INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Implied Risk-Neutral Distribution: Parametric approaches

Cláudio dos Santos Machado

Master in Financial Mathematics

Supervisors:
PhD João Pedro Vidal Nunes, Full Professor,
ISCTE-IUL

PhD João Pedro Bento Ruas, Assistant Professor,
ISCTE-IUL

September, 2025

Finance Department

Mathematics Department

Implied Risk-Neutral Distribution: Parametric approaches

Cláudio dos Santos Machado

Master in Financial Mathematics

Supervisors:
PhD João Pedro Vidal Nunes, Full Professor,
ISCTE-IUL

PhD João Pedro Bento Ruas, Assistant Professor,
ISCTE-IUL

September, 2025

"Success is not final, failure is not fatal: it is the courage to continue that counts."

– Unknown author

Acknowledgment

The completion of this thesis would not have been possible without the support, guidance, and encouragement of many people, to whom I am deeply grateful.

First and foremost, I would like to express my sincere appreciation to my professors and supervisors, Professor Doutor João Pedro Ruas and, most especially, Professor Doutor João Pedro Nunes. Their expertise, insightful feedback, and constant encouragement have been fundamental throughout this journey. Their dedication and availability have greatly enriched my academic and personal development, and I feel truly privileged to have worked under their supervision.

I am profoundly thankful to my family — my mother Ana, my father Hélder, and my brother Miguel — for their unconditional love, patience, and unwavering belief in me. They have been my greatest source of strength, providing the foundation upon which I could pursue this academic path. To them, I owe more than words can express.

To my colleagues and dear friends, Afonso, José, and Sérgio, I am grateful for their friendship, encouragement, and for all the shared moments of study, discussion, and laughter that made this journey not only more bearable but also memorable. Their companionship has been an invaluable part of this experience.

Finally, I would like to express my deepest gratitude to my girlfriend Joana. Her constant support, patience, and understanding have been a continuous source of motivation. Her encouragement during the most challenging times and her belief in me have been essential to the completion of this work. This achievement is as much hers as it is mine.

Resumo

Este trabalho analisa a estimação da distribuição de probabilidades de risco neutro implícita ao preço de mercado das opções financeiras, recorrendo a duas abordagens distintas: um modelo paramétrico baseado em misturas de distribuições lognormais e um modelo semi-não paramétrico baseado em funções hipergeométricas confluentes. Ambas as metodologias procuram representar de forma flexível características empíricas das distribuições implícitas, nomeadamente a assimetria, a curtose e a forma das caudas, distinguindo-se pela sua estrutura funcional e pelas propriedades de estimação que oferecem. A aplicação empírica incide sobre opções mensais do índice S&P 500 no período de janeiro a maio de 2022, caracterizado por elevada incerteza macroeconómica e geopolítica. A avaliação do desempenho é conduzida para diferentes maturidades, segundo dois critérios: (i) a capacidade de reproduzir os preços observados das opções, medida pelo Erro Percentual Absoluto Médio (MAPE), e (ii) a aptidão para replicar os momentos estatísticos da distribuição. Os resultados obtidos evidenciam as vantagens e limitações relativas de cada abordagem.

Palavras-chave: Densidade neutra ao risco; Mistura de lognormais; Funções hipergeométricas; Opções sobre o S&P 500.

Abstract

This work studies the estimation of option-implied risk-neutral distributions (RNDs) using two approaches: a parametric model based on mixtures of lognormal distributions and a semi-nonparametric model based on confluent hypergeometric functions. Both methodologies aim to capture key features of option-implied distributions such as skewness, kurtosis, and tail behavior, while differing in their functional structure and estimation properties. These models were applied to S&P 500 index monthly options between January and May 2022, a period characterized by heightened macroeconomic and geopolitical uncertainty. The empirical evaluation is conducted for the next three SPX option maturities, and performance is assessed according to two criteria: (i) the fit to observed option prices measured by the Mean Absolute Percentage Error (MAPE), and (ii) the ability to reproduce distributional moments. The results highlight the relative strengths and limitations of the two approaches.

Keywords: Risk-neutral density; Option pricing; Mixture of lognormals; Hypergeometric functions; S&P 500 options.

Contents

CHAPTER 1

Introduction

Option prices contain forward-looking information about the distribution of future asset values (Breeden and Litzenberger, 1978). This information is summarised in the *implied risk-neutral distribution* (RND), which describes the probabilities, under the risk-neutral measure, assigned by market participants to possible future prices (Bahra, 1997).
The purpose of this thesis is to study the estimation of option-implied risk-neutral distributions using two approaches: (i) a *parametric* model based on mixtures of lognormal distributions, and (ii) a *semi-nonparametric* model based on confluent hypergeometric functions, which specifies a functional form for the integral of the density constructed from combinations of the confluent hypergeometric function ${}_1F_1$. Both methods aim to capture key empirical features of option-implied distributions, including skewness, kurtosis, and tail shape, but they differ in structure and estimation properties. The empirical evaluation is carried out on S&P 500 (SPX7) monthly options from January to May 2022, a period marked by elevated uncertainty linked to monetary policy changes and geopolitical developments. Estimating a RND requires inferring the distribution from a finite and sometimes noisy set of option prices. The available methods vary in how they address this problem. Non-parametric techniques make few assumptions about the functional form but may be sensitive to data gaps or irregularities. Parametric approaches impose a specific structure, which can improve stability and ensure consistent behavior in the tails, at the cost of reduced flexibility. Mixtures of lognormals, as proposed by Bahra (1997), can approximate a variety of shapes with few parameters. Combining two or more components allows for skewness, excess kurtosis, and even multimodality, making the approach adaptable to markets with heterogeneous expectations. The parameters of each component can be linked to distinct market scenarios, adding interpretative value. An alternative is to use functional forms derived from confluent hypergeometric functions, as developed by Abadir and Rockinger (2003). This family of functions includes several well-known distributions as special cases and ensures non-negativity and proper integration by construction. With relatively few shape parameters, it can match a broad range of empirical patterns and has been shown to perform well in recovering tail behavior and higher-order moments.
This thesis implements and compares these two approaches using daily RND estimates for maturities of 1, 2, and 3 months. The comparison is based on:

- the fit to observed option prices, measured by the Mean Absolute Percentage Error (MAPE);

- the ability to reproduce distributional moments (mean, variance, skewness, kurtosis); and
- the stability of implied variance over time.

The remainder of the work reviews the literature on RND estimation, outlines the theoretical basis of the two parametric methods, describes the empirical setup, and presents the comparative results. The thesis is organized as follows. Chapter 2 reviews the main contributions in the literature on RND estimation, with emphasis on both non-parametric and parametric approaches. Chapter 4 and Chapter 5 develop the theoretical framework of the two parametric methods analyzed in this thesis: the mixture of lognormals and the hypergeometric functional form. Chapter 6.1 describes the dataset of S&P 500 index options and the procedures employed for estimation. Chapter 6.4 presents the empirical results, comparing the methods in terms of pricing accuracy, higher-order moments, and the stability of implied variance, and discusses the implications of the findings in relation to the existing literature. Finally, Chapter 7 summarizes the main results, highlights the contributions of the thesis, and points to possible directions for future research.

CHAPTER 2

Literature Review

The extraction of option-implied risk-neutral densities (RNDs) from market prices is based on the result of Breeden and Litzenberger (1978), which shows that, under certain conditions, the second derivative of the European call price with respect to the strike yields the discounted RND of the underlying asset price at maturity (Bahra, 1997; Figlewski, 2018). This relation allows the recovery of state prices from observed option prices without making assumptions on investors' preferences, given market completeness and absence of arbitrage.

2.1. Theoretical Foundations

In an arbitrage-free and complete market, option prices can be expressed using a risk-neutral probability measure Q. Under Q, the present value of any contingent claim is equal to the discounted expected payoff (Harrison and Kreps, 1979). For an European call option on the asset S, with maturity $T(\geq t)$, strike X, and risk-free rate r, the price at time t can be written as

$$c_t\left(S_t, X, T\right) = e^{-r(T-t)} \int_X^{\infty} \left(S_T - X\right) q(S_T)\, dS_T, \tag{2.1}$$

where $q(S_T)$ is the risk-neutral density. Differentiating once with respect to X gives the negative of the risk-neutral survival function, and differentiating twice gives the discounted risk-neutral density:

$$\frac{\partial^2 c_t\left(S_t, X, T\right)}{\partial X^2} = e^{-r(T-t)}\, q(X). \tag{2.2}$$

This is the Breeden–Litzenberger relation, which forms the basis for extracting $q(\cdot)$ from observed option prices. In practice, the continuum of strikes required by the theory is not available, so interpolation or functional approximation methods are used to construct a smooth option price function from discrete market data (Bahra, 1997).

2.2. Lognormal Model

The Black and Scholes (1973) model assumes that the underlying asset follows a geometric Brownian motion with constant drift and volatility, which implies that the logarithm of the terminal price is normally distributed (Bahra, 1997). The terminal RND of the underlying price is therefore lognormal. If the underlying price S_t evolves according to

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t, \tag{2.3}$$

with constant volatility σ and drift μ, then applying Itô's Lemma shows that $\ln S_T$ is normally distributed with mean $\ln S_t + \left(\mu - \frac{\sigma^2}{2}\right)(T-t)$ and variance $\sigma^2(T-t)$. Under the risk-neutral measure, μ is replaced by the risk-free rate r minus the dividend yield q, giving a lognormal density for S_T. Empirical option prices, however, often deviate from Black–Scholes predictions, producing volatility smiles and skews. These indicate that the RND departs from the lognormal form. Extensions to the lognormal model include mixtures of lognormals (Melick and Thomas, 1997; Bahra, 1997), jump–diffusion models (Bates, 1996), and stochastic volatility models (Heston, 1993).

2.3. Parametric and Nonparametric Methods

Bahra (1997) describes several estimation approaches:

(1) Discrete risk-neutral histograms based on finite-difference approximations of the Breeden–Litzenberger formula.
(2) Direct interpolation of the call price function.
(3) Interpolation of the implied volatility smile (Shimko, 1993).
(4) Fitting an assumed option pricing model to data.
(5) Fitting an assumed parametric form for the RND directly to observed option prices, including the two-lognormal mixture, applied to equity, interest rates, and foreign exchange options.

Nonparametric methods do not impose a specific functional form on the risk-neutral density (RND). Instead, they rely directly on the information contained in option prices. Some examples are kernel-based estimators (Aït-Sahalia and Lo, 1998), implied binomial trees (Rubinstein, 1994), and maximum entropy approaches (Buchen and Kelly, 1996; Stutzer, 1996). A related strand of work develops smoothing techniques, such as the Smoothed Implied Volatility Smile (SML) proposed by Bliss and Panigirtzoglou (2002), which fits a spline to the implied volatility–delta relationship and then transforms it back into option prices in order to recover the corresponding RND.

2.4. Hypergeometric Function–Based Density Functionals

Abadir and Rockinger (2003) propose a semi-nonparametric approach based on confluent hypergeometric functions. This specifies a functional form for the integral of the density, constructed from combinations of the ${}_1F_1$ confluent hypergeometric function. The family includes distributions such as normal, gamma, inverse gamma, Weibull, Pareto, and their mixtures, ensuring non-negativity and integration to unity. The hypergeometric representation allows the parameters to control location, scale, skewness, and tail behavior. It can be estimated with relatively small samples, making it suitable for markets with limited strike coverage.

