Calibration risk under probabilistic parameter dependencies and model output effects

Gianluca Fusai[a,b,*], Marina Marena[c], Ioannis Kyriakou[a]

[a] *Bayes Business School, City St George's, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK*
[b] *Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone 18, 28100 Novara, Italy*
[c] *Dipartimento di Scienze Economico-Sociali e Matematico-Statistiche, Università di Torino, Corso Unione Sovietica, 218 bis, 10134 Torino, Italy*

Abstract

We propose a novel regression-based framework to model calibration risk in asset price models. Traditionally, models are calibrated to liquid contract quotes by minimizing error. Calibration risk relates to uncertainty in parameter estimates (inputs), which is transferred to other contracts (outputs), but is often overlooked. Our probabilistic modelling of parameter estimates and their dependencies allows systematic detection and alleviation of this and its effects on ultimate outputs. Studying the global sensitivity of contracts' values to model parameters enables us to rank them by influence of their knowledge on maximizing the increase in likelihood of profit or loss of an investor's position.

Keywords: Finance, asset price models, calibration risk, model outputs, global sensitivity analysis

1. Introduction

Significant losses suffered by financial institutions in recent decades due to mispricing of financial contracts have prompted regulators to explicitly address the model risk, which pertains to the use of generally *inadequate models*. Early academic research by Derman (1996), Green and Figlewski (1999), Hull and Suo (2002) and Cont (2006) broadly interpreted model inadequacy as encompassing model uncertainty/misspecification, incorrect implementation, or erroneous estima-

*Corresponding author.
Email addresses: gianluca.fusai.1@city.ac.uk & gianluca.fusai@uniupo.it (Gianluca Fusai), marina.marena@unito.it (Marina Marena), ioannis.kyriakou@city.ac.uk (Ioannis Kyriakou)

tion of the unobservable model parameters—what is known as the calibration risk. In the same vein, the Basel Committee on Banking Supervision (2013, 2019a) highlighted that, for complex products, financial institutions must explicitly assess the necessity of making valuation adjustments to account for the model risk arising from potentially inaccurate valuation methodologies but also from the use of the potentially incorrect calibration parameters in the valuation model. Besides, in their empirical study across various models, Schoutens et al. (2004) showed that, despite achieving high precision in calibrations to the same market quotes for liquid derivatives, extrapolating (marking-to-model), for example, values of exotic products can result in significant inaccuracies. Indeed, calibration risk poses a significant challenge with exotic instruments and structured deals, and accurately addressing this issue is certainly nontrivial, warranting further investigation.

Our paper seeks to advance this line of research. We introduce an innovative regression-based framework for modelling the calibration risk and studying aspects of it. More specifically, we present a methodology that conceptualizes calibration as a nonlinear least-squares regression problem and capitalizes on it. This endows us with sufficient tractability allowing us to apply normal asymptotic results to the estimator of the unknown parameter vector and, subsequently, to the model value of exotic contracts and the associated error from the originally calibrated driving model. This is crucial for financial institutions which, in accordance with the Basel Committee on Banking Supervision (2013, 2019a), must ensure that the valuation of their positions meets the standards of prudential soundness and regulatory capital requirements, by reflecting the risk associated with using unobservable (and possibly spurious) calibration parameters in their valuation model.

In building a reliable and interpretable model, it is important to be able to detect and address collinearity. Although an extensive literature in applied statistics provides valuable suggestions and guidelines for data analysts to diagnose the presence of collinearity among estimated parameters (e.g., Belsley et al., 2005, Fox, 1984, Theil, 1971, Farrar and Glauber, 1967), it so far seems that none have been picked up for use when calibrating nonlinear option pricing models, despite parameter identification being a well-known issue (e.g., see Coleman and Liu, 1999 and Guillaume

and Schoutens, 2012). Collinearity can cause statistical problems, such as inflated variance of the estimated parameters and poor predictions of the calibrated model, since small changes in the data can lead to large changes in the results and model misconception. As our first contribution, through variance decomposition analysis, we are able to segment the variance into components, allowing us to robustly infer parameters that are not identifiable.

This leads to our second key contribution. By leveraging the asymptotic normality of the model parameter vector estimator, we introduce a delta approximation for the distribution of the product value dependent on it. This is achieved through the combined tractability of standard optimization in a nonlinear regression model with expressions for various product values. The implementation is particularly fast and provides user-oriented insights, such as confidence intervals that incorporate information from the original sample used for calibration, including the point estimates of the parameters and their variability as captured in the estimated covariance matrix. This is key in the model validation process and can reveal significant errors, for example due to non-identifiable parameters, which may ultimately lead financial positions to fall outside a bank's predetermined thresholds of acceptability (Federal Reserve, 2011). In such cases, a recalibration is warranted, which we pursue by pre-specifying critical parameters based on available information, effectively lowering dimensionality and refining the model calibration.

Our third contribution is inspired by the seminal work of Baucells and Borgonovo (2013). We widen their scope by integrating an analytical method for sensitivity analysis into our framework, enabling us to determine the ranking of parameter influence. This is essential for banks that apply models to complex products and are interested, as part of model validation, to assess the impact of changes in inputs (such as parameter values) on model outputs to ensure they remain within an acceptable range (Federal Reserve, 2011). As we show later in our paper, a standard local approach, based on the so-called Greeks, is not sufficient as it fails to account for correlations among the unknown model input parameters. In fact, varying several inputs simultaneously can provide evidence of unexpected interactions, particularly if these are complex and not intuitively clear (Federal Reserve, 2011); our approach sheds light in this direction. In the spirit of Baucells and Borgonovo (2013), we determine the most important parameter as the one that, if known,

would result in the largest expected maximum increase in the probability of profit or loss of a position. This relates to a critical implication of model inadequacy for a financial institution, as pointed out in Green and Figlewski (1999), namely causing contracts to be sold for too low a price or purchased for too high a price (see also Hull and Suo, 2002).

With the core principles of our methodology in place, we then advance to an extensive empirical implementation by calibrating to EURO STOXX 50 implied volatility surfaces. It turns out that while complex asset price models provide a better fit to liquid market data, they introduce greater uncertainty when valuing illiquid derivative securities, highlighting the necessity of caution when simply relying on the in-sample fit. According to the Basel Committee on Banking Supervision (2019b), in the absence of a transparent price from a liquid market, valuation must rely on models and can be highly sensitive to inputs and subject to estimation error uncertainty. To this end, we decompose the variance of input model parameter estimates into components, which can be computed straightforwardly regardless of the underlying model. This allows us to assess the truthfulness of their values and tackle parameter identifiability challenges.

We examine a selection of widespread contracts and models with or without jumps and stochastic volatility, which can be easily expanded, as needed, owing to the generality of the theoretical building block of our framework and the availability of efficient numerical pricing techniques (see Section 4.2). We find that, even with good in-sample performance, calibration risk may still be substantial. Our proposed calibration reduction strategy effectively lowers condition numbers, indices, and variance proportions below critical thresholds, resolving near dependencies and reducing calibration risk without practically compromising the in-sample fit. Extensive testing suggests no significant evidence against the assumed normality of the model parameter[1] estimates, nor against the resulting normality of contract values[2], under stressed conditions of sample size (even smaller than typical) and large variance of non-normal residuals. In addition, we find that ignoring parameter estimates' dependencies, as in conventional local sensitivity analysis, can lead to misspecified probability of profit or loss and misconceptions regarding parameter importance

[1] These are transformed, where necessary, see Section 3.1.

[2] In the quite unlikely, based on evidence presented in the paper, opposite case, a solution based on nonparametric bootstrapping is possible, as briefly mentioned in Section 2.1, with further details available upon request.

in this context. This applies particularly to contracts with barrier clauses, which we observe to be mostly susceptible to calibration error uncertainty.

The outline of the paper is as follows. Starting with background information on the problem in Section 2, we then introduce our regression-based framework for detecting and modelling the calibration risk, discussing how this translates to model outputs and how it can be used to identify the important parameters. In Section 3, we present a general class of parametric models and proceed with calibration using real data, discussing the reliability of our model parameter estimates and how to mitigate the calibration risk. Section 4 delves into the second part of our empirical analysis focused on model outputs, particularly the valuation of contracts with exotic payoff structures. We examine the effect of calibration risk and the sensitivity of product values to parameters' estimates accounting for their probabilistic dependencies. Section 5 concludes the paper. A rich collection of accompanying materials is provided as an appendix.

2. Modelling calibration risk

In this section, we present our systematic framework for identifying and modelling calibration risk. We start with a brief overview of the pertinent literature.

Over the years, the significance of calibration error in option pricing has been prominently highlighted in the literature. Boyle and Ananthanarayanan (1977) and Lo (1988) studied the uncertainty in time series-based parameter estimation. The standard error of the estimated parameters could be determined by utilizing the asymptotic properties of the maximum likelihood (ML) estimator. Their focus was on the transition density of the log-return process under the real measure, rather than directly on the model option price under the risk-neutral measure. Once the ML estimators of the model parameters were obtained, they could be utilized to evaluate derivative contract prices.

Time series estimation has been extensively studied, with improvements such as Bayesian methods (Jacquier and Jarrow, 2000, Gupta and Reisinger, 2014), filtering techniques (Elliott et al., 1997), and finite sample bias-corrected ML estimation via indirect inference (Phillips and Yu, 2009). However, these approaches are sometimes constrained to plain vanilla options within

the Black–Scholes framework, leaving the impact of uncertainty on inferred model quantities, such as the value of products with exotic payout structures in non-Gaussian settings, quite under-explored. Also, the Markov Chain Monte Carlo method for high-dimensional parameters faces well-known computational challenges. Building on the more standard nonlinear least-squares regression, this paper takes a different view to quantifying the variability of the estimation error using a cross-section of options with varying strikes and maturities on given days.

2.1. Regression analysis of calibration

For a given parametric model with true parameter vector $\boldsymbol{\theta}^* \in \mathbb{R}^k$, let $h(\boldsymbol{X}, \boldsymbol{\theta}^*)$ be a market quantity such as the fair value of a financial instrument or the implied volatility. The vector $\boldsymbol{X}$ represents a collection of various market and contract-specific variables, such as the current asset price, the dividend yield, the maturity time, the strike price and others.

It is general practice (e.g., see Cui et al., 2017, Baschetti et al., 2024, Amici et al., 2025) that, once we pick our preferred model, we calibrate it to liquid market data of, usually, plain vanilla options via some nonlinear cross-sectional regression procedure. From an econometric perspective, the calibration of a model to market quantities is represented by a general nonlinear regression model of the form

$$y_i = h(\boldsymbol{X}_i, \boldsymbol{\theta}^*) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\{y_i\}$ are the market observed quantities, such as prices or implied volatilities, and h the corresponding model quantities. The zero-mean error terms $\{\varepsilon_i\}$ with variance σ^2 account for noise in the market data due to factors such as bid-ask spreads, non-synchronous trading, and rounding of market quotes. Capturing heteroskedasticity in errors is also possible (see Greene, 2012); we account for that using suitable weights in the model estimation (see Huang and Wu, 2004 and later Section 3.1 in this paper).

Given a sample size n, and for an error functional defined by the weighted norm $\|\cdot\|$, model calibration involves selecting the vector $\widehat{\boldsymbol{\theta}}_n$ as an estimate of $\boldsymbol{\theta}^*$ that minimizes $\|y - h(\boldsymbol{X}, \boldsymbol{\theta})\|$. Typical objective function is the L^2-distance, which corresponds to the nonlinear least-squares

regression (NLLS) model:

$$S(\boldsymbol{\theta}) = \|y - h(\boldsymbol{X}, \boldsymbol{\theta})\|_2 = \sum_{i=1}^{n} (y_i - h(\boldsymbol{X}_i, \boldsymbol{\theta}))^2 . \tag{1}$$

The NLLS estimator

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} S(\boldsymbol{\theta}) \tag{2}$$

allows us to establish several important statistical properties for the estimator. While the resulting estimator is not guaranteed to be the most efficient, except in the case of normal disturbances, consistency remains the primary requirement outside the linear case. Assuming that (i) the parameter space $\boldsymbol{\Theta}$ containing the true parameter vector is compact and $\boldsymbol{\theta}^*$ is its interior point; (ii) for any vector $\boldsymbol{\theta}$ in this parameter space, $\operatorname{plim}(1/n) S(\boldsymbol{\theta}) = q(\boldsymbol{\theta})$, which is a continuous and differentiable function[3]; and (iii) $q(\boldsymbol{\theta})$ has a unique minimum at $\boldsymbol{\theta}^*$, then the NLLS regression estimator $\widehat{\boldsymbol{\theta}}_n$ is consistent (Greene, 2012, Chapter 7).

Once this is established, we may rely on asymptotic normality as our basis for statistical inference. Let $\boldsymbol{H}$ be the $n \times k$ Jacobian matrix at the optima

$$H(i,j) = \frac{\partial h}{\partial \theta_j}\left(\boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}_n\right), \quad i = 1, \ldots, n, \quad j = 1, \ldots, k. \tag{3}$$

If $\operatorname{plim}(1/n) \boldsymbol{H}'\boldsymbol{H} = \boldsymbol{Q}$, where $\boldsymbol{Q}$ is a positive definite matrix, then

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* \overset{asy}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{4}$$

where $\boldsymbol{\Sigma} = (\sigma^2/n)\boldsymbol{Q}^{-1}$. A consistent estimator of the covariance matrix is

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\sigma}_n^2 \left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}, \tag{5}$$

[3]Writing $\operatorname{plim} x_n = c$ means that x_n converges in probability to c, i.e., $\lim_{n\to\infty} P(|x_n - c| \geq \varepsilon) = 0$ for any $\varepsilon > 0$.

where $\widehat{\sigma}_n^2$ is the variance of the residuals of the regression, i.e.,

$$\widehat{\sigma}_n^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{\varepsilon}_i^{\,2}, \quad \widehat{\varepsilon}_i = y_i - h\left(\boldsymbol{X}_i, \widehat{\boldsymbol{\theta}}_n\right).$$

The asymptotic normality arises because, under the assumed conditions, the NNLS estimator behaves like the ML estimator in large samples, and the quadratic approximation of the objective function around the true parameter ensures that the estimator's distribution converges to normality; asymptotic normality can also be guaranteed assuming heteroskedasticity in errors (see White and Domowitz, 1984). Reflecting on the above, the NLLS estimator is asymptotically unbiased and its variance depends on the inverse of the outer product of the Jacobian matrix at the optima, scaled by the sample variance of the estimated residuals. The calibrated parameters are affected by a sampling error related to the size of $(\boldsymbol{H}'\boldsymbol{H})^{-1}$. Establishing finite sample properties can be challenging, except when the errors are normally distributed, in which case (4) holds for finite samples[4]. Confidence intervals are straightforward to compute and are included in many nonlinear regression routines.

2.2. Diagnosing collinearity in calibration

Collinearity in a nonlinear regression model affects the reliability of parameter estimates. Small residuals, measured by $\widehat{\sigma}_n^2$, do not guarantee low overall estimation error due to potential inflation of the $(\boldsymbol{H}'\boldsymbol{H})^{-1}$ factor. Indeed, the calibration problem is typically ill-posed: the solution may not be unique or stable (e.g., see Cont and Tankov, 2006 and He et al., 2006). In this section, we discuss diagnosing collinearity in nonlinear least-squares regression estimation based on condition indices and variance decompositions. Variance decomposition analysis entails decomposing the total variance of an estimator into multiple distinct components and, to the best of our knowledge, has not previously been applied to the calibration of nonlinear financial models. The variance

[4]Our extensive empirical analysis in Appendix B shows significant evidence in support of normality also under relatively small finite samples. Notably, this parametric approach is not restrictive within our framework. As an alternative, we can use the bootstrap method (see Shao and Tu, 1995, Chapter 8.1) to determine the joint distribution of parameter estimators. This allows us to obtain the sampling distribution of a statistic by resampling, without relying on a parametric model assumption, and can also be insightful for flexible sample sizes. A tailored description can be provided upon request.

decomposition proportions identify groups of variates involved in near dependencies, and the extent to which the dependencies degrade the regression.

To explore the ill-posedness of calibration, we first compute the condition number of the Jacobian defined in (3), which represents our data matrix. The condition number of a matrix is the ratio of its maximum to minimum singular values, defined in the next paragraph; a higher condition number indicates that the matrix is closer to rank deficiency. If the columns of the Jacobian are highly collinear, the model parameters cannot be accurately identified, meaning different parameters' combinations can yield similar error function values at the optima. In such cases, even if the model fits the market quotes well, it cannot be reliably used to evaluate other model quantities.

We employ the diagnostic procedure by Belsley et al. (2005) based on singular-value decomposition to detect the presence of one or more near dependencies among the columns of the Jacobian matrix and to assess the degree to which each such dependency degrades the parameter estimates. To this end, we re-write (5) as $\widehat{\sigma}_n^2 \left(\boldsymbol{H}^{'}\boldsymbol{H}\right)^{-1} = \widehat{\sigma}_n^2 \left(\boldsymbol{V}\boldsymbol{D}^{-2}\boldsymbol{V}'\right)$, based on $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{'}$, such that $\boldsymbol{U}^{'}\boldsymbol{U} = \boldsymbol{V}^{'}\boldsymbol{V} = \boldsymbol{I}_k$, from which the variance of the j-th parameter estimate decomposes into the sum of components, each associated with one and only of the k singular values[5] λ_l, $l = 1, \ldots, k$, of $\boldsymbol{H}$ from the leading diagonal of $\boldsymbol{D}$:

$$\mathrm{V}ar\left(\widehat{\theta}_{n,j}\right) = \widehat{\sigma}_n^2 \sum_{l=1}^{k} \frac{v_{j,l}^2}{\lambda_l^2}, \quad j = 1, \ldots, k. \tag{6}$$

The variance decomposition proportions $\boldsymbol{\Pi} = \{\pi_{l,j}\}_{l,j}$ are given by

$$\pi_{l,j} = \frac{v_{j,l}^2/\lambda_l^2}{\sum_{q=1}^{k} v_{j,q}^2/\lambda_q^2}, \quad j = 1, \ldots, k, \quad l = 1, \ldots, k,$$

with each $\pi_{l,j}$ representing the proportion of variance of the j-th regression coefficient (in our context, model parameter) associated with the l-th component of its decomposition in (6). Finally,

[5]These are scaled to have unit column lengths using the Euclidean norm of each column to avoid artificial ill-conditioning caused by the scale of $\boldsymbol{H}$.

when diagnosing collinearity, it is standard to check high condition indices $\boldsymbol{\zeta} = \{\zeta_j\}_j$ given by

$$\zeta_j = \frac{\max_l \lambda_l}{\lambda_j}, \quad j = 1, \ldots, k,$$

associated with large proportions of variance of two or more coefficients.

If two or more elements in the j-th row of matrix $\mathbf{\Pi}$ are relatively large and the associated condition index ζ_j is also large, this signals that near dependencies are influencing the regression estimates. Variance estimates degrade if a near dependency dominates in explaining the variance of a given subset of estimated parameters. Degrading collinearity can be detected by a high condition index, typically ranging from 10 to 30, combined with high variance decomposition proportions, such as above 0.5, of two or more parameter estimates. The simultaneous occurrence of several large ζ_j points to the existence of more than one near dependency. This allows us to detect calibration risk caused by the inability to uniquely identify the unknown model parameters. We later practically demonstrate this issue and address it through a reduced calibration strategy in Section 3.2.

2.3. Probabilistic approach to calibration error, sensitivity analysis, and parameter importance

Practitioners typically focus solely on solving for $\widehat{\boldsymbol{\theta}}_n$ in (2), often overlooking its randomness and the consequences in subsequent applications. They compute a new quantity of interest $\widehat{U}_n = U\left(\widehat{\boldsymbol{\theta}}_n\right)$, such as the value of a financial product, considering that as true, ignoring that $\widehat{\boldsymbol{\theta}}_n$ is merely an estimate of $\boldsymbol{\theta}^*$. The error

$$U\left(\widehat{\boldsymbol{\theta}}_n\right) - U\left(\boldsymbol{\theta}^*\right) \tag{7}$$

is given little attention, if any, despite its significance (see Federal Reserve, 2011 and further discussion in the introduction of this paper). Indeed, even if the estimated parameter vector provides a very good fit to the data, resulting in a small variance of the residuals, the difference (7) depends on $(\boldsymbol{H}'\boldsymbol{H})^{-1}$ from (5), which, as we will show next, can be large, implying large variability of the plug-in estimate $U\left(\widehat{\boldsymbol{\theta}}_n\right)$.

