zero means and unit variances and I_4 denotes the (4×4) identity matrix. Given that $\Sigma(t, X(t))$ is a function of $X(t)$, the transition density of equation (4.16) will not be Gaussian, so that we have a quasi-optimal Kalman filter. However, as explained in Duan and Simonato (1999) and in Ewald and Zou (2021), the use of this quasi-optimal filter yields an approximate quasi-likelihood function with which parameter estimation can be efficiently carried out.

The state-space system of equations (4.15)-(4.16) can be directly used in the Kalman filter recursion, which we briefly recall hereafter. Using $\theta_F = \{A, B, \widehat{A}, \widehat{B}, \Omega_0, \Omega_1, \sigma_{\varepsilon_1}, \ldots, \sigma_{\varepsilon_W}\}$ as the set of parameters, we compute the one-step ahead prediction and variance of $X(t)$ for a step size h, conditional to the information at time t:

$$\tilde{X}(t+h|t) = \widehat{A}h + \left[I_4 + \widehat{B}h\right]\tilde{X}(t|t), \tag{4.17}$$

$$P(t+h|t) = \left[I_4 + \widehat{B}h\right]P(t|t)\left[I_4 + \widehat{B}h\right]^\top + \Sigma(t, \tilde{X}(t)|t),$$

where

$$\Sigma(t, \tilde{X}(t)|t) = \Omega_0 + \Omega_1\tilde{v}(t|t).$$

To finish the Kalman filter we need the updating equations:

$$\tilde{X}(t|t) = \tilde{X}(t|t-h) + P(t|t-h)\beta(\tau)^\top V(t|t-h)^{-1}e(t),$$
$$P(t|t) = P(t|t-h) - P(t|t-h)\beta(\tau)^\top V(t|t-h)^{-1}\beta(\tau)P(t|t-h).$$

where $e(t) = y(t) - y(t|t-h)$ is the forecast error. We compute then the one-period-ahead prediction and variance of the logarithm of futures prices as:

$$y(t+h|t) = \alpha(\tau) + \beta(\tau)\tilde{X}(t+h|t), \tag{4.18}$$
$$V(t+h|t) = \beta(\tau)P(t+h|t)\beta(\tau)^\top + Q, \tag{4.19}$$

where Q is a $(H \times H)$ diagonal matrix with entries $\sigma^2_{\varepsilon_i}$, $i = 1, \ldots, H$. The estimated parameter vector θ^*_F solves

$$\theta^*_F = \arg\max_\theta \sum_{t=1}^{N} \log L(\theta|e(t), V(t|t-h)),$$

where N represents length of the time series of futures prices and the day-t log-likelihood is equal to

$$\log L(\theta|e(t), V(t|t-h)) = -\frac{1}{2}\left(N\log 2\pi + \left[\log\det\left(V(t|t-h)\right) + e(t)^\top V^{-1}(t|t-h)e(t)\right]\right). \tag{4.20}$$

4.4.2 Futures prices and bond yields

So far, we have shown how to estimate the model using only futures prices. Ideally, as explained in Schwartz (1997), the parameters θ_F of the state vector should be estimated simultaneously

from a time series cross-sectional data of futures prices and bond yields. In this section, we illustrate an estimation framework based on a state-space representation that jointly models futures prices and bond yields, allowing for effective estimation of the model parameters.

Let $R(t, \tau, X(t))$ denote the time t continuously compounded yield on a zero-coupon bond of maturity τ with price $P(t, \tau, X(t))$:

$$R(t, \tau, X(t)) = -\frac{1}{\tau} \ln P(t, \tau, X(t)). \tag{4.21}$$

Similarly to our approach for futures prices, we assume that yields for different maturities are observed with errors of unknown magnitudes. Using the bond pricing formula in equation (4.13), the yield to maturity can be written, after the addition of a measurement error to equation (4.21), as:

$$R(t, \tau, X(t)) = -\frac{1}{\tau} \left[\gamma(\tau) + \zeta(\tau) X(t)\right] + \psi(t), \tag{4.22}$$

where $\psi(t)$ is an error term with zero mean and standard deviation σ_ψ.

Given that at each time t we observe H futures and K bond yields with different maturities, equations (4.15) and (4.22) can be stacked to obtain the following representation:

$$y(t) = \begin{pmatrix} \ln F^{(1)}(t, \tau_1, X(t)) \\ \vdots \\ \ln F^{(H)}(t, \tau_H, X(t)) \\ R^{(1)}(t, \tau_1, X(t)) \\ \vdots \\ R^{(K)}(t, \tau_K, X(t)) \end{pmatrix} = \begin{pmatrix} \alpha(\tau) \\ \gamma(\tau) \end{pmatrix} + \begin{pmatrix} \beta(\tau) \\ \zeta(\tau) \end{pmatrix} X(t) + \begin{pmatrix} \varepsilon^{(1)}(t) \\ \vdots \\ \varepsilon^{(H)}(t) \\ \psi^{(1)}(t) \\ \vdots \\ \psi^{(K)}(t) \end{pmatrix}, \tag{4.23}$$

where $\gamma(\tau) = \begin{pmatrix} -\frac{\gamma(\tau_1)}{\tau_1} & \dots & -\frac{\gamma(\tau_K)}{\tau_K} \end{pmatrix}^\top$, $\zeta(\tau) = \begin{pmatrix} -\frac{\zeta_1(\tau_1)}{\tau_1} & -\frac{\zeta_2(\tau_1)}{\tau_1} & -\frac{\zeta_3(\tau_1)}{\tau_1} & -\frac{\zeta_4(\tau_1)}{\tau_1} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{\zeta_1(\tau_K)}{\tau_K} & -\frac{\zeta_2(\tau_K)}{\tau_K} & -\frac{\zeta_3(\tau_K)}{\tau_K} & -\frac{\zeta_4(\tau_K)}{\tau_K} \end{pmatrix}$ and $\psi^{(j)}(t)$ represents the measurement error related to bond yield j at time t with zero mean and standard deviation σ_{ψ_j}, $j = 1, \dots, K$. Moreover, given that now we also observe bond yields, we need to consider the parameter set $\theta_{FB} = \{A, B, \widehat{A}, \widehat{B}, \Omega_0, \Omega_1, \sigma_{\varepsilon_1}, \dots, \sigma_{\varepsilon_H}, \sigma_{\psi_1}, \dots, \sigma_{\psi_K}\}$. Note that θ_{FB} contains the same parameters as θ_F along with the standard deviations σ_{ψ_j}, $j = 1, \dots, K$. In addition, we adjust Q of the observation equation (4.19) as a $(H+K)\times(H+K)$ diagonal matrix with entries $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{\psi_j}$, $i = 1, \dots, H$, $j = 1, \dots, K$.

4.4.3 Positivity of the state volatility process

The Euler discretization scheme in equation (4.16) does not guarantee the positivity of the CIR process for the variance process. We address this issue by modifying equation (4.16) to utilize an alternative discretization approach. This adjustment ensures the positivity of the variance state process without enforcing restrictive parameter conditions. We follow the numerical approach developed in Kelly and Lord (2023), which is to apply a Lamperti transformation to the CIR process and to numerically approximate the related process.

Let us consider the equation for the variance process v_t in (4.6):

$$dv(t) = \widehat{\kappa}_4(\widehat{\mu}_4 - v(t))dt + \sqrt{v(t)}\Omega_1^{1/2}{}_{(4)}dZ(t), \tag{4.24}$$

where $\Omega_1^{1/2}{}_{(4)}$ denotes the fourth-row of the Cholesky decomposition of Ω_1. By using the Lamperti transform $m(t) = \sqrt{v(t)}$, after an application of Itô's formula we obtain

$$dm(t) = (\nu m(t)^{-1} - \rho m(t))dt + \sum_{i=1}^{4} \gamma_i dZ^{(i)}(t), \tag{4.25}$$

where $\nu = \frac{1}{8}(4\widehat{\kappa}_4\widehat{\mu}_4 - \sum_{i=1}^4 (2\gamma_i)^2)$, $\rho = \widehat{\kappa}_4/2$ and $\gamma_i = \frac{1}{2}\Omega_1^{1/2^{(i)}}{}_{(4)}$, $\Omega_1^{1/2^{(i)}}{}_{(4)}$ denoting the i-th entry of

$\Omega_1^{1/2}{}_{(4)}$, $i = 1, 2, 3, 4$. Following Kelly and Lord (2023), let us consider the mesh $\{t_0, t_1, \ldots, t_M\}$ consisting of $M + 1$ equally spaced points with step size h on the interval $[0, T]$, where $t_0 = 0$ and $t_M = T$. Thus, $t_n = nh$ with $n = 0, \ldots, M$ with $h = T/M$. We approximate (4.25) using the Lie-Trotter splitting method (see Chapter 2 of Hairer et al. (2006)):

$$m(t+h) = e^{-\rho h}\left(\sqrt{m(t)^2 + 2\nu h} + \sum_{i=1}^{4} \gamma_i \Delta Z^{(i)}(t+h)\right),$$

where $\Delta Z^{(i)}(t+h) = Z^{(i)}(t+h) - Z^{(i)}(t) \sim N(0, h)$. Then, since $v(t+1) = m^2(t+1)$, we compute

$$v(t+h) = e^{-2\rho h}\left(\sqrt{v(t) + 2\nu h} + \sum_{i=1}^{4} \gamma_i \Delta Z(t+h)^{(i)}\right)^2,$$

or, equivalently,

$$v(t+h) = e^{-\widehat{\kappa}_4 h}\left(v(t) + 2\nu h + 2\sqrt{v(t) + 2\nu h}\left(\sum_{i=1}^{4} \gamma_i \Delta Z^{(i)}(t+h)\right) + \left(\sum_{i=1}^{4} \gamma_i \Delta Z^{(i)}(t+h)\right)^2\right)^2,$$

which is non-negative for any $\nu > 0$. We finally proceed to derive the first (conditional) moment as in equation (16) of Kelly and Lord (2023):

$$v(t+h|t) = e^{-\widehat{\kappa}_4 h}\left(v(t) + h\left(2\nu + \sum_{i=1}^{4} \gamma_i^2\right)\right) = e^{-\widehat{\kappa}_4 h}\left(v(t) + h\widehat{\kappa}_4\widehat{\mu}_4\right), \tag{4.26}$$

which we use as a state transition equation for the variance process of the *SRV-4f* model. We note that the Kalman filtering procedure for the benchmark models remains unchanged.

4.5 Empirical results

Having introduced the models, the data and the estimation methodology, we proceed to describe the calibration results to market data, both in-sample and out-of-sample.

4.5.1 Data description

We use daily observations of crude oil futures prices sourced from Thomson Reuters Eikon Datastream. In particular, we consider light crude oil futures prices quoted on the New York Mercantile Exchange (NYMEX), as discussed in Hughen (2010). Specifically, the underlying asset of the futures prices is the West Texas Intermediate (WTI) crude oil spot price, quoted in USD per barrel (USD/BBL)[1].

Each futures contract is defined by its delivery month, and its trading terminates three business days before the first day of its delivery month. The maturity date for each futures contract is set to the midweek of its delivery month. We select 11 time series of futures prices, organized by their time to maturity: F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8, F_9, F_{10}, and F_{11}. Here, F_n represents the contract that is the n-th month closest to maturity. The sampling period spans 24 years, from January 2000 to April 2024, encompassing a total of 6,333 business days. In Figure 4.1, we display the time series of crude oil futures' daily prices for the maturities we considered. We note that on April 20, 2020, both the crude oil spot price and the price of the one-month maturity futures contract F_1 were quoted at negative values. This unprecedented event occurred as the COVID-19 pandemic caused a sharp decline in global petroleum demand, while U.S. crude oil inventories surged. As a result, we opted to exclude F_1 from the analysis.

Following Schwartz (1997), the bond yield data we used consist of U.S. Treasury Par Yield Curve Rates and were gathered from the U.S. Department of the Treasury dataset. We consider the bond yields for two maturities, 3 months and 6 months, denoted as R_3 and R_6, respectively. In Figure 4.2, we display the time series for these bond yields.

[1]BBL is an abbreviation for oilfield barrel.

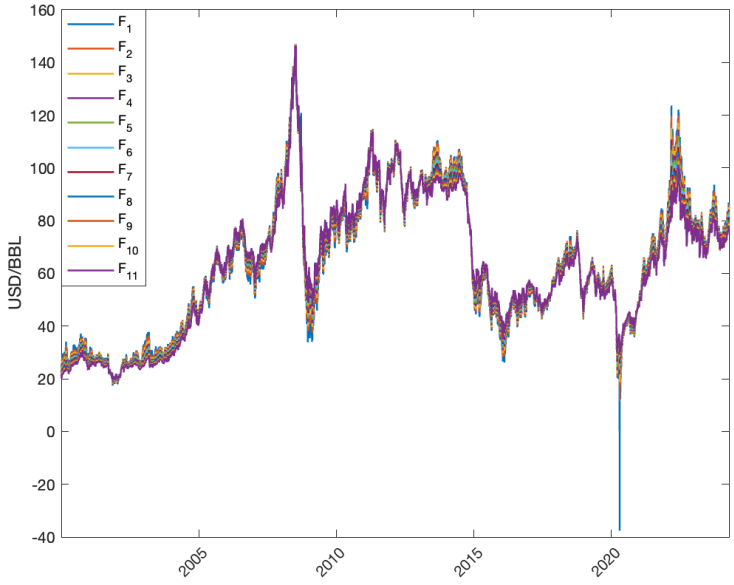


Figure 4.1: Crude oil futures daily prices (dollars per barrel) for several maturities from January 3, 2000 to April 14, 2024.