CHAPTER 3

General Idea

This chapter introduces the general framework adopted to estimate the risk-neutral distribution (RND) implied by observed option prices. The motivation for recovering the RND stems from its role in representing the market's forward-looking view of the distribution of an asset's future value, under the assumption of no arbitrage. While implied volatilities capture only the dispersion of returns, the full RND provides a more complete description, including asymmetry and tail behaviour. As discussed by Figlewski (2018), the RND synthesises both investor beliefs and risk premia, and provides a compact way to extract market-based information on expectations and uncertainty. The implied RND is relevant in several applications, including derivative pricing, risk management, and financial stability analysis. It allows for pricing any European-style derivative as a discounted expected payoff under the RND. Additionally, it enables estimation of moments (e.g., skewness, kurtosis), crash probabilities, and Value-at-Risk, based not on historical data but on current market conditions (Bahra, 1997). Because of its information content, the RND is also used by policymakers to assess market-perceived uncertainty and by academics to infer risk aversion or the pricing kernel through comparison with physical return distributions (Jackwerth, 2000; Figlewski, 2018). The link between option prices and the risk-neutral distribution was formally established by Breeden and Litzenberger (1978), who showed that, under differentiability and no-arbitrage conditions, the second derivative of an European call price with respect to strike yields the state price density. Building on this foundation, and under the standard risk-neutral valuation framework, the price of an European call option can be written as:

$$\begin{aligned} c_t\left(S_t, X, T\right) &= e^{-r\tau} \int_0^{+\infty} \left(S_T - X\right)^+ q\left(t, S_t; T, S_T\right) dS_T, \\ &\approx e^{-r\tau} \int_X^{+\infty} \left(S_T - X\right) \hat{q}\left(t, S_t; T, S_T; \underline{\theta}\right) dS_T, \end{aligned} \tag{3.1}$$

and for an European put:

$$\begin{aligned} p_t\left(S_t, X, T\right) &= e^{-r\tau} \int_0^{+\infty} \left(X - S_T\right)^+ q\left(t, S_t; T, S_T\right) dS_T, \\ &\approx e^{-r\tau} \int_0^{X} \left(X - S_T\right) \hat{q}\left(t, S_t; T, S_T; \underline{\theta}\right) dS_T, \end{aligned} \tag{3.2}$$

where $\tau = T - t$, r is the continuously compounded risk-free rate, $q(\cdot)$ is the true (unobserved) RND, and $\hat{q}(\cdot; \underline{\theta})$ is a parametric approximation defined by the parameter vector $\underline{\theta} \in \mathbb{R}^p$. The goal is to estimate $\underline{\theta}$ for each trading day by minimising the distance between observed option prices and those implied by the model. This leads to the nonlinear least squares problem:

$$\min_{\underline{\theta}} \left\{ \sum_{i=1}^{n} \left[c_t^m(S_t, X_i, T) - c_t(S_t, X_i, T; \underline{\theta})\right]^2 + \sum_{i=1}^{n} \left[p_t^m(S_t, X_i, T) - p_t(S_t, X_i, T; \underline{\theta})\right]^2 \right\}, \tag{3.3}$$

where c_t^m and p_t^m denote observed market prices, and c_t, p_t are model-implied prices from $\hat{q}$. Since option quotes are available only for a discrete set of strikes and are affected by bid–ask spreads and other noise, the estimated density must satisfy:

- **Normalization:**

$$\int_0^{+\infty} \hat{q}(t, S_t; T, S_T; \underline{\theta})\, dS_T = 1,$$

- **Non-negativity:**

$$\hat{q}(t, S_t; T, S_T; \underline{\theta}) \geq 0, \quad \forall S_T > 0.$$

The functional form of $\hat{q}$ determines the balance between flexibility and tractability. This thesis considers two parametric families: (i) a mixture of lognormal distributions (Bahra, 1997), widely used for its intuitive structure and ability to approximate skewed and leptokurtic shapes; and (ii) densities based on the confluent hypergeometric functions of Abadir and Rockinger (2003), offering a general, theoretically grounded framework capable of producing smooth densities with controlled tails, even when strike coverage is limited. Increasing the number of parameters can improve in-sample fit but risks overfitting and numerical instability. As Figlewski (2018) emphasises, extracting an RND from finite and noisy option data is inherently ill-posed, and overly flexible models may capture noise rather than the underlying distribution. The estimation strategy here seeks to balance accuracy and parsimony, estimating parameters separately for each maturity and day to analyse the time variation in the implied distribution and related risk measures.

CHAPTER 4

Mixture of Lognormals

The mixture of lognormals (MLN) is proposed by Bahra (1997) and Melick and Thomas (1997) to deal with complex and often multimodal distributions that cannot be modeled using a single lognormal distribution. This model is particularly flexible in capturing the non-normality and skewness that are often observed in financial returns by assuming a flexible form for the risk-neutral density (RND) function. Unlike traditional models such as the Black and Scholes (1973), which specify a dynamic for the underlying asset price, the MLN model assumes a distributional form for the RND function, whose parameters are estimated to best fit the observed option prices. The RND provides insights into how the market perceives potential future asset prices in a world where everyone is risk-neutral. The MLN model, by blending multiple lognormal distributions, is better suited to describe real-world phenomena like the volatility smile seen in option prices, a feature not captured by the Black-Scholes (1973) model. Essentially, the MLN model's ability to combine various distributions gives it an edge, allowing it to more accurately reflect different market scenarios for the underlying asset price. It is well established that the time-t price of an European-style call option with underlying asset price S, with strike X and maturity at time $T(\geq t)$ can be expressed as the discounted sum of all expected future payoffs:

$$c_t(S_t, X, T) = e^{-r\tau} \int_X^\infty q(S_T)(S_T - X)dS_T, \tag{4.1}$$

where $q(S_T) \equiv q(t, S_t; T; S_T)$ is the risk-neutral density function, r is the risk-free (constant) instant rate and $\tau := T - t$. Bahra (1997) suggests that any functional form for the risk-neutral density function, $q(S_T)$, can be used in equation (4.1). The parameters of the chosen density function can be recovered through numerical optimization, which involves minimizing the difference between prices obtained through the model and market prices. As the observed price distributions of several underlying assets are in the vicinity of a normal distribution, Bahra (1997) chooses the framework suggested by Ritchey (1990), which assumes that $q(S_T)$ is the weighted sum of k lognormal density functions. In other words, we can express $q(S_T)$ as:

$$q(S_T, \theta) = \sum_{i=1}^{k} \left[\omega_i L(\alpha_i, \beta_i, S_T)\right], \tag{4.2}$$

where $L(\alpha_i, \beta_i, S_T)$ represents the lognormal density function for the i-th component, with parameters α_i and β_i, and $w_i(\geq 0)$ is the corresponding weight that is such that

$$\sum_{i=1}^{k} w_i = 1, \tag{4.3}$$

while the lognormal density function is given by

$$L(\alpha_i, \beta_i, S_T) := e^{-\frac{(lnS_T - \alpha_i)^2}{2\beta_i^2}} \frac{1}{\beta_i\sqrt{2\pi}S_T}, \tag{4.4}$$

if it is assumed that $dS_t = \mu_i S_t\, dt + \sigma_i S_t\ dW_t$.
Combining equations (4.2) to (4.4) with equation (4.1) , then

$$c_t(S_t, X, \tau) = e^{-r\tau} \int_X^\infty (S_T - X) \sum_{i=1}^{k} \omega_i L(\alpha_i, \beta_i, S_T) dS_T. \tag{4.5}$$

From equation (4.5), we can obtain a closed-form formula for the price of the European-style call option, which can be written as:

PROPOSITION 4.1. *Aproximating the RND through the mixture of k lognormal densities as given by equation (4.2), the time-t price of an European-style call on an asset with time-t price S_t, strike X and maturity at time $T(\geq t)$ is equal to:*

$$c_t(S_t, X, \tau) = e^{-r\tau} \sum_{i=1}^{k} \left\{\omega_i e^{\alpha_i + \frac{1}{2}\beta_i^2} \Phi\left(d_{1,i}\right)\right\} - e^{-r\tau} X \sum_{i=1}^{k} \left\{\omega_i \Phi\left(d_{2,i}\right)\right\}, \tag{4.6}$$

where

$$d_{1,i} = \frac{-ln(X) + \alpha_i + \beta_i^2}{\beta_i}, \tag{4.7}$$

and

$$d_{2,i} = d_{1,i} - \beta_i, \tag{4.8}$$

with $\Phi(\cdot)$ representing the cumulative probability function of the univariate standard normal distribution.

PROOF. Equation (4.5) can be rewritten as follows:

$$\begin{aligned} c_t(S_t, X, \tau) &= e^{-r\tau} \sum_{i=1}^{k} \int_X^\infty (S_T - X)\omega_i L(\alpha_i, \beta_i, S_T) dS_T \\ &= e^{-r\tau} \sum_{i=1}^{k} \left\{ \int_X^\infty S_T \omega_i L(\alpha_i, \beta_i, S_T) dS_T - X \int_X^\infty \omega_i L(\alpha_i, \beta_i, S_T) dS_T \right\}. \end{aligned} \tag{4.9}$$

We shall start by examining the initial integral on the left-hand side of equation (4.9). Drawing upon the formula for the lognormal density function given in equation (4.4), the first integral can be reformulated as:

$$\begin{aligned}\sum_{i=1}^{k}\int_{X}^{\infty} S_T\omega_i L(\alpha_i,\beta_i,S_T)dS_T &= \sum_{i=1}^{k}\int_{X}^{\infty} S_T\omega_i e^{-\frac{(lnS_T-\alpha_i)^2}{2\beta_i^2}}\frac{1}{\beta_i\sqrt{2\pi}S_T}dS_T\\ &= \sum_{i=1}^{k}\int_{X}^{\infty}\frac{\omega_i}{\beta_i\sqrt{2\pi}}e^{-\frac{(lnS_T-\alpha_i)^2}{2\beta_i^2}}dS_T. \quad (4.10)\end{aligned}$$

A variable substitution facilitates a transition from lognormal distributions to normal distributions. By letting $\gamma := ln(S_T)$, we obtain:

$$S_T = e^{\gamma} \implies dS_T = e^{\gamma}d\gamma, \quad (4.11)$$

where γ is distributed as a mixture of normal distributions. Applying the transformation to equation (4.10),

$$\sum_{i=1}^{k}\int_{X}^{\infty} S_T\omega_i L(\alpha_i,\beta_i,S_T)dS_T = \sum_{i=1}^{k}\int_{ln(X)}^{\infty}\frac{\omega_i}{\beta_i\sqrt{2\pi}}e^{\gamma-\frac{(\gamma-\alpha_i)^2}{2\beta_i^2}}d\gamma, \quad (4.12)$$

we can reformulate the exponents within the exponential terms by employing the completion of the square method as follows:

$$\begin{aligned}\gamma-\frac{(\gamma-\alpha_i)^2}{2\beta_i^2} &= \frac{-\left(\gamma^2-2\alpha_i\gamma+\alpha_i^2-2\beta_i^2\gamma\right)}{2\beta_i^2} && (4.13)\\ &= \frac{-\left[\gamma^2-2\gamma\left(\alpha_i+\beta_i^2\right)+\alpha_i^2+2\alpha_i\beta_i^2+\beta_i^4\right]+2\alpha_i{\beta_i}^2+\beta_i^4}{2\beta_i^2} && (4.14)\\ &= \frac{-\left[\gamma-(\alpha_i+\beta_i^2)\right]^2}{2\beta_i^2}+\alpha_i+\frac{1}{2}\beta_i^2. && (4.15)\end{aligned}$$