The appeal of our method stems from the analytical tractability of the (asymptotic) normality of the parameters' estimator. This allows us to derive the error distribution based on information from the original sample via the point estimate $\widehat{\boldsymbol{\theta}}_n$ and the estimated covariance matrix $\widehat{\boldsymbol{\Sigma}}$. Suppose that $U : \mathbb{R}^k \to \mathbb{R}$ has a gradient $\boldsymbol{\Delta} = \boldsymbol{\nabla} U(\boldsymbol{\theta}^*)$ at $\boldsymbol{\theta}^* \in \mathbb{R}^k$ and (4) holds, then

$$U\left(\widehat{\boldsymbol{\theta}}_n\right) - U(\boldsymbol{\theta}^*) \overset{asy}{\sim} \mathcal{N}\left(0, \boldsymbol{\Delta}'\boldsymbol{\Sigma}\boldsymbol{\Delta}\right), \tag{8}$$

provided that $\boldsymbol{\Delta}$ is not equal to the zero vector (delta method; see also Lehmann, 1999, Chapter 2). Since $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^*$, we can be almost certain that $\widehat{\boldsymbol{\theta}}_n$ is very close to $\boldsymbol{\theta}^*$ when n is sufficiently large. In a small neighborhood, a differentiable function is nearly linear, hence normality is maintained. The plug-in estimate $\widehat{U}_n$ is asymptotically unbiased and has variance

$$\mathrm{V}ar(\widehat{U}_n) = \sigma^2 \boldsymbol{\Delta}' \left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta}, \tag{9}$$

which can be estimated by replacing $\boldsymbol{\Delta}$ with $\widehat{\boldsymbol{\Delta}}_n = \boldsymbol{\nabla} U\left(\widehat{\boldsymbol{\theta}}_n\right)$, σ^2 with $\widehat{\sigma}_n^2$, and $\boldsymbol{\Sigma}$ with $\widehat{\boldsymbol{\Sigma}}$. The confidence interval for $\widehat{U}_n$ can also be computed. To gauge the accuracy of the estimate, based on expression (9) we need to consider two elements: the variance of the residuals σ^2 and the factor $\boldsymbol{\Delta}'(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{\Delta}$, which we will focus on in Section 4.2.

In finance, it is common to use local sensitivity measures that focus on how a model output, such as a contract value, responds to changes around a reference value of a model input, such as a model parameter. However, as noted by Saltelli et al. (2008), a fatal limitation of this approach is its intrinsic local nature, which fails to account for correlations among the model inputs. For this, a global sensitivity method that incorporates all probabilistic input-output information is necessary (see Hazen and Huang, 2006). For a comprehensive overview of available local and global methods, readers may refer to Borgonovo and Plischke (2016). Baucells and Borgonovo (2013) examine the distance between cumulative distributions, also referred to as risk profiles, rather than densities (e.g., see Borgonovo, 2007), which enables to relate the sensitivity measure to the probabilities of exceeding or falling short of a threshold. The delta method fits with the sensitivity analysis of our model output and is the first (analytical) development in mapping model

inputs/parameters to model outputs/financial product values.

Let F be the distribution of the model output $\widehat{U}_n$. Additionally, F_l is the cumulative distribution function of $\widehat{U}_n$ knowing that the l-th component $\widehat{\theta}_{n,l}$ of the input parameter vector $\widehat{\boldsymbol{\theta}}_n$ is fixed at value θ_l, for $l = 1, \ldots, k$. Averaging the distance between F and F_l over all possible realizations of $\widehat{\theta}_{n,l}$ quantifies the effect of knowing the l-th component. This is linked to the probability of exceeding or failing to meet a target level. To conceptualize this within our context, let's consider the position of a market maker selling options. The fair value of the contract, denoted U^*, is unknown to the market maker. If they are able to sell the derivative for a price $\widehat{U}_n \geq U^*$, they will make a profit. The unconditional probability of this is $1 - F(U^*)$. If they knew that $\widehat{\theta}_{n,l} = \theta_l$, the conditional probability of making a profit would be $1 - F_l(U^*)$. The quantity $\Delta\pi_l^s = \sup_u (F(u) - F_l(u))$ provides the highest possible increase in the probability of making a profit (*success, s*) upon knowing the particular parameter. Similarly, the market maker's fear is tied to the risk of selling the derivative at loss, i.e., below the fair value. The corresponding highest possible increase in the probability of a loss (*failure, f*) upon conditioning on $\widehat{\theta}_{n,l}$ is $\Delta\pi_l^f = \sup_u (F_l(u) - F(u))$. The highest possible increase in the probability of profit or loss is given by Kuiper's (1960) distance

$$d^{\text{Ku}}(F, F_l) = \Delta\pi_l^s + \Delta\pi_l^f, \quad 0 \leq d^{\text{Ku}}(F, F_l) \leq 1. \tag{10}$$

The most important parameter is the one that yields the largest expected maximum increase in the probability of profit or loss

$$\beta_l^{\text{Ku}} = E\left[d^{\text{Ku}}(F, F_l)\right]. \tag{11}$$

By virtue of (4) and (8), the distribution of the vector with first k components $\widehat{\boldsymbol{\theta}}_n$ and $(k+1)$-th component $\widehat{U}_n$ is multivariate normal with $(k+1) \times (k+1)$ covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{\Delta} \\ \boldsymbol{\Delta}'\boldsymbol{\Sigma} & \boldsymbol{\Delta}'\boldsymbol{\Sigma}\boldsymbol{\Delta} \end{bmatrix}, \tag{12}$$

where the gradient vector $\boldsymbol{\Delta}$ collects the partial derivatives of the contract value with respect to

the model parameters. The conditional distribution F_l is also normal with mean and variance given, respectively, by

$$m_l = \widehat{U}_n + \frac{\Omega_{l,k+1}}{\Omega_{l,l}}\left(\theta_l - \widehat{\theta}_{n,l}\right), \quad s_l^2 = s^2 - \frac{\Omega_{l,k+1}^2}{\Omega_{l,l}}, \tag{13}$$

where $s^2 = \Omega_{k+1,k+1} = \boldsymbol{\Delta}'\boldsymbol{\Sigma}\boldsymbol{\Delta}$. Based on this, we derive next an explicit expression for β_l^{Ku}.

Theorem 1. *The expected maximum increase in the probability of profit or loss due to the conditioning parameter θ_l is given by*

$$\beta_l^{\mathrm{Ku}} = \int_{\mathbb{R}} d^{\mathrm{Ku}}\left(F, F_l\right) \frac{\phi\left(\frac{\theta_l - \widehat{\theta}_{n,l}}{\sqrt{\Omega_{l,l}}}\right)}{\sqrt{\Omega_{l,l}}} d\theta_l, \tag{14}$$

where $\phi(z) = \exp\left(-z^2/2\right)/\sqrt{2\pi}$ and

$$d^{\mathrm{Ku}}\left(F, F_l\right) = F(u_1) - F_l(u_1) + F_l(u_2) - F(u_2), \tag{15}$$

with

$$u_{1,2} = \frac{m_l s^2 - \widehat{U}_n s_l^2 \pm \sqrt{D}}{s^2 - s_l^2} \tag{16}$$

and

$$D = \left(m_l s^2 - \widehat{U}_n s_l^2\right)^2 - \left(s_l^2 - s^2\right)\left(\widehat{U}_n s_l^2 - m_l s^2 - 2 s^2 s_l^2 \ln \frac{s_l}{s}\right).$$

Proof. By normality, the unconditional distribution F and the conditional distribution F_l have continuous and differentiable densities. Therefore, $\Delta\pi^s$ and $\Delta\pi^f$ are attained at the points that satisfy the equation

$$\frac{\partial\left(F(u) - F_l(u)\right)}{\partial u} = \frac{\phi\left(\frac{u - \widehat{U}_n}{s}\right)}{s} - \frac{\phi\left(\frac{u - m_l}{s_l}\right)}{s_l} = 0$$

with solutions u_1 and u_2 given in (16). We then get that the second partial derivatives at the

stationary points are

$$\left.\frac{\partial^2\left(F(u)-F_l(u)\right)}{\partial u^2}\right|_{u=u_{1,2}} = \mp 2\frac{\phi\left(\frac{u_{1,2}-\widehat{U}_n}{s}\right)}{s}\frac{\left(u_{1,2}-\widehat{U}_n\right)^2}{2s^2} \pm 2\frac{\phi\left(\frac{u_{1,2}-m_l}{s_l}\right)}{s_l}\frac{\left(u_{1,2}-m_l\right)^2}{2s_l^2},$$

and

$$d^{\mathrm{Ku}}\left(F,F_l\right) = \int_{-\infty}^{u_1}\left(\frac{\phi\left(\frac{u-\widehat{U}_n}{s}\right)}{s}-\frac{\phi\left(\frac{u-m_l}{s_l}\right)}{s_l}\right)du + \int_{-\infty}^{u_2}\left(\frac{\phi\left(\frac{u-m_l}{s_l}\right)}{s_l}-\frac{\phi\left(\frac{u-\widehat{U}_n}{s}\right)}{s}\right)du$$

from which (15) follows. Finally, (14) follows by taking the expectation with respect to the distribution of the estimator $\widehat{\theta}_{n,l}$ which is normal. ■

3. Empirical analysis—Part I: Model inputs, calibration ill-posedness and risk

In what follows, we will focus on examples from prevalent asset price models for illustrative purposes. These range from the basic Black–Scholes model to the Merton jump diffusion (members of the wider Lévy class), the Heston model (a stochastic volatility model), and their combined version, the Bates model. By examining these, we can gain insights into the broader applicability of our approach and demonstrate its robustness, versatility, and effectiveness in diverse market settings. It is important to note that this choice is without loss of generality, as the construction of our method does not rely on specific underlying model assumptions. We are operating under the risk-neutral measure, and, for calibration purposes, we adopt a general option pricing formula for plain vanilla options (see Appendix A).

We begin with a brief description of the general model and its nested variants under consideration. Our selection of stochastic environments is based on the empirical findings of Bakshi et al. (1997), which indicate that jumps and stochastic volatility are key factors in pricing short-term options. Consequently, we adopt the Bates model that incorporates asset price jumps and

stochastic variance. The asset return dynamics is given by

$$\begin{aligned} \frac{dS_t}{S_t} &= (r-q)\,dt + \sqrt{v_t}\left(\rho dW_t^{(1)} + \sqrt{1-\rho^2}dW_t^{(2)}\right) + J_t dN_t, \\ dv_t &= \kappa\,(\eta - v_t)\,dt + \vartheta\sqrt{v_t}dW_t^{(1)}, \end{aligned}$$

where $r > 0$ is the continuously compounded risk-free interest rate, $q > 0$ the continuous dividend yield paid by the asset, $W^{(1)}$ and $W^{(2)}$ are independent standard Brownian motions, and v_t the variance process. We also have that $\rho \in (-1, 1)$ is the instantaneous correlation between the asset return and variance processes, $\eta > 0$ the long-run mean variance, $\kappa > 0$ the speed of mean reversion, and $\vartheta > 0$ the volatility of the variance process. The jumps arrive according to an independent time-homogeneous Poisson process N with intensity $\lambda_J > 0$. The percentage jump sizes are independent with an assumed identical lognormal distribution with parameters $\mu_J \in \mathbb{R}$ and $\sigma_J > 0$, such that $\ln(1+J) \sim \mathcal{N}(\ln(1+\mu_J) - \sigma_J^2/2, \sigma_J^2)$ (for instance, see Bakshi et al., 1997). If the variance is assumed to be constant, the model simplifies to the Merton jump diffusion, and, when referring to it, we denote the (time-constant) volatility as σ. Further simplifying by setting the jump intensity $\lambda_J = 0$ yields the basic Black–Scholes model with volatility σ. Alternatively, when $\lambda_J = 0$ but $\vartheta \neq 0$, we arrive at the Heston model, which is arguably the most widely used stochastic volatility model.

3.1. Data and in-sample results

We choose a dataset which comprises EUREX settlement prices of European options on the EURO STOXX 50 index. Specifically, we use Bloomberg quotes on the third Wednesday of each month over a period of a year from January to December 2022, that is, a total of 12 trading days. To avoid poor liquidity issues, following Detlefsen and Härdle (2007) we consider out-of-the-money call and put option prices within the moneyness-maturity region $\{(M,T)\colon 5/250 \leq T \leq 1.1,\ 0.8 - 0.1T \leq M \leq 1.2 + 0.1T\}$, where T is the option maturity time and $M = K/S_0$ determines the moneyness for strike price K and initial asset value S_0. This selection yields the following sample sizes (number of contracts) for 12 different calibration dates: $n \in \{95,\ 151,\ 151,\ 126,\ 164,\ 172,\ 138,\ 168,\ 150,\ 118,\ 216,\ 188\}$. The daily time series for the

underlying asset is given by the closing quotes of the EURO STOXX 50 index. We approximate the risk-free interest rates by interpolating EURIBOR on the maturities of available options. The risk-free rates are derived from Bloomberg term structures of different trading days. Dividends are extracted from the forward prices computed for each contract month as outlined in the CBOE VIX documentation[6].

We estimate the unknown model parameters using the quadratic error function (1). We consider absolute or percentage errors for prices or implied volatilities, which fits within the NLLS framework and general market practice. Specifically,

$$\mathrm{MSE}_h = \sum_{i=1}^{n} w_i \left(h_i\left(\boldsymbol{\theta}\right) - y_i\right)^2 \text{ and } \mathrm{PMSE}_h = \sum_{i=1}^{n} w_i \left(\frac{h_i\left(\boldsymbol{\theta}\right)}{y_i} - 1\right)^2, \tag{17}$$

where MSE (PMSE) refers to the (percentage) mean squared error, $h \in \{p, IV\}$, p is the price, and $IV = \mathrm{BS}^{-1}(p)$ the model implied volatility from the inverted Black–Scholes formula for the volatility parameter. Optimal normalized weights $\{w_i\}$ are introduced to balance the importance of options across different moneyness and maturity levels. These weights are estimated using nonparametric kernel regression, as in Huang and Wu (2004), for each moneyness and maturity level, allowing for the possible presence of heteroskedasticity in the error terms.

The normal distribution with the assigned covariance matrix estimate (5) holds for the unconstrained NLLS problem; however, some parameters are constrained. For example, the volatility σ and the jump arrival intensity λ_J must be positive and the correlation coefficient in the range $(-1, 1)$. Normality of the constrained estimators cannot generally be ensured as in the unconstrained case[7] (e.g., see Wang, 1996 and Andrews, 1999). To detour the constraints, we use suitable transformations of the relevant parameters and perform the calibration using dummy parameters. For example, positivity of σ is enforced by estimating $\ln \sigma$ in lieu of σ. Similarly, $\operatorname{artanh} \rho$ can be used to satisfy the requirement of $|\rho| < 1$. By virtue of the delta method, the normal distribution is valid for both the transformed and the original parameter.

[6] https://cdn.cboe.com/api/global/us_indices/governance/VIX_Methodology.pdf.

[7] If the true parameter vector is in the interior of the constrained parameter space and if the constrained and unconstrained parameter spaces have the same dimension, then the constrained NLLS estimator has the same asympotic distribution as the unconstrained (see Judge et al., 1985, p. 207).

We calibrate each model based on the different error functionals (17) using a nonlinear least-squares solver for 30 different sets of initial values keeping the best result. Beyond common practice, Table 1 provides evidence supporting the MSE_{IV} metric, which generally yields low condition numbers of the Jacobian (thus, it is less prone to collinearity, see Section 2.2) as well as the lowest minimum, maximum and average number across dates for all models[8]. Based on our results, the models are ranked with Bates being the best, followed by Heston, Merton and the Black–Scholes model. This is expected, as Bates has the largest number of parameters, leading to the best fit, in a least-squares sense, in reproducing the implied volatility surface. Heston and Merton are both nested variants of it. However, as discussed in Gilli and Schumann (2012), small mean squared errors do not necessarily imply accurate parameter estimation, hence this does not constitute reassuring news. We discuss this further in the next section.

	Condition numbers Heston				Condition numbers Merton				Condition numbers Bates			
Month	MSE_p	PMSE_p	MSE_{IV}	PMSE_{IV}	MSE_p	PMSE_p	MSE_{IV}	PMSE_{IV}	MSE_p	PMSE_p	MSE_{IV}	PMSE_{IV}
1	42	15	14	19	74	87	139	52	115	119	286	134
2	20	5	5	7	195	96	134	145	677	167	202	101
3	19	7	7	10	205	85	120	124	289	142	153	194
4	13	6	5	6	242	189	112	196	224	351	286	205
5	27	8	8	11	167	218	151	121	120	1348	73	377
6	16	6	6	9	199	154	89	161	118	2022	72	278
7	22	7	7	9	179	116	121	141	86	101	165	157
8	18	8	7	11	169	84	116	110	144	58	170	151
9	23	7	7	9	162	112	121	114	241	151	116	165
10	25	14	12	16	138	80	91	83	105	73	350	247
11	23	9	9	11	65	106	41	62	90	46	45	52
12	26	7	7	8	154	108	120	104	181	50	96	43
min	13	5	5	6	65	80	41	52	86	46	45	43
mean	23	8	8	11	162	120	113	118	199	386	168	175
max	42	15	14	19	242	218	151	196	677	2022	350	377

Table 1: *Condition numbers.* The table reports the condition numbers of the Jacobian at the optima of each objective function MSE_p, PMSE_p, MSE_{IV} and PMSE_{IV} for each model and each of 12 calibration dates. Minimum, maximum and average metrics are also reported at the bottom of the table.

3.2. Parameter variance decomposition

As discussed in Section 2.2, collinearity leads to unstable regression estimates with typically high standard errors. We note that the estimated variance of the residuals (i.e., the $\widehat{\sigma}_n^2$ estimate) remains unaffected, consequently the variance estimates given by (6) can still be small if $\widehat{\sigma}_n^2$ is

[8]Additional results in Appendix D include the averages of each error metric across all dates on the upper panel of Table D.1, while the lower panel showcases the case of MSE_{IV} in fitting the volatility surface on the individual dates. Table D.2 reports the estimated parameters for each of the 12 calibration dates for the MSE_{IV} metric.

sufficiently small. The following variance decomposition analysis allows us to disentangle the effect of $\widehat{\sigma}_n^2$ from the ill-conditioned Jacobian matrix at the optima, identify the most affected parameters, and implement a reduced calibration strategy by fixing critical parameters *a priori*.

In Tables 2–4, we report for each model the condition indices of the Jacobian, scaled to have unit column length (see Section 2.2), and the variance decomposition proportions on the last calibration date, which is representative of the prevailing pattern for each model. The Black–Scholes model with single parameter σ has no identification issue, but naturally struggles to provide a reasonable fit, for all objective functions, leading to the largest in-sample error.

Heston model calibration						
Condition indices		Proportions of variance				
		$Var(\widehat{v_0})$	$Var(\widehat{\kappa})$	$Var(\widehat{\eta})$	$Var(\widehat{\vartheta})$	$Var(\widehat{\rho})$
ζ_1	1	0.0052	0.0138	0.0202	0.0134	0.0317
ζ_2	1	0.0791	0.0002	0.0413	0.0134	0.0440
ζ_3	3	0.2189	0.0086	0.0936	0.0574	0.6646
ζ_4	3	0.2998	0.0288	0.5960	0.0934	0.2530
ζ_5	7	0.3970	0.9485	0.2489	0.8223	0.0067

Table 2: *Heston model parameter variance decomposition.* The table reports the proportions of variance, $\mathbf{\Pi} = \{\pi_{l,j}\}_{l=1,j=1}^{5,5}$, and condition indices, $\boldsymbol{\zeta} = \{\zeta_j\}_{j=1}^{5}$, for the Heston model parameter estimators $\widehat{v_0}$, $\widehat{\kappa}$, $\widehat{\eta}$, $\widehat{\vartheta}$ and $\widehat{\rho}$. The calibration shown is based on minimization of MSE_{IV} on the last (twelfth) date. A combination of condition index in excess of value 30 and variance proportion(s) above the threshold 0.5 signals near dependencies.

.

Full Merton model calibration					
Condition indices		Proportions of variance			
		$Var(\widehat{\sigma})$	$Var(\widehat{\lambda_J})$	$Var(\widehat{\mu_J})$	$Var(\widehat{\sigma_J})$
ζ_1	1	0.0009	0.0001	0.0000	0.0003
ζ_2	2	0.0233	0.0000	0.0000	0.0010
ζ_3	8	0.0249	0.0040	0.0005	0.0287
ζ_4	120	0.9509	0.9960	0.9994	0.9699
Reduced Merton model calibration					
Condition indices		Proportions of variance			
		$Var(\widehat{\sigma})$	$Var(\widehat{\lambda_J})$	$Var(\widehat{\mu_J})$	$Var(\widehat{\sigma_J})$
ζ_1	1	0.0635	–	0.0089	0.0094
ζ_2	2	0.8788	–	0.0099	0.0169
ζ_3	9	0.0577	–	0.9812	0.9738

Table 3: *Merton model parameter variance decomposition.* The table reports the proportions of variance and condition indices for the Merton model parameter estimators $\widehat{\sigma}$, $\widehat{\lambda_J}$, $\widehat{\mu_J}$ and $\widehat{\sigma_J}$. The calibration shown is based on minimization of MSE_{IV} on the last (twelfth) date. The top panel focuses on the fully calibrated model. We note that a combined condition index exceeding 30 with variance proportion(s) above 0.5 signal near dependencies. For this, the bottom panel showcases the reduced model calibration with parameter λ_J preset at 0.983.