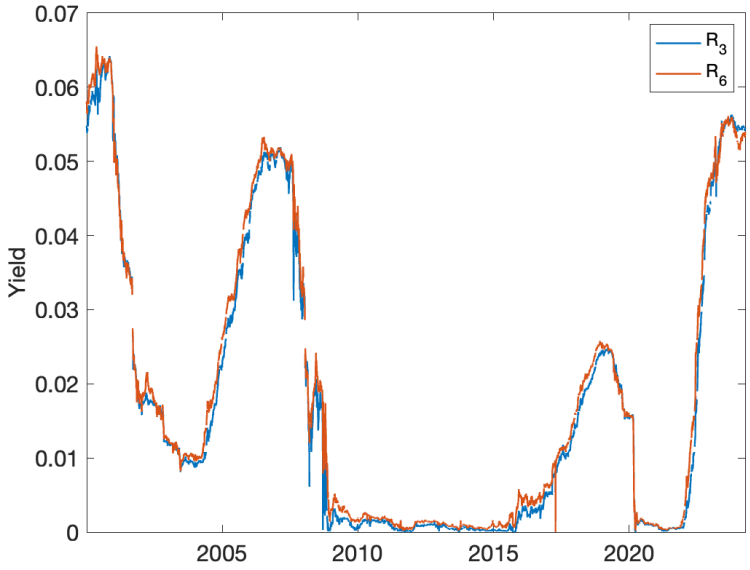


Figure 4.2: Bond yield daily quotes for two maturities from January 3, 2000 to April 14, 2024.

4.5.2 In-sample analysis

Table 4.1 presents the results of the estimation of the *SRV-4f* model. First, we perform an estimation using only futures data, applying equations (4.15), (4.17), and (4.26) with even maturities, specifically F_2, F_4, F_6, F_8, and F_{10} (so that futures with odd maturities are left

for the out-of-sample analysis). As already mentioned, we include bond yield data in the estimation, using equations (4.17), (4.23), and (4.26). In this joint estimation, bond yields R_3 and R_6 are incorporated alongside the futures prices. In both estimations, the average return on the spot commodity, $\widehat{\mu}_1$, and the average convenience yield, $\widehat{\mu}_2$, are positive and significant at standard levels, and, the speed of adjustment coefficients are significant.

As expected, the correlation between the first two state variables - the spot price and convenience yield - is positive. The correlation between the spot price and volatility is also positive and significant. Thus, several correlations among the state variables are significant, contrary to the assumptions made in Yan (2002). Only the correlation between the interest rate and volatility is not significant at any conventional level. Furthermore, it is noteworthy that the significant correlation between the spot price and interest rate change sign when bond yields are included in the estimation. This suggests that the value of a futures contract is sensitive to the interest rate used in its calculation.

In Table 4.2, we report in-sample results for the different models we considered. The information criteria suggest that the four-factor models, with the exception of the model in Yan (2002), provide a better fit to futures term structure than one, two and three factor models. Moreover, the proposed *SRV-4f* model, when estimated on solely futures data, outperforms the other specifications.

Figure 4.3 presents the filtered state variables of the *SRV-4f* model corresponding to the estimation in Table 4.1. In this figure, we also include the spot price, which is observed in the market, as well as the other state variables computed using different models or formulas. In particular, the convenience yield has been estimated using the formula in Javaheri et al. (2003), while the instantaneous interest rate has been estimated using the Vasicek model. Finally, the volatility of the logarithm of the spot price has been filtered using the (annualized) volatility estimate of the GARCH(1,1) model on log-spot prices, excluding the negative price value from April 20, 2020 (see subsection 4.5.1).

From the first plot in Figure 4.3, we observe that the first state variable closely follows the observed crude oil spot price. In the second and third plots, the filtered estimates of the

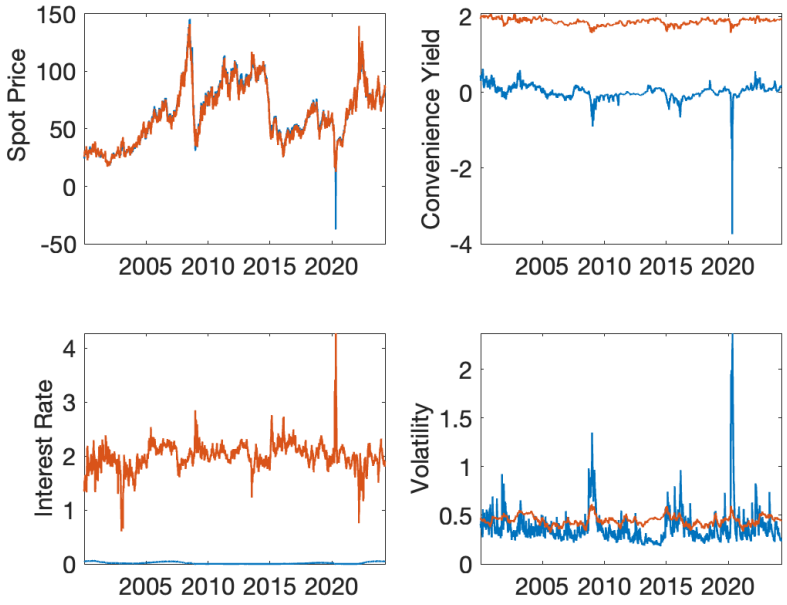


Figure 4.3: Filtered (red) and observable (blue) state variables.

convenience yield and interest rate are not as accurate when compared to the estimates from the models or functions we considered. Finally, the volatility estimate is centered around the mean of the GARCH estimate, but with lower variation.

Figure 4.4 mirrors Figure 4.3 but it represents results from the joint esimation procedure, using crude oil futures F_2, F_4, F_6, F_8, and F_{10} in conjunction with bond yields R_3 and R_6. In this case, we can see that the filtered estimate of the interest rate is closer to the Vasicek model estimate. The convenience yield estimate is also more aligned with the one from Javaheri et al. (2003), resulting in a more accurate representation of the spot price drift term dynamics.

In Figure 4.5, we display the prediction errors $e(t)$ for both the estimations based solely on futures and the estimation incorporating futures and bond yields. As we may see, the errors are centered around zero and exhibit volatility clustering, indicating changing levels of uncertainty over time.

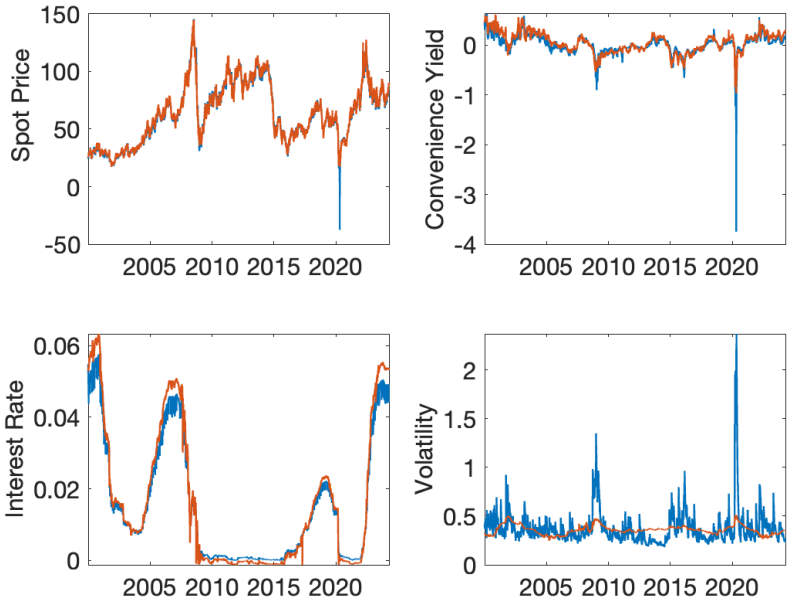


Figure 4.4: Filtered (red) and observable (blue) state variables.

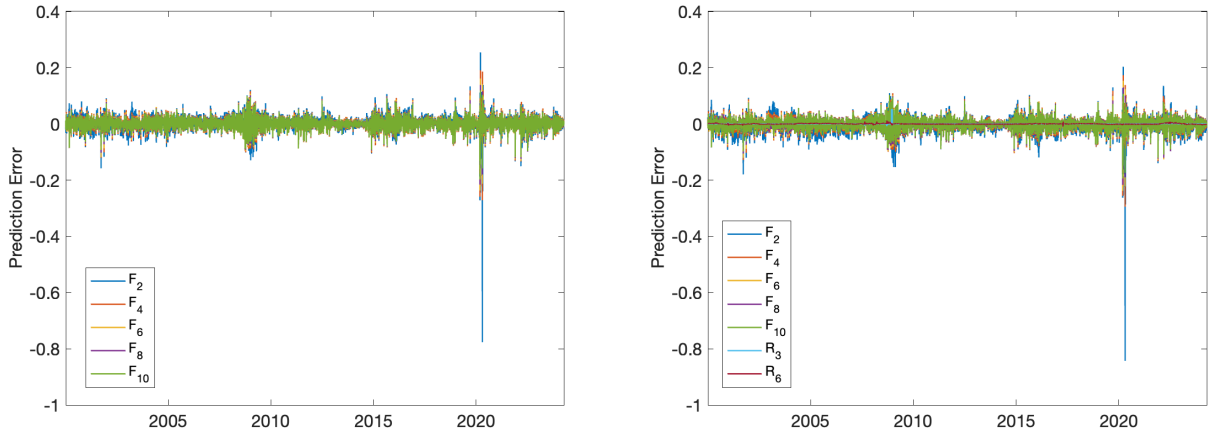


Figure 4.5: Prediction errors using futures (left panel) and using both futures and bond yields (right panel). The left panel shows errors from the estimation based solely on futures, while the right panel reflects errors from the joint estimation.

4.5.3 Out-of-sample analysis

As is standard in the reference literature (see Schwartz (1997) and Schöne and Spinler (2017)), we evaluate futures prices for maturities not included in the estimation. We consider futures prices corresponding to odd maturities, that is, F_3, F_5, F_7, F_9, and F_{11}, and we evaluate them using the parameters from Table 4.1. More precisely, after filtering the state variables and obtaining the coefficients in the futures pricing equation (4.8), we can price the futures that were excluded from the estimation, that is F_3, F_5, F_7, F_9, and F_{11}. We then compare models using standard metrics, namely root mean square error (RMSE), mean absolute percentage error (MAPE). For a futures contract with maturity T, the error metrics are defined as follows:

$$\text{RMSE}(T) = \sqrt{\sum_{t=1}^{N} \frac{\left(y(t,T) - \tilde{y}(t,T,\tilde{X}(t))\right)^2}{N}},$$
$$\text{MAPE}(T) = \frac{1}{N}\sum_{i=1}^{N} \frac{|y(t,T) - \tilde{y}(t,T,\tilde{X}(t))|}{y(t,T)},$$

where N is the length of the time series of futures prices, $y(t,T)$ is the observed log futures price at time t with maturity T and $\tilde{y}(t,T,\tilde{X}(t)) = \ln F(t,T-t,\tilde{X}(t))$ is the corresponding predicted value, with $\tilde{X}(t)$ representing the filtered state variable. The results are reported in Table 4.3, where we also include the predictive log-likelihood, which is the log-likelihood function evaluated on out-of-sample data using in-sample parameters (for more details, see González-Rivera et al. (2004)). For the *SRV-4f* model, we highlight performance for both the futures-only estimation and the joint estimation involving futures and bonds. In the case of joint estimation, when calculating the predictive likelihood, we used only the futures data, while the model parameters were estimated using both futures and bond data.

The results show the superior performance of the *SRV-4f* model across the different futures maturities, in both the futures-only estimation and the joint estimation with futures and bonds. Specifically, for the futures-only estimation, the RMSE and MAPE futures pricing errors are

lower for almost all maturities compared to other models, and the predictive likelihood value is higher. When estimated jointly with bond data, the *SRV-4f* model achieves the lowest average errors for shorter maturities, likely due to the inclusion of short-term bonds in the estimation process.