Hence, equation (4.12) becomes

$$\begin{aligned}\sum_{i=1}^{k}\int_{X}^{\infty} S_T\omega_i L(\alpha_i,\beta_i,S_T)dS_T &= \sum_{i=1}^{k}\int_{\ln(X)}^{\infty}\frac{\omega_i}{\beta_i\sqrt{2\pi}}e^{-\frac{\left[\gamma-\left(\alpha_i+\beta_i^2\right)\right]^2}{2\beta_i^2}+\alpha_i+\frac{1}{2}\beta_i^2}d\gamma && (4.16)\\ &= \sum_{i=1}^{k}\left\{\omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2}\int_{\ln(X)}^{\infty}\frac{1}{\beta_i\sqrt{2\pi}}e^{-\frac{\left[\gamma-\left(\alpha_i+\beta_i^2\right)\right]^2}{2\beta_i^2}}d\gamma\right\}. && (4.17)\end{aligned}$$

Since the integrand of equation (4.17) is the density function of a normal random variable, with variance β_i^2 and mean $(\alpha_i + \frac{1}{2}\beta_i^2)$, then:

$$\begin{aligned}\sum_{i=1}^{k}\int_{X}^{\infty} S_T\omega_i L(\alpha_i,\beta_i,S_T)dS_T &= \sum_{i=1}^{k}\left\{\omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2}\left[1-\int_{-\infty}^{\ln(X)}\frac{1}{\beta_i\sqrt{2\pi}}e^{-\frac{\left[\gamma-\left(\alpha_i+\beta_i^2\right)\right]^2}{2\beta_i^2}}d\gamma\right]\right\}\\ &= \sum_{i=1}^{k}\left\{\omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2}\left[1-\Phi\left(\frac{\ln(X)-(\alpha_i+\beta_i^2)}{\beta_i}\right)\right]\right\}\\ &= \sum_{i=1}^{k}\left\{\omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2}\left[\Phi\left(\frac{-\ln(X)+(\alpha_i+\beta_i^2)}{\beta_i}\right)\right]\right\}, && (4.18)\end{aligned}$$

and equation (4.9) becomes

$$c_t(S_t, X, \tau) = e^{-r\tau} \sum_{i=1}^{k} \left\{ \omega_i e^{\alpha_i + \frac{1}{2}\beta_i^2} \left[\Phi \left(\frac{-\ln(X) + (\alpha_i + \beta_i^2)}{\beta_i} \right) \right] \right\}$$
$$- e^{-r\tau} \sum_{i=1}^{k} \left\{ X \int_X^{\infty} \omega_i L(\alpha_i, \beta_i, S_T) dS_T \right\}. \quad (4.19)$$

Taking into account the remaining integral,

$$\int_X^{\infty} \sum_{i=1}^{k} \omega_i L(\alpha_i, \beta_i, S_T) dS_T = \sum_{i=1}^{k} \int_X^{\infty} \omega_i e^{-\frac{(lnS_T - \alpha_i)^2}{2\beta_i^2}} \frac{1}{\beta_i \sqrt{2\pi} S_T} dS_T$$
$$= \sum_{i=1}^{k} \omega_i \int_X^{\infty} \frac{1}{S_T \beta_i \sqrt{2\pi}} e^{-\frac{(lnS_T - \alpha_i)^2}{2\beta_i^2}} dS_T. \quad (4.20)$$

Employing the following substitution,

$$\gamma_i = \frac{[ln(S_T) - \alpha_i]}{\beta_i} \quad (4.21)$$

which implies

$$dS_T = S_T \beta_i d\gamma_i, \quad (4.22)$$

then equation (4.20) becomes

$$\int_X^{\infty} \sum_{i=1}^{k} \omega_i L(\alpha_i, \beta_i, S_T) dS_T = \sum_{i=1}^{k} \omega_i \int_{\frac{ln(X) - \alpha_i}{\beta_i}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\gamma_i^2}{2}} d\gamma_i$$
$$= \sum_{i=1}^{k} \omega_i \left[1 - \int_{-\infty}^{\frac{ln(X) - \alpha_i}{\beta_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\gamma_i^2}{2}} d\gamma_i \right]$$
$$= \sum_{i=1}^{k} \omega_i \left[1 - \Phi \left(\frac{ln(X) - \alpha_i}{\beta_i} \right) \right]$$
$$= \sum_{i=1}^{k} \omega_i \Phi \left(\frac{-ln(X) + \alpha_i}{\beta_i} \right). \quad (4.23)$$

Finally, the closed-form solution for the European-style call, using the mixture of lognormal functions, can be written as:

$$c_t(S_t, X, \tau) = e^{-r\tau} \sum_{i=1}^{k} \left\{ \omega_i e^{\alpha_i + \frac{1}{2}\beta_i^2} \left[\Phi \left(\frac{-\ln(X) + (\alpha_i + \beta_i^2)}{\beta_i} \right) \right] \right\}$$
$$- e^{-r\tau} X \sum_{i=1}^{k} \omega_i \left[\Phi \left(\frac{-\ln(X) + \alpha_i}{\beta_i} \right) \right]$$
$$= e^{-r\tau} \sum_{i=1}^{k} \left\{ \omega_i e^{\alpha_i + \frac{1}{2}\beta_i^2} \left[\Phi \left(d_{1,i} \right) \right] \right\} - e^{-r\tau} X \sum_{i=1}^{k} \omega_i \left[\Phi \left(d_{2,i} \right) \right], \quad (4.24)$$

with

$$d_{1,i} = \frac{-ln(X) + \alpha_i + \beta_i^2}{\beta_i}, \tag{4.25}$$

and,

$$d_{2,i} = d_{1,i} - \beta_i. \tag{4.26}$$

□

Bahra (1997) argues that, in many markets, options are usually traded within a narrow range of exercise prices. This situation places a cap on the number of distributional parameters we can extract from the available data. Hence, Bahra (1997) recommends, a two lognormal mixture framework approach. Given $k = 2$, equation (4.24) translates into the subsequent closed-form equation:

$$\begin{aligned} c_t(S_t, X, \tau) = e^{-r\tau} \Bigg\{ & w\Big[e^{\alpha_1 + \frac{1}{2}\beta_1^2}\Phi(d_1) - X\Phi(d_2)\Big] \\ & + (1-w)\Big[e^{\alpha_2 + \frac{1}{2}\beta_2^2}\Phi(d_3) - X\Phi(d_4)\Big]\Bigg\}, \end{aligned} \tag{4.27}$$

where

$$d_1 = \frac{-ln(X) + \alpha_1 + \beta_1^2}{\beta_1}, \tag{4.28}$$

$$d_2 = d_1 - \beta_1, \tag{4.29}$$

$$d_3 = \frac{-ln(X) + \alpha_2 + \beta_2^2}{\beta_2}, \tag{4.30}$$

and

$$d_4 = d_3 - \beta_2. \tag{4.31}$$

The parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \omega$ are estimated through numerical optimization. However, when analyzing a dataset composed of European calls and puts, certain complexities arise, notably the potential for singularities. These challenges necessitate a methodical approach. Therefore, our analytical focus will be meticulously restricted to out-of-the- -money call and put options (in accordance with the VIX calculation methodology, only options with non-zero bid prices are employed). To streamline our analysis and ensure consistency, we will utilize the put-call parity theorem,

$$c_t(S_t, X, \tau) - p_t(S_t, X, \tau) = S_t - Xe^{-r\tau}, \tag{4.32}$$

which allows us to coherently transform European-style put prices into their equivalent European call option counterparts. In this way, we can determine the five parameters by solving the minimization problem with one constraint:

$$\begin{aligned}
&\min_{\alpha_1,\alpha_2,\beta_1,\beta_2,\omega} \sum_{i=1}^{n} \left[c_t(S_t, X_i, \tau) - \hat{c}_i\right]^2 \\
&\text{s.t.} \quad S_t e^{(r-q)\tau} = w e^{\alpha_1+\frac{1}{2}\beta_1^2} + (1-w)e^{\alpha_2+\frac{1}{2}\beta_2^2},
\end{aligned} \tag{4.33}$$

where $c_t(S_t, X_i, \tau)$ is given by equation (4.27) and $\hat{c}_i$ correspond to the observed market values for each option. The constraint (4.33) is imposed because it dictates that, in the absence of arbitrage, the projected forward price of the underlying asset must be align with its actual forward price, this is the no-arbitrage condition, see Appendix A. We employed Matlab's optimization solver, fmincon, to estimate the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \omega$ that minimize equation (4.33).

Extending our analytical scope, we also implement a three lognormal mixture into our framework. For $k = 3$, equations (4.24) to (4.26) yield the corresponding closed-form solutions:

$$\begin{aligned}
c_t(S_t, X, \tau) = e^{-r\tau}\Big\{ & w_1 \left[e^{\alpha_1+\frac{1}{2}\beta_1^2}\Phi(d_1) - X\Phi(d_2)\right] + w_2 \left[e^{\alpha_2+\frac{1}{2}\beta_2^2}\Phi(d_3) - X\Phi(d_4)\right] \\
& + (1-w_1-w_2)\left[e^{\alpha_3+\frac{1}{2}\beta_3^2}\Phi(d_5) - X\Phi(d_6)\right]\Big\},
\end{aligned} \tag{4.34}$$

where

$$d_1 = \frac{-ln(X) + \alpha_1 + \beta_1^2}{\beta_1}, \tag{4.35}$$

$$d_2 = d_1 - \beta_1, \tag{4.36}$$

$$d_3 = \frac{-ln(X) + \alpha_2 + \beta_2^2}{\beta_2}, \tag{4.37}$$

$$d_4 = d_3 - \beta_2, \tag{4.38}$$

$$d_5 = \frac{-ln(X) + \alpha_3 + \beta_3^2}{\beta_3}, \tag{4.39}$$

and

$$d_6 = d_5 - \beta_3. \tag{4.40}$$

Following the same reasoning as for $k = 2$, the parameters $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \omega_1, \omega_2$ can be estimated by solving a minimization problem with one constraint:

$$\min_{\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2,\beta_3,\omega_1,\omega_2} \sum_{i=1}^{n} \left[c_t(S_t, X_i, \tau) - \hat{c}_i\right]^2 \tag{4.41}$$

$$\text{s.t.} \quad S_t e^{(r-q)\tau} = w_1 e^{\alpha_1+\frac{1}{2}\beta_1^2} + w_2 e^{\alpha_2+\frac{1}{2}\beta_2^2} + (1-w_1-w_2)e^{\alpha_3+\frac{1}{2}\beta_3^2}, \tag{4.42}$$

where $c_t(S_t, X_i, \tau)$ is given by equation (4.34) and $\hat{c}_i$ correspond to the observed market values for each option. As previously elucidated, the constraint imposed is congruent with the no-arbitrage condition. We employed Matlab's optimization solver, fmincon, to estimate the parameters $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \omega_1, \omega_2$ that minimize equation (4.34). The results will be discussed in a dedicated chapter.