For the Heston model, we always have condition indices below the threshold of 30. Nonetheless, even within this model, on all dates we observe some weak near dependency between the speed of mean reversion κ and the volatility ϑ of the variance process, since the 5-th row of matrix $\mathbf{\Pi}$, corresponding to the highest condition index, reports high variance proportions for both parameters' estimates. This is not surprising, as both parameters affect the variance of the variance

Full Bates model calibration									
Condition indices		Proportions of variance							
		$Var(\widehat{v_0})$	$Var(\widehat{\kappa})$	$Var(\widehat{\eta})$	$Var(\widehat{\vartheta})$	$Var(\widehat{\rho})$	$Var(\widehat{\lambda_J})$	$Var(\widehat{\mu_J})$	$Var(\widehat{\sigma_J})$
ζ_1	1	0.0022	0.0004	0.0006	0.0000	0.0000	0.0001	0.0000	0.0002
ζ_2	1	0.0005	0.0006	0.0005	0.0094	0.0032	0.0000	0.0000	0.0001
ζ_3	3	0.0085	0.0014	0.0038	0.0065	0.0239	0.0001	0.0001	0.0014
ζ_4	3	0.1184	0.0033	0.0067	0.0014	0.0028	0.0000	0.0000	0.0002
ζ_5	5	0.0004	0.0001	0.0086	0.2847	0.0554	0.0001	0.0002	0.0001
ζ_6	11	0.0287	0.0404	0.0342	0.0589	0.0278	0.0120	0.0008	0.0770
ζ_7	26	0.0077	0.9436	0.8002	0.2657	0.0253	0.0204	0.0005	0.0775
ζ_8	96	0.8335	0.0101	0.1454	0.3734	0.8616	0.9673	0.9983	0.8435
Reduced Bates model calibration									
Condition indices		Proportions of variance							
		$Var(\widehat{v_0})$	$Var(\widehat{\kappa})$	$Var(\widehat{\eta})$	$Var(\widehat{\vartheta})$	$Var(\widehat{\rho})$	$Var(\widehat{\lambda_J})$	$Var(\widehat{\mu_J})$	$Var(\widehat{\sigma_J})$
ζ_1	1	0.0322	–	0.0168	0.0004	0.0001	–	0.0025	0.0026
ζ_2	1	0.0096	–	0.0124	0.0196	0.0081	–	0.0010	0.0000
ζ_3	2	0.1947	–	0.0749	0.0025	0.0264	–	0.0066	0.0100
ζ_4	3	0.6323	–	0.3960	0.0050	0.0071	–	0.0005	0.0000
ζ_5	4	0.0065	–	0.0943	0.4114	0.0631	–	0.0319	0.0040
ζ_6	17	0.1246	–	0.4057	0.5612	0.8952	–	0.9575	0.9834

Table 4: *Bates model parameter variance decomposition.* The table reports the proportions of variance and condition indices for the Bates model parameter estimators $\widehat{v_0}$, $\widehat{\kappa}$, $\widehat{\eta}$, $\widehat{\vartheta}$, $\widehat{\rho}$, $\widehat{\lambda_J}$, $\widehat{\mu_J}$ and $\widehat{\mu_J}$. The calibration shown is based on minimization of MSE_{IV} on the last (twelfth) date. The top panel focuses on the fully calibrated model. We note that a combined condition index exceeding 30 with variance proportion(s) above 0.5 signal near dependencies. For this, the bottom panel showcases the reduced model calibration with parameters κ and λ_J preset at 1.879 and 0.182, respectively.

process. That said, this is not sufficient evidence to raise concerns about degradation of parameter estimates, since the condition indices are safely below the threshold on all dates.

The Merton model exhibits a large condition index, well above the threshold of 30, associated with variance proportions above 0.90 for all parameter estimates across all calibration dates (we report only the last one; refer to the upper panel of Table 3). This signals a strong near dependency among all parameters. More specifically, on 11 out of 12 dates, we observe that the jump arrival intensity λ_J and the mean jump size μ_J have the largest variance proportions, exceeding 0.99. Bates, having a richer, thus more flexible, structure than Heston and Merton due to its combination of stochastic volatility and jumps, demonstrates a better in-sample fit. However, this comes at the cost of overfitting, leading to limited predictive power, and of non-uniqueness of solution with consequent complex identifiability challenges. On most dates, we detect two high condition indices of different magnitudes. The dominant dependency, with the highest condition index, involves parameters v_0, ρ, λ_J, μ_J and σ_J (see the 8-th row in the upper panel of Table 4); the jump process parameters λ_J, μ_J and σ_J have the largest variance proportions. A second dependency with a lower condition index is detected involving parameters κ and η (see the 7-th row in the upper panel of Table 4). Overall, we observe the following pattern: a strong dependency associated with the jump parameters (inherited from the Merton model component) and a weaker dependency associated

with the stochastic volatility parameters κ (inherited from the Heston model component) and η.

Our results based on market data align with the findings of Gilli and Schumann (2012) based on artificial data, casting serious doubts on standard calibration of these models and the reliability of the parameter estimates, especially when evaluating other contracts. Even if a selected model appears to perform well in terms of in-sample error, this should not be a cause of optimism. On the contrary, the calibration risk could still be significant and should not be ignored, as implied by the variance decomposition proportions. Since adding more parameters is only beneficial up to a certain extent, there is a need for an efficient parameter reduction method. Previous attempts have involved implementing parameter subset selection (see Guillaume and Schoutens, 2010) by fixing beforehand some parameters, which, driven by intuition, may have a similar effect on the model's behaviour (e.g., in producing implied volatility profiles). However, this is not straightforward in the face of a large parameter set due to the complex interplay of the parameters in producing all kinds of stylized empirical facts, i.e., it is impossible to find unique, meaningful parameter values within their ranges, and this can lead to a deceptively good model calibration.

In this paper, we propose a structured approach to identifying critical parameters, which create collinearity, based on a parameter estimates' variance decomposition, and subsequently adopt a reduction strategy guided by this. More specifically, we fix certain parameters and recalibrate the models to obtain updated estimates for the remaining free parameters. Here, we used fixing at the average across all calibration dates[9]. In the Merton model, which exhibits a strong near dependency, we set one parameter, the jump arrival intensity λ_J, which is critically involved in the dependency, to 0.983. For the Bates model, we predetermine two parameters: λ_J, of strong dependency associated with jumps, and the speed of mean reversion κ, of weaker dependency associated with stochastic volatility, at 0.182 and 1.879, respectively. We then repeat the calibration exercise on each date to re-estimate the remaining free parameters (see Table D.3 in Appendix D) and obtain a revised variance decomposition under the reduced parameter set, as shown in the lower panels of Tables 3 and 4 for the last date. The outcome is a drastic

[9]Fixing at the average value should not be seen as a limitation. To ensure this did not result in imprecise remaining parameter estimates, we repeated with different fixings (median, 5th percentile, 95th percentile). Ultimately, the average turned out to give us the lowest variance of the residuals.

reduction in the condition numbers (compare Tables 1 and 5), as well as in the condition indices and variance proportions, bringing them below their thresholds on all dates. For the Merton model, the condition index drops from 120 to 9 (last date), with the (unreported) in-sample fit remaining practically unchanged. Similarly, for the Bates model, the condition index is reduced to at most 17, rendering the issue of near dependency no longer critical. Our approach therefore effectively tempers calibration risk.

Condition numbers												
Month	1	2	3	4	5	6	7	8	9	10	11	12
Reduced Merton	9	8	7	8	9	8	8	8	9	9	9	9
Reduced Bates	22	10	9	7	16	11	17	5	17	18	15	17

Table 5: *Condition numbers under reduced models.* The table reports the condition numbers of the Jacobian at the optima of the objective function MSE_{IV} for the reduced calibrations on the last (twelfth) date of the Merton model (preset parameter $\lambda_J = 0.983$) and Bates model (preset parameters $\kappa = 1.879$ and $\lambda_J = 0.182$).

It is worth noting that the variance decomposition-recalibration procedure can be made iterative, which becomes more meaningful for higher-dimensional parameters. Nevertheless, in our current study, a single iteration was sufficient, as confirmed by our variance decomposition results before and after the reduction.

3.3. Statistical analysis of model inputs

Our approach relies on the asymptotic normality (4) of the parameter estimator; however, the sample sizes are finite. This raises the question of whether potential small-sample bias from this configuration could influence our results. To address this concern, we designed and conducted a simulated calibration experiment, followed by a series of statistical tests. In the interest of space, we defer all details and results to Appendix B.

4. Empirical analysis—Part II: Model outputs and evolving calibration risk

In the previous section, we examined the in-sample error and addressed uniqueness considerations of model calibration. However, a model's scope extends beyond capturing market features in an arbitrage-free manner; it is also utilized to extrapolate the value of instruments not directly priced in the market. It would be of visceral importance to understand how the estimation uncertainty based on liquidly traded options transfers to other contracts. To this end, we consider

a battery of contracts, including Asian, lookback, different types of barrier options, and a cliquet option.

4.1. Payout structures

We describe next the payoffs that will form the basis of our analysis of calibration risk. A floating-strike lookback call option has payoff given by the difference between the terminal asset price and the lowest asset price level throughout the option's lifespan $G_{\mathrm{LC},T} = S_T - \min_{0\le t\le lT} S_t$, where $t = 1, \ldots, lT$ for l trading days in a year (under daily monitoring) and T years to maturity. We also consider knock-out call options, including down-and-out and up-and-out for fixed barrier levels L and H, respectively, with payoffs $G_{\mathrm{DOC},T} = (S_T - K)^+ \mathbf{1}_{\left\{\min_{0\le t\le lT} S_t > L\right\}}$ and $G_{\mathrm{UOC},T} = (S_T - K)^+ \mathbf{1}_{\left\{\max_{0\le t\le lT} S_t < H\right\}}$, where $z^+ = \max(z, 0)$ and K is the fixed strike price. Knock-in call options of the down-and-in and up-and-in type have payoffs $G_{\mathrm{DIC},T} = (S_T - K)^+ \mathbf{1}_{\left\{\min_{0\le t\le lT} S_t \le L\right\}}$ and $G_{\mathrm{UIC},T} = (S_T - K)^+ \mathbf{1}_{\left\{\max_{0\le t\le lT} S_t \ge H\right\}}$. A cliquet option has payoff $G_{\mathrm{cliquet},T} = \min(C_{\mathrm{glob}}, \max[F_{\mathrm{glob}}, \sum_{t=1}^{lT} \min\{C_{\mathrm{loc}}, \max(F_{\mathrm{loc}}, R_t)\}])$, where $R_t = (S_t - S_{t-1})/S_{t-1}$ is the return over the subperiod $[t-1, t]$. The returns are then floored by F_{loc} and capped by C_{loc}. A global floor F_{glob} and cap C_{glob} may also apply. Finally, a prevalent payout structure of Asian options is $G_{\mathrm{Asian},T} = \left(\frac{1}{lT}\sum_{t=1}^{lT} S_t - K\right)^+$, based on the arithmetic average of the underlying asset values recorded over the lifetime of the contract.

4.2. Statistical analysis of model outputs

In all our experiments, we assume that all contracts mature after a year. We evaluate them analytically, where possible, or by Monte Carlo simulation (10^6 trials).[10] For the sake of implementation, we generally set $K = S_0 = 100$ and $r = q = 0$. In addition, for the down-type (resp. up-type) barrier options, we use $L = 0.8S_0$ (resp. $H = 1.2S_0$); for the cliquet option, we set $C_{\mathrm{glob}} = \infty$, $F_{\mathrm{glob}} = -0.05$, $C_{\mathrm{loc}} = 0.08$ and $F_{\mathrm{loc}} = -0.03$; the monitoring frequency is assumed to be daily, that is, $l = 252$.

[10] For Asian options (all models), we use the method of Fusai and Kyriakou (2016). For lookbacks and barriers, we follow Fusai et al. (2016) under the Black–Scholes and Merton models; for cliquet options, under the same two models, we use exact simulation (e.g., see Cont and Tankov, 2004, Algorithm 6.2). Under the Heston and Bates models, we use the simulation approach of Kyriakou et al. (2024) for lookback and cliquet options, and the Hilbert transform method of Zeng and Kwok (2014) for barrier options.

In this part, we aim to investigate how the calibration risk affects the distribution of ultimate model option prices. For this, we calculate confidence intervals for the derivatives' prices based on (8), which involves efficiently calculating the partial derivatives of each contract with respect to the model parameters. When explicit pricing expressions are available, we use finite differences; otherwise, we employ the adjoint method of Giles and Glasserman (2006), which is particularly effective in cases of high-dimensional parameters.

Considering our study in Section 3.2, we examine both full and reduced calibrations of the Bates and Merton models for illustrative purposes, while the Heston model is only fully calibrated as there were no identifiability concerns. For such multi-parameter models, small partial derivatives can still be associated with large parameter risk if the elements of the matrix $(\boldsymbol{H}'\boldsymbol{H})^{-1}$ are large, and vice versa. According to expression (9), the overall effect depends on the magnitude and sign of the elements of the gradient vector $\boldsymbol{\Delta}$ and the matrix $(\boldsymbol{H}'\boldsymbol{H})^{-1}$, which together can either decrease or increase the variance of the estimator. The variance of the residuals σ^2 alone may not be sufficient to determine the width of the confidence interval.

We produce 90% confidence intervals for the normalized price estimates for all calibration dates to facilitate comparisons. Let's consider a given contract within the Black–Scholes and Heston model frameworks. In this situation, the variance of the residuals, which is significantly larger for the former, dominates the variance of a price estimator over the combined effect of the elements of $(\boldsymbol{H}'\boldsymbol{H})^{-1}$ and $\boldsymbol{\Delta}$. This results in generally wider confidence intervals under the Black–Scholes model (see Figure D.1 in Appendix D) compared to the Heston model in Figure 1. Under the Bates model, the smaller variance of the residuals can be offset by the larger elements of $(\boldsymbol{H}'\boldsymbol{H})^{-1}$, leading to unreasonably ample confidence intervals. This is evident in the fully calibrated Bates model (see Figure D.2) and is also quite pronounced in the Merton model (see Figure D.3). A reduced calibration for both models decreases the size of $(\boldsymbol{H}'\boldsymbol{H})^{-1}$, significantly ameliorating and streamlining the confidence intervals we obtain in Figures 2 and 3. Finally, when comparing different contracts within a given model setting, the main driver becomes the gradient vector $\boldsymbol{\Delta}$. For the contracts in question, this has the strongest effect on barrier options, resulting in the largest confidence intervals. In contrast, Asians and cliquets exhibit the least uncertainty due to

the tapering effect inherent in their payoffs.

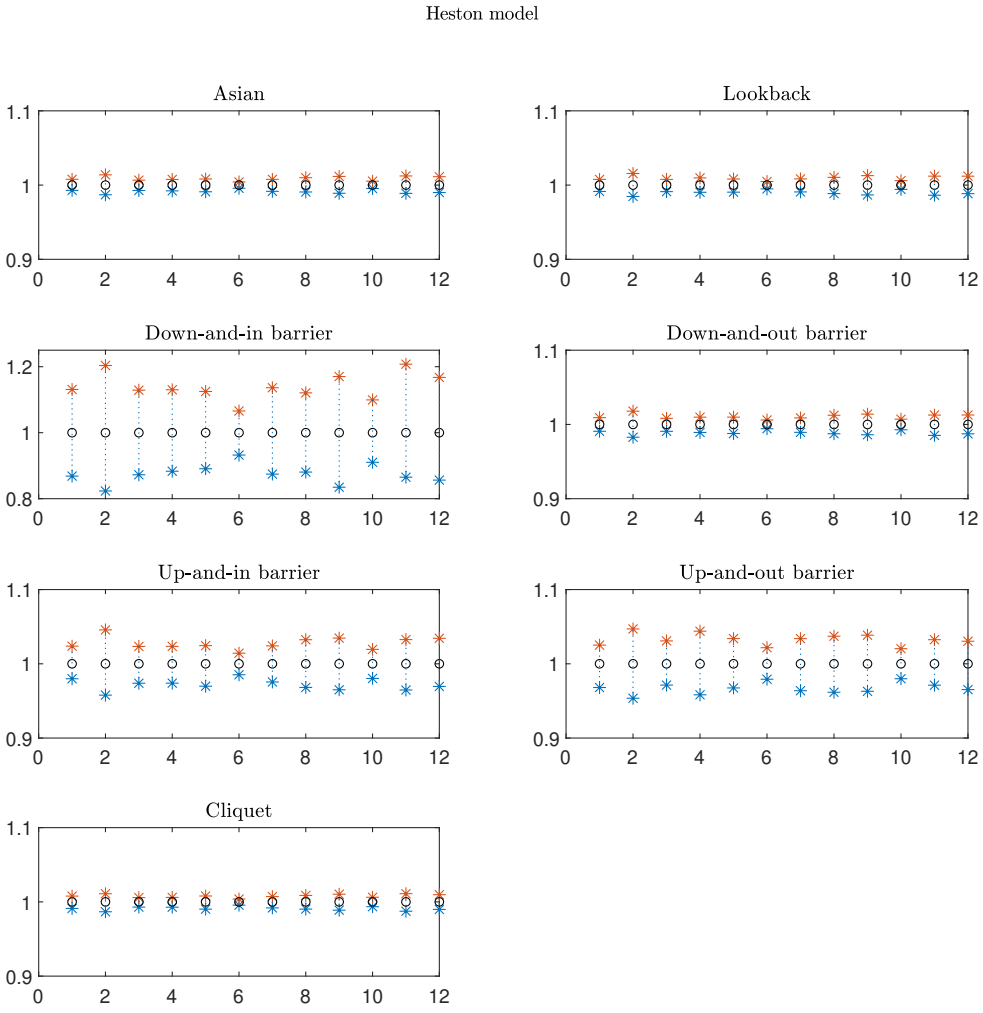

Figure 1: *Confidence intervals: Heston model.* Normalized 90% confidence intervals for option prices in the Heston model based on full calibration on each of 12 months. Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

Lastly, we conduct a supplementary investigation to verify the validity of (8). Building on the normality analysis of model inputs for various sample sizes in Appendix B, we extend it to our model outputs. More specifically, Table C.1 in Appendix C presents the p-values from the one-sample Kolmogorov–Smirnov test, which show no significant evidence against the null hypothesis of normality for the distribution of, for example, Asian and down-and-out barrier call option prices in the Heston and reduced Merton and Bates models. Supporting illustrations C.1–C.6 can also be found there.

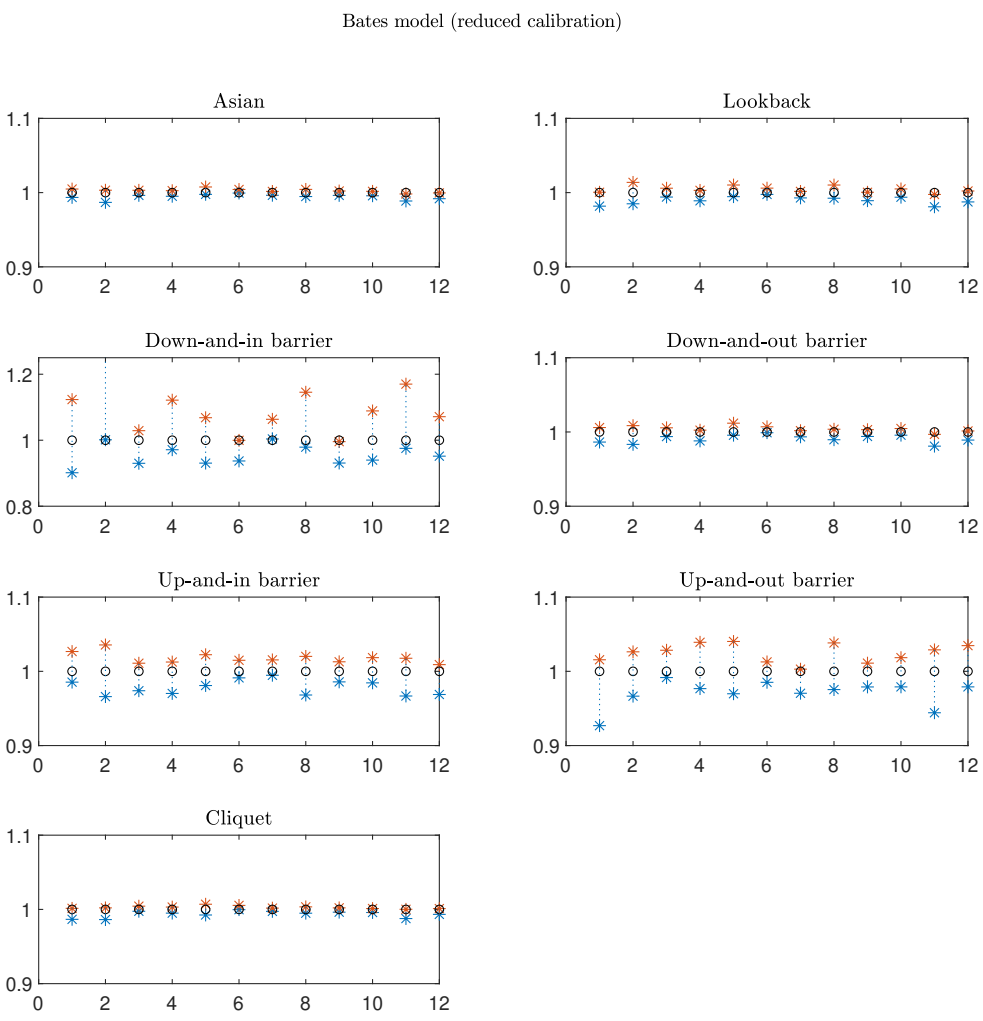

Figure 2: *Confidence intervals: Bates model (reduced calibration).* Normalized 90% confidence intervals for option prices in the Bates model based on reduced calibration on each of 12 months (preset parameters $\kappa = 1.879$ and $\lambda_J = 0.182$). Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

4.3. Analysis of probabilistic dependence of parameter estimates on financial positions

When evaluating financial products, it is crucial to be able to understand the sensitivity of their values to different probabilistic inputs/parameters and the implications for the profitability of a given position. As discussed in Section 2.3, when inputs are correlated, the sensitivity method must be global, accounting for the probabilistic dependencies.

