

Feng, Yu et al.

Article

Quantifying the model risk inherent in the calibration and recalibration of option pricing models

Risks

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Feng, Yu et al. (2021) : Quantifying the model risk inherent in the calibration and recalibration of option pricing models, Risks, ISSN 2227-9091, MDPI, Basel, Vol. 9, Iss. 1, pp. 1-20,
<https://doi.org/10.3390/risks9010013>

This Version is available at:

<https://hdl.handle.net/10419/258103>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

Quantifying the Model Risk Inherent in the Calibration and Recalibration of Option Pricing Models

Yu Feng ¹, Ralph Rudd ², Christopher Baker ², Qaphela Mashalaba ², Melusi Mavuso ² and Erik Schlögl ^{1,2,3,*} 

- ¹ Quantitative Finance Research Centre, University of Technology Sydney, Broadway, NSW 2007, Australia; Yu.Feng-5@student.uts.edu.au
- ² The African Institute for Financial Markets and Risk Management (AIFMRM), University of Cape Town, Rondebosch 7701, South Africa; contact@ralphrudd.com (R.R.); chris.baker2@yahoo.co.uk (C.B.); qzmashalaba@yahoo.com (Q.M.); mvsmel001@myuct.ac.za (M.M.)
- ³ Department of Statistics, Faculty of Science, University of Johannesburg, Auckland Park 2006, South Africa
- * Correspondence: Erik.Schlogl@uts.edu.au

Abstract: We focus on two particular aspects of model risk: the inability of a chosen model to fit observed market prices at a given point in time (calibration error) and the model risk due to the recalibration of model parameters (in contradiction to the model assumptions). In this context, we use relative entropy as a pre-metric in order to quantify these two sources of model risk in a common framework, and consider the trade-offs between them when choosing a model and the frequency with which to recalibrate to the market. We illustrate this approach by applying it to the seminal Black/Scholes model and its extension to stochastic volatility, while using option data for Apple (AAPL) and Google (GOOG). We find that recalibrating a model more frequently simply shifts model risk from one type to another, without any substantial reduction of aggregate model risk. Furthermore, moving to a more complicated stochastic model is seen to be counterproductive if one requires a high degree of robustness, for example, as quantified by a 99% quantile of aggregate model risk.

Keywords: model risk; option pricing; relative entropy; model calibration; stochastic volatility



Citation: Feng, Yu, Ralph Rudd, Christopher Baker, Qaphela Mashalaba, Melusi Mavuso, and Erik Schlogl. 2021. Quantifying the Model Risk Inherent in the Calibration and Recalibration of Option Pricing Models. *Risks* 9: 13. <https://doi.org/10.3390/risks9010013>

Received: 19 October 2020
Accepted: 23 December 2020
Published: 4 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The renowned statistician George E. P. Box (see [Box and Draper \(1987\)](#)) wrote that “essentially, all models are wrong, but some are useful”. This is certainly true in finance, where many models and techniques that have been extensively empirically invalidated remain in widespread use, not just in academia, but also (perhaps especially) among practitioners. At times, the way models are used directly contradicts the model assumptions: as observed, market prices change, parameters in option pricing models, which are assumed to be time-invariant, are recalibrated, often on a daily basis. Incorrect models and model misuse represent a source of risk that is being increasingly recognised—this is called “model risk”. As a paper by the Board of Governors of the Federal Reserve System put it in 2011 (see [Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency \(2011\)](#)), “the use of models invariably presents model risk, which is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports”.

In broad terms, one could identify four general classes of model risk that are inherent to the way mathematical models are used in finance, for example, in (but not limited to) option pricing applications:

- Parameter uncertainty (and sensitivity to parameters)—let us call this “Type 0” model risk for short. If model parameters need to be statistically estimated, then they will only be known up to some level of statistical confidence, and this parameter uncertainty induces uncertainty regarding the correctness of the model outputs. Examples

of where this type of risk is considered explicitly in the literature include [Löffler \(2003\)](#), [Bannör and Scherer \(2013\)](#) and [Kerkhof et al. \(2010\)](#).