CHAPTER 5

Mixture of hypergeometric functions

5.1. An introduction to hypergeometric functions

In this section, we will briefly present the fundamental theoretical concepts that describe a class of functions. One of the most versatile and extensively studied among them is the hypergeometric function, a generalization of the geometric series. The geometric series itself draws its nomenclature from the geometric interpretation of its initial terms, ranging from lines to squares and cubes, and extending to higher-dimensional "hypercube" terms. The study of hypergeometric functions, however, goes beyond the geometric series to offer a more comprehensive analytical tool, one that encapsulates a wide variety of simpler functions like exponential, logarithmic, normal, gamma, inverse gamma, Weibull or Pareto.

The historical underpinnings of the hypergeometric function can be traced back to the work of Wallis that first extended the ordinary geometric series to its hypergeometric counterpart. His formulation was later adapted into what is now known as the Pochhammer symbol, a notation widely used for representing rising factorials. Following this foundational work on hypergeometric functions, eminent mathematicians like Leonhard Euler, Alexandre-Theophile Vandermonde and Carl Friedrich Gauss made significant contributions to expanding the function's mathematical properties and broadening its applications.

The generalized hypergeometric function is defined by the power series expansion:

$$ {}_pF_q\begin{pmatrix}\alpha_1, ..., \alpha_p; & \\ & z \\ \beta_1, ..., \beta_q; & \end{pmatrix} \equiv \sum_{j=0}^{\infty} \frac{\prod_{k=1}^{p}(\alpha_k)_j}{\prod_{k=1}^{q}(\beta_k)_j} \frac{z^j}{j!}, \tag{5.1} $$

where $(\alpha)_j \equiv (\alpha)(\alpha+1)...(\alpha+j-1) = \frac{\Gamma(\alpha+j)}{\Gamma(\alpha)}$, $\Gamma(\nu) \in \mathbb{R}$ is the gamma function and $-\beta \notin \mathbb{N} \cup \{0\}$.

Within the framework of generalized hypergeometric functions, there exists a particular subclass of functions that will be the focus of our detailed investigation. This specific function is designated as the confluent hypergeometric function or Kummer's function, denoted by ${}_1F_1$:

$$ {}_1F_1(\alpha_1; \beta_1; z) \equiv \sum_{j=0}^{\infty} \frac{\alpha_j}{\beta_j} \frac{z^j}{j!}. \tag{5.2} $$

Two key factors that will be beneficial throughout our analysis are the first derivative of ${}_1F_1$, expressed as:

$$\frac{d\ \left[{}_1F_1(\alpha,\beta,z)\right]}{dz}=\frac{\alpha}{\beta}\left[{}_1F_1(\alpha+1,\beta+1,z)\right], \tag{5.3}$$

and the asymptotic representation of Kummer's function for $z\in\mathbb{R}$ (Abramowitz and Stegun, 1972):

$${}_1F_1(\alpha,\beta,z)=\begin{cases}\frac{\Gamma(\beta)}{\Gamma(\beta-\alpha)}\left|z\right|^{-a}(1+O(\frac{1}{2})) & as\ z\to-\infty\\ \frac{\Gamma(\beta)}{\Gamma(\alpha)}\left|z\right|^{a-c}e^{z}(1+O(\frac{1}{2})) & as\ z\to\infty.\end{cases} \tag{5.4}$$

Some special cases of the Krummer function are well-known in the literature and used in many contexts of option pricing.

5.2. Special cases related to the Kummer function

A fundamental illustration of the transformation of a function into a Kummer function is provided by the exponential function,

$$e^{z}\equiv\ {}_1F_1(a;a;z), \tag{5.5}$$

where a is arbitrary. Another one is the gamma function:

$$\Gamma(\nu,z)\equiv\int_0^z e^{-x}x^{\nu-1}dx=\frac{z^{\nu}}{\nu}\ {}_1F_1(\nu;\nu+1;z), \tag{5.6}$$

with $-\nu\notin\cup\{0\}$. Additionaly,

$$\Phi(z)\equiv\int_{-z}^{z}\frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}dx=\frac{1}{2}+\frac{z}{\sqrt{2\pi}}\ {}_1F_1\left(\frac{1}{2};\frac{3}{2};-\frac{z^2}{2}\right), \tag{5.7}$$

where $\Phi(\cdot)$ represents the cumulative probability function of the univariate standard normal distribution. Finally,

$$I_{\nu}(z)\equiv\sum_{j=0}^{\infty}\frac{\left(\frac{z}{2}\right)^{2j+\nu}}{j!\Gamma(j+1+\nu)}=\frac{1}{\Gamma(\nu+1)}\left(\frac{z}{2}\right)^{\nu}\ {}_1F_1\left(\nu+1;2\nu+1;-\frac{z^2}{4}\right), \tag{5.8}$$

where $I_{\nu}(\cdot)$ is the modified Bessel function of the first kind of order ν and $\Gamma(\cdot)$ is the gamma function.

5.3. Option values

The mixture of hypergeometric functions (MHF) is proposed by Abadir and Rockinger (2003) for the estimation of the RND function, and is based on a mixture of confluent hypergeometric functions. This methodological approach eschews the need for preselecting a rigid functional structure for the corresponding RND function. It employs an adaptable equation, formulated to encompass a wide array of density functions including, but not restricted to, normal, gamma, inverse gamma, Weibull, and Pareto distributions, as well as composite permutations of these. Abadir and Rockinger (2003) argue that the use of the confluent hypergeometric function stands out for its ability to capture specialized subsets of both incomplete gamma and normal distributions, as well as composites of the two.

This feature aims the density function with a dual advantage: it provides a more precise estimator for limited sample sizes compared to entirely non-parametric models, while simultaneously allowing greater methodological adaptability than conventional parametric techniques, as it places no constraints on the form of the functional representation.
Abadir and Rockinger (2003) introduce a closed-form formula for the forward price of the European-style call option, given by a mixture of two confluent hypergeometric functions:

$$\begin{aligned} C(X) \equiv e^{r\tau}c_t(S_t, X, \tau) = c_1{+}c_2X + \mathbb{1}_{X>m_1}a_1(z-m_1)^{b_1}\ {}_1F_1(a_2; a_3; b_2(z-m_1)^{b_3}) \\ + (a_4)\ {}_1F_1(a_5; a_6; b_4(X-m_2)^2), \end{aligned} \tag{5.9}$$

where $a_3, a_6 \notin \mathbb{N} \cup \{0\}$, $b_2, b_4 \in \mathbb{R}^-$, and for $X \in [X_l, X_u] \subseteq \mathbb{R}$. The indicator function ($\mathbb{1}.$) is required to represent a component of the density with bounded support. It is also sufficient for keeping the function real-valued for b_1 and b_3.

5.4. Transition density

In an arbitrage-free market, as outlined in the previous chapter, there is a RND that allows that, at time-t, the price of an European-style call option with underlying asset price S, with strike X and maturity at time $T(\geq t)$ can be expressed as the discounted sum of all expected future payoffs, is given in equation (4.1).
As observed by Breeden and Litzenberger (1978), taking the derivative of the integral in order to the strike X yields:

$$\frac{dc_t(S_t, X, T)}{dX} = -e^{r\tau}\int_X^\infty q(S_T)dS_T = -e^{-r\tau}\left[1 - Q(X)\right], \tag{5.10}$$

where $Q(\cdot)$ is the cumulative distribution function associated to the risk-neutral density function $q(\cdot)$. The second derivative can be readily derived from equation (5.10) as:

$$\left.\frac{d^2c_t(S_t, X, T)}{dX^2}\right|_{X=S_T} = e^{-r\tau}q(S_T). \tag{5.11}$$

Using, instead, the forward price $C(X)$, allows us to express equation (5.11) in the form:

$$\left.\frac{d^2C(X)}{dX^2}\right|_{X=S_T} = q(S_T), \tag{5.12}$$

where $q(S_T)$ is the risk-neutral density function.
As shown in the previous equation, to obtain the risk-neutral density function we can proceed by taking the second derivative of equation (4.1). The process begins by calculating the first derivative of the function in order to z:

$$\begin{aligned} \frac{dC(z)}{dz} = c_2 + \mathbb{1}_{z>m_1}a_1(z-m_1)^{b_1-1}\Big[(b_1)_1F_1(a_2; a_3; b_2(z-m_1)^{b_3}) \\ + \frac{a_2}{a_3}b_2b_3(z-m_1)^{b_3}{}_1F_1(a_2+1; a_3+1; b_2(z-m_1)^{b_3})\Big] \\ + 2a_4\frac{a_5}{a_6}b_4(z-m_2)_1F_1(a_5+1; a_6+1; b_4(z-m_2)^2), \end{aligned} \tag{5.13}$$

and the second derivative is, therefore, given by:

$$
\begin{aligned}
\frac{d^2C(z)}{dz^2} &= \mathbb{1}_{z>m_1} a_1 (z-m_1)^{b_1-2} \Bigg[(b_1)(b_1-1) {}_1F_1(a_2; a_3; b_2(z-m_1)^{b_3}) \\
&+ \frac{a_2}{a_3} b_2 b_3 (2b_1 + b_3 - 1)(z-m_1)^{b_3} {}_1F_1(a_2+1; a_3+1; b_2(z-m_1)^{b_3}) \\
&+ \frac{a_2(a_2+1)}{a_3(a_3+1)} b_2^2 b_3^2 (z-m_1)^{2b_3} {}_1F_1(a_2+2; a_3+2; b_2(z-m_1)^{b_3}) \Bigg] \\
&+ 2a_4 \frac{a_5}{a_6} b_4 \Bigg[{}_1F_1(a_5+1; a_6+1; b_4(z-m_2)^2) \\
&+ 2\frac{a_5+1}{a_6+1} b_4 (z-m_2)^2 {}_1F_1(a_5+2; a_6+2; b_4(z-m_2)^2) \Bigg].
\end{aligned}
\tag{5.14}
$$

By substituting the obtained results into equation (5.12),

$$
\begin{aligned}
q(z) &= \mathbb{1}_{z>m_1} a_1 (z-m_1)^{b_1-2} \Bigg[(b_1)(b_1-1) {}_1F_1(a_2; a_3; b_2(z-m_1)^{b_3}) \\
&+ \frac{a_2}{a_3} b_2 b_3 (2b_1 + b_3 - 1)(z-m_1)^{b_3} {}_1F_1(a_2+1; a_3+1; b_2(z-m_1)^{b_3}) \\
&+ \frac{a_2(a_2+1)}{a_3(a_3+1)} b_2^2 b_3^2 (z-m_1)^{2b_3} {}_1F_1(a_2+2; a_3+2; b_2(z-m_1)^{b_3}) \Bigg] \\
&+ 2a_4 \frac{a_5}{a_6} b_4 \Bigg[{}_1F_1(a_5+1; a_6+1; b_4(z-m_2)^2) \\
&+ 2\frac{a_5+1}{a_6+1} b_4 (z-m_2)^2 {}_1F_1(a_5+2; a_6+2; b_4(z-m_2)^2) \Bigg],
\end{aligned}
\tag{5.15}
$$

where $a_3, a_6 \notin \mathbb{N} \cup \{0\}$ and $b_2, b_4 \in \mathbb{R}^-$. The variable z is related to S_T $(z = \alpha + \beta S_T)$, and will be discussed in greater detail throughout the chapter.