In this section, we put into practice the importance measure using Kuiper's metric based on segregation of the unconditional cumulative distribution F of the contract value and the con-

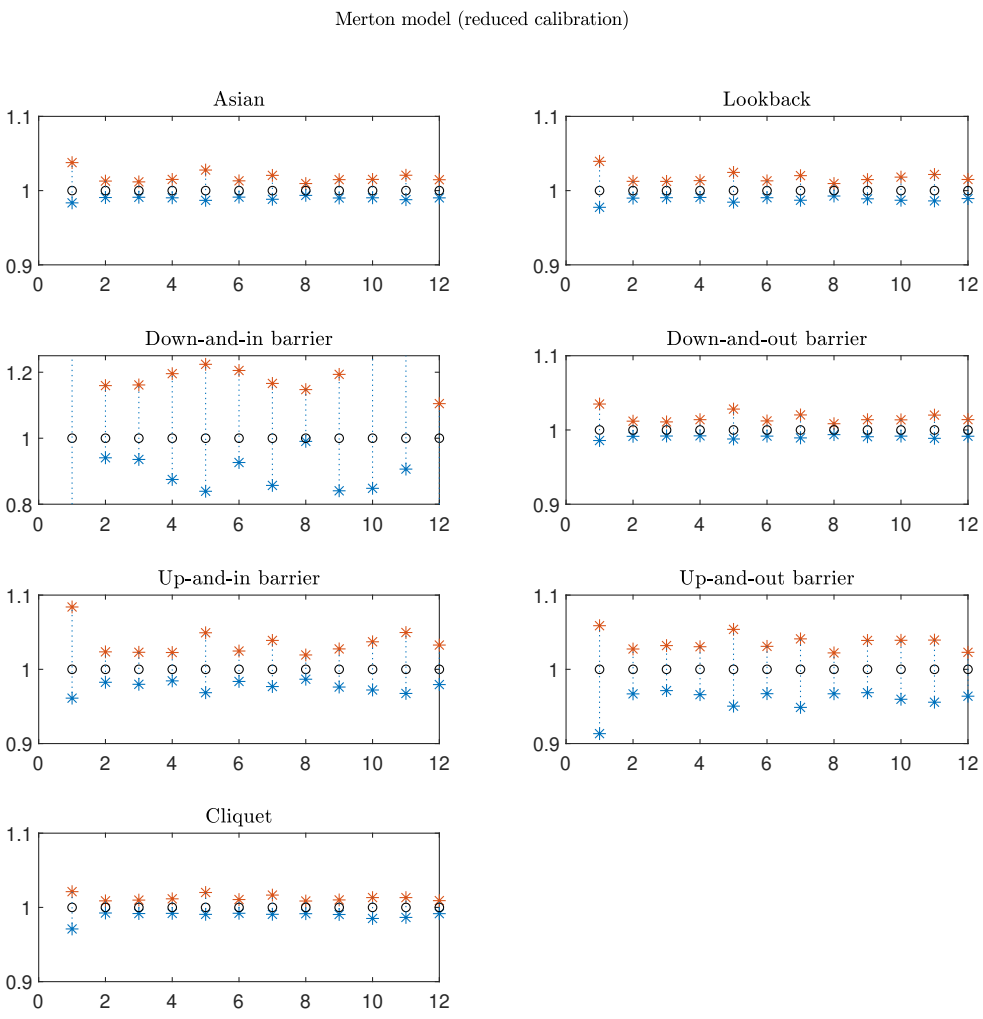

Figure 3: *Confidence intervals: Merton model (reduced calibration).* Normalized 90% confidence intervals for option prices in the Merton model based on reduced calibration (preset parameter $\lambda_J = 0.983$) on each of 12 months. Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

ditional distributions (risk profiles) F_l obtained by fixing the l-th component of the parameter vector over its defining domain. Taking the expectation over the conditioning parameter makes the separation unconditional, providing a measure of the effect of knowing the l-th parameter (see expression 14). The importance measures of the different model parameters with respect to various contract values are reported in Table 6. We also include local sensitivities in our analysis and discuss how their inadequacy, due to overlooking probabilistic information, can lead to misjudgements. For illustrative purposes, we focus the spotlight on the Heston model and a single

contract, while specific references to other contracts and models are made later. In selecting the contract, we consult the confidence intervals in Figure 1. A narrow confidence interval naturally does not raise particular concerns. However, cases like those of barrier options, which show largely varied confidence intervals on different calibration dates, can be quite worrisome regarding the accuracy of their values. In such situations, it is crucial to identify the most influential parameter, warranting extra caution when determining it.

	Asian	Lookback	DI	DO	UI
Heston model calibration					
$\widehat{v_0}$	0.170	0.212	0.111	0.313	0.177
$\widehat{\kappa}$	0.269	0.086	0.451	0.330	0.277
$\widehat{\eta}$	0.091	0.246	0.404	0.093	0.079
$\widehat{\vartheta}$	0.136	0.056	0.179	0.206	0.076
$\widehat{\rho}$	0.128	0.250	0.267	0.158	0.603
Reduced Merton model calibration					
$\widehat{\sigma}$	0.335	0.449	0.391	0.297	0.224
$\widehat{\mu_J}$	0.201	0.155	0.155	0.207	0.707
$\widehat{\sigma_J}$	0.183	0.156	0.160	0.181	0.513
Full Bates model calibration					
$\widehat{v_0}$	0.917	0.910	0.963	0.856	0.727
$\widehat{\kappa}$	0.817	0.815	0.828	0.822	0.768
$\widehat{\eta}$	0.850	0.846	0.868	0.850	0.767
$\widehat{\vartheta}$	0.826	0.822	0.846	0.822	0.733
$\widehat{\rho}$	0.234	0.344	0.144	0.250	0.292
$\widehat{\lambda_J}$	0.281	0.308	0.136	0.258	0.307
$\widehat{\mu_J}$	0.250	0.307	0.135	0.244	0.305
$\widehat{\sigma_J}$	0.235	0.318	0.142	0.244	0.267
Reduced Bates model calibration					
$\widehat{v_0}$	0.093	0.207	0.088	0.239	0.134
$\widehat{\eta}$	0.069	0.050	0.395	0.083	0.095
$\widehat{\vartheta}$	0.344	0.361	0.526	0.354	0.403
$\widehat{\rho}$	0.268	0.428	0.455	0.285	0.397
$\widehat{\mu_J}$	0.163	0.320	0.291	0.194	0.203
$\widehat{\sigma_J}$	0.170	0.301	0.338	0.197	0.205

Table 6: *Importance measures.* The table reports the importance measures based on Kuiper's distance, β_i^{Ku} (see expression 14), for Heston, Merton and Bates models' parameters with respect to the prices of Asian, lookback, down-and-in (DI), down-and-out (DO), and up-and-in (UI) barrier call options. We analyze fully calibrated Heston, and reduced Merton and Bates models, and also present the fully calibrated Bates model (the fully calibrated Merton model has been omitted due to poor quality of results).

Heston model parameter estimates' correlation matrix					
	$\widehat{v_0}$	$\widehat{\kappa}$	$\widehat{\eta}$	$\widehat{\vartheta}$	$\widehat{\rho}$
$\widehat{v_0}$	1.000	-0.827	0.549	-0.770	0.145
$\widehat{\kappa}$	-0.827	1.000	-0.892	0.825	-0.205
$\widehat{\eta}$	0.549	-0.892	1.000	-0.623	0.230
$\widehat{\vartheta}$	-0.770	0.825	-0.623	1.000	0.241
$\widehat{\rho}$	0.145	-0.205	0.230	0.241	1.000
Down-and-in option price sensitivities					
	$\partial U/\partial v_0$	$\partial U/\partial \kappa$	$\partial U/\partial \eta$	$\partial U/\partial \vartheta$	$\partial U/\partial \rho$
	-1.40	-0.04	1.41	0.05	0.87

Table 7: *Probabilistic dependencies.* The top panel of the table reports the correlation matrix based on (5) of Heston model's parameter estimators. With a focus on the case of the down-and-in barrier option, the local sensitivities with respect to the model parameters can be found along the bottom row. Note that positive (negative) products of sensitivities with correlations for a given parameter imply an option price change in the same (opposite) direction with the changing parameter in question but also, implicitly, the correlated parameter estimates along the given row.

Let's therefore consider the case of the down-and-in barrier call option in the Heston model.

From the top panel of Table 6, the parameter estimate $\widehat{\kappa}$ yields the largest importance measure of 0.45. In investigating further, we consider, for instance, from Table D.2 the parameter point estimate for the first calibration date which is 5.309 with a standard error of 0.631. Interestingly, the local first-order price sensitivity with respect to κ is the lowest at -0.04 (see the bottom panel of Table 7), which traditionally implies the least sensitivity with respect to changes in this parameter, contrary to our expectations based on the corresponding importance measure. We choose to stress the parameter estimate such that $\widehat{\kappa} \in \{\text{mean} + \{0, \pm 4\} \times \text{standard error}\} = \{2.7844, 5.3090, 7.8336\}$, while adjusting the remaining parameters to their conditional means. The resulting option price estimates are 0.33, 0.26 and 0.19. At first glance, the decreasing pattern is plausible given the reported first-order sensitivity with respect to κ. However, allowing for parameter uncertainty, as advocated in this paper, reveals not only the individual effect of the parameter but also its dependencies with the others, which should not be neglected. From the parameter estimates' correlations reported in Table 7, changes in κ result in a mixture of changes in the other parameters (consider the signs of correlation-sensitivity products along a row for given a parameter), which do contribute to the overall price decrease.

We repeat the same exercise, this time by stressing parameter ρ, which has a point estimate of -0.546 and a standard error of 0.015. The corresponding first-order price sensitivity is positive and fairly high at 0.87. The resulting option prices are 0.22, 0.25 and 0.26. While this pattern aligns with our expectations, it does not reflect the magnitude of the local sensitivity, which suggests a strong individual effect of ρ. By accounting for parameter interaction effects, the price pattern becomes justifiable, consistent with the lower importance measure of ρ at 0.27. Although ρ still has an effect, the global sensitivity statistic suggests that this is not as strong as initially thought. Compared to knowing ρ, knowledge of κ has a greater incremental impact on the probability of higher value (or lower cost) for an investor who owns (or buys) a positive amount of the contract, but also on the probability of higher cost (or lower value) for an investor who buys (or sells) it.

To illustrate the influence of each parameter on the distribution of a contract value and enhance interpretability, we present a collection of swarm charts in Figure D.4 in Appendix D. These plots display the various predictors/influencing parameters as categorical features along

the vertical axes, while the corresponding Kuiper's distance values, $d^{\mathrm{Ku}}(F, F_l)$ (see expression 10), are distributed along the horizontal axes. Specifically, the plots demonstrate the density of values: densely packed areas indicate high concentration/likelihood of similar values. In addition, "warmer" ("colder") colours indicate values significantly above (below) the mean value (see Table 6). The spread from left to right along the horizontal axis for each predictor reflects the variability in its importance. Predictors with tightly grouped points have a more consistent effect on the model's prediction of the contract value, whereas those with a wider spread of points have a more variable impact. The combined effect of location, spread and likelihood of a particular feature determines its relative importance.

For instance, consider the influence of the Heston model parameters on the down-and-in barrier option. In the top-middle sub-figure, the Kuiper distance values range from 0 to approximately 0.6. Predictors with data points extending further to the right have instances of high influence. Parameters κ and η compete for importance: κ is more concentrated at higher values and exhibits less spread, while η is more dispersed in both directions but more concentrated on the left. Ultimately, κ emerges as the winning predictor. In contrast, v_0 shows a dense cluster around 0.1 with minimal spread from left to right, yielding the lowest importance score. The remaining predictors, ρ and ϑ, also have their dense clusters, but tend to spread towards lower values, particularly ϑ, which explains their intermediate ranking.

A few additional observations are worth noting. Aside from barrier options, the other contract values under the Heston model are less affected by probabilistic dependencies, as reflected in the generally lower Kuiper metric values in Table 6. Nonetheless, κ can still be identified as a main driver of uncertainty. In the Merton model with parameter λ_J preset, σ typically stands out as the most significant parameter. In the Bates model with κ additionally preset, ϑ and ρ compete for dominance as the most influential parameters. It is also worth mentioning that for the fully calibrated Merton and Bates models the sensitivity analysis does not clearly pinpoint the truly most important parameter due to the very close resulting Kuiper distances (see, for instance, full Bates model's reports in Table 6). As discussed in Section 3.2, this highlights the need for a preliminary variance decomposition to uncover near-parameter dependencies, followed

by a reduced calibration.

5. Conclusions

A model may provide a better fit to available market data than another. However, this can be misleading, as different fitted model parameters can exhibit different levels of uncertainty, which then propagate to other model outputs inferred from them. In this paper, we introduce a methodical approach to detecting and addressing the calibration risk, coupled with a probabilistic model that accounts for parameter estimates' dependencies, and its effect on ultimate contract values. Our paradigm leverages on regression and normal modelling, providing analytical tractability while remaining theoretically sound, intuitive and interpretable, and empirically validated.

We calibrate to EURO STOXX 50 implied volatility surfaces, providing empirical evidence of calibration risk. Through a variance decomposition analysis, we detect near dependencies among parameter estimates that raise non-uniqueness and identifiability challenges, which we effectively address by pre-specifying the affected parameters. We examine the validity of core normality for both inputs and outputs, which can be replaced with different approaches, such as nonparametric bootstrapping, if required.

Our findings show that better in-sample model fit, such as in the Bates model, is not sufficient for ensuring well-posedness and should not be considered a safeguard. In fact, this can be detrimental when used to derive model outputs; in our exercise, we observe that contracts with barrier provisions are mostly impacted. In addition, we develop a method to identify the key sources of uncertainty. This focuses on how sensitive output values are to each probabilistic input parameter, while accounting for dependencies among them, as opposed to traditional local sensitivity measures, with implications for the profitability of a given position.

Our framework can guide decision-makers in financial institutions, regulators in applying prudent valuation principles, market participants, and risk managers concerned with parameter estimation risk. From a methodological standpoint, future research will explore alternative calibration methods and their impact on the relevant risk and computational efficiency. From a finance perspective, we will also focus on profitability at portfolio level and exposure to parameter (e.g., see

Torri et al., 2019) but also individual security value uncertainty[11].

References

Amici, G., L. Ballotta, and P. Semeraro (2025). Multivariate additive subordination with applications in finance. *European Journal of Operational Research 321*(3), 1004–1020.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *The Journal of Finance 52*(5), 2003–2049.

Baschetti, F., G. Bormetti, and P. Rossi (2024). Deep calibration with random grids. *Quantitative Finance 24*(9), 1263–1285.

Basel Committee on Banking Supervision (2013). Regulation (EU) No 575/2013 on prudential requirements for credit institutions and investment firms (CRR). Technical report, Bank for International Settlements.

Basel Committee on Banking Supervision (2019a). CAP50 Prudent Valuation Guidance. Technical report, Bank for International Settlements.

Basel Committee on Banking Supervision (2019b). SRP30 Risk Management. Technical report, Bank for International Settlements.

Baucells, M. and E. Borgonovo (2013). Invariant probabilistic sensitivity analysis. *Management Science 59*(11), 2536–2549.

Belsley, D. A., E. Kuh, and R. E. Welsch (2005). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Volume 571. New Jersey: John Wiley & Sons.

Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety 92*(6), 771–784.

Borgonovo, E. and E. Plischke (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research 248*(3), 869–887.

Boyle, P. P. and A. L. Ananthanarayanan (1977). The impact of variance estimation in option valuation models. *Journal of Financial Economics 5*(3), 375–387.

Coleman, T. F. and J. Liu (1999). An interior Newton method for quadratic programming. *Mathematical Programming 85*(3), 491–523.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance 16*(3), 519–547.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. CRC Financial Mathematics Series. Florida: Chapman & Hall.

Cont, R. and P. Tankov (2006). Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization 45*(1), 1–25.

Cui, Y., S. del Baño Rollin, and G. Germano (2017). Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research 263*(2), 625–638.

[11] Preliminary results, available upon request, have been omitted here for brevity.

Derman, E. (1996, April). Model risk. Technical report, Goldman Sachs, New York. Quantitative Strategies Research Notes.

Detlefsen, K. and W. K. Härdle (2007). Calibration risk for exotic options. *The Journal of Derivatives 14*(4), 47–63.

Elliott, R. J., D. B. Madan, and C. H. Lahaie (1997). Filtering derivative security evaluations from market prices. In *Proceedings of the Isaac Newton Workshop in Financial Mathematics*, pp. 141–162.

Farrar, D. E. and R. R. Glauber (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics 49*(1), 92–107.

Federal Reserve (2011). Supervisory Guidance on Model Risk Management. Technical report, Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, SR Letter 11-7.

Feng, L. and V. Linetsky (2008). Pricing discretely monitored barrier options and defaultable bonds in Lévy process models: a fast Hilbert transform approach. *Mathematical Finance 18*(3), 337–384.

Fox, J. (1984). *Linear Statistical Models and Related Methods: With Applications to Social Research*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.

Fusai, G., G. Germano, and D. Marazzina (2016). Spitzer identity, Wiener-Hopf factorization and pricing of discretely monitored exotic options. *European Journal of Operational Research 251*(1), 124–134.

Fusai, G. and I. Kyriakou (2016). General optimized lower and upper bounds for discrete and continuous arithmetic Asian options. *Mathematics of Operations Research 41*(2), 531–559.

Giles, M. and P. Glasserman (2006). Smoking adjoints: fast Monte Carlo Greeks. *Risk 19*(1), 88–92.

Gilli, M. and E. Schumann (2012). Calibrating option pricing models with heuristics. In A. Brabazon, M. O'Neill, and D. Maringer (Eds.), *Natural Computing in Computational Finance: Volume 4*, pp. 9–37. Berlin, Heidelberg: Springer Berlin Heidelberg.

Green, T. C. and S. Figlewski (1999). Market risk and model risk for a financial institution writing options. *The Journal of Finance 54*(4), 1465–1499.

Greene, W. H. (2012). *Econometric Analysis* (Seventh ed.). New Jersey: Prentice Hall.

Guillaume, F. and W. Schoutens (2010). Use a reduced Heston or reduce the use of Heston? *Wilmott Journal 2*(4), 171–192.

Guillaume, F. and W. Schoutens (2012). Calibration risk: Illustrating the impact of calibration risk under the Heston model. *Review of Derivatives Research 15*(1), 57–79.

Gupta, A. and C. Reisinger (2014). Robust calibration of financial models using Bayesian estimators. *The Journal of Computational Finance 17*(4), 3–36.

Hazen, G. B. and M. Huang (2006). Parametric sensitivity analysis using large-sample approximate Bayesian posterior distributions. *Decision Analysis 3*(4), 208–219.

He, C., J. S. Kennedy, T. F. Coleman, P. A. Forsyth, Y. Li, and K. R. Vetzal (2006). Calibration and hedging under jump diffusion. *Review of Derivatives Research 9*(1), 1–35.

Huang, J.-Z. and L. Wu (2004). Specification analysis of option pricing models based on time-changed Lévy processes. *The Journal of Finance 59*(3), 1405–1439.

Hull, J. and W. Suo (2002). A methodology for assessing model risk and its application to the implied volatility function model. *Journal of Financial and Quantitative Analysis 37*(2), 297–318.

Jacquier, E. and R. Jarrow (2000). Bayesian analysis of contingent claim model error. *Journal of Econometrics 94*(1), 145–180.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985). *The Theory and Practice of Econometrics* (Second ed.). Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.

Kuiper, N. H. (1960). Tests concerning random points on a circle. In *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, Volume 63 of *Series A*, pp. 38–47.

Kyriakou, I., R. Brignone, and G. Fusai (2024). Unified moment-based modeling of integrated stochastic processes. *Operations Research 72*(4), 1630–1653.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer Texts in Statistics. New York, NY: Springer New York.

Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory 4*(2), 231–247.

Phillips, P. C. B. and J. Yu (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In T. Mikosch, J. P. Kreiß, R. Davis, and T. Andersen (Eds.), *Handbook of Financial Time Series*, pp. 497–530. Berlin, Heidelberg: Springer.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis. The Primer*. England: John Wiley & Sons.

Schoutens, W., E. Simons, and J. Tistaert (2004). A perfect calibration! Now what? In *Wilmott Magazine* (March 2004 ed.)., pp. 66–78. Wilmott.

Shao, J. and D. Tu (1995). *The Jackknife and Bootstrap* (First ed.). Springer Series in Statistics. New York: Springer.

Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley & Sons.

Torri, G., R. Giacometti, and S. Paterlini (2019). Sparse precision matrices for minimum variance portfolios. *Computational Management Science 16*(3), 375–400.

Wang, J. (1996). Asymptotics of least-squares estimators for constrained nonlinear regression. *The Annals of Statistics 24*(3), 1316–1326.

White, H. and I. Domowitz (1984). Nonlinear regression with dependent observations. *Econometrica 52*(1), 143–161.