4.6 Conclusions

We present a novel four-factor model for commodity prices that includes convenience yield, interest rate, and volatility of logarithm of the spot price. This particular combination has not been explored in existing research, even though these factors play a crucial role in influencing commodity futures prices. Consistent with previous empirical findings, the proposed model allows for time-varying correlations and time-varying risk premiums. We develop a novel estimation framework using the Kalman filter, which enables joint estimation of the term structure of NYMEX crude oil futures prices and U.S. Treasury bond yields. Moreover, we ensure the positivity of the discretized equation for variance process by leveraging an discretization scheme for square-root processes.

Based on an out-of-sample analysis, our model is statistically preferred over well-known two-factor, three-factor, and even four-factor model specifications that also include stochastic jump factors. Incorporating bond data in the estimation significantly improves the out-of-sample accuracy of short-term futures pricing compared to using futures data alone. Several extensions are possible for our model. One possibility is to add a jump factor to the spot price and examine the effect of the jump risk premium on futures pricing. Another approach is to specify the instantaneous covariance matrix as a function of directly relevant fundamental factors, such as statistical measures of the degree of backwardation. These topics are left for futures research.

Table 4.1: Model estimation results for crude-oil futures.

Parameter	θ_F: Only futures	θ_{FB}: Futures and bonds
$\widehat{\mu}_1$	1.670 (0.543)	0.622 (0.160)
μ_2	4.962 (0.631)	-0.151 (0.002)
$\widehat{\mu}_2$	1.989 (0.601)	0.244 (0.107)
μ_3	2.027 (0.561)	0.317 (4.000e-2)
$\widehat{\mu}_3$	2.051 (0.527)	-0.053 (0.013)
μ_4	-2.763 (0.221)	1.577 (8.000e-3)
$\widehat{\mu}_4$	0.227 (0.015)	0.095 (0.004)
κ_2	0.222 (0.022)	1.075 (0.011)
$\widehat{\kappa}_2$	-0.723 (0.129)	1.516 (0.267)
κ_3	2.712 (0.129)	0.030 (4.000e-4)
$\widehat{\kappa}_3$	4.711 (0.398)	-0.022 (0.006)
κ_4	1.557 (0.085)	4.945 (0.027)
$\widehat{\kappa}_4$	2.075 (0.329)	1.097 (0.221)
s_{12}	-0.055 (0.008)	0.045 (0.002)
s_{22}	0.019 (0.004)	0.249 (0.038)
s_{13}	-0.171 (0.018)	0.001 (1.000e-4)
s_{23}	0.102 (0.018)	0.001 (1.000e-4)
s_{33}	0.772 (0.053)	0.006 (3.000e-4)
ρ_{12}	0.998 (0.056)	0.572 (0.009)
ρ_{13}	0.444 (0.113)	-0.445 (0.020)
ρ_{14}	0.725 (0.082)	0.385 (0.017)
ρ_{23}	0.827 (0.124)	-0.933 (0.012)
ρ_{24}	0.583 (0.081)	0.330 (0.043)
ρ_{34}	-0.001 (2.000e-4)	0.568 (0.317)
σ_{22}	0.375 (0.016)	0.309 (0.002)
σ_{33}	0.595 (0.146)	0.015 (0.001)
σ_{44}	0.194 (0.015)	0.155 (0.007)
σ_{ε_2}	0.011 (0.001)	0.017 (0.001)
σ_{ε_4}	1.582e-5 (6.453e-6)	4.000e-3 (7.145e-5)
σ_{ε_6}	1.000e-3 (1.498e-5)	6.881e-5 (4.807e-6)
σ_{ε_8}	6.948e-5 (1.227e-5)	2.000e-4 (2.530e-6)
$\sigma_{\varepsilon_{10}}$	3.539e-5 (1.764e-5)	2.000e-3 (6.936e-5)
σ_{ψ_3}		7.535e-6 (1.964e-6)
σ_{ψ_6}		2.000e-3 (3.467e-5)

The estimation data includes crude oil futures F_2, F_4, F_6, F_8, and F_{10}. The standard errors, reported in parentheses, are computed by inverting the negative Hessian matrix evaluated at the optimum parameter values.

Table 4.2: In-sample model estimation results for crude-oil futures F_2, F_4, F_6, F_8, and F_{10}.

Model	Log.Lik.	AIC	BIC
SCH-1f	66.582	−133.150	−133.070
SCH-2f	85.473	−170.920	−170.810
SCH-3f	92.550	−185.070	−184.930
HU-3f	100.550	−201.030	−200.790
YAN-4f	82.786	−165.530	−165.350
SS-4f	109.670	−192.050	−191.770
SRV-4f	135.791	−271.512	−271.245

Table 4.3: Out-of-sample RMSE and MAPE pricing errors (in percentage) for valuing futures contracts for the considered models, along with the out-of-sample predictive log-likelihood in equation (4.20), computed using the $F_3, F_5, F_7, F_9, F_{11}$ futures. *SRV-4f* (θ_F) and *SRV-4f* (θ_{FB}) denote the estimations using only futures and using both futures and bonds, respectively.

Futures	*SCH-1f*	*SCH-2f*	*SCH-3f*	*HU-3f*	*YAN-4f*	*SS-4f*	*SRV-4f* (θ_F)	*SRV-4f* (θ_{FB})
RMSE(3)	4.588	2.095	1.332	1.717	2.241	1.577	1.568	1.436
RMSE(5)	2.964	1.097	1.256	1.477	1.120	1.195	1.118	1.064
RMSE(7)	2.667	0.891	1.041	1.118	0.994	0.956	0.904	0.906
RMSE(9)	3.339	0.923	0.788	0.768	0.846	0.789	0.773	0.773
RMSE(11)	4.324	1.163	0.913	0.721	0.715	0.708	0.686	0.745
Mean	3.576	1.234	1.066	1.160	1.183	1.045	1.010	0.985
MAPE(3)	0.873	0.440	0.254	0.287	0.522	0.314	0.319	0.303
MAPE(5)	0.555	0.224	0.235	0.272	0.231	0.240	0.227	0.227
MAPE(7)	0.489	0.189	0.204	0.224	0.204	0.196	0.188	0.190
MAPE(9)	0.655	0.183	0.166	0.164	0.176	0.165	0.165	0.165
MAPE(11)	0.889	0.233	0.192	0.149	0.154	0.148	0.148	0.159
Mean	0.692	0.254	0.210	0.219	0.257	0.213	0.209	0.209
Predictive Log.Lik.	68.631	86.761	93.987	107.420	83.555	112.160	138.862	126.680

Appendix A

Proofs of Chapter 2

A.1 Strict stationarity and geometric ergodicity

A.1.1 Proof of Lemma 2.1

Proof. We need to prove that for each $k \in \mathbb{N}$, the k-step transition probability density of $\{\boldsymbol{v}_t\}$ is strictly positive over the set of states Ω defined below. This ensures that if the set Ω has positive Lebesgue measure, then $\{\boldsymbol{v}_t\}$ is irreducible w.r.t. the Lebesgue measure restricted to Ω, that is the measure ψ, see Meyn and Tweedie (1993).

We begin by studying the set of states that are reachable in one step from an initial state $\boldsymbol{x} = (x_1, x_2)^T \in \mathcal{D}$, which we denote by $G_{x,1}$.

Given an initial value $\boldsymbol{x}$ and using equations (2.2) and Assumption 1, we derive the set of points $\boldsymbol{y} = (y_1, y_2)^T$ that can be reached by the process $\{\boldsymbol{v}_t\}$ in one step starting from $\boldsymbol{x}$ as

$$G_{x,1} = \{\boldsymbol{y} \in \mathcal{D} : \boldsymbol{y} \geq_{vec} \boldsymbol{\omega} + \boldsymbol{\beta}\boldsymbol{x}\},$$

where $\geq_{vec}$ denotes component-wise inequality. We proceed deriving the reachable set in two-steps, $G_{x,2}$ which includes the points $\boldsymbol{z} = (z_1, z_2)^T$ that can be reached by the process $\{\boldsymbol{v}_t\}$ in

two-steps starting from $\boldsymbol{x}$, i.e.,

$$G_{x,2} = \{\boldsymbol{z} \in \mathcal{D} : \boldsymbol{z} \geq_{vec} \boldsymbol{\omega} + \boldsymbol{\beta}\boldsymbol{y}, \quad \boldsymbol{y} \in G_{x,1}\}.$$

Similarly, the reachable set in k-steps starting from $\boldsymbol{x}$ can be recursively obtained as follows

$$G_{x,k} = \{\boldsymbol{z} \in \mathcal{D} : \boldsymbol{z} \geq_{vec} \boldsymbol{\omega} + \boldsymbol{\beta}\boldsymbol{y}, \quad \boldsymbol{y} \in G_{x,k-1}\}.$$

By induction and using Proposition 1.5.31 Hubbard and Hubbard (2009) we can derive a more compact expression for $G_{x,k}$, that is:

$$G_{x,k} = \left\{\boldsymbol{z} \in \mathcal{D} : \boldsymbol{z} \geq_{vec} \left(\boldsymbol{I}_2 - \boldsymbol{\beta}^k\right)\left(\boldsymbol{I}_2 - \boldsymbol{\beta}\right)^{-1}\boldsymbol{\omega} + \boldsymbol{\beta}^k\boldsymbol{x}\right\}, \tag{A.1}$$

where $\boldsymbol{I}_2$ denotes a 2×2 identity matrix and $(\cdot)^{-1}$ denotes matrix inversion.

We can also define the limit set Ω that can be reached with an infinite number of steps as

$$\Omega = \left\{\boldsymbol{z} \in \mathcal{D} : \boldsymbol{z} \geq_{vec} \left(\boldsymbol{I}_2 - \boldsymbol{\beta}\right)^{-1}\boldsymbol{\omega}\right\}. \tag{A.2}$$

We note that the set Ω does not depend on $\boldsymbol{x}$. Now that we have derived the reachable sets in any step, we proceed studying the one-step transition probability. Given an initial value $\boldsymbol{x} \in \mathcal{D}$ and $\boldsymbol{y} \in G_{x,1}$, the one-step transition probability is given by

$$P\left(\boldsymbol{v}_1 \leq \boldsymbol{y} | \boldsymbol{v}_0 = \boldsymbol{x}\right) = P\left(\boldsymbol{\omega} + \boldsymbol{\beta}\boldsymbol{v}_0 + \boldsymbol{\alpha}\boldsymbol{A}_1 \leq \boldsymbol{y} | \boldsymbol{v}_0 = \boldsymbol{x}\right), \tag{A.3}$$

where $\boldsymbol{A}_1 = (A_{1,1}, A_{2,1})^T$ with $A_{1,1} = \left(W_{1,1} - \gamma_1\sqrt{v_{1,0}}\right)^2$ and $A_{2,1} = \left(W_{2,1} - \gamma_2\sqrt{v_{2,0}}\right)^2$.

Using the independence between $A_{1,1}$ and $A_{2,1}$, we can compute the transition probability density function of $\boldsymbol{A}_t$ given $\boldsymbol{v}_0$ as follows

$$p_{\boldsymbol{A}|\boldsymbol{v}_0}(a_1, a_2|\boldsymbol{v}_0 = \boldsymbol{x}) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi a_i}} e^{-(a_i+\kappa_i)/2} \cosh(\sqrt{\kappa_i a_i}) \quad a_i > 0,$$

which is the product of two independent non-central chi-squared probability density functions with one degree of freedom and non-centrality parameters $\kappa_1 = \gamma_1^2 x_1$ and $\kappa_2 = \gamma_2^2 x_2$.

Using the linear transformation $\boldsymbol{M}_t = \boldsymbol{\alpha}\boldsymbol{A}_t$ and based on (A.3) we have

$$P\left(\boldsymbol{v}_1 \leq \boldsymbol{y}|\boldsymbol{v}_0 = \boldsymbol{x}\right) = P(\boldsymbol{M}_1 \leq \boldsymbol{y} - \boldsymbol{\omega} - \boldsymbol{\beta}\boldsymbol{v}_0|\boldsymbol{v}_0 = \boldsymbol{x}) = \iint\limits_{D_2 \times D_1} p_{\boldsymbol{M}|\boldsymbol{v}_0}(\boldsymbol{m}|\boldsymbol{v}_0 = \boldsymbol{x})\mathrm{d}\boldsymbol{m},$$

where $D_1 = (0, y_1 - \omega_1 - \beta_{11}x_1 - \beta_{12}x_2]$, $D_2 = (0, y_2 - \omega_2 - \beta_{21}x_1 - \beta_{22}x_2]$ and $p_{\boldsymbol{M}|\boldsymbol{v}_0}$ is the one-step transition probability density function of $\boldsymbol{M}_t$ given a starting value $\boldsymbol{v}_0$, which is computed as follows

$$p_{\boldsymbol{M}|\boldsymbol{v}_0}\left(\boldsymbol{m}|\boldsymbol{v}_0 = \boldsymbol{x}\right) = p_{\boldsymbol{A}|\boldsymbol{v}_0}\left(\boldsymbol{\alpha}^{-1}\boldsymbol{m}\middle|\boldsymbol{v}_0 = \boldsymbol{x}\right) \cdot \det(\boldsymbol{\alpha}^{-1}), \tag{A.4}$$

where the Jacobian determinant $\det(\boldsymbol{\alpha}^{-1})$ is not null by assumption.