5.5. Parameter restrictions

The formula providing the closed-form solution for the price of the European-style call option, as outlined in equation (5.9), encompasses a total of twelve parameters $(m_1, m_2, a_1, a_2, a_3, a_4, a_5, a_6, b_1, b_2, b_3, b_4)$ that require an accurate estimation. Abadir and Rockinger (2003) underline the presence of certain inherent restrictions that must be imposed. It is through an analysis of these restrictions that we are poised to effectively reduce the initial twelve parameters to a more manageable subset of seven $(m_1, m_2, a_2, a_3, b_2, b_3, b_4)$. This subset, consequently, becomes the focus of our estimation efforts as we delve further into our study.

The first two constraints are related to the RND and its integration, and can be described as ensuring that

$$\int_{z_\ell}^{z_u} q(z)\, dz = 1, \tag{5.16}$$

where $z_\ell \in \mathbb{R}$ and $z_u \in \mathbb{R}$ denote, respectively, the lower and upper integration limits.

Substituting $q(z)$ with the relationship expressed in equation (5.12), we can rewrite the previous equation as:

$$\int_{z_l}^{z_u} \frac{d^2C(z)}{dz^2} dz = 1 \iff \left.\frac{dC(z)}{dz}\right|_{z_l}^{z_u} = 1 \iff \left.\frac{dC(z)}{dz}\right|_{z=z_u} - \left.\frac{dC(z)}{dz}\right|_{z=z_l} = 1, \tag{5.17}$$

which implies that

$$\left.\frac{dC(z)}{dz}\right|_{z=z_u} = 0, \tag{5.18}$$

and

$$\left.\frac{dC(z)}{dz}\right|_{z=z_l} = -1. \tag{5.19}$$

From equation (5.19) we obtain[1] the constraint associated with the parameter c_2:

$$c_2 = -1 + a_4\sqrt{-b_4\pi}. \tag{5.20}$$

And, from equation (5.18) we obtain[1] the constraint associated with the parameter a_4:

$$a_4 = \frac{1}{2\sqrt{-b_4\pi}} \left[1 - a_1(-b_2)^{-a_2} \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}\right]. \tag{5.21}$$

As X tends towards infinite, the call option price approaches zero, $c_t(S_t, \infty, T) = 0$. This outcome is inherent to call options that are in close proximity to being OTM (out of the money), as the probability of them transitioning into ITM (in the money) becomes infinitesimally low. Within the context of equation (5.9), the option value will possess a minimum value of c_1, yielding the subsequent simplification:

$$c_1 = -c_2 m_2. \tag{5.22}$$

The remaining constraints we are going to adopt will correspond to the recommendations outlined by Abadir and Rockinger (2003). These constraints are substantiated by the fact that, within the numerical applications they conducted, the restrictions implied by the joint hypothesis of those constraints could not be refuted through an F-test. We proceed by employing the subsequent assumptions:

$$b_1 = 1 + a_2 b_3, \tag{5.23}$$

$$a_5 = -\frac{1}{2}, \tag{5.24}$$

and

$$a_6 = \frac{1}{2}. \tag{5.25}$$

[1]Proof provided in the mathematical appendix.

The last constraint is imposed because it dictates that, in the absence of arbitrage, the projected forward price of the underlying asset must be aligned with its actual forward price — this is the no-arbitrage condition:[2]

$$E(z) = a_1 \frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)} (-b_2)^{-a_2} (m_1 - m_2) + m_2. \tag{5.26}$$

Throughout the entirety of the process undertaken thus far, we have been engaged with the arbitrary variable z, which traverses the entirety of the real number continuum $\mathbb{R}$. However, we are cognizant that in the context of asset valuation (whether pertaining to options or the underlying asset itself), the domain remains pertinent exclusively within the positive real number continuum $\mathbb{R}^+ \cup \{0\}$. Furthermore, in alignment with Abadir and Rockinger (2003), we recognize the criticality of acquiring estimates of commensurate magnitudes when dealing with non-linear estimations, with such parity playing a pivotal role in preserving numerical stability. Given our reliance on computational optimization techniques for parameter estimation, we undertake to enhance the process by subjecting the strike X to a linear rescaling procedure, engendered by the introduction of a linear transformation given by

$$z = \alpha + \beta S_T. \tag{5.27}$$

Effecting this transformation mandates a corresponding adaptation of the option prices through the utilization of the Jacobian associated with the transformation,

$$c(z) = \beta c_t(S_t, X, \tau). \tag{5.28}$$

To ensure a numerically stable estimation procedure, the variable z will lie within the interval $z \in [-3, 3]$. To determine the parameters α and β, we simply need to solve the following system of equations for each maturity:

$$\begin{cases} 3 = \alpha + \beta X_u \\ -3 = \alpha + \beta X_l \end{cases} \iff \begin{cases} \alpha = 3\left(\frac{X_l + X_u}{X_l - X_u}\right) \\ \beta = -\frac{6}{X_l - X_u} \end{cases} \tag{5.29}$$

being X_l (X_u) the minimum (maximum, respectivetly) strike available for the maturity T.

Considering equation (5.27), we can formulate the no-arbitrage condition as follows:

$$E(z) = \alpha + \beta E(S_T), \tag{5.30}$$

and, since

$$E(S_T) = e^{(r-q)\tau} S_t, \tag{5.31}$$

we obtain,

$$E(z) = \alpha + \beta e^{(r-q)\tau} S_t, \tag{5.32}$$

[2]Proof provided in the mathematical appendix.

By applying the aforementioned transformations, we derive the subsequent equations pertaining to the price of an European-style call option, RND, and constraints.

PROPOSITION 5.1. *Approximating the RND through the density functional form based on confluent hypergeometric functions, the time-t* forward price *of a European-style call option with underlying price S_t, strike X, and maturity $T \geq t$ is given by:*

$$
\begin{aligned}
\tilde{C}(X) = & -\frac{X_\ell - X_u}{6}\Bigg\{c_1 + c_2\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u}\right) \\
& + \mathbb{1}_{\frac{3(X_\ell+X_u)-6X}{X_\ell - X_u} > m_1}\, a_1\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_1\right)^{b_1} \\
& \times {}_1F_1\left(a_2; a_3; b_2\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_1\right)^{b_3}\right) \\
& + a_4\, {}_1F_1\left(a_5; a_6; b_4\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_2\right)^{2}\right)\Bigg\}.
\end{aligned} \tag{5.33}
$$

which, through the relationship of the forward price,

$$
c(S_t, X, T) = e^{-r\tau}\tilde{C}(X), \tag{5.34}
$$

leads us to the equation for the price of an European-style call option.

PROPOSITION 5.2. *Approximating the RND through the density functional form based on confluent hypergeometric functions, the time-t price of a European-style call option with underlying price S_t, strike X, and maturity $T \geq t$ is given by:*

$$
\begin{aligned}
c(S_t, X, \tau) = e^{-r\tau}\Bigg[& -\frac{X_\ell - X_u}{6}\Bigg\{c_1 + c_2\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u}\right) \\
& + \mathbb{1}_{\frac{3(X_\ell+X_u)-6X}{X_\ell - X_u} > m_1}\, a_1\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_1\right)^{b_1} \\
& \times {}_1F_1\left(a_2; a_3; b_2\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_1\right)^{b_3}\right) \\
& + a_4\, {}_1F_1\left(a_5; a_6; b_4\left(\frac{3(X_\ell + X_u) - 6X}{X_\ell - X_u} - m_2\right)^{2}\right)\Bigg\}\Bigg].
\end{aligned} \tag{5.35}
$$

And considering equation (5.15), we can derive the expression for the risk-neutral density function pertaining to the European-style call option as follows:

$$
\begin{aligned}
q(S_T) = -\frac{6}{X_l - X_u}\Bigg\{ & \mathbb{1}_{\frac{3(X_l+X_u)-6X}{X_l - X_u} > m_1}\, a_1\left(\frac{3(X_l + X_u) - 6X}{X_l - X_u} - m_1\right)^{b_1 - 2} F_1 \\
& + 2a_4\frac{a_5}{a_6}b_4F_2\Bigg\},
\end{aligned} \tag{5.36}
$$

where

$$F_1 = \Bigg[(b_1)(b_1-1)\,{}_1F_1\left(a_2; a_3; b_2\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_1\right)^{b_3}\right)$$
$$+\frac{a_2}{a_3}b_2b_3(2b_1+b_3-1)\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_1\right)^{b_3}$$
$$\times\,{}_1F_1\left(a_2+1; a_3+1; b_2\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_1\right)^{b_3}\right)$$
$$+\frac{a_2(a_2+1)}{a_3(a_3+1)}b_2^2b_3^2\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_1\right)^{2b_3}$$
$$\times\,{}_1F_1\left(a_2+2; a_3+2; b_2\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_1\right)^{b_3}\right)\Bigg], \tag{5.37}$$

and

$$F_2 = \Bigg[{}_1F_1\left(a_5+1; a_6+1; b_4\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_2\right)^{2}\right)$$
$$+2\frac{a_5+1}{a_6+1}b_4\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_2\right)^{2}$$
$$\times\,{}_1F_1\left(a_5+2; a_6+2; b_4\left(\frac{3(X_l+X_u)-6X}{X_l-X_u}-m_2\right)^{2}\right)\Bigg]. \tag{5.38}$$

Taking into account the following constraints, the number of parameters requiring estimation for computing the theoretical price, as defined in equation (5.9), is reduced to seven. The derivations for the restrictions on a_4 and a_1 are provided in Appendix B:

$$b_1 = 1 + a_2b_3, \tag{5.39}$$

$$a_5 = -\frac{1}{2}, \tag{5.40}$$

$$a_6 = \frac{1}{2}, \tag{5.41}$$

$$a_1 = \frac{\Gamma(a_3-a_2)(-b_2)^{a_2}\left[e^{(r-q)\tau}S_t - m_2\right]}{(m_1-m_2)\Gamma(a_3)}, \tag{5.42}$$

and

$$a_4 = \frac{1}{2\sqrt{-b_4\pi}}\left[1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}\right]. \tag{5.43}$$

These parameters can be determined by solving a minimization problem:

$$\min_{m_1,m_2,a_2,a_3,b_2,b_3,b_4}\sum_{i=1}^{n}\left[c(S_t,X_i,T)-\hat{c}_i\right]^2, \tag{5.44}$$

where $c(S_t, X_i, T)$ is given by equation (5.35) and $\hat{c}_i$ correspond to the observed market values for each option. As previously elucidated, the five constraints are also imposed as defined by equations (5.39) through (5.43). We employed Matlab's optimization solver, fmincon, to estimate the parameters $m_1, m_2, a_2, a_3, b_2, b_3, b_4$ that minimize equation (5.44). The results will be discussed in a dedicated chapter.

CHAPTER 6

Empirical Analysis

6.1. Data and Methodology

The empirical study uses daily prices of S&P 500 (SPX) index options from the Chicago Board Options Exchange (CBOE), covering the period January 3, 2022 to May 31, 2022. Only trading days are included.

The analysis focuses on a consistent maturity structure by using the *SPX7* dataset, which contains standard monthly index options expiring on the third Friday of each month. The raw data set is refined to include only observations that satisfy the following criteria:

- Retain out–of–the–money (OTM) contracts — calls with strikes above the forward price and puts with strikes below.
- Exclude contracts with fewer than seven calendar days to expiration.
- Remove quotes with zero bids, very wide bid–ask spreads, or missing implied volatilities.

To streamline our analysis and ensure consistency, we will utilize the put-call parity (4.32) that allows us to coherently transform European-style put prices into their equivalent European call option counterparts.