Zeng, P. and Y. K. Kwok (2014). Pricing barrier and Bermudan style options under time-changed Lévy processes: fast Hilbert transform approach. *SIAM Journal on Scientific Computing 36*(3), B450–B485.

Online supplementary materials

Appendix A. Option pricing by fast Hilbert transform

We use the Hilbert transform method (see Feng and Linetsky, 2008) to compute fast and accurately the model call option prices corresponding to different strikes K and maturity times T. The pricing formula is given by

$$\begin{aligned} p\left(K,T,\boldsymbol{\theta}^{*}\right) &= \frac{S_0 e^{-qT}}{2}-\frac{Ke^{-rT}}{2} \\ &\quad +\frac{ie^{-rT}}{2}\mathcal{H}\left\{e^{-iu\ln(K/S_0)}\left(S_0\phi\left(u-i,T,\boldsymbol{\theta}^{*}\right)-K\phi\left(u,T,\boldsymbol{\theta}^{*}\right)\right)\right\}(0), \end{aligned} \tag{A.1}$$

where $i=\sqrt{-1}$ and, associated with a given model with parameter vector $\boldsymbol{\theta}^{*}$,

$$\phi\left(u,T,\boldsymbol{\theta}\right)=E\left(e^{iu\ln(S_T/S_0)}\right)$$

is the characteristic function of the log-price increment $\ln\left(S_T/S_0\right)$.

For the Black–Scholes, Merton, Heston, and Bates models in Section 3 of the paper, this is given by

$$\begin{aligned} \phi_{\text{BS}}\left(u,T,\sigma\right) &= \exp\left(iu\left(r-q\right)T-\frac{1}{2}u^2\sigma^2T\right), \\ \phi_{\text{Merton}}\left(u,T,\sigma,\lambda_J,\mu_J,\sigma_J\right) &= \exp\left\{iu\left(r-q-\frac{\sigma^2}{2}-\lambda_J\mu_J\right)T-\frac{1}{2}u^2\sigma^2T\right. \\ &\quad \left.+\lambda_JT\left(\left(1+\mu_J\right)^{iu}e^{\frac{1}{2}iu(iu-1)\sigma_J^2}-1\right)\right\}, \\ \phi_{\text{Heston}}\left(u,T,v_0,\rho,\eta,\kappa,\vartheta\right) &= \exp\left\{iu\left(r-q\right)T\right. \\ &\quad +\frac{\eta\kappa}{\vartheta^2}\left(\left(\kappa-\rho\vartheta iu-d\left(u\right)\right)T-2\ln\frac{1-g\left(u\right)e^{-d(u)T}}{1-g\left(u\right)}\right) \\ &\quad \left.+\frac{v_0}{\vartheta^2}\frac{\left(\kappa-\rho\vartheta iu-d\left(u\right)\right)\left(1-e^{-d(u)T}\right)}{1-g\left(u\right)e^{-d(u)T}}\right\}, \\ \phi_{\text{Bates}}\left(u,T,v_0,\rho,\eta,\kappa,\vartheta,\lambda_J,\mu_J,\sigma_J\right) &= \phi_{\text{Heston}}\left(u,T,v_0,\rho,\eta,\kappa,\vartheta\right) \\ &\quad \times\exp\left\{-\lambda_J\mu_JiuT+\lambda_JT\left(\left(1+\mu_J\right)^{iu}e^{\frac{1}{2}iu(iu-1)\sigma_J^2}-1\right)\right\}, \end{aligned}$$

where

$$d(u) = \sqrt{(\rho\vartheta ui - \kappa)^2 + \vartheta^2 (iu + u^2)} \text{ and } g(u) = \frac{\kappa - \rho\vartheta iu - d(u)}{\kappa - \rho\vartheta iu + d(u)}.$$

The Hilbert transform $\mathcal{H}f(u)$ in (A.1) is computed by the discrete transform

$$\sum_{m=-M}^{M} f(mh) \frac{1 - \cos(\pi(u - mh)/h)}{\pi(u - mh)/h}$$

for some truncation level M and discretization step size h. The required first-order partial derivatives for the computation of the Jacobian matrix (see equation 3 in the paper) can be obtained easily by differentiating equation (A.1).

Appendix B. Statistical analysis of model inputs: simulated calibration experiment and results

Our approach is based on the asymptotic normality of the parameter estimator (refer to Section 2.1 in the paper); however, the finite sample sizes raise the question of whether potential small-sample bias might affect our results. To investigate this, we designed and conducted a simulated calibration experiment, complemented by various illustrations and statistical tests, which we detail next.

Specifically, we assumed five sample sizes: very small ($n = 12$); moderately small ($n = 95$); medium-sized ($n = 153$, which is the average sample size across all original calibration dates—see Section 3.1 in the paper); large ($n = 216$, matching the largest sample size in Section 3.1); and very large ($n = 1000$). For each n, we assumed ranges of option maturity times and strike prices consistent with those observed in the market (see Section 3.1). We then proceeded with the following steps.

Step 1. We computed n model implied volatilities for our models, based on the different strikes and maturities, and the parameter estimates from the calibration which yielded the largest variance of residuals, $\widehat{\sigma}^2$.

Step 2. We perturbed the n model implied volatilities from Step 1 by adding to them n independent Student's t disturbances with: degrees of freedom (i) low $\nu = 5$; (ii) high $\nu = 200$ (essentially,

normal distribution), and a scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu-2)/\nu}$ for $\nu > 2$. This resulted in ten different hypothetical cases based on the aforementioned sample sizes n ($\times 5$) and degrees of freedom ν ($\times 2$), combined with the largest variance of residuals. Of these, the pair $(n, \nu) = (12, 5)$ corresponds to the most challenging calibration and potentially, eventually, the most reassuring one for the validity of the normality of the estimator; the case $(n, \nu) = (1000, 200)$ is the least stressing; the remaining cases lie in between, with $n = 153$ being the average/typical case.

Step 3. We calibrated our models anew to the synthetic n implied volatilities, obtaining an estimate of the model parameter vector $\widehat{\boldsymbol{\theta}}$ and the asymptotic covariance matrix $\widehat{\boldsymbol{\Sigma}}$.

Step 4. We repeated Steps 2–3 $J = 500$ times, ultimately resulting in independent replications $(\widehat{\boldsymbol{\theta}}^{(j)}, \widehat{\boldsymbol{\Sigma}}^{(j)})$, $j = 1, \ldots, J$.

Figures B.1–B.6 display the marginal and bivariate distributions of parameter estimates for the Heston and reduced Merton and Bates models under both normal and Student's t disturbances, based on the simulation outcomes above. In all three models, the plots under both kinds of disturbances provide evidence supporting multivariate normality, with subtle differences between them. Overall, the bivariate relationships exhibit elliptical contours, on several occasions particularly elongated and symmetrically centered around the mean, indicating linear correlations reinforcing the multivariate normality. Under normal disturbances, the contours are smoother and more regular, reflecting stable relationships between the parameters and suggesting fewer outliers or extreme values; under Student's t disturbances, the contours are slightly more irregular, indicating greater variability or potential outliers, though the overall structure of normality is still present. The marginal distributions, along the diagonal, provide further support for normality. While some slight deviations from perfect normality are visible in a few histograms (in peak height and tail behaviour), these are within the range expected of simulation and still generally fit a normal distribution. The correlations are also reflected in the orientation of the ellipses, with certain pairs showing moderate correlations, while others, such as (μ_J, σ_J) in the Merton model, display weak or no correlation, as indicated by more circular contours.

Moving forward to a more rigorous statistical analysis of the results, we tested the multivariate

normality of the parameter vector $\widehat{\boldsymbol{\theta}}$ for each model using Royston's multivariate normality test, and the normality of the marginal distributions of its elements using the one-sample Kolmogorov–Smirnov test. Table B.1 presents the associated p-values for two sample sizes: $n = 12$ (very small) and $n = 95$ (moderately small). The reported p-values suggest acceptance of the null hypothesis of normality at the 10% significance level, with most parameters displaying even higher significance. These findings hold true regardless of the disturbances considered. Encouragingly, even in the most challenging case of $(n, \nu) = (12, 5)$, the results favour normality, with improvements as n increases, regardless of ν. Additional results for larger n are available upon request.

As an additional analysis, we examined: (i) the sample covariance matrix of the estimated parameter vectors $\{\widehat{\boldsymbol{\theta}}^{(j)}\}_{j=1}^{J}$, which provides an unbiased estimate of the true covariance matrix, and (ii) the average of the generated covariance matrices $\{\widehat{\boldsymbol{\Sigma}}^{(j)}\}_{j=1}^{J}$. We compared (i) and (ii) by calculating their Frobenius norms and conducting Box's M test for homogeneity of covariance matrices. Our results, reported in Table B.2, for the Heston and reduced Merton and Bates models (under both types of disturbances) show minimal differences, even at $n = 12$. Statistical evidence supports the null hypothesis of covariance matrix equality, typically for $n \geq 95$ and in some cases even for $n = 12$, further corroborating the validity of the asymptotic normality of the NLLS estimator.

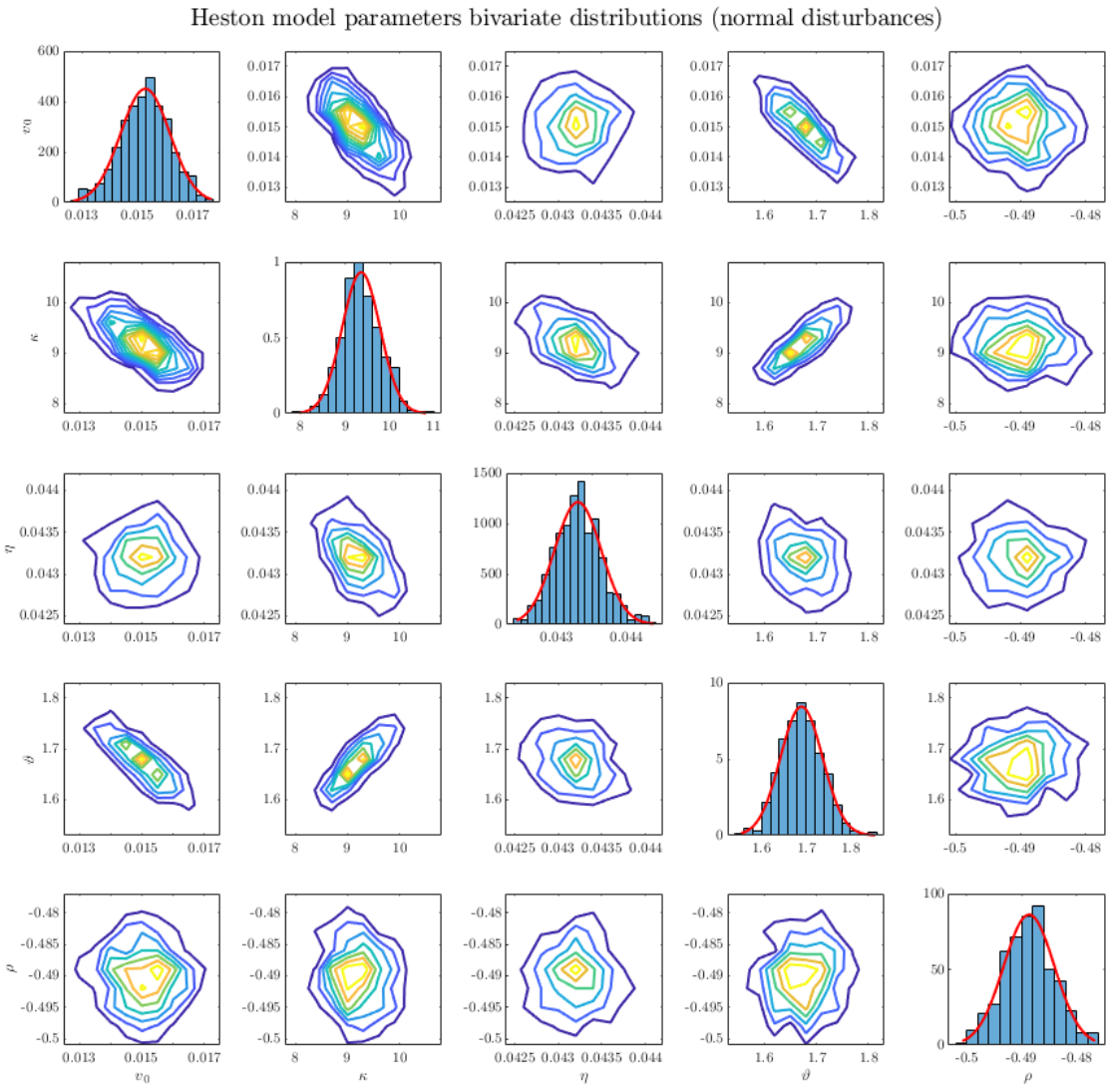

Figure B.1: *Marginal and bivariate distributions of parameter estimates: Heston model, case of normal disturbances.* The distributions are derived from the simulated calibration experiment, for calibration sample size $n = 95$ and normal disturbances.

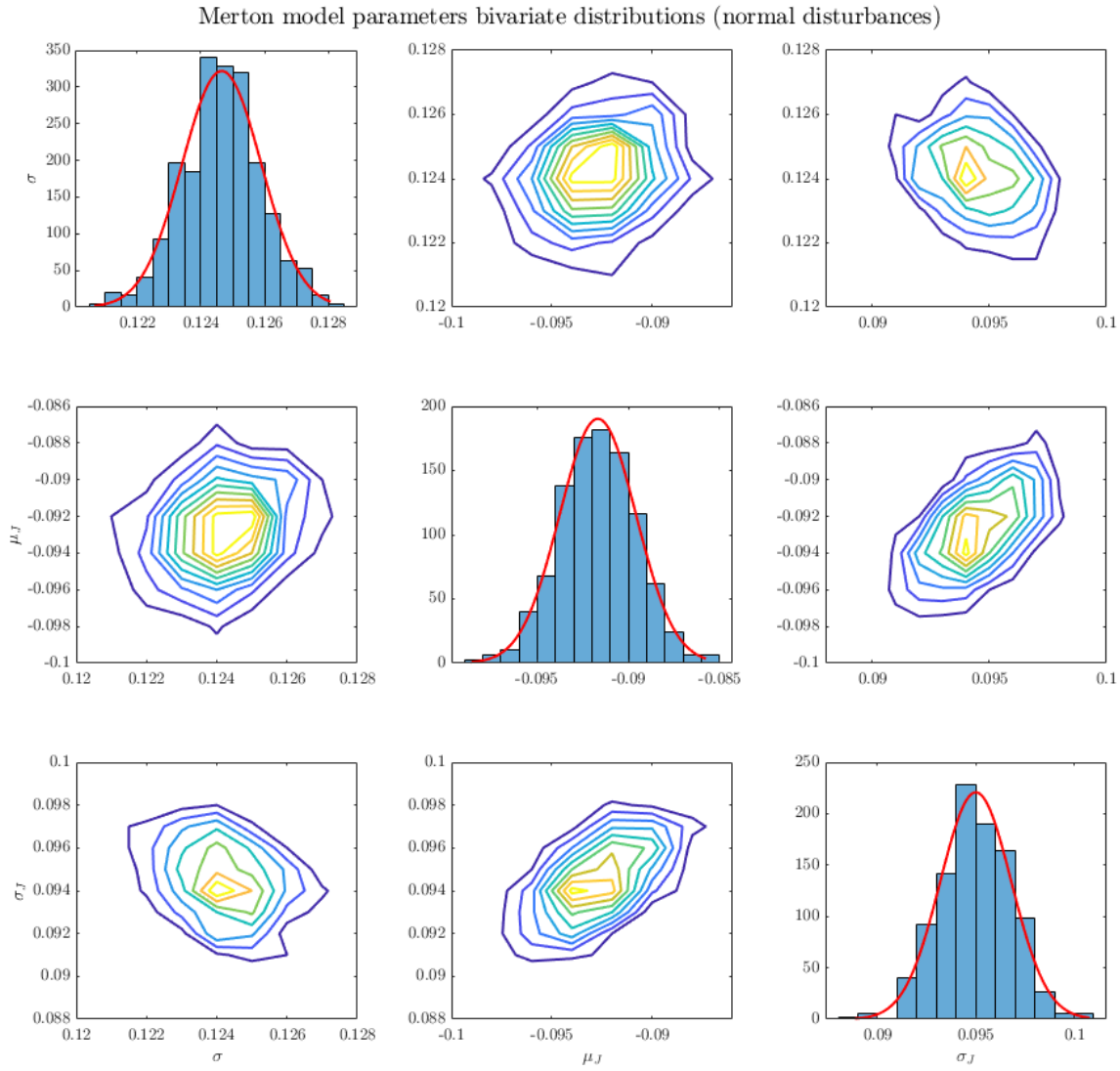

Figure B.2: *Marginal and bivariate distributions of parameter estimates: Reduced Merton model, case of normal disturbances.* See Figure B.1 for additional notes.

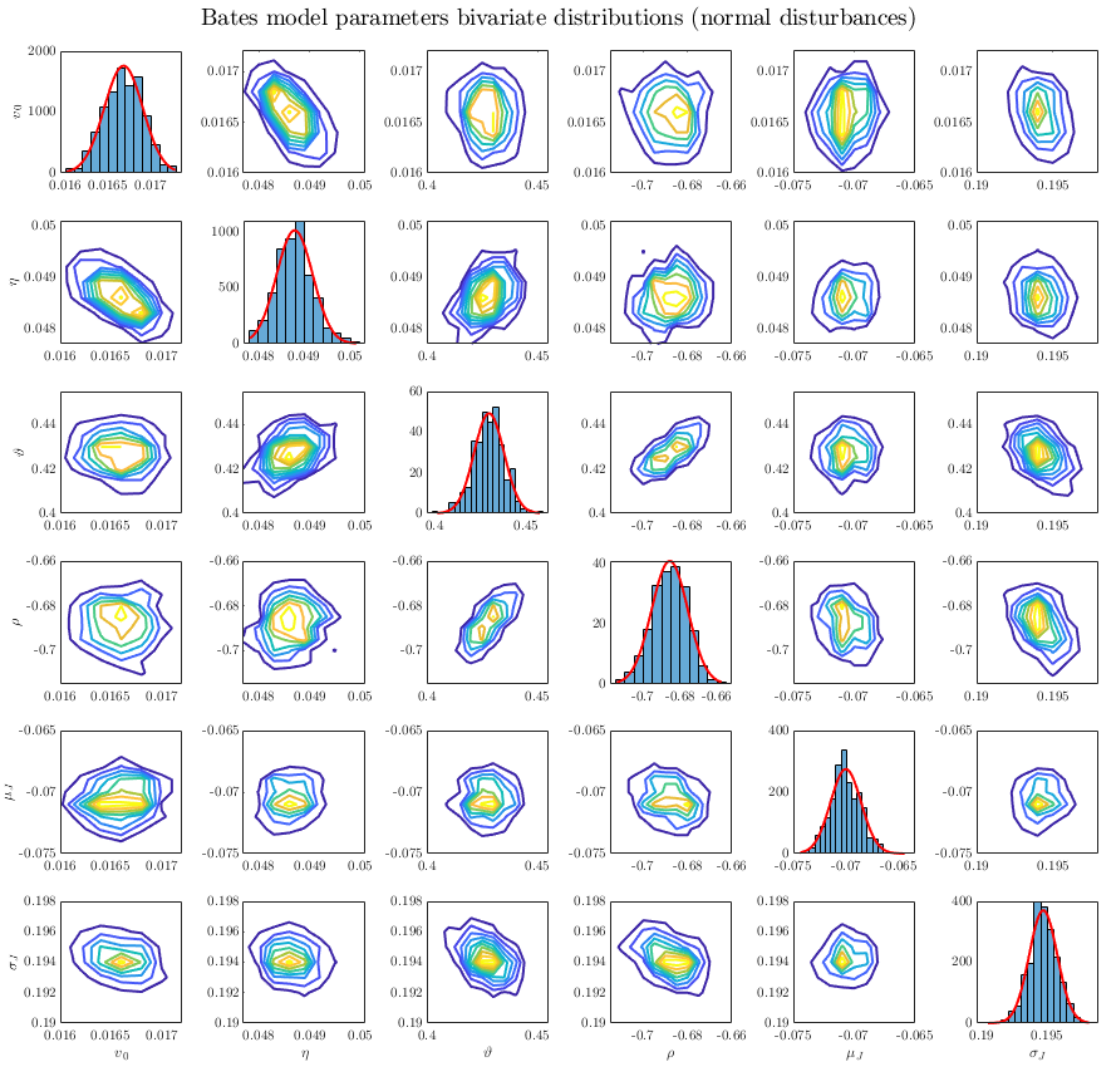

Figure B.3: *Marginal and bivariate distributions of parameter estimates: Reduced Bates model, case of normal disturbances.* See Figure B.1 for additional notes.