According to Definition 2.2, it remains to prove that for any set A that satisfies $\psi(A) > 0$ there exists $k \in \mathbb{N}$ such that $P^k(\boldsymbol{x}, A) > 0$ for all $\boldsymbol{x} \in \mathcal{D}$. Since $\psi(A) > 0$ the intersection between A and Ω has positive Lebesgue measure. Moreover, since Ω is the limit of the $G_{x,k}$ sequence of sets, there exists a set $A^* \subset A$, with $\psi(A^*) > 0$, which is included in $G_{x,k}$ for every $\boldsymbol{x}$ and for k sufficiently large. Moreover, for each integer $k \geq 2$, given an initial state $\boldsymbol{x}$, the k-step transition probability density function of reaching state $\boldsymbol{z}$ is given by the Chapman-Kolmogorov equation (Meyn and Tweedie (1993), Theorem 3.4.2)

$$p^k_{\boldsymbol{M}|\boldsymbol{v}_0}(\boldsymbol{m}|\boldsymbol{v}_0 = \boldsymbol{x}) = \iint\limits_{G_{x,k-1}} p^{k-1}_{\boldsymbol{M}|\boldsymbol{v}_0}(\boldsymbol{y}|\boldsymbol{v}_0 = \boldsymbol{x}) \cdot p_{\boldsymbol{M}|\boldsymbol{v}_1}(\boldsymbol{m}|\boldsymbol{v}_1 = \boldsymbol{y})d\boldsymbol{y},$$

where $\boldsymbol{m} = \boldsymbol{z} - \boldsymbol{\omega} - \boldsymbol{\beta}\boldsymbol{v}_1$ for $\boldsymbol{z} \in G_{x,k}$ and $p_{\boldsymbol{M}|\boldsymbol{v}_1}(\boldsymbol{m}|\boldsymbol{v}_1 = \boldsymbol{y})$ denotes the transition probability

density function of $\boldsymbol{M}_t$ given $\boldsymbol{v}_1$. By mathematical induction, we deduce that $p^k_{\boldsymbol{M}|\boldsymbol{v}_0}(\boldsymbol{m}|\boldsymbol{v}_0 = \boldsymbol{x})$ is positive for every $\boldsymbol{z} \in A^*$ so that $P^k(\boldsymbol{x}, A^*) > 0$ and $P^k(\boldsymbol{x}, A) > 0$.

Having proved ψ-irreducibility, we now proceed to study the aperiodicity of the Markov chain $\{\boldsymbol{v}_t\}$. The aperiodicity is proved if $d = 1$ according to Definition 2.3. By contradiction, if we assume a period $d \geq 2$, then there exists a d-cycle of ψ-positive disjoint sets $E_0, \ldots, E_{d-1}$ such that $P(\boldsymbol{x}, E_{i+1}) = 1$ for every $\boldsymbol{x} \in E_i$ with $i = 0, \ldots, d-1$ and such that $V = \left[\bigcup_{i=0}^{d-1} E_i\right]^c$ has null ψ measure. This implies that when starting our chain $\{\boldsymbol{v}_t\}$ at any initial state $\boldsymbol{x} \in E_0$ we have $P(\boldsymbol{x}, E_1) = 1$ so that $\psi(E_1 \cap G_{x,1}) > 0$ and that $\psi(E_i \cap G_{x,1}) = 0$ for every $i \neq 1$. Then, similarly at a second step, we would have $P(\boldsymbol{y}, E_2) = 1$ for $\boldsymbol{y} \in E_1$ implying $\psi(E_2 \cap G_{x,2}) > 0$ and $\psi(E_i \cap G_{x,2}) = 0$ for every $i \neq 2$ and for initial states $\boldsymbol{x} \in E_0$ such that our chain is in E_1 after one step. But in our case, we know that the intersection $G_{x,1} \cap G_{x,2}$ has non-null ψ measure since $G_{x,1}$ and $G_{x,2}$ are unbounded sets of the type $[a, \infty) \times [b, \infty)$ for some $a, b \in \mathbb{R}_+$ with non-empty intersection with Ω. Therefore, $\psi(G_{x,1} \cap G_{x,2}) > 0$. Furthermore,

$$\psi(G_{x,1} \cap G_{x,2}) = \psi\left((G_{x,1} \cap G_{x,2}) \cap \left(\bigcup_{i=0}^{d-1} E_i \cup V\right)\right) = \sum_{i=0}^{d-1} \psi(G_{x,1} \cap G_{x,2} \cap E_i) > 0.$$

Hence, there exists an integer $i = 0, \ldots, d-1$ such that $\psi(G_{x,1} \cap G_{x,2} \cap E_i) > 0$. This contradicts the fact that $\psi(E_i \cap G_{x,1}) = 0$ for every $i \neq 1$ and $\psi(E_i \cap G_{x,2}) = 0$ for every $i \neq 2$. □

A.1.2 Proof of Proposition 2.1

Proof. First, let us define

$$\bar{r} = \max\left(\sqrt{\omega_1^2 + \omega_2^2}, \underline{\Omega}\right), \tag{A.5}$$

where $\underline{\Omega} = \frac{\max[(1-\beta_{22})\omega_1 + \beta_{12}\omega_2, \quad \beta_{21}\omega_1 + (1-\beta_{11})\omega_2]}{((1-\beta_{11})(1-\beta_{22}) - \beta_{12}\beta_{21})}$ represents the the maximum coordinate value of the closest point to the origin of the set Ω, as defined in (A.2). It can be easily shown that $\underline{\Omega} > 0$. Indeed, we have

$$(\boldsymbol{I}_2 - \boldsymbol{\beta})^{-1}\boldsymbol{\omega} = \frac{1}{\det(\boldsymbol{I}_2 - \boldsymbol{\beta})} \begin{bmatrix} 1-\beta_{22} & \beta_{12} \\ \beta_{21} & 1-\beta_{11} \end{bmatrix} \boldsymbol{\omega},$$

where

$$\begin{aligned} \det(\boldsymbol{I}_2 - \boldsymbol{\beta}) &= 1 - \operatorname{trace}(\boldsymbol{\beta}) + \det(\boldsymbol{\beta}) \\ &= 1 - (\lambda_1 + \lambda_2) + \lambda_1\lambda_2 \\ &= (1-\lambda_1)(1-\lambda_2) > 0, \end{aligned}$$

with λ_1 and λ_2 denoting the eigenvalues of $\boldsymbol{\beta}$, which are guaranteed strictly smaller than one in absolute value by $\rho(\boldsymbol{\beta}) < 1$.

For proving that the compact set $C = \{\boldsymbol{x} \in \mathcal{D} : \|\boldsymbol{x}\| \leq r\}$, for any $r > \bar{r}$, is small we use Proposition 2.11 of Nummelin (1984). We note that by letting $r > \bar{r}$ we ensure C to be non-empty and $C \in \mathcal{B}^+$, given that C has μ_{Leb}-positive intersection with Ω. The proposition states that a set $C \in \mathcal{B}^+$ is small, if there is a set $B \in \mathcal{B}^+$, such that for all $A \in \mathcal{B}^+$ with $A \subseteq B$,

$$\inf_{\boldsymbol{x} \in C} \sum_{n=0}^{L} P^n(\boldsymbol{x}, A) > 0,$$

for some non-negative integer L.

We prove that the condition above holds for some $L > 1$. The first step is to find the set B with $\psi(B) > 0$ which implies that $B \in \mathcal{B}^+$ as ψ is the irreducibility measure. Let us define $B = (d_1, \infty) \times (d_2, \infty)$ with $(d_1, d_2)^T = (\boldsymbol{I}_2 - \boldsymbol{\beta})^{-1}\boldsymbol{\omega} + \boldsymbol{\beta} \cdot (r, r)^T$ so that $B \subseteq G_{x,k}$ for each $\boldsymbol{x} \in C$ and $k > 0$.

By the ψ-irreducibility proved in Lemma 2.1, for $\boldsymbol{x} \in C$, we have a positive probability of visiting all sets $A \subseteq B$ with $\psi(A) > 0$, i.e., $P^k(\boldsymbol{x}, A) > 0$ for some $k > 0$. Lastly, the proof is concluded since we know that a positive and real-valued continuous function on C has a

positive infimum on C. □

A.1.3 Proof of Theorem 2.1

Proof. We first show that the so-called drift condition is satisfied with the function $Q(\boldsymbol{v}) = \|\boldsymbol{v}\|$. Then, we can extend geometric ergodicity property of the variance process $\{\boldsymbol{v}_t\}$ to the return process thanks to Carrasco and Chen (2002).

For verifying the drift condition we use the m-step criterion introduced by Tjøstheim (1990). Specifically, we verify Conditions (3.7) and (3.8) in Tjøstheim (1990) using a small set $C = \{\boldsymbol{x} \in \mathcal{D} : \|\boldsymbol{x}\| \leq r\}$ for $r > \bar{r}$ sufficiently large with complement C^c and a non-negative test function Q:

$$\mathbb{E}\left[Q(\boldsymbol{v}_t)\mathbb{1}_{(Q(\boldsymbol{v}_t)\in C^c)}\middle|\boldsymbol{v}_{t-1} = \boldsymbol{x}\right] \leq M, \quad \boldsymbol{x} \in C, \tag{A.6}$$

$$R \cdot \mathbb{E}\left[Q(\boldsymbol{v}_t)|\boldsymbol{v}_{t-1} = \boldsymbol{x}\right] \leq Q(\boldsymbol{x}) - \epsilon, \quad \boldsymbol{x} \in C^c, \tag{A.7}$$

for some $M < \infty$, $\epsilon > 0$ and some $R > 1$.

Given a standard Gaussian vector $\begin{bmatrix} Z_{1,t} & Z_{2,t} \end{bmatrix}^T$, we have the expression

$$\mathbb{E}[\boldsymbol{v}_t|\boldsymbol{v}_{t-1}] = \boldsymbol{\omega} + \boldsymbol{\alpha}\mathbb{E}\left[\begin{bmatrix} Z_{1,t-1}^2 & Z_{2,t-1}^2 \end{bmatrix}^T\right] + \boldsymbol{B}\boldsymbol{v}_{t-1} = \boldsymbol{d} + \boldsymbol{B}\boldsymbol{v}_{t-1}, \tag{A.8}$$

where $\boldsymbol{B}$ is given in (2.6) and $\boldsymbol{d} = \boldsymbol{\omega} + \begin{bmatrix} \alpha_{11} + \alpha_{12} & \alpha_{21} + \alpha_{22} \end{bmatrix}^T$. Iterating m times this formula, we get

$$\mathbb{E}[\boldsymbol{v}_t|\boldsymbol{v}_{t-m}] = \boldsymbol{B}^m\boldsymbol{v}_{t-m} + \sum_{j=0}^{m-1}\boldsymbol{B}^j\boldsymbol{d}, \tag{A.9}$$

By leveraging the non-negativity, subadditivity, and submultiplicativity properties of the Euclidean norm, and under the assumption that $\rho(\boldsymbol{B}) < 1$ (as imposed in Lemma 2.1), Gelfand's formula ensures that $\|\boldsymbol{B}^m\| < 1$ if m is large enough. Consequently, we obtain:

$$\mathbb{E}\left[Q(\boldsymbol{v}_t)|\boldsymbol{v}_{t-m} = \boldsymbol{x}\right] \leq \|\boldsymbol{B}^m\|\, Q(\boldsymbol{x}) + c, \tag{A.10}$$

where $c = \|\boldsymbol{d}\| \sum_{j=0}^{m-1} \left\|\boldsymbol{B}^j\right\| < \infty$ is independent of $\boldsymbol{x}$.

Using inequality (A.10), we can easily verify condition (A.6), since $\mathbb{E}\left[Q(\boldsymbol{v}_t)|\boldsymbol{v}_{t-1} = \boldsymbol{x}\right] \leq M$, for $\boldsymbol{x} \in C$ and Q is nonnegative. Let us prove inequality (A.7). We can find a constant $R > 1$, such that $R\,\|\boldsymbol{B}^m\| < 1$, so that after multiplying inequality (A.10) by R and by adding and substracting $\|\boldsymbol{x}\|$ we obtain

$$R \cdot \mathbb{E}\left[Q(\boldsymbol{v}_t)|\boldsymbol{v}_{t-1} = \boldsymbol{x}\right] \leq \|\boldsymbol{x}\| + \left(R\,\|\boldsymbol{B}^m\| - 1\right)\|\boldsymbol{x}\| + Rc, \quad \boldsymbol{x} \in C^c,$$

which satisfies (A.7) for $r > \bar{r}$ large enough.