6.2. Estimation Procedure

Implied risk–neutral densities (RNDs) are estimated on a daily basis for the next three SPX option maturities. Three parametric families are considered:

(1) Mixture of two lognormals (MLN2);
(2) Mixture of three lognormals (MLN3);
(3) Hypergeometric functional form (MHF).

The MLN2 and MLN3 specifications follow the approach of Bahra (1997) and are calibrated by minimizing the sum of squared pricing errors subject to constraints that ensure valid densities and strictly positive mixture weights. For the MHF specification, following Abadir and Rockinger (2003) and Bu and Hadri (2007), three parameters are fixed ex ante,

$$a_2 = 4, \quad a_3 = 6, \quad b_3 = 1,$$

while two others are initialized over a coarse grid:

$$b_2 \in [-20, 0] \text{ (20 points)}, \qquad m_1 \in [0.02, 1] \text{ (50 points)}.$$

The remaining parameters b_4 and m_2 are initialized using a moment–matching step based on the normalized strike transformation of Abadir and Rockinger (2003). For each (b_2, m_1)

grid pair, a local constrained optimization (`fmincon`) minimizes the root mean integrated squared error (RMISE) between observed and model prices, subject to the admissibility conditions

$$b_2 \leq 0, \qquad b_4 \leq 0, \qquad \theta \in \mathbb{R}.$$

Finally, the optimal parameter set for each maturity is used to compute the implied RND analytically from the fitted option pricing function.

6.3. Estimated Parameters

The following tables and figures summarize the results obtained, including the estimated parameters for each model and their variation across maturities.

Maturity	τ	α_1	α_2	β_1	β_2	ω
18-Nov-2022	0.8733	8.2693	8.5510	0.2240	0.0776	0.3243
16-Dec-2022	0.9500	8.2396	8.5574	0.2124	0.0826	0.3097
20-Jan-2023	1.0459	8.2511	8.5607	0.2417	0.0852	0.3342
17-Mar-2023	1.1993	8.2633	8.5707	0.2640	0.0877	0.3760
16-Jun-2023	1.4486	8.2449	8.5864	0.2759	0.0994	0.3906
15-Dec-2023	1.9472	8.0281	8.5845	0.1541	0.1440	0.2367
20-Dec-2024	2.9637	7.9774	8.6082	0.1076	0.1881	0.2414
19-Dec-2025	3.9609	7.9748	8.6220	0.0548	0.2290	0.2422

TABLE 1. Example of MLN2 parameter estimates.

	α_1	α_2	β_1	β_2	ω
Min	6.7018	-9017.5119	-0.0949	-124.5536	0.1723
Max	9.8477	9.7234	0.4177	3.5958	1.0000
Mean	8.3806	-8190.1462	0.2000	-113.1277	0.9765
Std	0.1099	2602.7146	0.0855	35.9441	0.1012

TABLE 2. Summary statistics of MLN2 parameter estimates considering all maturities and days.

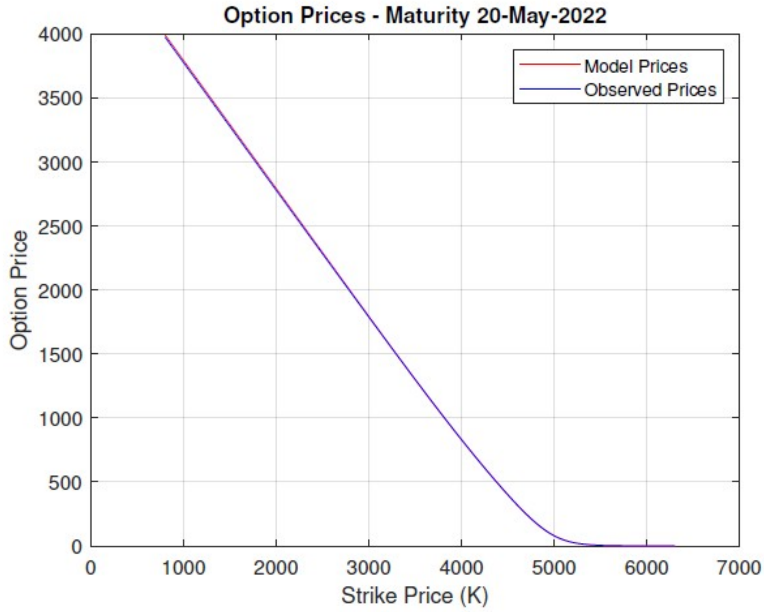

FIGURE 1. Option prices from the MLN2 model compared with observed market prices on 03-Jan-2022 (maturity 20-May-2022).

Maturity	τ	α_1	α_2	α_3	β_1	β_2	β_3	w_1	w_2	w_3
18-Nov-2022	0.8733	7.9998	8.5563	8.3877	0.3453	0.0687	0.1818	0.0851	0.5871	0.3207
16-Dec-2022	0.9500	8.3780	8.5611	7.9601	0.1927	0.0710	0.3731	0.3299	0.5826	0.0806
20-Jan-2023	1.0459	8.5660	7.9546	8.3801	0.0752	0.3819	0.1996	0.5756	0.0875	0.3297
17-Mar-2023	1.1993	7.9791	8.3938	8.5758	0.3866	0.2085	0.0798	0.1094	0.3361	0.5455
16-Jun-2023	1.4486	8.4553	8.2440	8.5781	0.3088	0.2946	0.1036	0.0000	0.3428	0.6351
15-Dec-2023	1.9472	8.3667	7.9288	8.6128	0.2542	0.5161	0.1072	0.3246	0.1439	0.5197
20-Dec-2024	2.9637	8.0893	8.6948	8.5504	0.4825	0.0864	0.1987	0.3683	0.2377	0.3944
19-Dec-2025	3.9609	8.2098	8.5538	8.7785	0.4897	0.1495	0.0841	0.4653	0.2877	0.2167

TABLE 3. Example of MLN3 parameter estimates.

	α_1	α_2	α_3	β_1	β_2	β_3	w_1	w_2	w_3
Min	2.3778	3.0280	6.4659	-0.3547	-5.7224	0.0000	0.0000	0.0000	0.0000
Max	9.5625	11.7596	8.9075	4.2505	3.9496	3.5334	0.9597	0.8821	0.9757
Mean	8.0901	8.3862	8.3134	0.3226	0.1443	0.1799	0.2672	0.4243	0.3021
Std	0.4684	0.2873	0.2307	0.4340	0.3110	0.1697	0.1637	0.1336	0.1755

TABLE 4. Summary statistics of MLN3 parameter estimates considering all maturities and days.

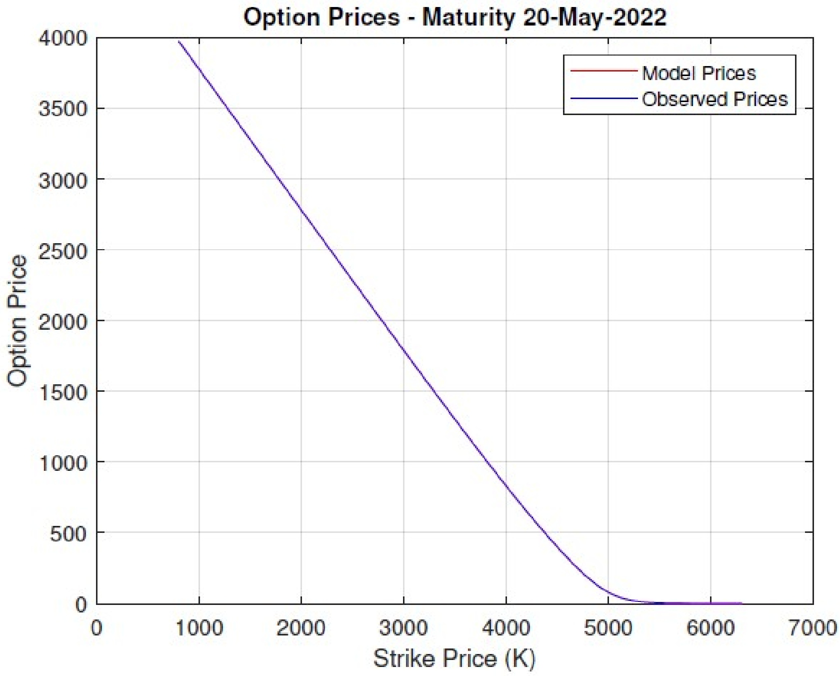

FIGURE 2. Option prices from the MLN3 model compared with observed market prices on 03-Jan-2022 (maturity 20-May-2022).

Maturity	τ	a_2	a_3	b_2	b_3	b_4	m_1	m_2
18-Nov-2022	0.8733	8.4581	171.6232	-53.4853	-355.5140	-1.0706	34.0753	0.7376
16-Dec-2022	0.9500	72.2571	171.6231	-44.4731	-303.8135	-0.9388	64.8236	0.9152
20-Jan-2023	1.0459	143.7857	-25.8771	-0.7489	-11.3164	-0.7945	65.1454	0.9529
17-Mar-2023	1.1993	80.8802	71.6074	-6024.2754	-145.5373	-0.9142	490146.256	0.4674
16-Jun-2023	1.4486	77.0817	34.5585	-191.4326	-3.1102	-0.5249	211155.7538	0.8478
15-Dec-2023	1.9472	2.8244	170.9099	-14.7444	-319.7637	-0.4083	40.4465	0.9401
20-Dec-2024	2.9637	60.0000	171.0067	-26.3039	-96.7484	-9.4197	24.0993	0.1691
19-Dec-2025	3.9609	31.6540	171.2526	-22.0600	-216.4418	-0.3127	43.5419	0.1238

TABLE 5. Example of MHF parameter estimates.

	a_2	a_3	b_2	b_3	b_4	m_1	m_2
Min	-130.25	-168.55	-6024.28	-874.26	-10.07	-281.87	-0.45
Max	197.84	171.62	-0.00	11.57	-0.02	490146.26	15.57
Mean	53.32	131.63	-94.95	-160.68	-1.41	2319.92	1.04
Std	53.54	79.18	531.64	152.54	1.23	29987.80	1.90

TABLE 6. Summary statistics of MHF parameter estimates considering all maturities and days.

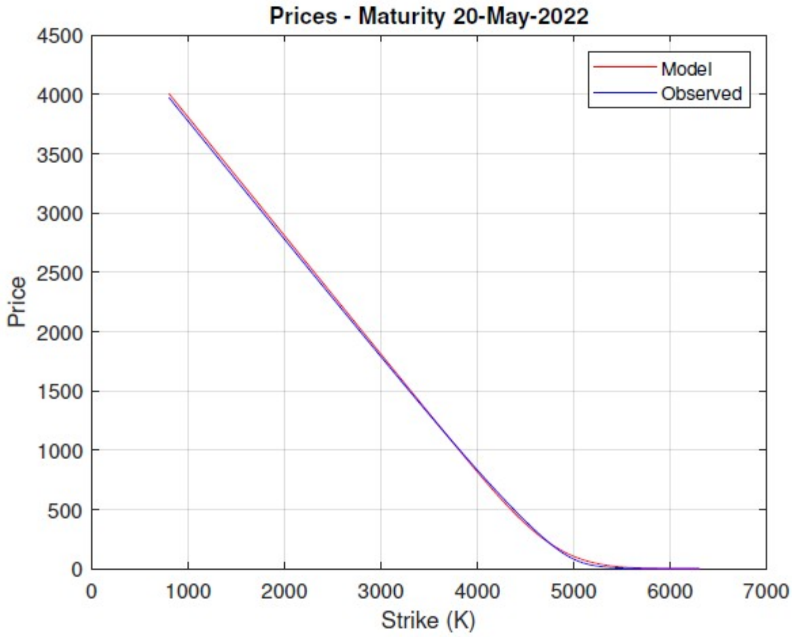

FIGURE 3. Option prices from the MHF model compared with observed market prices on 03-Jan-2022 (maturity 20-May-2022).