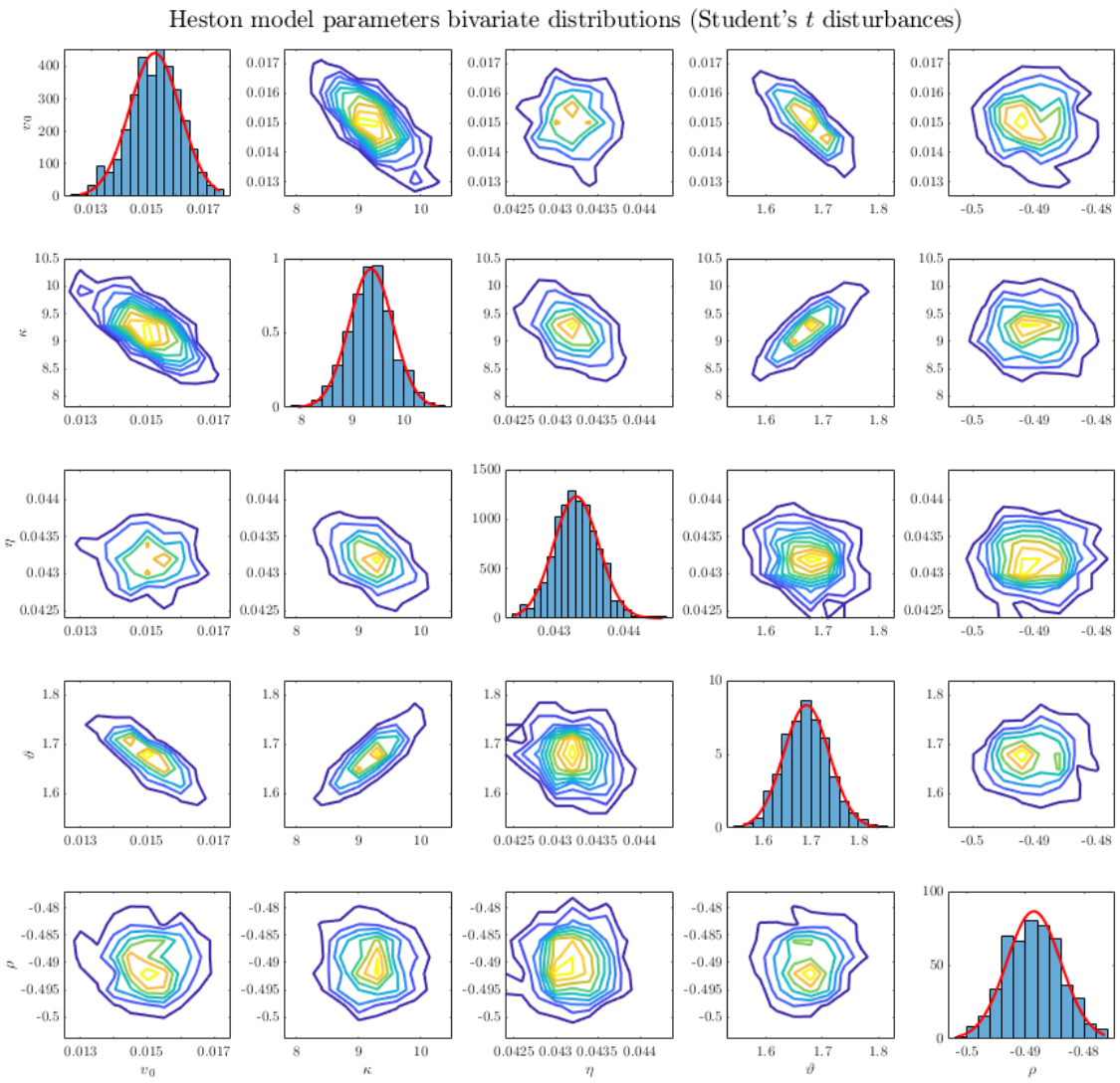

Figure B.4: *Marginal and bivariate distributions of parameter estimates: Heston model, case of Student's t disturbances.* The distributions are derived from the simulated calibration experiment, for calibration sample size $n = 95$ and Student's t disturbances for $\nu = 5$ degrees of freedom and scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu-2)/\nu}$.

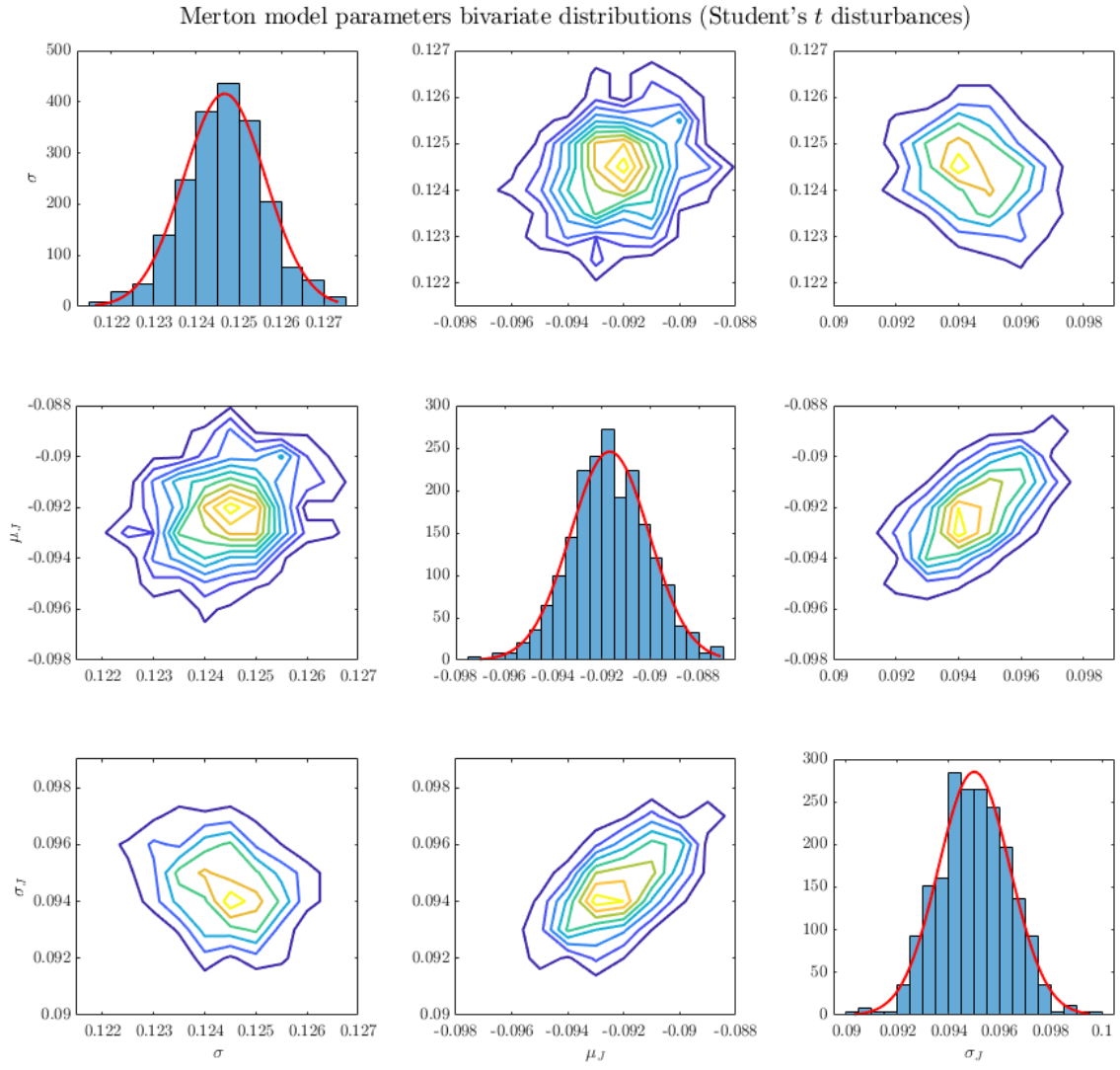

Figure B.5: *Marginal and bivariate distributions of parameter estimates: Reduced Merton model, case of Student's t disturbances.* See Figure B.4 for additional notes.

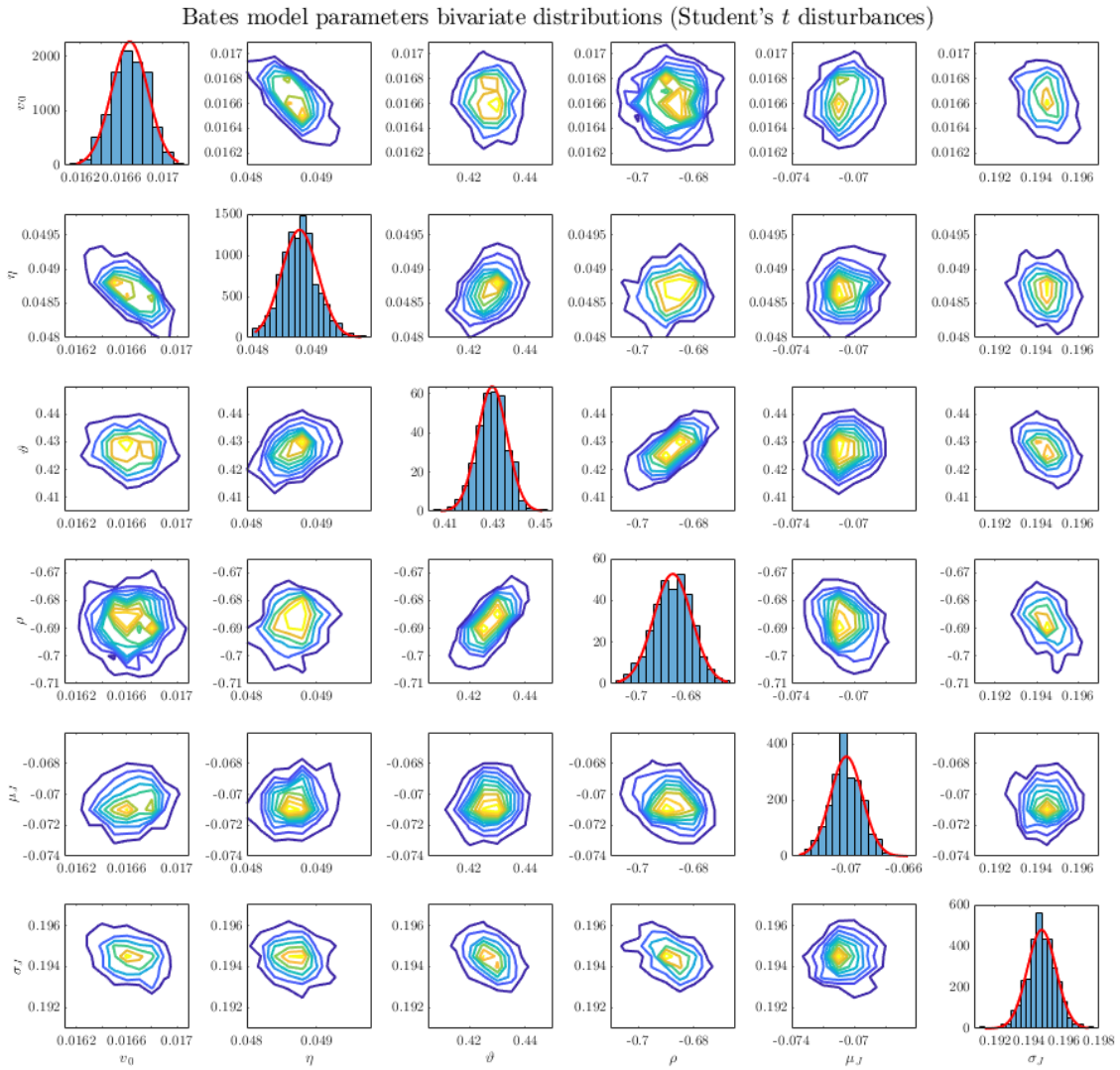

Figure B.6: *Marginal and bivariate distributions of parameter estimates: Reduced Bates model, case of Student's t disturbances.* See Figure B.4 for additional notes.

Marginal parameter distributions. Case of normal disturbances ($\nu = 200$)							
n	Model	Parameter p-value					
	Heston	v_0	κ	η	ϑ	ρ	
12		0.62	0.12	0.78	0.21	0.77	
95		0.99	0.54	0.83	0.98	0.85	
	Reduced Merton	σ	μ_J	σ_J			
12		0.48	0.39	0.39			
95		0.77	0.98	0.98			
	Reduced Bates	v_0	η	ϑ	ρ	μ_J	σ_J
12		0.58	0.50	0.78	0.27	0.16	0.15
95		0.65	0.93	0.85	0.57	0.32	0.64
Joint parameter distribution. Case of normal disturbances ($\nu = 200$)							
n	Model	Royston's statistic	p-value				
95	Heston	2.95	0.40				
	Reduced Merton	0.66	0.88				
	Reduced Bates	7.21	0.24				
Marginal parameter distributions. Case of Student's t disturbances ($\nu = 5$)							
n	Model	Parameter p-value					
	Heston	v_0	κ	η	ϑ	ρ	
12		0.24	0.32	0.79	0.36	0.76	
95		0.83	0.33	0.89	0.55	0.84	
	Reduced Merton	σ	μ_J	σ_J			
12		0.83	0.56	0.40			
95		0.82	0.81	0.48			
	Reduced Bates	v_0	η	ϑ	ρ	μ_J	σ_J
12		0.76	0.50	0.79	0.50	0.31	0.27
95		0.58	0.91	0.85	0.58	0.33	0.21
Joint parameter distribution. Case of Student's t disturbances ($\nu = 5$)							
n	Model	Royston's statistic	p-value				
95	Heston	4.78	0.18				
	Reduced Merton	0.56	0.90				
	Reduced Bates	7.06	0.25				

Table B.1: *One-sample Kolmogorov–Smirnov test for normality and Royston's multivariate normality test of estimated parameter vectors.* The table reports the tests' p-values for the parameter estimates of the Heston, reduced Merton (with preset parameter $\lambda_J = 0.983$) and reduced Bates (with preset parameters $\kappa = 1.879$ and $\lambda_J = 0.182$) models. The distributions are derived from the simulated calibration experiment, for calibration sample size n and Student's t disturbances for degrees of freedom $\nu = 5$ and $\nu = 200$ (normal) and scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu-2)/\nu}$.

Case of normal disturbances ($\nu = 200$)				*Case of Student's t disturbances* ($\nu = 5$)		
			Heston model			
n	Statistic	p-value	Frobenius norm	Statistic	p-value	Frobenius norm
12	25.405	0.046	4.29E-03	20.562	0.155	1.09E-02
95	13.959	0.534	2.09E-03	17.012	0.324	8.13E-02
153	14.357	0.504	3.44E-03	10.539	0.788	1.99E-01
216	6.803	0.964	4.09E-04	8.723	0.894	2.24E-02
			Reduced Merton model			
n	Statistic	p-value	Frobenius norm	Statistic	p-value	Frobenius norm
12	13.170	0.041	1.55E-05	27.301	0.167	2.57E-04
95	10.303	0.114	1.37E-05	21.963	0.410	1.67E-04
153	7.228	0.302	1.65E-05	19.731	0.546	5.53E-04
216	1.600	0.953	3.32E-07	12.509	0.927	1.03E-05
			Reduced Bates model			
n	Statistic	p-value	Frobenius norm	Statistic	p-value	Frobenius norm
12	38.424	0.012	2.20E-03	10.196	0.118	8.18E-06
95	34.250	0.036	4.67E-04	7.086	0.315	8.88E-06
153	25.711	0.224	3.58E-04	6.059	0.419	8.43E-06
216	12.528	0.927	1.71E-05	1.610	0.952	1.98E-07

Table B.2: *Box's M test for homogeneity of covariance matrices.* The Frobenius norm of the difference of the two matrices is defined as the square root of the sum of the absolute squares of its elements. This distance is a measure of how "far apart" the matrices are from each other in terms of the sum of the squared differences of their elements. The table reports the tests' p-values for the Heston and reduced Merton and Bates models. The distributions are derived from the simulated calibration experiment, for calibration sample sizes $n \in \{12, 95, 153, 216\}$ and Student's t disturbances for degrees of freedom $\nu = 5$ and $\nu = 200$ (normal) and scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu - 2)/\nu}$.

Appendix C. Statistical analysis of model outputs: supplementary results

	Asian option			Down-and-out barrier option		
n	Heston	Reduced Merton	Reduced Bates	Heston	Reduced Merton	Reduced Bates
12	0.33	0.28	0.53	0.89	0.28	0.51
95	0.96	0.93	0.51	0.91	0.91	0.53
153	0.92	0.96	0.88	0.98	0.92	0.63
216	0.92	0.95	0.87	0.92	0.92	0.94
1000	0.98	0.92	0.89	0.94	0.98	1.00

Table C.1: *One-sample Kolmogorov–Smirnov test for normality of option price distributions.* The table reports the tests' p-values for the Asian and down-and-out barrier option price distributions in the Heston, reduced Merton (preset parameter $\lambda_J = 0.983$) and reduced Bates (preset parameters $\kappa = 1.879$ and $\lambda_J = 0.182$) models. The distributions are derived from the simulated calibration experiment in Appendix B, for calibration sample sizes $n \in \{12, 95, 153, 216, 1000\}$ and Student's t disturbances for degrees of freedom $\nu = 5$ and scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu-2)/\nu}$.

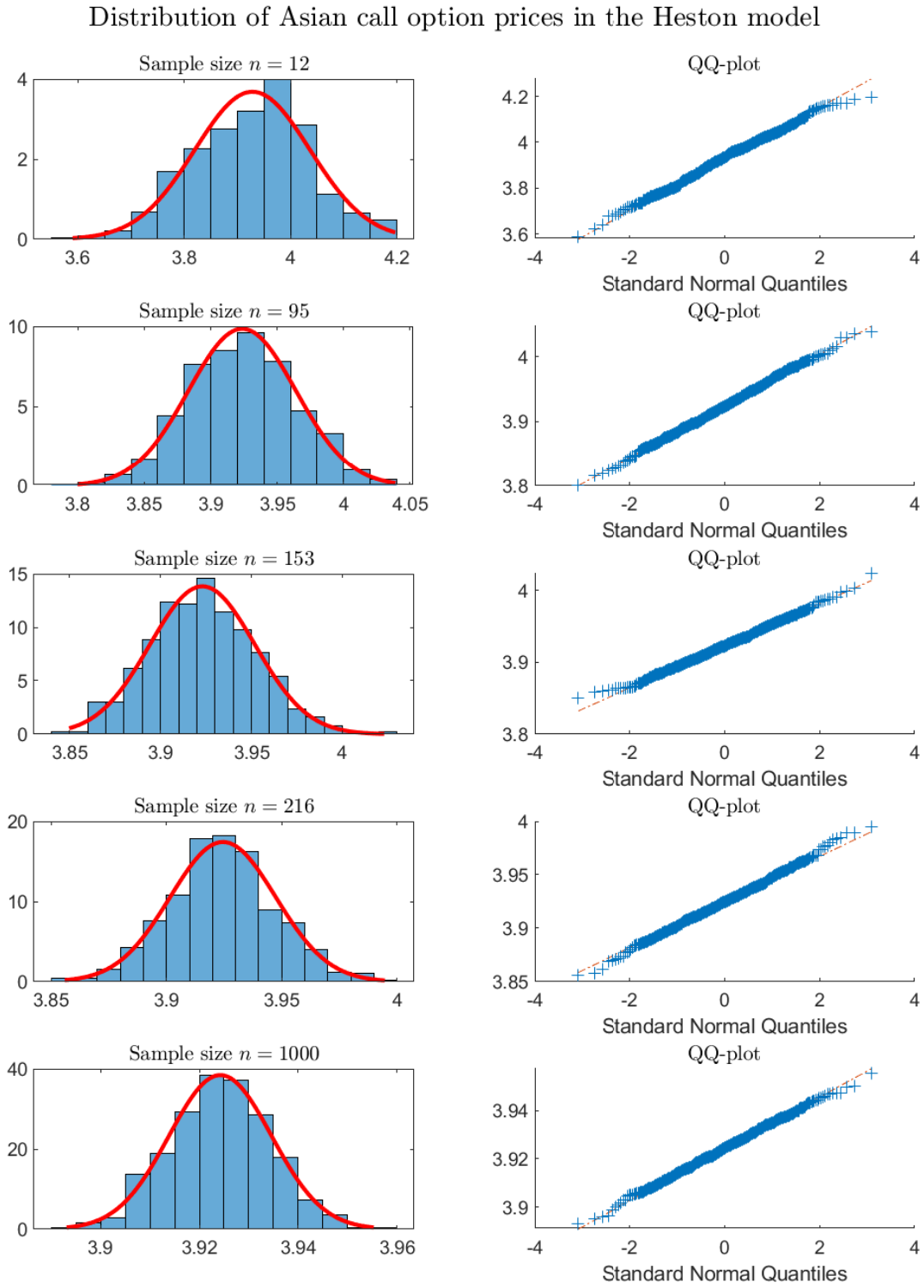

Figure C.1: *Asian call option price distributions in the Heston model.* The distributions are derived from the simulated calibration experiment in Appendix B, for calibration sample sizes $n \in \{12, 95, 153, 216, 1000\}$ and Student's t disturbances for degrees of freedom $\nu = 5$ and scale parameter $sc = \sqrt{\widehat{\sigma}^2(\nu-2)/\nu}$.