Now that we have proved the conditions (A.6)-(A.7), according to Tjøstheim (1990) we can state that $\{\boldsymbol{v}_t\}$ is geometrically ergodic and if it is initialized from its invariant measure π, then it is strictly stationary. Finally, the geometric ergodicity can be extended to the Markov chain $\{(R_t, \boldsymbol{v}_t)\}$ by Proposition 4(i) of Carrasco and Chen (2002), so that $\{R_t\}$ is strictly stationary and geometrically ergodic. □

A.2 Continuous-time limit

A.2.1 Proof of Proposition 2.2

Proof. Following the discretization scheme of Nelson (1990), we consider the stochastic difference equations (2.1) and (2.2) as we partition time more and more finely. We allow the

parameters of the system $\boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ to depend on the time partition $h > 0$ and make both the drift term of equation (2.1) and the variance of $Z_{1,t}$ and $Z_{2,t}$ proportional to h, so that

$$ {}_hR_{kh} = r_h + \lambda_h({}_hv_{1,kh} + {}_hv_{2,kh}) + \sqrt{{}_hv_{1,kh}} \cdot {}_hZ_{1,kh} + \sqrt{{}_hv_{2,kh}} \cdot {}_hZ_{2,kh}, $$

where ${}_hv_{1,kh}$ and ${}_hv_{2,kh}$ denote a sequence of discrete times that depend on both h and the (discrete) time index kh, $k = 0, 1, 2, \ldots$ and ${}_hZ_{i,kh} \overset{i.i.d.}{\sim} N(0, h)$, $i = 1, 2$, with ${}_hZ_{1,kh} \perp {}_hZ_{2,kh}$. After computing the square and using matrix notation, equation (2.2) is rewritten as follows

$$ {}_h\boldsymbol{v}_{kh} = \boldsymbol{\omega}_h + (\boldsymbol{\beta}_h + \boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h){}_h\boldsymbol{v}_{(k-1)h} + h^{-1/2}\boldsymbol{\alpha}_h \begin{bmatrix} {}_hZ^2_{1,(k-1)h} \\ {}_hZ^2_{2,(k-1)h} \end{bmatrix} - 2h^{-1}\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h \begin{bmatrix} {}_hZ_{1,(k-1)h}\sqrt{{}_hv_{1,(k-1)h}} \\ {}_hZ_{2,(k-1)h}\sqrt{{}_hv_{2,(k-1)h}} \end{bmatrix}, $$

where ${}_h\boldsymbol{v}_{kh} = \begin{bmatrix} {}_hv_{1,kh} & {}_hv_{2,kh} \end{bmatrix}^T$, $\boldsymbol{\omega}_h = \begin{bmatrix} \omega_{1,h} & \omega_{2,h} \end{bmatrix}^T$,

$$ \boldsymbol{\beta}_h = \begin{bmatrix} \beta_{11,h} & \beta_{12,h} \\ \beta_{21,h} & \beta_{22,h} \end{bmatrix}, \quad \boldsymbol{\alpha}_h = \begin{bmatrix} \alpha_{11,h} & \alpha_{12,h} \\ \alpha_{21,h} & \alpha_{22,h} \end{bmatrix}, \quad \boldsymbol{\gamma}_h \begin{bmatrix} \gamma_{1,h} & 0 \\ 0 & \gamma_{2,h} \end{bmatrix}, $$

where, for all h, Assumption 2.1 ensures that the process ${}_h\boldsymbol{v}_{kh}$ remains positive with probability one.

We define continuous time processes $\{{}_hR_t\}$ and $\{{}_h\boldsymbol{v}_t\}$ as

$$ {}_hR_t = {}_hR_{kh}, \quad {}_h\boldsymbol{v}_t = {}_h\boldsymbol{v}_{kh}, \quad kh \leq t < (k+1)h. $$

We derive the drift per unit of time, conditioned on information at time $(k-1)h$, for the return as

$$ \mathbb{E}_{kh}\left[h^{-1}{}_hR_{kh}\right] = r_h + \lambda_h({}_hv_{1,kh} + {}_hv_{2,kh}), \tag{A.11} $$

and for the variance process as

$$\mathbb{E}_{kh}\left[h^{-1}({}_h\boldsymbol{v}_{(k+1)h} - {}_h\boldsymbol{v}_{kh})\right] = h^{-1}\boldsymbol{\omega}_h + h^{-1}(\boldsymbol{\beta}_h + \boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h - \boldsymbol{I}_2){}_h\boldsymbol{v}_{kh} + h^{-1/2}\begin{bmatrix}\alpha_{11,h} + \alpha_{12,h}\\ \alpha_{21,h} + \alpha_{22,h}\end{bmatrix}, \tag{A.12}$$

where $\mathbb{E}_{kh}[\cdot] = \mathbb{E}[\cdot|M_{kh}]$ and M_{kh} denotes the $\sigma-$algebra generated by ${}_hR_0, \ldots, {}_hR_{(k-1)h}$ and with ${}_h\boldsymbol{v}_0, \ldots, {}_h\boldsymbol{v}_{kh}$. For the drift per unit of time to converge the following limits must exist and be finite

$$\lim_{h\to 0} h^{-1}r_h = r, \quad \lim_{h\to 0} h^{-1}\lambda_{i,h} = \lambda_i, \quad \lim_{h\to 0} h^{-1}\boldsymbol{\omega}_h = \boldsymbol{\omega}, \tag{A.13}$$

$$\lim_{h\to 0} h^{-1/2}\begin{bmatrix}\alpha_{11,h} + \alpha_{12,h}\\ \alpha_{21,h} + \alpha_{22,h}\end{bmatrix} = \begin{bmatrix}\alpha_{11} + \alpha_{12}\\ \alpha_{21} + \alpha_{22}\end{bmatrix}, \quad \lim_{h\to 0} h^{-1}(\boldsymbol{I}_2 - \boldsymbol{\beta}_h - \boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h) = \boldsymbol{\theta}. \tag{A.14}$$

Similarly, by letting ${}_h\mu_{kh} = r_h + \lambda_h({}_hv_{1,kh} + {}_hv_{2,kh})$ for convenience, we can derive the second conditional moment per unit of time for the return:

$$\mathbb{E}_{kh}\left[h^{-1}R_{kh}^2\right] = {}_h\mu_{kh}^2 + {}_hv_{1,kh} + {}_hv_{2,kh}, \tag{A.15}$$

the conditional covariance per unit of time:

$$\begin{aligned}
\mathbb{E}_{kh}\left[h^{-1}{}_hR_{kh}({}_h\boldsymbol{v}_{(k+1)h}-{}_h\boldsymbol{v}_{kh})\right] &= h^{-1}\mathbb{E}_{kh}\Bigg[({}_h\mu_{kh}+{}_hZ_{1,kh}\sqrt{{}_hv_{1,kh}}+{}_hZ_{2,kh}\sqrt{{}_hv_{2,kh}}) \qquad \text{(A.16)}\\
&\times\Bigg(\boldsymbol{\omega}_h+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2){}_h\boldsymbol{v}_{kh}+h^{-1/2}\boldsymbol{\alpha}_h\begin{bmatrix}{}_hZ_{1,kh}^2\\ {}_hZ_{2,kh}^2\end{bmatrix}\\
&-2h^{-1}\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h\begin{bmatrix}{}_hZ_{1,kh}\sqrt{{}_hv_{1,kh}}\\ {}_hZ_{2,kh}\sqrt{{}_hv_{2,kh}}\end{bmatrix}\Bigg)\Bigg]\\
&=\boldsymbol{\omega}_h\cdot{}_h\mu_{kh}\cdot{}_h\boldsymbol{v}_{kh}+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2){}_h\boldsymbol{v}_{khh}\mu_{kh}\\
&+h^{1/2}{}_h\mu_{kh}\begin{bmatrix}\alpha_{11,h}+\alpha_{12,h}\\ \alpha_{21,h}+\alpha_{22,h}\end{bmatrix}-2h^{-1}\boldsymbol{\alpha}_h\boldsymbol{\gamma}_{hh}\boldsymbol{v}_{kh},
\end{aligned}$$

and the second conditional moment of the variance per unit of time:

$$\mathbb{E}_{kh}\left[h^{-1}({}_h\boldsymbol{v}_{(k+1)h}-{}_h\boldsymbol{v}_{kh})({}_h\boldsymbol{v}_{(k+1)h}-{}_h\boldsymbol{v}_{kh})^T\right]= \tag{A.17}$$

$$= h^{-1}\mathbb{E}_{kh}\left[4h^{-2}\left(\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h\begin{bmatrix}{}_hZ_{1,kh}\sqrt{{}_hv_{1,kh}}\\ {}_hZ_{2,kh}\sqrt{{}_hv_{2,kh}}\end{bmatrix}\right)\left(\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h\begin{bmatrix}{}_hZ_{1,kh}\sqrt{{}_hv_{1,kh}}\\ {}_hZ_{2,kh}\sqrt{{}_hv_{2,kh}}\end{bmatrix}\right)^T\right.$$

$$+\left(\boldsymbol{\omega}_h+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2)_h\boldsymbol{v}_{kh}+\boldsymbol{\alpha}_hh^{-1/2}\begin{bmatrix}{}_hZ_{1,kh}^2\\ {}_hZ_{2,kh}^2\end{bmatrix}\right)$$

$$\left.\times\left(\boldsymbol{\omega}_h+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2)_h\boldsymbol{v}_{kh}+\boldsymbol{\alpha}_hh^{-1/2}\begin{bmatrix}{}_hZ_{1,kh}^2\\ {}_hZ_{2,kh}^2\end{bmatrix}\right)^T+\boldsymbol{\xi}_0\right]$$

$$= 4h^{-2}\begin{bmatrix}\alpha_{11,h}^2\gamma_{1,hh}^2v_{1,kh}+\alpha_{12,h}^2\gamma_{2,hh}^2v_{2,kh} & \alpha_{11,h}\alpha_{21,h}\gamma_{1,hh}^2v_{1,kh}+\alpha_{12,h}\alpha_{22,h}\gamma_{2,hh}^2v_{2,kh}\\ \alpha_{11,h}\alpha_{21,h}\gamma_{1,hh}^2v_{1,kh}+\alpha_{12,h}\alpha_{22,h}\gamma_{2,hh}^2v_{2,kh} & \alpha_{21,h}^2\gamma_{1,hh}^2v_{1,kh}+\alpha_{22,h}^2\gamma_{2,hh}^2v_{2,kh}\end{bmatrix}$$

$$+h^{-1}\left(\boldsymbol{\omega}_h+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2)_h\boldsymbol{v}_{kh}+3h^{1/2}\begin{bmatrix}\alpha_{11,h}+\alpha_{12,h}\\ \alpha_{21,h}+\alpha_{22,h}\end{bmatrix}\right)$$

$$\times\left(\boldsymbol{\omega}_h+(\boldsymbol{\beta}_h+\boldsymbol{\alpha}_h\boldsymbol{\gamma}_h^T\boldsymbol{\gamma}_h-\boldsymbol{I}_2)_h\boldsymbol{v}_{kh}+3h^{1/2}\begin{bmatrix}\alpha_{11,h}+\alpha_{12,h}\\ \alpha_{21,h}+\alpha_{22,h}\end{bmatrix}\right)^T,$$

where $\boldsymbol{\xi}_0$ is defined as a (2×1) vector containing terms of equation (A.17) that satisfy $\mathbb{E}_{kh}[\boldsymbol{\xi}_0]=\mathbf{0}$.