6.4. Comparative Analysis

6.4.1. Fit to Option Prices.

Model accuracy is evaluated using the mean absolute percentage error (MAPE) and the root mean integrated squared error (RMISE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{c_i - \hat{c}_i}{c_i}\right| \times 100, \quad \text{RMISE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{c}_i - c_i)^2},$$

where c_i and $\hat{c}_i$ are observed and model–implied prices, respectively.

Method	**MAPE (%)**	**RMISE**	R^2
MLN2	409.05	32.31	0.99827
MLN3	20.67	1.20	0.99997
MHF	55.79	23.12	0.99907

TABLE 7. Average measures of goodness of fit to option prices.

6.4.2. Higher–Order Moments.

Skewness and kurtosis are computed from the estimated RNDs:

$$\text{Skew} = \frac{\int (S_T - \mu)^3\, q(S_T)\, dS_T}{\sigma^3}, \qquad \text{Kurtosis} = \frac{\int (S_T - \mu)^4\, q(S_T)\, dS_T}{\sigma^4}.$$

Method	Skewness	Kurtosis
MLN2	0.36	2.95
MLN3	1.14	4.11
MHF	-0.03	3.04

Table 8. Average estimates of higher-order moments for the implied distributions.

Model performance is evaluated in terms of pricing accuracy and the distributional properties of the implied densities. Table 7 reports the average measures of goodness of fit, and Table 8 presents the corresponding higher-order moments.

With respect to option price fitting, the mixture of three lognormals produces the lowest mean absolute percentage error and root mean integrated squared error, with an R^2 close to one. This is consistent with the findings of Bahra (1997), who shows that increasing the number of components in mixture specifications enhances flexibility in matching the cross-section of option prices. The mixture of two lognormals generates higher errors, which indicates that two components are not sufficient to represent the variation observed in market data. The hypergeometric functional form results in error measures between those of the two lognormal mixtures, reflecting its design that emphasizes parametric control of shape and tails rather than flexibility in fitting option prices.

The results on higher-order moments indicate that the mixture of three lognormals produces positive skewness and kurtosis above three, which corresponds to distributions that are asymmetric and leptokurtic, in line with the empirical patterns reported in Figlewski (2018). The mixture of two lognormals yields lower skewness and kurtosis values close to the Gaussian benchmark, reflecting the more limited flexibility of this specification. The hypergeometric functional form produces skewness close to zero and kurtosis near three, which correspond to symmetric distributions with Gaussian-like tails.

Taken together, the results highlight differences in how the specifications represent option prices and the shape of the implied distribution. The mixture of three lognormals provides the best in-sample fit and reproduces the empirical regularities in skewness and kurtosis, the mixture of two lognormals offers a parsimonious but less accurate benchmark, and the hypergeometric functional form provides analytical tractability and tail control but does not match the cross-section of option prices as closely as the three-component mixture. This comparison illustrates the broader trade-off documented in the literature between flexibility in price replication and structural control over the distributional form (Bahra, 1997; Figlewski, 2018).

CHAPTER 7

Conclusions

This dissertation examined the estimation of risk–neutral densities (RNDs) from option prices using three models: a mixture of two lognormal distributions, a mixture of three lognormal distributions, and a mixture of hypergeometric functions. The study combined a theoretical review with an empirical application to the S&P 500 (SPX) options market. The theoretical part outlined the role of RND in the risk–neutral valuation framework, based on the link between option prices and the state price density, Figlewski (2018). It reviewed the mixture of two lognormal distributions model (Bahra, 1997), the extension to a mixture of three lognormal distributions, and the mixture of hypergeometric functions specification (Abadir and Rockinger, 2003; Bu and Hadri, 2007). The empirical analysis used daily SPX7 option data from January 3, 2022 to May 31, 2022. The data set was filtered to include only out–of–the–money options and, for each trading day and for the next three SPX option maturities, the parameters of each model were estimated by minimizing the sum of squared pricing errors under theoretical constraints. The performance of the model was evaluated using the mean absolute percentage error (MAPE), the mean integrated squared error (RMISE) and the coefficient of determination (R^2), as well as the skewness and kurtosis of the implied distributions. The results were:

- The mixture of three lognormal distributions produced the lowest MAPE and RMISE and the highest R^2 across the sample, and generated higher skewness and kurtosis, in line with Bu and Hadri (2007).
- The mixture of two lognormal distributions achieved a high R^2 but larger pricing errors.
- The mixture of hypergeometric functions produced intermediate values for the fit measures and moments.

The findings indicate that adding a third lognormal component improves the fit between observed and model prices and changes the shape of the implied distribution in a way consistent with earlier studies (Figlewski, 2018). The mixture of hypergeometric functions offers flexibility in functional form but did not improve the fit in this dataset.
The study is limited by the short sample period, the focus on in–sample results, and the restriction to three models. Further research could:

- Extend the analysis to different market conditions and longer periods.
- Assess out–of–sample performance.
- Examine parameter stability over time.
- Compare with non–parametric and semi–parametric approaches.

- Apply the methods to other markets or asset classes.

APPENDIX A

Proof of equation (4.33)

In this appendix, we will derive the formulas for the underlying condition of equation (4.33). Recalling equation (4.33), the underlying condition could be rewritten as:

$$\begin{aligned}\mathbb{E}[S_T \mid \mathcal{F}_t] &= \int_0^\infty S_T \sum_{i=1}^{k} \omega_i L(\alpha_i, \beta_i, S_T)\, dS_T \\ &= \sum_{i=1}^{k} \left\{ \int_0^\infty S_T\, \omega_i L(\alpha_i, \beta_i, S_T)\, dS_T \right\} \qquad (A.1) \\ &= \sum_{i=1}^{k} \int_0^\infty S_T \omega_i e^{-\frac{(lnS_T-\alpha_i)^2}{2\beta_i^2}} \frac{1}{\beta_i\sqrt{2\pi}S_T} dS_T \\ &= \sum_{i=1}^{k} \int_0^\infty \frac{\omega_i}{\beta_i\sqrt{2\pi}} e^{-\frac{(lnS_T-\alpha_i)^2}{2\beta_i^2}} dS_T. \qquad (A.2)\end{aligned}$$

By applying the same change of variables as in the transition from equation (4.11) to equation (4.15), equation (A.2) can be rewritten as follows:

$$\begin{aligned}\mathbb{E}[S_T \mid \mathcal{F}_t] &= \sum_{i=1}^{k} \int_{-\infty}^{\infty} \frac{\omega_i}{\beta_i\sqrt{2\pi}} e^{\left(-\frac{\left[\gamma-(\alpha_i+\beta_i^2)\right]^2}{2\beta_i^2}+\alpha_i+\frac{1}{2}\beta_i^2\right)} d\gamma \\ &= \sum_{i=1}^{k} \left\{ \omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2} \int_{-\infty}^{\infty} \frac{1}{\beta_i\sqrt{2\pi}} e^{\left(-\frac{\left[\gamma-(\alpha_i+\beta_i^2)\right]^2}{2\beta_i^2}\right)} d\gamma \right\}. \qquad (A.3)\end{aligned}$$

Since the integrand of equation (A.4) is the density function of a normal random variable, with variance β_i^2 and mean $(\alpha_i + \frac{1}{2}\beta_i^2)$, then:

$$\mathbb{E}[S_T \mid \mathcal{F}_t] = \sum_{i=1}^{k} \left\{ \omega_i e^{\alpha_i+\frac{1}{2}\beta_i^2} \right\}. \qquad (A.4)$$

Under the risk–neutral measure with a constant risk–free rate r and continuous dividend yield q, the forward price relation is given by,

$$\mathbb{E}[S_T \mid \mathcal{F}_t] = S_t e^{(r-q)\tau}, \qquad (A.5)$$

In the special case $k = 2$, let $w_1 = w$ and $w_2 = 1 - w$, so that the weights sum to one. The forward constraint then becomes

$$S_t e^{(r-q)\tau} = w\, e^{\alpha_1+\frac{1}{2}\beta_1^2} + (1-w)\, e^{\alpha_2+\frac{1}{2}\beta_2^2}. \qquad (A.6)$$

APPENDIX B

Constraints on the parameters c_2, a_4, and a_1

In this appendix, we will derive the constraints associated with the parameters c_2, a_4, and a_1. The derivation proceeds by recalling the equation (5.20) and assuming that $z_l < m_1$, the previous equation becomes:

$$\begin{aligned} & c_2 + 2a_4\frac{a_5}{a_6}b_4(z_l - m_2)\,{}_1F_1(a_5+1; a_6+1; b_4(z_l-m_2)^2) = -1 \\ \iff\ & c_2 = -1 - 2a_4\frac{a_5}{a_6}b_4(z_l - m_2)\,{}_1F_1(a_5+1; a_6+1; b_4(z_l-m_2)^2). \end{aligned} \tag{B.1}$$

Using the asymptotic representation of Kummer's function, as specified by equation (5.4), under the condition $b_4(z_l - m_2)^2 \to -\infty$, we subsequently arrive at the following expression:

$$c_2 = -1 - 2a_4\frac{a_5}{a_6}(-b_4)(m_2 - z_l)\frac{\Gamma(a_6+1)}{\Gamma(a_6-a_5)}(-b_4)^{-a_5-1}(m_2-z_l)^{-2a_5-2}. \tag{B.2}$$

By applying the constraints delineated in equations (5.39) to (5.41), one arrives at the subsequent expression for the parameter c_2:

$$c_2 = -1 + a_4\frac{\Gamma(\frac{3}{2})}{\Gamma(1)}(-b_4)^{\frac{1}{2}}(m_2 - z_l)^0 = -1 + a_4\sqrt{-b_4\pi}. \tag{B.3}$$

Now, recalling equation (5.13) and assuming that $z_u > m_1$, the previous equation becomes:

$$\begin{aligned} & c_2 + a_1(z_u - m_1)^{b_1-1}\Big[(b_1)\,{}_1F_1(a_2; a_3; b_2(z_u-m_1)^{b_3}) \\ & + \frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}\,{}_1F_1(a_2+1; a_3+1; b_2(z_u-m_1)^{b_3})\Big] \\ & + 2a_4\frac{a_5}{a_6}b_4(z_u-m_2)\,{}_1F_1(a_5+1; a_6+1; b_4(z_u-m_2)^2) = 0 \\ \iff\ & c_2 = -a_1(z_u - m_1)^{b_1-1}\Big[(b_1)\,{}_1F_1(a_2; a_3; b_2(z_u-m_1)^{b_3}) \\ & + \frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}\,{}_1F_1(a_2+1; a_3+1; b_2(z_u-m_1)^{b_3})\Big] \\ & - 2a_4\frac{a_5}{a_6}b_4(z_u-m_2)\,{}_1F_1(a_5+1; a_6+1; b_4(z_u-m_2)^2). \end{aligned} \tag{B.4}$$