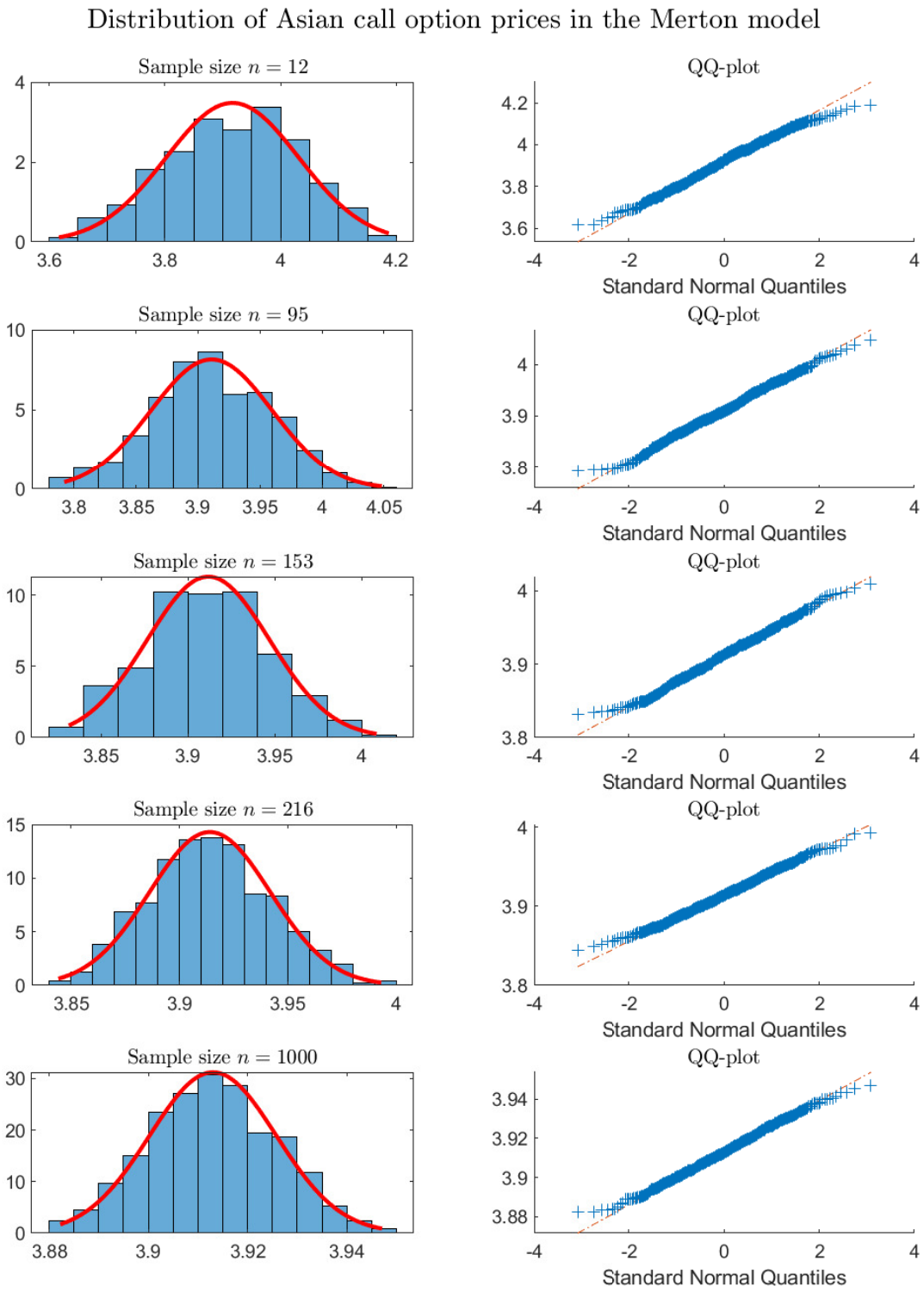

Figure C.2: *Asian call option price distributions in the reduced Merton model.* See Figure C.1 for additional notes.

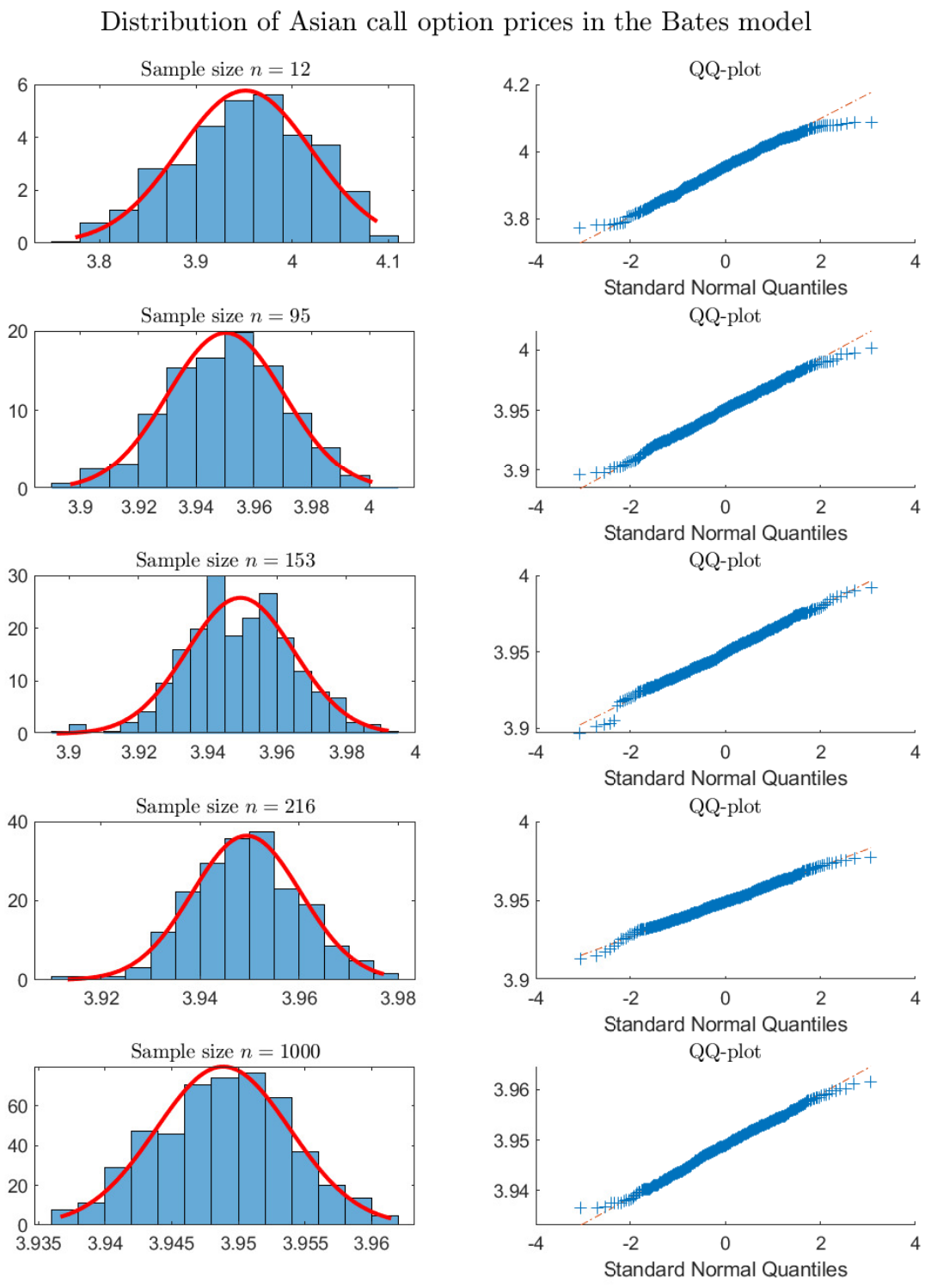

Figure C.3: *Asian call option price distributions in the reduced Bates model.* See Figure C.1 for additional notes.

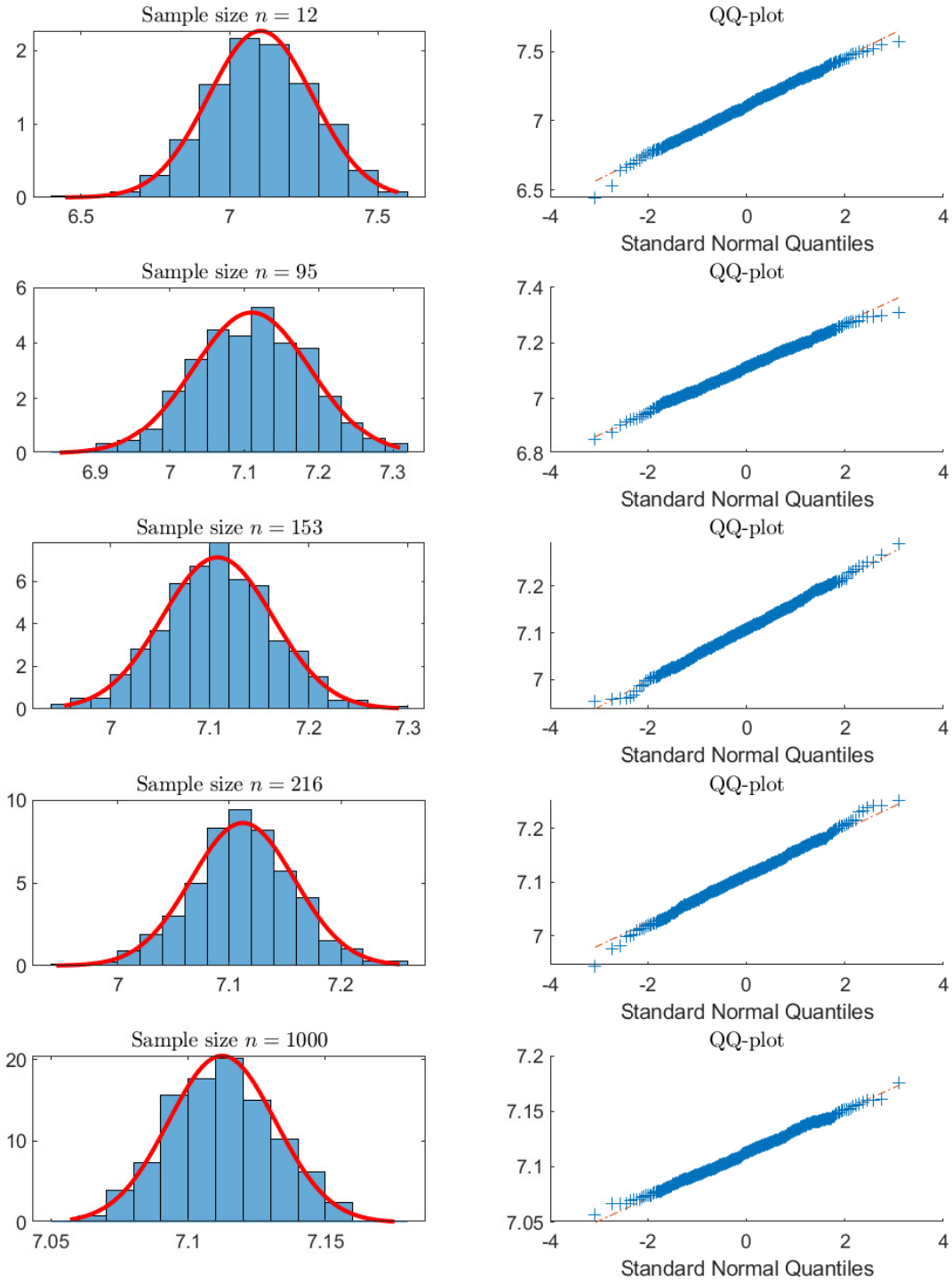

Figure C.4: *Down-and-out (DO) barrier call option price distributions in the Heston model.* See Figure C.1 for additional notes.

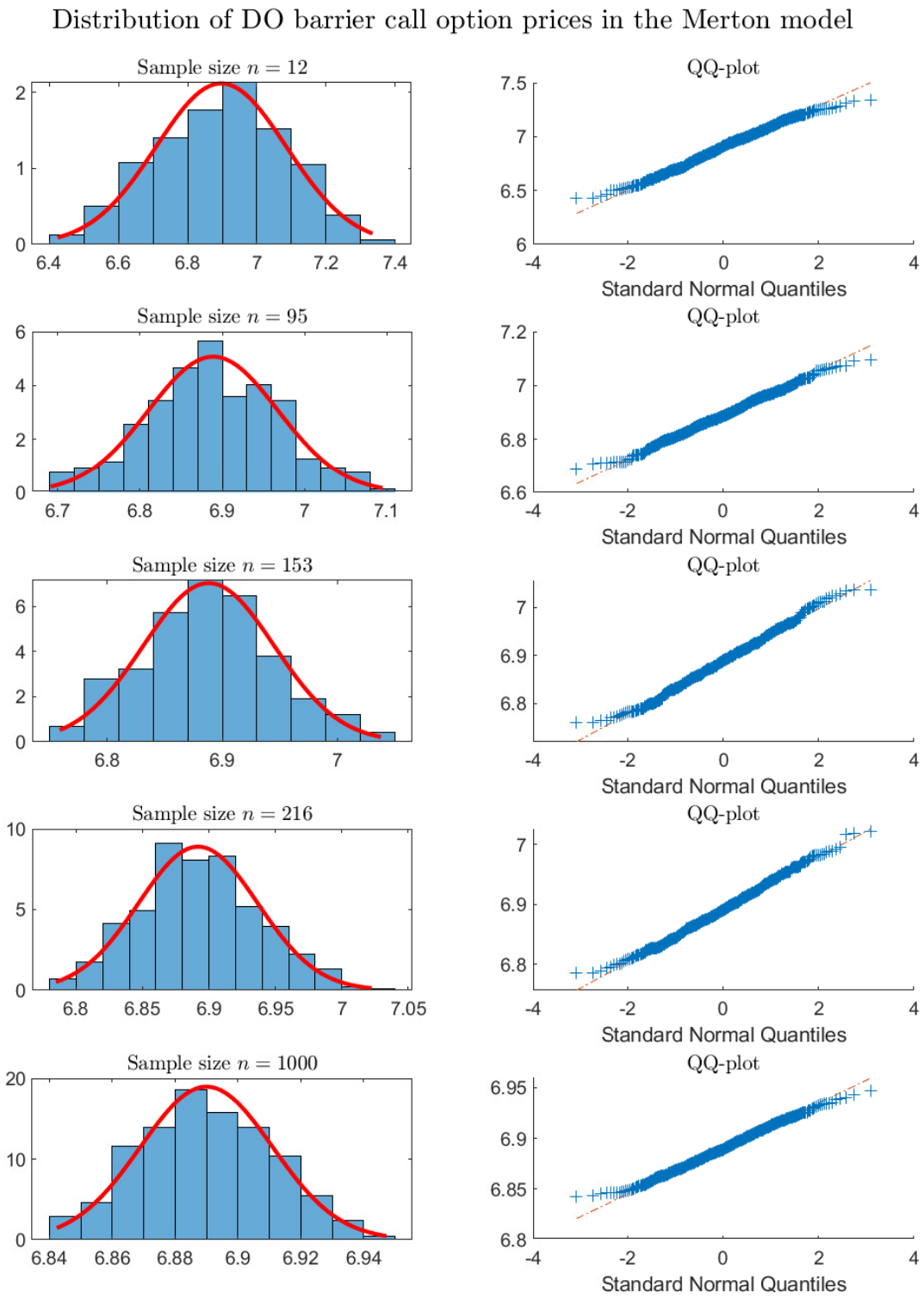

Figure C.5: *DO barrier call option price distributions in the reduced Merton model.* See Figure C.1 for additional notes.

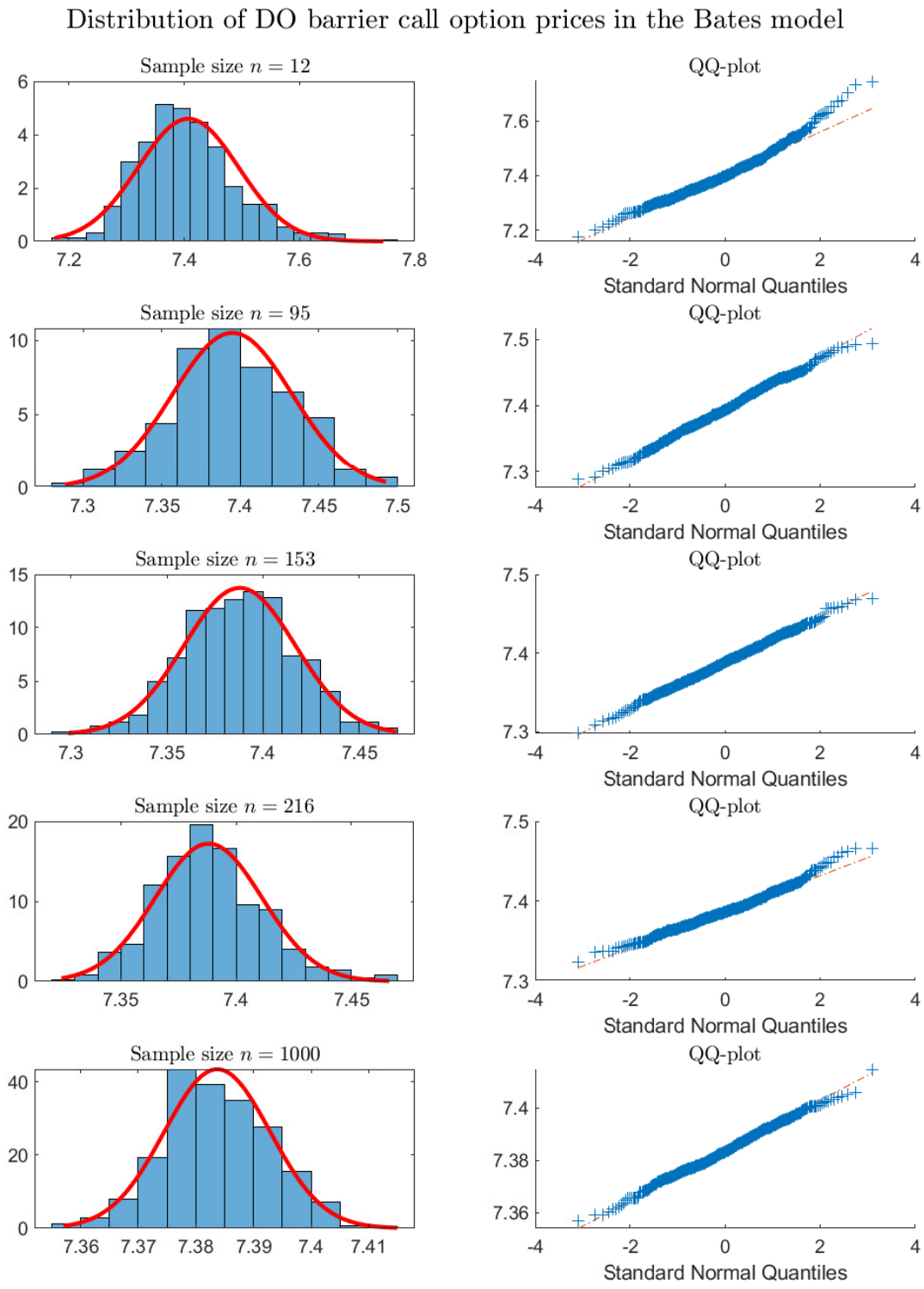

Figure C.6: *DO barrier call option price distributions in the reduced Bates model.* See Figure C.1 for additional notes.

Appendix D. General supplementary results

Average in-sample error metrics				
	MSE_p	MSE_{IV}	PMSE_p	PMSE_{IV}
Black–Scholes	138.40	2.42E-03	3.47E-01	4.46E-02
Merton	12.96	2.11E-04	4.38E-02	4.05E-03
Heston	0.81	4.99E-05	1.37E-02	8.31E-04
Bates	0.45	7.98E-06	1.68E-03	1.62E-04
MSE_{IV} *at different calibration dates*				
Month	Black–Scholes	Merton	Heston	Bates
1	1.62E-03	2.54E-04	2.40E-05	2.02E-06
2	2.45E-03	1.69E-04	9.65E-05	3.31E-05
3	3.11E-03	1.49E-04	3.38E-05	5.82E-06
4	2.56E-03	2.09E-04	5.06E-05	9.05E-06
5	1.65E-03	3.93E-04	3.85E-05	4.86E-06
6	2.09E-03	1.99E-04	1.69E-05	3.54E-06
7	2.16E-03	2.93E-04	3.91E-05	3.06E-06
8	3.81E-03	1.01E-04	5.89E-05	1.04E-05
9	2.62E-03	1.64E-04	6.41E-05	4.40E-06
10	2.13E-03	1.01E-04	1.23E-05	3.59E-06
11	2.50E-03	3.24E-04	9.44E-05	9.31E-06
12	2.30E-03	1.83E-04	7.02E-05	6.62E-06

Table D.1: *Error metrics.* The top panel reports the average MSE_p, MSE_{IV}, PMSE_p, PMSE_{IV} across 12 calibration dates for each model. The bottom panel spotlights the case of MSE_{IV}, which is the lowest metric, for each date and model.