Assuming that

$$\lim_{h\to0}h^{-1/2}\boldsymbol{\gamma}_h=\boldsymbol{\gamma}, \tag{A.18}$$

and substituting (A.13)-(A.14) into equations (A.15)-(A.17) we obtain

$$\mathbb{E}_{kh}\left[h^{-1}{}_hR_{kh}^2\right] = {}_hv_{1,kh} + {}_hv_{2,kh} + o(1), \tag{A.19}$$

$$\mathbb{E}_{kh}\left[h^{-1}{}_hR_{kh}({}_h\boldsymbol{v}_{(k+1)h} - {}_h\boldsymbol{v}_{kh})\right] = -2\boldsymbol{\alpha}\boldsymbol{\gamma}\cdot {}_h\boldsymbol{v}_{kh} + \boldsymbol{o}(1), \tag{A.20}$$

$$\mathbb{E}_{kh}\left[h^{-1}({}_h\boldsymbol{v}_{(k+1)h} - {}_h\boldsymbol{v}_{kh})({}_h\boldsymbol{v}_{(k+1)h} - {}_h\boldsymbol{v}_{kh})^T\right] = \tag{A.21}$$
$$4\begin{bmatrix} \alpha_{11}^2\gamma_1^2{}_hv_{1,kh} + \alpha_{12}^2\gamma_2^2{}_hv_{2,kh} & c_{kh} \\ c_{kh} & \alpha_{21}^2\gamma_1^2{}_hv_{1,kh} + \alpha_{22}^2\gamma_2^2{}_hv_{2,kh} \end{bmatrix} + \boldsymbol{o}(1),$$

where $c_{kh} = \alpha_{11}\alpha_{21}\gamma_1^2{}_hv_{1,kh} + \alpha_{12}\alpha_{22}\gamma_2^2{}_hv_{2,kh}$. We can easily find $\{r_h, \lambda_h, \boldsymbol{\omega}_h, \boldsymbol{\alpha}_h, \boldsymbol{\gamma}_h, \boldsymbol{\beta}_h\}$ sequences that satisfy (A.13)-(A.14) and (A.18), for instance, $r_h = r\cdot h$, $\lambda_h = \lambda\cdot h$, $\boldsymbol{\omega}_h = \boldsymbol{\omega}\cdot h, \boldsymbol{\alpha}_h = \sqrt{h}\boldsymbol{\alpha}, \boldsymbol{\gamma}_h = \sqrt{h}\boldsymbol{\gamma}, \boldsymbol{\beta}_h = \boldsymbol{I}_2 - h^{3/2}\boldsymbol{\alpha}\left(\boldsymbol{\gamma}^T\boldsymbol{\gamma}\right) - \boldsymbol{\theta}h$. Let us denote the instantaneous drift function, i.e., the right hand side limit for h tending to zero of equations (A.11)-(A.12), as b:

$$b(R_t, v_{1,t}, v_{2,t}) = \begin{bmatrix} r + \lambda(v_{1,t} + v_{2,t}) \\ \omega_1 + \alpha_{11} + \alpha_{12} - \theta_{11}v_{1,t} - \theta_{12}v_{2,t} \\ \omega_2 + \alpha_{21} + \alpha_{22} - \theta_{21}v_{1,t} - \theta_{22}v_{2,t} \end{bmatrix}, \tag{A.22}$$

and denote the instantaneous covariance matrix, i.e., as the right hand side limit for h tending to zero of equations (A.19)-(A.21), as a:

$$a(R_t, v_{1,t}, v_{2,t}) = \tag{A.23}$$
$$\begin{bmatrix} v_{1,t} + v_{2,t} & -2(\alpha_{11}\gamma_1v_{1,t} + \alpha_{12}\gamma_2v_{2,t}) & -2(\alpha_{21}\gamma_1v_{1,t} + \alpha_{22}\gamma_2v_{2,t}) \\ -2(\alpha_{11}\gamma_1v_{1,t} + \alpha_{12}\gamma_2v_{2,t}) & 4(\alpha_{11}^2\gamma_1^2v_{1,t} + \alpha_{12}^2\gamma_2^2v_{2,t}) & 4(\alpha_{11}\alpha_{21}\gamma_1^2v_{1,t} + \alpha_{12}\alpha_{22}\gamma_2^2v_{2,t}) \\ -2(\alpha_{21}\gamma_1v_{1,t} + \alpha_{22}\gamma_2v_{2,t}) & 4(\alpha_{11}\alpha_{21}\gamma_1^2v_{1,t} + \alpha_{12}\alpha_{22}\gamma_2^2v_{2,t}) & 4(\alpha_{21}^2\gamma_1^2v_{1,t} + \alpha_{22}^2\gamma_2^2v_{2,t}) \end{bmatrix}.$$

Using Theorem 2.1 of Nelson (1990), the limits in (A.13)-(A.14) and (A.18), and definitions (A.22) and (A.23), we have that the processes ${}_hR_{kh}$ and ${}_h\boldsymbol{v}_{kh}$ converge to continuous time

processes R_t and $\boldsymbol{v}_t$, solutions of the following system of stochastic differential equations:

$$\begin{aligned}
dR_t &= (r + \lambda(v_{1,t} + v_{2,t}))dt + \sqrt{v_{1,t}}dW_{1,t} + \sqrt{v_{2,t}}dW_{2,t}, \\
dv_{1,t} &= (\omega_1 + \alpha_{11} + \alpha_{12} - \theta_{11}v_{1,t} - \theta_{12}v_{2,t})dt + 2\alpha_{11}\gamma_1\sqrt{v_{1,t}}dW_{3,t} + 2\alpha_{12}\gamma_2\sqrt{v_{2,t}}dW_{4,t}, \\
dv_{2,t} &= (\omega_2 + \alpha_{21} + \alpha_{22} - \theta_{21}v_{1,t} - \theta_{22}v_{2,t})dt + 2\alpha_{21}\gamma_1\sqrt{v_{1,t}}dW_{3,t} + 2\alpha_{22}\gamma_2\sqrt{v_{2,t}}dW_{4,t},
\end{aligned}$$

where the Brownian motions $W_{i,t}$, $i = 1, \ldots, 4$, are independent from the initial values $(R_0, v_{1,0}, v_{2,0})$ and from each other, except for

$$\mathbb{E}\left[dW_{1,t}dW_{3,t}\right] = -\text{sign}(\gamma_1)dt, \quad \mathbb{E}\left[dW_{2,t}dW_{4,t}\right] = -\text{sign}(\gamma_2)dt.$$

□

A.3 Risk neutral dynamics

A.3.1 Proof of Proposition 2.3

Proof. The equivalent martingale measure $\mathbb{Q}$ must satisfy

$$\mathbb{E}\left[e^{R_t}\frac{dQ_t}{dP_t}\middle|\mathcal{F}_{t-1}\right] = e^r. \tag{A.24}$$

From equations (2.1) and (2.14) we obtain

$$e^{r+\lambda(v_{1,t}+v_{2,t})}\frac{\mathbb{E}\left[e^{(\Lambda+1)(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}{\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}})Z_{2,t}}\middle|\mathcal{F}_{t-1}\right]} = e^r. \tag{A.25}$$

By taking logarithms we can rewrite (A.25) as

$$\lambda(v_{1,t}+v_{2,t})+\ln\left(\frac{\mathbb{E}\left[e^{(\Lambda+1)(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}{\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}\right)=0. \tag{A.26}$$

The conditional expectations in (A.26) can be easily computed since $Z_{1,t}$ and $Z_{2,t}$ are normally distributed and independent, we obtain

$$\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]=e^{\frac{1}{2}\Lambda^2(v_{1,t}+v_{2,t})}, \tag{A.27}$$

$$\mathbb{E}\left[e^{(\Lambda+1)(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]=e^{\frac{1}{2}(\Lambda+1)^2(v_{1,t}+v_{2,t})}. \tag{A.28}$$

Therefore, by using (A.27) and (A.28) we obtain

$$\lambda(v_{1,t}+v_{2,t})+\ln\left(\frac{e^{\frac{1}{2}(\Lambda+1)^2(v_{1,t}+v_{2,t})}}{e^{\frac{1}{2}\Lambda^2(v_{1,t}+v_{2,t})}}\right)=0, \tag{A.29}$$

which yields

$$\left(\lambda+\frac{1}{2}\right)(v_{1,t}+v_{2,t})+\Lambda(v_{1,t}+v_{2,t})=0. \tag{A.30}$$

By collecting terms we obtain

$$\left(\lambda+\frac{1}{2}+\Lambda\right)(v_{1,t}+v_{2,t})=0, \tag{A.31}$$

and, by equating the coefficients of $v_{1,t}$ and $v_{2,t}$ to zero we obtain equation (2.16). □

A.3.2 Proof of Proposition 2.4

Proof. First, we denote by $\mathbb{E}^{\mathbb{Q}}$ the expectation operator under the risk-neutral measure $\mathbb{Q}$, so that the risk-neutral (conditional) moment generating function of the process $\{\sqrt{v_{1,t}}Z_{1,t}+$

$\sqrt{v_{2,t}}Z_{2,t}\}$ is given by

$$\mathbb{E}^{\mathbb{Q}}\left[e^{\phi\left(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t}\right)}\middle|\mathcal{F}_{t-1}\right]=\mathbb{E}\left[e^{\phi(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\frac{dQ_t}{dP_t}\middle|\mathcal{F}_{t-1}\right]. \tag{A.32}$$

Using the Radon-Nikodym derivative given in equation (2.14) we can write the conditional expectation in (A.32) as

$$\mathbb{E}\left[e^{\phi(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\frac{dQ_t}{dP_t}\middle|\mathcal{F}_{t-1}\right]=\frac{\mathbb{E}\left[e^{(\phi+\Lambda)(\sqrt{v_{1,t}}Z_{1,t}\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}{\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}. \tag{A.33}$$

Since the innovations $Z_{1,t}$ and $Z_{2,t}$ are independent and based on (A.27)-(A.28) we finally have

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[e^{\phi\left(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t}\right)}\middle|\mathcal{F}_{t-1}\right]&=\frac{\mathbb{E}\left[e^{(\phi+\Lambda)(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}{\mathbb{E}\left[e^{\Lambda(\sqrt{v_{1,t}}Z_{1,t}+\sqrt{v_{2,t}}Z_{2,t})}\middle|\mathcal{F}_{t-1}\right]}\\&=e^{\frac{v_{1,t}+v_{2,t}}{2}(\phi+\Lambda)^2-\frac{1}{2}\Lambda^2(v_{1,t}+v_{2,t})}.\end{aligned} \tag{A.34}$$

It can be seen that $\frac{1}{2}\phi^2 v_{1,t}+\phi\Lambda v_{1,t}$ is the exponent of the normal (conditional) moment generating function with mean $\Lambda v_{1,t}$ and variance $v_{1,t}$, so that we denote the risk-neutral shocks as $\hat{Z}_{1,t}\sim N(\Lambda v_{1,t},1)$ and $\hat{Z}_{2,t}\sim N(\Lambda v_{2,t},1)$.

Therefore, under the risk-neutral measure, the return process in (2.1) can be written as

$$R_t=r+\lambda(v_{1,t}+v_{2,t})+\sqrt{v_{1,t}}\hat{Z}_{1,t}+\sqrt{v_{2,t}}\hat{Z}_{2,t}. \tag{A.35}$$

However, to express the normal shocks as mean-zero innovation we set

$$Z^*_{1,t}=\hat{Z}_{1,t}-\Lambda\sqrt{v_{1,t}},\quad Z^*_{2,t}=\hat{Z}_{2,t}-\Lambda\sqrt{v_{2,t}}, \tag{A.36}$$

so that $Z^*_{i,t}\sim N(0,1)$ under the risk-neutral measure, $i=1,2$. Then, by using (2.16) we can

rewrite equation (A.35) as

$$R_t = r - \frac{1}{2}(v_{1,t} + v_{2,t}) + \sqrt{v_{1,t}} Z^*_{1,t} + \sqrt{v_{2,t}} Z^*_{2,t}.$$

Lastly, under the risk-neutral measure and based on equation (A.36), the recursions in (2.2) are written as

$$\begin{bmatrix} v_{1,t+1} \\ v_{2,t+1} \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} \left(Z^*_{1,t} - \gamma^*_1 \sqrt{v_{1,t}}\right)^2 \\ \left(Z^*_{2,t} - \gamma^*_2 \sqrt{v_{2,t}}\right)^2 \end{bmatrix}, \tag{A.37}$$

where $\gamma^*_1 = \lambda + \frac{1}{2} + \gamma_1$ and $\gamma^*_2 = \lambda + \frac{1}{2} + \gamma_2$, which concludes the proof. □

A.4 Moment generating function

A.4.1 Proof of Proposition 2.5

Proof. To obtain the recursive equations for $A_t(\phi)$, $B_{1,t}(\phi)$ and $B_{2,t}(\phi)$ for $t < T$, we use the tower property of the conditional expectation

$$f_t(\phi) = \mathbb{E}\left[f_{t+1}(\phi)|\mathcal{F}_t\right], \tag{A.38}$$

so that by using equation (3.32) we can express (B.1) as

$$f_t(\phi) = \mathbb{E}\left[e^{\phi \log S_{t+1} + A_{t+1}(\phi) + B_{1,t+1}(\phi) v_{1,t+2} + B_{2,t+1}(\phi) v_{2,t+2}} \middle| \mathcal{F}_t\right]. \tag{A.39}$$

If we substitute equations (2.1) and (2.2) into (B.2) we obtain

$$f_t(\phi) = \mathbb{E}\Bigg[e^{\phi\left(\log S_t + r + \lambda(v_{1,t+1}+v_{2,t+1}) + \sqrt{v_{1,t+1}}Z_{1,t+1} + \sqrt{v_{2,t+1}}Z_{2,t+1}\right) + A_{t+1}(\phi)}$$
$$\times\, e^{B_{1,t+1}(\phi)\left(\omega_1 + \beta_{11}v_{1,t+1} + \beta_{12}v_{2,t+1} + \alpha_{11}\left(Z_{1,t+1} - \gamma_1\sqrt{v_{1,t+1}}\right)^2 + \alpha_{12}\left(Z_{2,t+1} - \gamma_2\sqrt{v_{2,t+1}}\right)^2\right)}$$
$$\times\, e^{B_{2,t+1}(\phi)\left(\omega_2 + \beta_{21}v_{1,t+1} + \beta_{22}v_{2,t+1} + \alpha_{21}\left(Z_{1,t+1} - \gamma_1\sqrt{v_{1,t+1}}\right)^2 + \alpha_{22}\left(Z_{2,t+1} - \gamma_2\sqrt{v_{2,t+1}}\right)^2\right)} \Bigg| \mathcal{F}_t \Bigg].$$