By substituting the parameter c_2 in the preceding equation with that from equation (B.1), we can rewrite the equation in terms of the parameter a_4.

$$\begin{aligned}
&-1-2a_4\frac{a_5}{a_6}b_4(z_l-m_2)_1F_1(a_5+1;a_6+1;b_4(z_l-m_2)^2)\\
&=-a_1(z_u-m_1)^{b_1-1}\Bigg[(b_1)_1F_1(a_2;a_3;b_2(z_u-m_1)^{b_3})\\
&+\frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}{}_1F_1(a_2+1;a_3+1;b_2(z_u-m_1)^{b_3})\Bigg]\\
&-2a_4\frac{a_5}{a_6}b_4(z_u-m_2)_1F_1(a_5+1;a_6+1;b_4(z_u-m_2)^2)\\
\iff a_4=\frac{A}{B},
\end{aligned}\tag{B.5}$$

being

$$\begin{aligned}
A=1-a_1(z_u-m_1)^{b_1-1}&\Bigg[b_1\,{}_1F_1\left(a_2;a_3;b_2(z_u-m_1)^{b_3}\right)\\
&+\frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}{}_1F_1\left(a_2+1;a_3+1;b_2(z_u-m_1)^{b_3}\right)\Bigg]
\end{aligned}\tag{B.6}$$

and

$$\begin{aligned}
B=2\frac{a_5}{a_6}b_4&\Bigg[(z_u-m_2)\,{}_1F_1\left(a_5+1;a_6+1;b_4(z_u-m_2)^2\right)\\
&-(z_l-m_2)\,{}_1F_1\left(a_5+1;a_6+1;b_4(z_l-m_2)^2\right)\Bigg]
\end{aligned}\tag{B.7}$$

We are now positioned to treat equations (B.6) and (B.7) on an individual basis for the purpose of elucidating the constraint associated with the parameter a_4. The term A can subsequently be reduced to the following expression:

$$\begin{aligned}
A=1-a_1(z_u-m_1)^{b_1-1}&\left[(b_1)_1F_1(a_2;a_3;b_2(z_u-m_1)^{b_3})\right.\\
&\left.+\frac{a_2}{a_3}b_2b_3((z_u-m_1)^{b_3}))_1F_1(a_2+1;a_3+1;b_2(z_u-m_1)^{b_3}\right]\\
&=1-a_1\left(z_u-m_1\right)^{b_1-1}\left[b_1e^{b_2(z_u-m_1)^{b_3}}{}_1F_1\left(a_3-a_2;a_3;-b_2(zu-m_1)^{b_3}\right)\right]\\
&+\frac{a_2}{a_3}b_2b_3\left(z_u-m_1\right)^{b_3}e^{b_2(z_u-m_1)^{b_3}}{}_1F_1\left(a_3-a_2;a_3+1;-b_2(z_u-m_1)^{b_3}\right),
\end{aligned}\tag{B.8}$$

utilizing the asymptotic representation of Kummer's function, as specified by equation (5.4), we get:

$$
\begin{aligned}
&= 1 - a_1(z_u - m_1)^{b_1-1}\left[b_1 e^{b_2(z_u-m_1)^{b_3}}\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}\left(-b_2(z_u-m_1)^{b_3}\right)^{-a_2}e^{-b_2(z_u-m_1)^{b_3}}\right.\\
&\left.+\frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}e^{b_2(z_u-m_1)^{b_3}}\frac{\Gamma(a_3+1)}{\Gamma(a_3-a_2)}\left(-b_2(z_u-m_1)^{b_3}\right)^{-a_2-1}e^{-b_2(z_u-m_1)^{b_3}}\right]\\
&= 1 - a_1(z_u-m_1)^{b_1-1}b_1\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}\left(-b_2(z_u-m_1)^{b_3}\right)^{-a_2}\\
&- a_1(z_u-m_1)^{b_1-1}\frac{a_2}{a_3}b_2b_3(z_u-m_1)^{b_3}\frac{\Gamma(a_3+1)}{\Gamma(a_3-a_2)}\left(-b_2(z_u-m_1)^{b_3}\right)^{-a_2-1}\\
&= 1 - a_1\left(z_u-m_1\right)^{b_1-1}b_1\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}\left(-b_2\left(z_u-m_1\right)^{b_3}\right)^{-a_2} \qquad \text{(B.9)}\\
&- a_1\left(z_u-m_1\right)^{b_1-1}\frac{a_2}{a_3}b_2b_3\left(z_u-m_1\right)^{b_3}\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}\left(-b_2\left(z_u-m_1\right)^{b_3}\right)^{-a_2-1}\\
&= 1 - a_1\left(z_u-m_1\right)^{1+a_2b_3-a_2b_3-1}(1+a_2b_3)\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}(-b_2)^{-a_2}\\
&+ a_1\left(z_u-m_1\right)^{1+a_2b_3-a_2b_3-1}a_2b_3\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}(-b_2)^{-a_2}\\
&= 1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3-a_2)}
\end{aligned}
$$

And B can be expressed as follows:

$$
\begin{aligned}
B &= 2\frac{a_5}{a_6}b_4\left[(z_u-m_2)\,{}_1F_1(a_5+1;a_6+1;b_4(z_u-m_2)^2)\right.\\
&\quad\left.-(z_l-m_2)\,{}_1F_1(a_5+1;a_6+1;b_4(z_l-m_2)^2)\right]\\
&= -2\frac{a_5}{a_6}b_4\left[(z_u-m_2)e^{b_4(z_u-m_2)^2}\,{}_1F_1(a_6-a_5;a_6+1;-b_4(z_u-m_2)^2)\right.\\
&\quad\left.-(z_l-m_2)\,{}_1F_1(a_5+1;a_6+1;b_4(z_l-m_2)^2)\right]\\
&= -2\frac{a_5}{a_6}b_4\left[(z_u-m_2)\frac{\Gamma(a_6+1)}{\Gamma(a_6-a_5)}(-b_4(z_u-m_2)^2)^{-a_5-1}\right.\\
&\quad\left.-(z_l-m_2)\frac{\Gamma(a_6+1)}{\Gamma(a_6-a_5)}|b_4(z_l-m_2)^2|^{-a_5-1}\right]\\
&= -2\frac{a_5}{a_6}b_4\left[(z_u-m_2)\frac{\Gamma(a_6+1)}{\Gamma(a_6-a_5)}(-b_4(z_u-m_2)^2)^{-a_5-1}\right.\\
&\quad\left.-(z_l-m_2)\frac{\Gamma(a_6+1)}{a_6-a_5}(-b_4)^{-a_5-1}(m_2-z_l)^{-2a_5-2}\right]\\
&= 2b_4\left[(z_u-m_2)\frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma\left(1\right)}(-b_4(z_u-m_2)^2)^{-\frac{1}{2}}\right.\\
&\quad\left.-(z_l-m_2)\frac{\Gamma(\frac{3}{2})}{\Gamma(1)}(-b_4)^{-\frac{1}{2}}(m_2-z_l)^{-1}\right]\\
&= 2\sqrt{-b_4\pi} \qquad \text{(B.10)}
\end{aligned}
$$

The constrain a_4 is subsequently delineated as follows:

$$a_4 = \left[1 - a_1(-b_2)^{-a_2}\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}\right]\frac{1}{2\sqrt{-b_4\pi}} \tag{B.11}$$

The no-arbitrage condition can be described as follows, taking into account $a_5 = -\frac{1}{2}$ and $a_6 = \frac{1}{2}$:

$$\begin{aligned}
E(z) = &\left(z\frac{dc(z)}{dz} - c(z)\right)\Bigg|_{z_\ell}^{z_u} \\
&= a_1(z_u - m_1)^{b_1-1}\Bigg[(z_u b_1 + m_1 - z_u)\,{}_1F_1\big(a_2; a_3; b_2(z_u - m_1)^{b_3}\big) \\
&\quad + \frac{a_2}{a_3}\,z_u b_2 b_3 (z_u - m_1)^{b_3}\,{}_1F_1\big(a_2 + 1; a_3 + 1; b_2(z_u - m_1)^{b_3}\big)\Bigg] \\
&\quad - a_4\left[{}_1F_1\big(a_5; a_6; b_4(z - m_2)^2\big) - 2\frac{a_5}{a_6}\,z(z - m_2)\,{}_1F_1\big(a_5 + 1; a_6 + 1; b_4(z - m_2)^2\big)\right] \\
&= a_1\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}(b_2)^{-a_2}(z_u - m_1)^{b_1 - a_2 b_3 - 1}\big[(b_1 - a_2 b_3 - 1)z_u + m_1\big] \\
&\quad + 2a_4\frac{\Gamma(a_6)}{\Gamma(a_6 - a_5)}(-b_4)^{-a_5} m_2 \\
&= a_1\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}(b_2)^{-a_2} m_1 + 2a_4 m_2\sqrt{-b_4\pi} \\
&= a_1\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}(-b_2)^{-a_2}(m_1 - m_2) + m_2.
\end{aligned} \tag{B.12}$$

From equation A.4, equation B.12 becomes:

$$\begin{aligned}
S_t e^{(r-q)\tau} &= a_1\,\frac{\Gamma(a_3)}{\Gamma(a_3 - a_2)}(-b_2)^{-a_2}(m_1 - m_2) + m_2 \\
\Leftrightarrow \quad a_1 &= \frac{\Gamma(a_3 - a_2)(-b_2)^{a_2}\left[e^{(r-q)\tau}S_t - m_2\right]}{(m_1 - m_2)\Gamma(a_3)}.
\end{aligned} \tag{B.13}$$

References

Abadir, K. M., & Rockinger, M. (2003). Density functionals, with an option-pricing application. *Econometric Theory, 19*(5), 778–811.

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, New York.

Aït-Sahalia, Y., & Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance, 53*(2), 499–547.

Bahra, B. (1997). *Implied risk-neutral probability density functions from option prices: Theory and application*. Bank of England Working Paper No. 66.

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

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy, 81*(3), 637–654.

Bliss, R. R., & Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance, 26*(2–3), 381–422.

Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business, 51*(4), 621–651.

Buchen, P. W., & Kelly, M. (1996). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis, 31*(1), 143–159.

Bu, R., & Hadri, K. (2007). Option-implied densities: Properties and empirical performance. *Journal of Futures Markets, 27*(9), 859–886.

Figlewski, S. (2018). Risk neutral densities: A review. *Annual Review of Financial Economics, 10*(1), 201–230.

Harrison, J. M., & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory, 20*(3), 381–408.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies, 6*(2), 327–343.

Jackwerth, J. C. (2000). Recovering risk aversion from option prices and realized returns. *Review of Financial Studies, 13*(2), 433–451.

Jondeau, E., & Rockinger, M. (2000). Reading the smile: The message conveyed by methods which infer risk-neutral densities. *Journal of International Money and Finance, 19*(6), 885–915.

Jondeau, E. (2007). *Financial modeling under non-Gaussian distributions*. Springer.

Melick, W. R., & Thomas, C. P. (1997). Recovering an asset's implied PDF from option prices: An application to crude oil during the Gulf crisis. *Journal of Financial and*

Quantitative Analysis, 32(1), 91–115.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance, 49*(3), 771–818.

Shimko, D. (1993). Bounds of probability. *Risk, 6*(4), 33–37.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation. *Journal of Finance, 51*(5), 1633–1652.