Month	Black–Scholes	Merton				Heston					Bates							
	$\widehat{\sigma}$	$\widehat{\sigma}$	$\widehat{\lambda_J}$	$\widehat{\mu_J}$	$\widehat{\sigma_J}$	$\widehat{v_0}$	$\widehat{\kappa}$	$\widehat{\eta}$	$\widehat{\vartheta}$	$\widehat{\rho}$	$\widehat{v_0}$	$\widehat{\kappa}$	$\widehat{\eta}$	$\widehat{\vartheta}$	$\widehat{\rho}$	$\widehat{\lambda_J}$	$\widehat{\mu_J}$	$\widehat{\sigma_J}$
1	0.200	0.159	0.034	-0.998	0.400	0.019	5.309	0.054	1.140	-0.546	0.020	1.187	0.072	0.404	-0.566	0.046	-0.374	0.221
	(0.004)	(0.008)	(0.1)	(1.466)	(6.979)	(0.001)	(0.631)	(0.001)	(0.066)	(0.015)	(4.57E-04)	(0.162)	(0.005)	(0.018)	(0.021)	(0.035)	(0.26)	(0.197)
2	0.214	0.141	1.174	-0.106	0.079	0.040	10.000	0.043	1.632	-0.494	0.030	1.588	0.043	0.496	-0.545	0.280	-0.139	0.111
	(0.004)	(0.012)	(1.184)	(0.094)	(0.035)	(0.001)	(1.248)	(0.001)	(0.086)	(0.022)	(0.002)	(0.551)	(0.004)	(0.046)	(0.066)	(0.343)	(0.164)	(0.067)
3	0.219	0.118	1.599	-0.103	0.078	0.034	10.000	0.048	1.667	-0.622	0.027	3.343	0.041	0.658	-0.652	0.269	-0.143	0.132
	(0.005)	(0.013)	(1.067)	(0.062)	(0.023)	(0.001)	(0.805)	(0.001)	(0.055)	(0.014)	(0.001)	(0.356)	(0.001)	(0.03)	(0.041)	(0.092)	(0.054)	(0.023)
4	0.238	0.151	1.603	-0.111	0.077	0.057	9.948	0.052	1.469	-0.598	0.043	2.461	0.038	0.501	-0.632	0.282	-0.177	0.123
	(0.005)	(0.02)	(2.234)	(0.134)	(0.053)	(0.001)	(0.907)	(0.001)	(0.059)	(0.019)	(0.002)	(0.328)	(0.003)	(0.037)	(0.058)	(0.304)	(0.175)	(0.084)
5	0.202	0.129	1.741	-0.084	0.064	0.024	4.544	0.052	0.984	-0.510	0.025	2.424	0.054	0.547	-0.573	0.004	-1.455	0.104
	(0.003)	(0.033)	(4.115)	(0.171)	(0.057)	(0.001)	(0.432)	(0.001)	(0.034)	(0.016)	(1.70E-04)	(0.211)	(0.001)	(0.014)	(0.009)	(0.002)	(142.318)	(1376.548)
6	0.223	0.136	1.500	-0.110	0.081	0.041	7.189	0.052	1.260	-0.609	0.040	5.968	0.047	0.889	-0.637	0.006	-1.399	0.079
	(0.003)	(0.016)	(1.435)	(0.093)	(0.035)	(3.98E-04)	(0.366)	(4.99E-04)	(0.028)	(0.009)	(2.47E-04)	(0.292)	(3.53E-04)	(0.019)	(0.007)	(0.003)	(158.858)	(2015.686)
7	0.217	0.131	1.928	-0.084	0.073	0.031	7.791	0.051	1.344	-0.559	0.029	2.033	0.057	0.583	-0.591	0.130	-0.160	0.150
	(0.004)	(0.023)	(2.138)	(0.087)	(0.029)	(0.001)	(0.691)	(0.001)	(0.049)	(0.016)	(0.001)	(0.153)	(0.002)	(0.018)	(0.027)	(0.055)	(0.076)	(0.033)
8	0.229	0.130	0.996	-0.139	0.099	0.040	10.000	0.053	2.075	-0.593	0.026	2.927	0.037	0.572	-0.575	0.356	-0.174	0.143
	(0.005)	(0.008)	(0.497)	(0.063)	(0.027)	(0.001)	(0.959)	(0.001)	(0.089)	(0.013)	(0.001)	(0.497)	(0.002)	(0.043)	(0.053)	(0.118)	(0.059)	(0.028)
9	0.214	0.139	0.836	-0.116	0.094	0.029	10.000	0.046	1.764	-0.484	0.026	0.830	0.074	0.472	-0.526	0.093	-0.278	0.101
	(0.004)	(0.012)	(0.767)	(0.101)	(0.038)	(0.001)	(1.189)	(0.001)	(0.092)	(0.019)	(9.35E-04)	(0.2)	(0.01)	(0.015)	(0.01)	(0.027)	(0.051)	(0.045)
10	0.207	0.130	0.513	-0.165	0.132	0.026	8.293	0.046	1.563	-0.570	0.022	1.437	0.044	0.427	-0.503	0.150	-0.247	0.169
	(0.004)	(0.009)	(0.291)	(0.09)	(0.037)	(0.001)	(0.665)	(0.001)	(0.07)	(0.01)	(0.001)	(0.314)	(0.003)	(0.036)	(0.033)	(0.095)	(0.143)	(0.075)
11	0.195	0.134	0.547	-0.113	0.107	0.016	10.000	0.042	1.834	-0.457	0.018	1.692	0.054	0.472	-0.571	0.090	-0.142	0.223
	(0.003)	(0.009)	(0.404)	(0.084)	(0.029)	(0.001)	(1.245)	(0.001)	(0.113)	(0.018)	(3.28E-04)	(0.189)	(0.003)	(0.022)	(0.038)	(0.015)	(0.034)	(0.018)
12	0.202	0.139	0.706	-0.115	0.091	0.026	10.000	0.041	1.649	-0.459	0.024	1.296	0.050	0.446	-0.513	0.120	-0.167	0.167
	(0.004)	(0.01)	(0.701)	(0.105)	(0.039)	(0.001)	(1.085)	(0.001)	(0.082)	(0.016)	(4.04E-04)	(0.187)	(0.003)	(0.018)	(0.03)	(0.032)	(0.053)	(0.025)

Table D.2: *Calibration outcome.* Estimated parameters (standard errors in parentheses) of the Black–Scholes, Merton, Heston and Bates models for 12 calibration dates. Each calibration is based on minimization of MSE_{IV}.

Month	Reduced Merton			Reduced Bates					
	$\widehat{\sigma}$	$\widehat{\mu_J}$	$\widehat{\sigma_J}$	$\widehat{v_0}$	$\widehat{\eta}$	$\widehat{\vartheta}$	$\widehat{\rho}$	$\widehat{\mu_J}$	$\widehat{\sigma_J}$
1	0.122	-0.096	0.123	0.017	0.055	0.400	-0.730	-0.096	0.224
2	0.144	-0.120	0.075	0.031	0.042	0.522	-0.541	-0.186	0.096
3	0.127	-0.142	0.068	0.027	0.049	0.595	-0.623	-0.217	0.088
4	0.159	-0.151	0.067	0.043	0.039	0.479	-0.618	-0.253	0.064
5	0.137	-0.119	0.056	0.022	0.054	0.428	-0.769	-0.025	0.188
6	0.143	-0.148	0.069	0.035	0.053	0.644	-0.582	-0.189	0.101
7	0.142	-0.132	0.061	0.029	0.057	0.564	-0.614	-0.116	0.158
8	0.130	-0.138	0.102	0.027	0.044	0.541	-0.564	-0.287	0.073
9	0.136	-0.104	0.095	0.025	0.049	0.497	-0.574	-0.142	0.175
10	0.114	-0.108	0.124	0.021	0.040	0.444	-0.540	-0.193	0.199
11	0.124	-0.083	0.096	0.017	0.049	0.431	-0.684	-0.068	0.197
12	0.133	-0.094	0.090	0.023	0.043	0.449	-0.583	-0.098	0.182

Table D.3: *Updated calibration outcome.* Estimated parameters of the reduced Merton model (preset parameter $\lambda_J = 0.983$) and Bates model (preset parameters $\kappa = 1.879$ and $\lambda_J = 0.182$) for 12 calibration dates. Each calibration is based on minimization of MSE_{IV}.

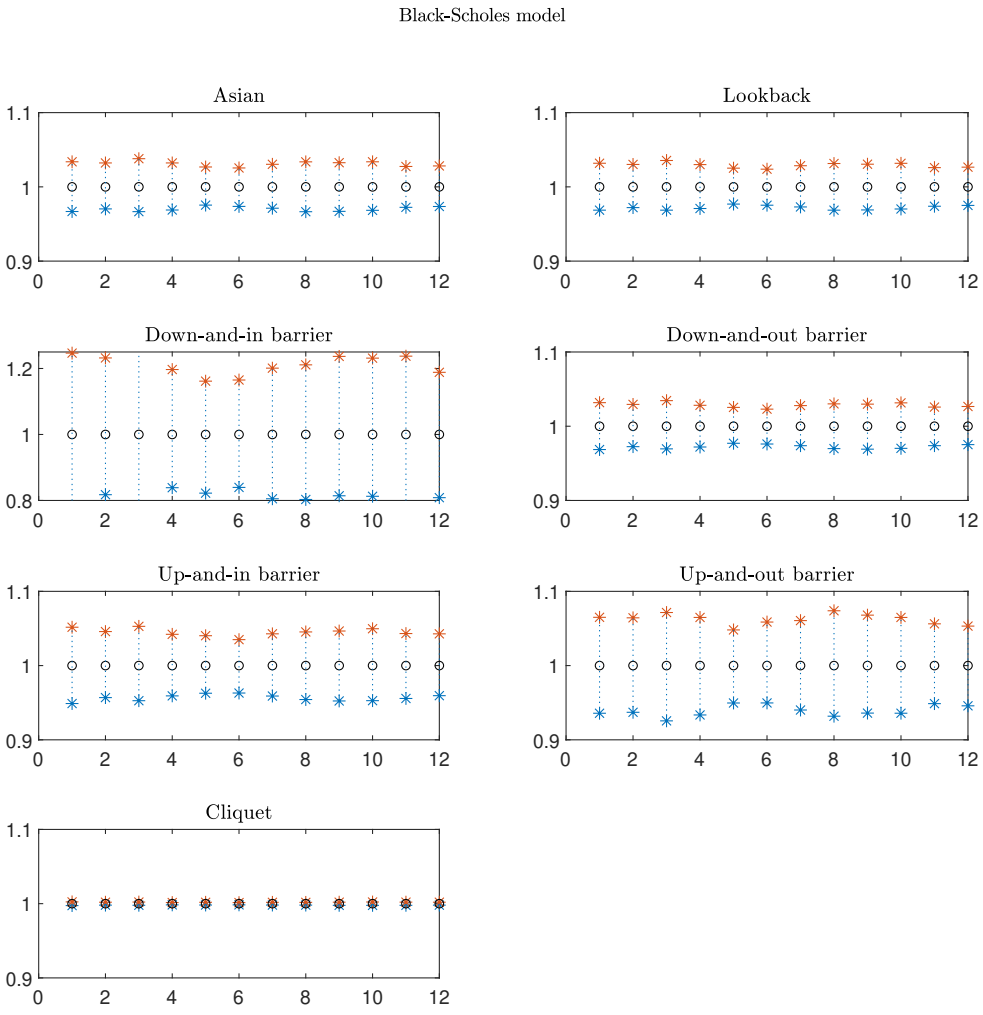

Figure D.1: *Confidence intervals: Black–Scholes model.* Normalized 90% confidence intervals for option prices in the Black–Scholes model based on calibration on each of 12 months. Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

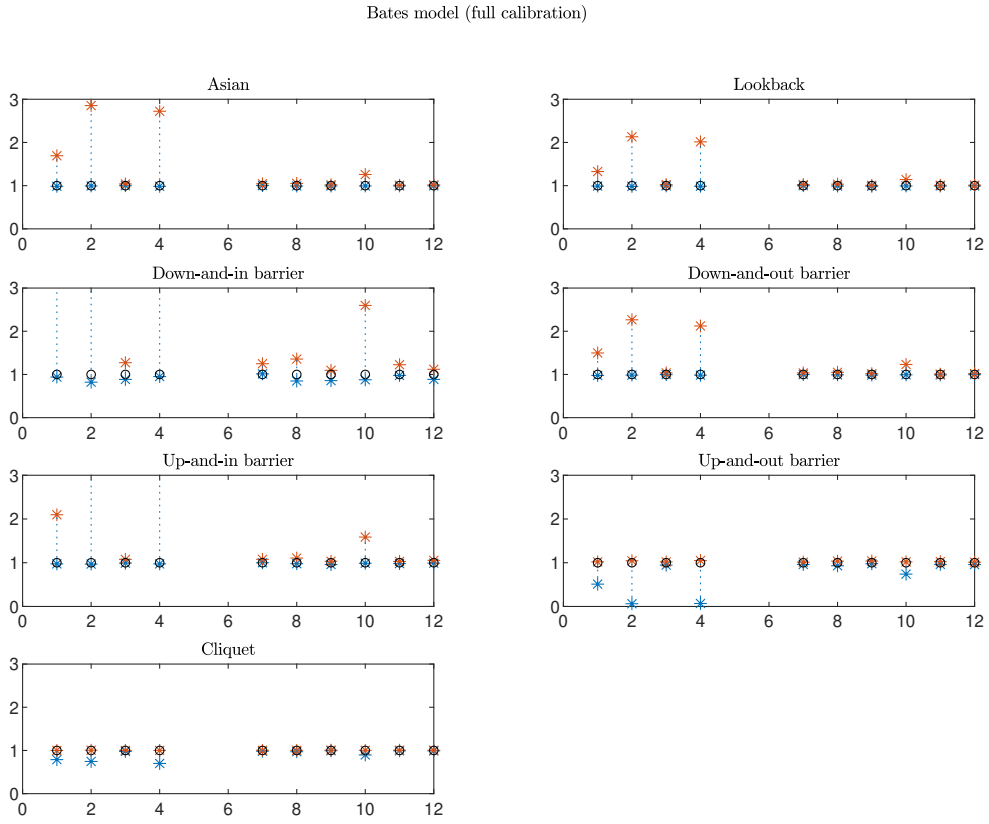

Figure D.2: *Confidence intervals: Bates model.* Normalized 90% confidence intervals for option prices in the Bates model based on full calibration on each of 12 months. Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

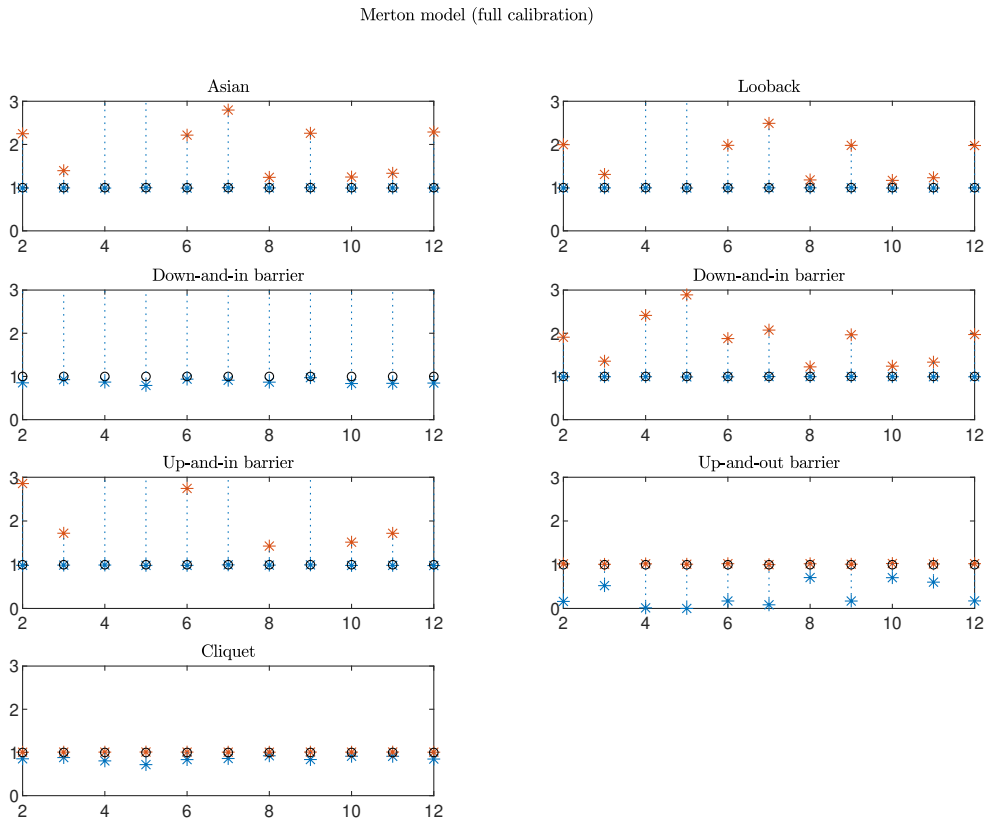

Figure D.3: *Confidence intervals: Merton model.* Normalized 90% confidence intervals for option prices in the Merton model based on full calibration on each of 12 months. Calibration is based on minimization of MSE_{IV}. Circles on the plots correspond to calibration on a particular day with whiskers extending to the lower and upper bounds of the confidence intervals.

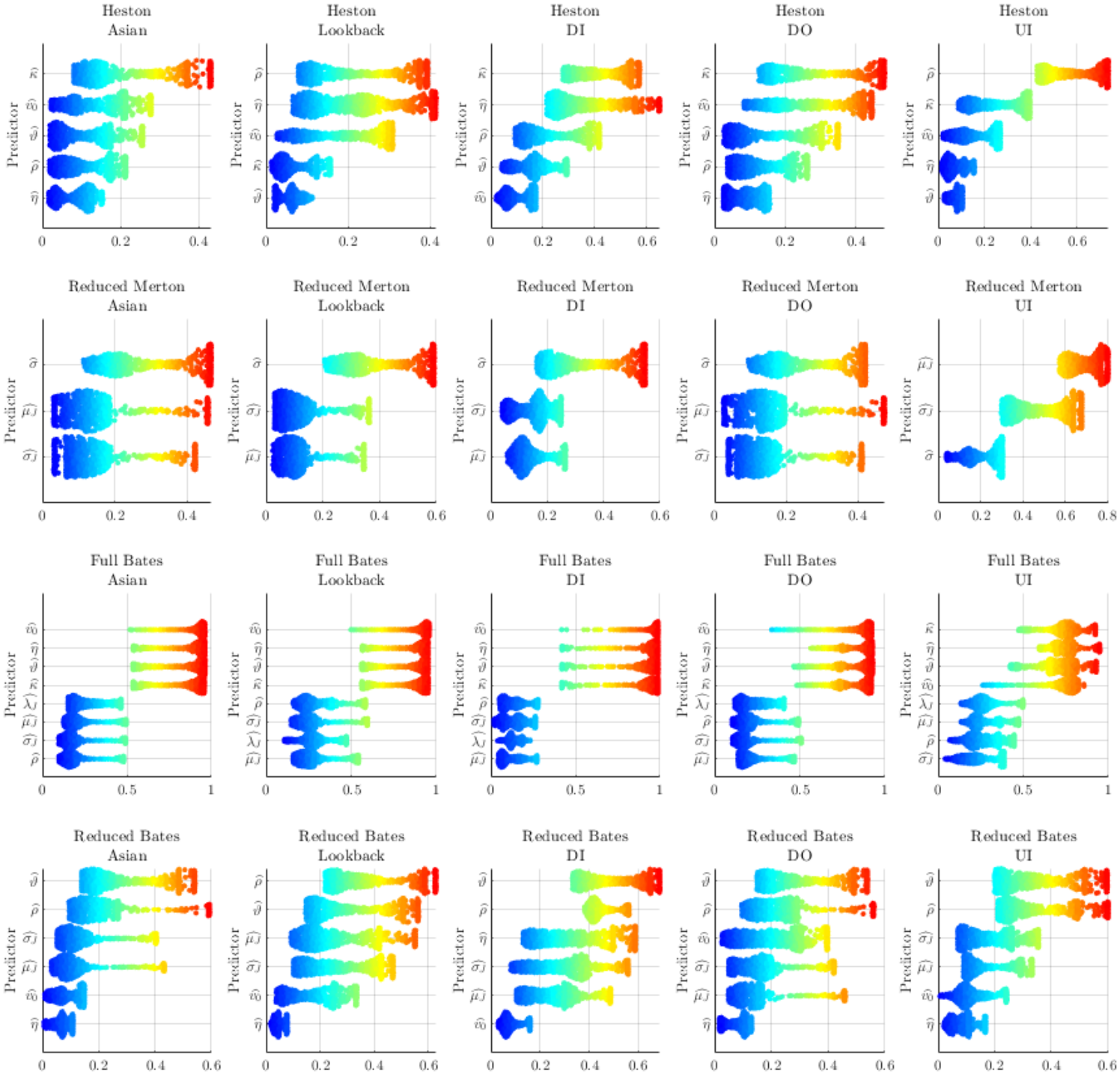

Figure D.4: *Kuiper's distance distributions.* Swarm charts visualizing the distribution of Kuiper's distance for various parameters (predictors) of the Heston, Merton, and Bates models with respect to the prices of Asian, lookback, down-and-in (DI), down-and-out (DO), and up-and-in (UI) barrier call options. We analyze fully calibrated Heston, and reduced Merton and Bates models, and also present the fully calibrated Bates model (the fully calibrated Merton model has been omitted due to poor quality).