After re-arranging terms and by completing squares and some algebra we obtain

$$f_t(\phi) = \mathbb{E}\Bigg[e^{\phi(\log S_t + r) + B_{1,t+1}(\phi)\omega_1 + B_{2,t+1}(\phi)\omega_2 + A_{t+1}(\phi) + \left(\phi\lambda + \beta_{11} + B_{1,t+1}\alpha_{11}\gamma_1^2\beta_{21} + B_{2,t+1}\alpha_{21}\gamma_1^2\right)v_{1,t+1}} \tag{A.40}$$
$$\times\, e^{(\alpha_{11}B_{1,t+1}(\phi) + \alpha_{21}B_{2,t+1}(\phi))Z_{1,t+1}^2 + \sqrt{v_{1,t+1}}2\gamma_1\left(\frac{\phi}{2\gamma_1} - \alpha_{11}B_{1,t+1}(\phi) - \alpha_{21}B_{2,t+1}(\phi)\right)Z_{1,t+1}}$$
$$\times\, e^{(\alpha_{12}B_{1,t+1}(\phi) + \alpha_{22}B_{2,t+1}(\phi))Z_{2,t+1}^2 + \sqrt{v_{2,t+1}}2\gamma_2\left(\frac{\phi}{2\gamma_2} - \alpha_{12}B_{1,t+1}(\phi) - \alpha_{22}B_{2,t+1}(\phi)\right)Z_{2,t+1}}$$
$$\times\, e^{\left(\phi\lambda + \beta_{12} + B_{1,t+1}\alpha_{12}\gamma_2^2\beta_{22} + B_{2,t+1}\alpha_{22}\gamma_2^2\right)v_{2,t+1}} \Bigg| \mathcal{F}_t \Bigg],$$

so that we can make use of the following result for a standard Normal random variable Z:

$$\mathbb{E}\left[e^{aZ^2 + bZ}\right] = e^{\frac{b^2}{2(1-2a)} - \frac{1}{2}\log(1-2a)}, \quad a < \frac{1}{2}. \tag{A.41}$$

Using (B.4), the independence between $Z_{1,t}$ and $Z_{2,t}$ and subsequently equating terms in both sides of (A.40) we obtain:

$$
\begin{aligned}
A_t(\phi) =& A_{t+1}(\phi) + \phi r + B_{1,t+1}(\phi)\omega_1 + B_{2,t+1}(\phi)\omega_2 - \frac{1}{2}\log\left(1 - 2\left(B_{1,t+1}(\phi)\alpha_{11} + B_{2,t+1}(\phi)\alpha_{21}\right)\right) \\
& - \frac{1}{2}\log\left(1 - 2\left(B_{2,t+1}(\phi)\alpha_{22} + B_{1,t+1}(\phi)\alpha_{12}\right)\right), \\
B_{1,t}(\phi) =& \phi\lambda + \left(\beta_{11} + \alpha_{11}\gamma_1^2\right) B_{1,t+1}(\phi) + \left(\beta_{21} + \alpha_{21}\gamma_1^2\right) B_{2,t+1}(\phi) \\
& + \frac{2\gamma_1^2\left(\frac{\phi}{2\gamma_1} - \alpha_{11}B_{1,t+1}(\phi) - \alpha_{21}B_{2,t+1}(\phi)\right)^2}{1 - 2(\alpha_{11}B_{1,t+1}(\phi) + \alpha_{21}B_{2,t+1}(\phi))}, \\
B_{2,t}(\phi) =& \phi\lambda + \left(\beta_{22} + \alpha_{22}\gamma_2^2\right) B_{2,t+1}(\phi) + \left(\beta_{12} + \alpha_{12}\gamma_2^2\right) B_{1,t+1}(\phi) \\
& + \frac{2\gamma_2^2\left(\frac{\phi}{2\gamma_2} - \alpha_{22}B_{2,t+1}(\phi) - \alpha_{12}B_{1,t+1}(\phi)\right)^2}{1 - 2\left(\alpha_{22}B_{2,t+1}(\phi) + \alpha_{12}B_{1,t+1}(\phi)\right)}.
\end{aligned}
\tag{A.42}
$$

We can finally use equations (A.42) to calculate the coefficients recursively starting with $A_T(\phi) = B_{1,T}(\phi) = B_{2,T}(\phi) = 0$. □

Appendix B

Proofs of Chapter 3

B.1 Moment generating function

B.1.1 Proof of Proposition 3.1

Proof. To obtain the recursive equations for for $A_t(\phi)$, $B_{1,t}(\phi)$ and $B_{2,t}(\phi)$ for $t < T$, we use the tower property of the conditional expectation

$$f(t, T; \phi) = \mathbb{E}\left[f(t+1, T; \phi)|\mathcal{F}_t\right], \tag{B.1}$$

so that by using equation (3.32) we can express (B.1) as

$$f(t, T; \phi) = \mathbb{E}\left[e^{\phi \log S_{t+1} + A_{t+1}(\phi) + B_{1,t+1}(\phi)(h_{t+2} - q_{t+2}) + B_{2,t+1}(\phi) q_{t+2}} \middle| \mathcal{F}_t\right]. \tag{B.2}$$

If we substitute equations (3.20)-(3.22) into (B.2) we obtain

$$f(t,T;\phi) = \mathbb{E}\Bigg[e^{\phi\left(\log S_t + r + \lambda h_{t+1} + \sqrt{h_{t+1}} Z_{t+1}\right) + A_{t+1}(\phi)} \\ \times e^{B_{1,t+1}(\phi)\left(\tilde{\beta}(h_{t+1} - q_{t+1}) + \alpha\left(Z_{t+1}^2 + \gamma_1^2 h_{t+1} - 2\gamma_1\sqrt{h_{t+1}} Z_{t+1}\right) - \alpha\gamma_1^2 q_{t+1}\right)} \\ \times e^{B_{2,t+1}(\phi)\left(\omega + \rho q_{t+1} + \varphi\left(Z_{t+1}^2 + \gamma_2^2 h_{t+1} - 2\gamma_2\sqrt{h_{t+1}} Z_{t+1}\right)\right)} \Bigg| \mathcal{F}_t \Bigg].$$

After re-arranging terms and performing some algebra, we obtain

$$f(t,T;\phi) = \mathbb{E}\Bigg[e^{\phi(\log S_t + r) + \omega B_{2,t+1}(\phi) + A_{t+1}(\phi)} \\ \times e^{\left(\phi\lambda + (\tilde{\beta} + \alpha\gamma_1^2) B_{1,t+1}(\phi) + \varphi\gamma_2^2 B_{2,t+1}(\phi)\right)(h_{t+1} - q_{t+1}) + (\phi\lambda + (\rho + \varphi\gamma_2^2) B_{2,t+1}(\phi)) q_{t+1}} \\ \times e^{(\alpha B_{1,t+1}(\phi) + \varphi B_{2,t+1}(\phi)) Z_{t+1}^2 - 2\sqrt{h_{t+1}} Z_{t+1}\left(\alpha\gamma_1 B_{1,t+1}(\phi) + \varphi\gamma_2 B_{2,t+1}(\phi) - \frac{\phi}{2}\right)} \Bigg| \mathcal{F}_t \Bigg], \tag{B.3}$$

so that we can make use of the following result for a standard Normal random variable Z:

$$\mathbb{E}\left[e^{aZ^2 + bZ} \right] = e^{\frac{b^2}{2(1-2a)} - \frac{1}{2}\log(1-2a)}, \quad a < \frac{1}{2}. \tag{B.4}$$

Using (B.4) and by equating both sides of (B.3) we obtain:

$$A_t(\phi) = A_{t+1}(\phi) + r\phi + B_{2,t+1}(\phi)\omega - \frac{1}{2}\log\left(1 - 2(\alpha B_{1,t+1}(\phi) + \varphi B_{2,t+1}(\phi))\right), \tag{B.5}$$

$$B_{1,t}(\phi) = B_{1,t+1}(\phi)(\tilde{\beta} + \alpha\gamma_1^2) + \lambda\phi + B_{2,t+1}(\phi)\varphi\gamma_2^2 + 2\frac{(\alpha\gamma_1 B_{1,t+1}(\phi) + \varphi\gamma_2 B_{2,t+1}(\phi) - \frac{\phi}{2})^2}{1 - 2(\alpha B_{1,t+1}(\phi) + \varphi B_{2,t+1}(\phi))},$$

$$B_{2,t}(\phi) = B_{2,t+1}(\phi)(\rho + \varphi\gamma_2^2) + \lambda\phi + 2\frac{(\alpha\gamma_1 B_{1,t+1}(\phi) + \varphi\gamma_2 B_{2,t+1}(\phi) - \frac{\phi}{2})^2}{1 - 2(\alpha B_{1,t+1}(\phi) + \varphi B_{2,t+1}(\phi))}.$$

We can use equations (B.5) to recursively calculate the coefficients starting from $A_T(\phi) = 0$, $B_{1,T}(\phi) = 0$ and $B_{2,T}(\phi) = 0$. □

B.1.2 Proof of Proposition 3.2

Proof. Similarly to the proof of Proposition 3.1 we can express (B.1) as

$$f(t,T;\phi)=\mathbb{E}\left[e^{\phi\log S_{t+1}+A_{t+1}(\phi)+B_{1,t+1}(\phi)(h_{t+2}-p_{t+2})+B_{2,t+1}(\phi)(p_{t+2}-q_{t+2})+B_{3,t+1}(\phi)q_{t+2}}\middle|\mathcal{F}_t\right]. \tag{B.6}$$

If we substitute equations (3.25)-(3.28) into (B.6) we obtain

$$\begin{aligned}f(t,T;\phi)=\mathbb{E}\Bigg[&e^{\phi\left(\log S_t+r+\lambda h_{t+1}+\sqrt{h_{t+1}}Z_{t+1}\right)+A_{t+1}(\phi)}\\&\times e^{B_{1,t+1}(\phi)\left(\rho_1(h_{t+1}-p_{t+1})+\alpha_1\left(Z_{t+1}^2+\gamma_1^2h_{t+1}-2\gamma_1\sqrt{h_{t+1}}Z_{t+1}\right)-\alpha_1\gamma_1^2p_{t+1}\right)}\\&\times e^{B_{2,t+1}(\phi)\left(\rho_2(p_{t+1}-q_{t+1})+\alpha_2\left(Z_{t+1}^2+\gamma_2^2h_{t+1}-2\gamma_2\sqrt{h_{t+1}}Z_{t+1}\right)-\alpha_2\gamma_2^2q_{t+1}\right)}\\&\times e^{B_{3,t+1}(\phi)\left(\omega+\rho_3q_{t+1}+\alpha_3\left(Z_{t+1}^2+\gamma_3^2h_{t+1}-2\gamma_3\sqrt{h_{t+1}}Z_{t+1}\right)\right)}\Bigg|\mathcal{F}_t\Bigg].\end{aligned}$$

After re-arranging terms and performing some algebra, we obtain

$$\begin{aligned}f(t,T;\phi)=\mathbb{E}\Bigg[&e^{\phi(\log S_t+r)+\omega B_{3,t+1}(\phi)+A_{t+1}(\phi)}\\&\times e^{\left(\phi\lambda+(\rho_1+\alpha_1\gamma_1^2)B_{1,t+1}(\phi)+\alpha_2\gamma_2^2B_{2,t+1}(\phi)+\alpha_3\gamma_3^2B_{3,t+1}(\phi)\right)(h_{t+1}-p_{t+1})}\\&\times e^{\left(\phi\lambda+(\rho_2+\alpha_2\gamma_2^2)B_{2,t+1}(\phi)+\alpha_3\gamma_3^2B_{3,t+1}(\phi)\right)(p_{t+1}-q_{t+1})}\\&\times e^{\left(\phi\lambda+(\rho_3+\alpha_3\gamma_3^2)B_{3,t+1}(\phi)\right)q_{t+1}+(\alpha_1B_{1,t+1}(\phi)+\alpha_2B_{2,t+1}(\phi)+\alpha_3B_{3,t+1}(\phi))Z_{t+1}^2}\\&\times e^{-2\sqrt{h_{t+1}}Z_{t+1}\left(\alpha_1\gamma_1B_{1,t+1}(\phi)+\alpha_2\gamma_2B_{2,t+1}(\phi)+\alpha_3\gamma_3B_{3,t+1}(\phi)-\frac{\phi}{2}\right)}\Bigg|\mathcal{F}_t\Bigg].\end{aligned} \tag{B.7}$$

Using (B.4) and by equating both sides of (B.7) we obtain:

$$A_t(\phi) = A_{t+1}(\phi) + r\phi + B_{3,t+1}(\phi)\omega - \frac{1}{2}\log\left(1 - 2(\alpha_1 B_{1,t+1}(\phi) + \alpha_2 B_{2,t+1}(\phi) + \alpha_3 B_{3,t+1}(\phi))\right), \tag{B.8}$$

$$\begin{aligned}
B_{1,t}(\phi) &= B_{1,t+1}(\phi)(\rho_1 + \alpha_1\gamma_1^2) + \lambda\phi + B_{2,t+1}(\phi)\alpha_2\gamma_2^2 + B_{3,t+1}(\phi)\alpha_3\gamma_3^2 \\
&\quad + 2\frac{(\alpha_1\gamma_1 B_{1,t+1}(\phi) + \alpha_2\gamma_2 B_{2,t+1}(\phi) + \alpha_3\gamma_3 B_{3,t+1}(\phi) - \frac{\phi}{2})^2}{1 - 2(\alpha_1 B_{1,t+1}(\phi) + \alpha_2 B_{2,t+1}(\phi) + \alpha_3 B_{3,t+1}(\phi))}, \\
B_{2,t}(\phi) &= B_{2,t+1}(\phi)(\rho_2 + \alpha_2\gamma_2^2) + \lambda\phi + B_{3,t+1}(\phi)\alpha_3\gamma_3^2 \\
&\quad + 2\frac{(\alpha_1\gamma_1 B_{1,t+1}(\phi) + \alpha_2\gamma_2 B_{2,t+1}(\phi) + \alpha_3\gamma_3 B_{3,t+1}(\phi) - \frac{\phi}{2})^2}{1 - 2(\alpha_1 B_{1,t+1}(\phi) + \alpha_2 B_{2,t+1}(\phi) + \alpha_3 B_{3,t+1}(\phi))}, \\
B_{3,t}(\phi) &= B_{3,t+1}(\phi)(\rho_3 + \alpha_3\gamma_3^2) + \lambda\phi + 2\frac{(\alpha_1\gamma_1 B_{1,t+1}(\phi) + \alpha_2\gamma_2 B_{2,t+1}(\phi) + \alpha_3\gamma_3 B_{3,t+1}(\phi) - \frac{\phi}{2})^2}{1 - 2(\alpha_1 B_{1,t+1}(\phi) + \alpha_2 B_{2,t+1}(\phi) + \alpha_3 B_{3,t+1}(\phi))}.
\end{aligned}$$

We can use equations (B.8) to recursively calculate the coefficients starting from $A_T(\phi) = 0$, $B_{1,T}(\phi) = 0$, $B_{2,T}(\phi) = 0$ and $B_{3,T}(\phi) = 0$. □

Bibliography

T. Adrian and J. Rosenberg. Stock returns and volatility: Pricing the short-run and long-run components of market risk. *Journal of Finance*, 8(6):2997–3030, 2008.

S. Alizadeh, M-J. Brandt, and Diebold F.X. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57(3):1047–1091, 2002.

L.V. Ballestra, E. D'Innocenzo, and A. Guizzardi. Score-driven modeling with jumps: An application to S&P500 returns and options. *Journal of Financial Econometrics*, 22(2):375–406, 2023.

L.V. Ballestra, E. D'Innocenzo, and A. Guizzardi. A new bivariate approach for modeling the interaction between stock volatility and interest rate: An application to S&P500 returns and options. *European Journal of Operational Research*, 314(3):1185–1194, 2024.

G. Bassett and R. Koenker. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

D. S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies*, 9(1):69–107, 1996.

D. S. Bates. Post-'87 crash fears in S&P500 futures options. *Journal of Econometrics*, 94(1-2): 181–238, 2000.

T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 3rd edition, 2009.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

T. Bollerslev and H. Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1):33–65, 2002.

G. Bormetti, F. Corsi, and A. Majewski. Smile from the past: A general option pricing framework with multiple volatility and leverage components. *Journal of Econometrics*, 187 (2):521–531, 2015.

J.B. Carlin, D.B. Dunson, A. Gelman, D.B. Rubin, H.S. Stern, and A. Vehtari. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2016.

M. Carrasco and X. Chen. Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39, 2002.

J. Casassus and P. Collin-Dufresne. Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance*, 60(5):2283–2331, 2005.

H-W Cheng, L-H Chang, C-L Lo, and J.T. Tsai. Empirical performance of component GARCH models in pricing VIX term structure and VIX futures. *Journal of Empirical Finance*, 72: 122–142, 2023.

P. Cheridito, D. Filipović, and R.L. Kimmel. Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, 83(1):123–170, 2007.

M. Chernov, A. Gallant, E. Ghysels, and G. Tauchen. Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1):225–257, 2003.

P. Christoffersen, K. Jacobs, C. Ornthanalai, and Y. Wang. Option valuation with long-run and short-run volatility components. *Journal of Financial Economics*, 90(3):272–297, 2008.

P. Christoffersen, K. Jacobs, and S.L. Heston. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 55 (12):1914–1932, 2009.

P. Christoffersen, C. Dorion, K. Jacobs, and Y. Wand. Volatility components, affine restrictions, and nonnormal innovations. *Journal of Business Economic Statistics*, 8(4):483–502, 2010.

P. Christoffersen, K. Jacobs, and C. Ornthanalai. Dynamic jump intensities and risk premiums: Evidence from S&P500 returns and options. *Journal of Financial Economics*, 106(3):447–472, 2012.

P. Christoffersen, B. Feunou, K. Jacobs, and N. Meddahi. The economic value of realized volatility: Using high-frequency returns for option valuation. *The Journal of Financial and Quantitative Analysis*, 49(3):663–697, 2014.

P.F. Christoffersen. Evaluating Interval Forecasts. *International Economic Review*, 39:841–862, 1998.

C. Conrad and O. Kleen. Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS model . *Journal of Applied Econometrics*, 35(1):19–45, 2020.

F. Corsi, N. Fusari, and D. La Vecchia. Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics*, 107(2):284–304, 2013.

Naranjo L. Cortazar, G. An n-factor Gaussian model of oil futures prices. *Journal of Futures Markets*, 26(3):243–268, 2006.

G. Cortazar and E.S. Schwartz. The valuation of commodity contingent claims. *Journal of Derivatives*, 1(4):27–39, 1994.

G. Cortazar and E.S. Schwartz. Implementing a stochastic model for oil futures prices. *Energy Economics*, 25(3):215–238, 2003.

P.S. de Magistris and S. Grassi. It's all about volatility of volatility: Evidence from a two-factor stochastic volatility model. *Journal of Empirical Finance*, 30:62–78, 2015.

S. Deng. Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes. *Working Paper, University of California Energy Institute*, 2000. Available at: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=26aa5225d37f464a4cb02560fac28824521d0f7a.

F.X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business Economic Statistics*, 20(1):134–144, 2002.

Z. Ding and C.W.J. Granger. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215, 1996.

J.-C. Duan and J.-G. Simonato. Estimating and testing exponential-affine term structure models by Kalman filter. *Review of Quantitative Finance and Accounting*, 13:111–135, 1999.

G.R. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57 (1):405–443, 2002.

D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4): 379–406, 1996.

D. Duffie, S. F. Gray, and P. Hoang. Volatility in energy prices. In R. Jameson, editor, *Managing Energy Price Risk*, pages 273–289. Risk Publications, 1999.

D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053, 2003.

G. Durham, J. Geweke, and G. Pulak. A comment on Christoffersen, Jacobs, and Ornthanalai (2012), "Dynamic jump intensities and risk premiums: Evidence from S&P500 returns and options". *Journal of Financial Economics*, 115(1):210–214, 2015.

R. Engle, E. Ghysels, and B. Sohn. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3):776–797, 2013.

R.F. Engle and G. Lee. A long-run and short-run component model of stock return volatility. In: Engle, R.F. and White, H., Eds., Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger. *Oxford University Press*, pages 475–497, 1999.

R.F. Engle and J.G. Rangel. The spline-garch model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21(3):1187–1222, 2008.

C. Ewald and Y. Zou. Analytic formulas for futures and options for a linear quadratic jump diffusion model with seasonal stochastic volatility and convenience yield: Do fish jump? *European Journal of Operational Research*, 294(2):801–815, 2021.

A. Eydeland and K. Wolyniec. *Energy and Power Risk Management*. Wiley, 1st edition, 2003.

J. P. Fouque, G. Papanicolaou, and R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, 2000.

J. P. Fouque, G. Papanicolaou, R. Sircar, and K. Solna. Short time-scale in S&P500 volatility. *Journal of Computational Finance*, 6(4):1–23, 2003.

J. P. Fouque, G. Papanicolaou, R. Sircar, and K. Solna. Multiscale stochastic volatility asymptotics. *SIAM Journal on Multiscale Modeling and Simulation*, 2(1):22–42, 2004.

J.P. Fouque and M.J. Lorig. A fast mean-reverting correction to Heston's stochastic volatility model. *SIAM Journal on Financial Mathematics*, 2(1):221–254, 2011.

R. Gallant, C.T. Hsu, and G. Tauchen. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics*, 81(4): 617–631, 1999.

H. Geman and V.-N. Nguyen. Soybean inventory and forward curve dynamics. *Management Science*, 51(7):1076–1091, 2005.

H. Ghanbari. Persistent and transient variance components in option pricing models with variance-dependent kernel. *Journal of Empirical Finance*, page 101531, 2024.

R. Gibson and E.S. Schwartz. Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, 45(3):959–976, 1990.

J. Gil-Pelaez. Note on the inversion theorem. *Biometrika*, 38(3-4):481–482, 1951.

G. González-Rivera, T.-H. Lee, and S. Mishra. Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4):629–645, 2004.

E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer, Heidelberg, 2006.

P. R. Hansen, A. Lunde, and J. M. Nason. Model confidence set. *Econometrica*, 79(2):453 – 497, 2011.

S.L. Heston and S. Nandi. A closed-form GARCH option valuation model. *The Review of Financial Studies*, 13(3):585–625, 2000.

S. Hikspoors and S. Jaimungal. Asymptotic pricing of commodity derivatives using stochastic volatility spot models. *Applied Mathematical Finance*, 15(5-6):449–477, 2008.

John H. Hubbard and Barbara Burke Hubbard. *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*. Matrix Editions, 4th edition, 2009.

W.K. Hughen. A maximal affine stochastic volatility model of oil prices. *Journal of Future Markets*, 30(2):101–133, 2010.

A. Javaheri, D. Lautier, and A. Galli. Filtering in finance. *Wilmott Magazine*, 5:67–83, 2003.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35—-45, 1960.

C. Kelly and G.J. Lord. An adaptive splitting method for the Cox-Ingersoll-Ross process. *Applied Numerical Mathematics*, 186:252–273, 2023.

P. H. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3:73–84, 1995.

R.H. Litzenberger and N. Rabinowitz. Backwardation in oil futures markets: Theory and empirical evidence. *Journal of Finance*, 50(5):1517–1545, 1995.

B. Lutz. *Pricing of Derivatives on Mean-Reverting Assets*. Springer, 1st edition, 2010.

C. Mellios, P. Six, and A.N. Lai. Dynamic speculation and hedging in commodity futures markets with a stochastic convenience yield. *European Journal of Operational Research*, 250 (2):493–504, 2016.

S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, 1993.

D.B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1-2): 7–38, 1990.

E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, 1984.

D.H. Oh and Y.-H. Park. GARCH option pricing with volatility derivatives. *Journal of Banking and Finance*, 146(2):106718, 2023.

A. Papapantoleon, E. Eberlein, and K. Glau. Analysis of fourier transform valuation formulas and applications. *Applied Mathematical Finance*, 17(3):211–240, 2010.

B. Routledge, D.J. Seppi, and C.S. Spatt. Equilibrium forward curves for commodities. *Journal of Finance*, 55(3):1297–1338, 2000.

P.A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.

E.S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, 52(3):923–973, 1997.

E.S. Schwartz and A.B. Trolle. Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies*, 22(11):4423–4461, 2009a.

E.S. Schwartz and A.B. Trolle. A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies*, 22(5):2007–2057, 2009b.

E.S. Schwartz and A.B. Trolle. Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies*, 22(11):4423–4461, 2009c.

M.F. Schöne and S. Spinler. A four-factor stochastic volatility model of commodity prices. *Review of Derivatives Research*, 20:135–165, 2017.

K. Tang. Time-varying long-run mean of commodity prices and the modeling of futures term structures. *Quantitative Finance*, 12(5):781–790, 2012.

S.J. Taylor and X. Xu. The term structure of volatility implied by foreign exchange options. *Journal of Financial and Quantitative Analysis*, 29(1):57–74, 1994.

D. Tjøstheim. Non-linear time series and Markov chains. *Advances in Applied Probability*, 22 (3):587 – 611, 1990.

Q. Wang, Z. Wang, Q. Zhang, and Y. Zhang. Option valuation via nonaffine dynamics with realized volatility. *Journal of Empirical Finance*, 77:101486, 2024.

X.S. Yan. Valuation of commodity derivatives in a new multi-factor model. *Review of Derivatives Research*, 5:251–271, 2002.