- Inability to fit a model to a full set of simultaneous market observations—this is “calibration error”, let’s call this “Type 1” model risk for short. “Calibration” in this context refers to choosing model parameters in such a way that prices for derivative financial instruments that are implied by the model match market prices at one given point in time. Thus, calibration error on single-time-point (a.k.a. “cross-sectional”) market data means that these data already contradict the model assumptions. The classical example of this is the Black/Scholes implied volatility smile.
- Change in parameters due to recalibration—let us call this “Type 2” model risk for short. Once one moves from one day to the next, this aspect of model risk becomes apparent: in order to again fit the market as closely as possible, it is common practice in the industry to recalibrate models. This recalibration results in model parameters (which the models assume to be fixed) changing from day-to-day, contradicting the model assumptions.
- The “true” dynamics of state variables do not match model dynamics—let us call this violation of model assumptions “Type 3” model risk. The classical example of this is the econometric rejection of the hypothesis that asset prices follow geometric Brownian motion, thus invalidating the key assumption in the seminal model of [Black and Scholes \(1973\)](#). This type of model risk is considered, for example, in [Kerkhof et al. \(2010\)](#), who also relate this to identification risk, which they define as risk that “arises when observationally indistinguishable models have different consequences for capital reserves”. [Boucher et al. \(2014\)](#) present a method for making value-at-risk more robust with respect to this source of model risk by “learning” from the results of model backtesting. Type 3 model risk would impact, in particular, the effectiveness of hedging strategies based on a model, for example [Detering and Packham \(2016\)](#) take the approach of measuring model risk that is based on the residual profit/loss from hedging in a misspecified model.

Note that there is a gradual transition between the different types of model risk and, depending on one’s modelling choices, to a certain extent one can trade off one type of model risk against another. For example,

- Less stringent requirements of an exact fit to market observations (Type 1) allows for less frequent recalibration (Type 2).
- Instead of different model dynamics (Type 3), one could consider a parameterised family of models (Type 2).
- Regime-switching models “legalise” changes in parameters, so Type 2 becomes more like Type 3.
- Adding parameters shifts model risk from Type 1 to Type 2 (or, to a certain extent, to Type 0).
- Adding state variables shifts model risk from Type 2 to Type 3.

[Glasserman and Xu \(2014\)](#) propose relative entropy as a consistent pre-metric by which to measure model risk from different sources. What matters in the application of mathematical models in finance is the probability distributions that the models imply, either under a “risk-neutral” probability measure (for applications to relative pricing of financial instruments) or the “physical” (a.k.a. “real-world”) probability measure (for risk management applications such as the calculation of expected shortfall). [Breuer and Csiszár \(2016\)](#) call this *distribution model risk*. Each type of model risk manifests itself as some form of ambiguity about the “true” probability measures that should be used for these purposes, and being able to quantify different types of model risk in a unified setting while using a pre-metric for the divergence between distributions (like relative entropy) allows for one to make an informed choice about the trade-offs between different sources of model risk. [Glasserman and Xu \(2014\)](#) postulate a “relative entropy budget” defining a set of models that are sufficiently close (in the sense of relative entropy) to a nominal reference model to be considered in an evaluation of model risk expressed as a “worst case” expectation—i.e.,

a worst-case price or a worst-case risk measure. However, they say little as to how one typically would obtain a specific number for this “relative entropy budget”. In a sense, we invert this problem by noting that higher relative entropy between model distributions indicates higher model risk, and propose a method to jointly evaluate model risk of two types, based on how this model risk manifests itself when option pricing models are calibrated and recalibrated to liquid market instruments.

We focus on the model risk that is inherent in the calibration and recalibration (i.e., in the above terminology, Types 1 and 2) of option pricing models, and to illustrate our approach we consider the models of [Black and Scholes \(1973\)](#) and [Heston \(1993\)](#), thus comparing the most classical option pricing model with its popular extension incorporating stochastic volatility. Clearly, if (as is often the case in practice) one solely focuses on calibration error, more frequent recalibration will always be preferred to less, and [Heston \(1993\)](#) preferred to [Black and Scholes \(1973\)](#). The latter point is due to the fact that the Heston model has more parameters than Black/Scholes (four as opposed to one) and an additional, unobserved state variable (the volatility), thus providing more freedom to fit option prices across different strikes.

We quantify calibration and recalibration risk in both models applied to equity option data, and also explore the trade-off between these two types of model risk, finding that there is no longer a trivial answer to the question which model and recalibration frequency should be preferred when these two sources of model risk are considered in a unified framework.

The rest of the paper is organised, as follows. Section 2 introduces a framework for the joint evaluation of model risk due to calibration error and model recalibration. Section 3 discusses the numerical implementation of the method. Section 4 presents the results that were obtained by applying this method to option price data, and Section 5 concludes.

2. Quantifying Model Risk by Relative Entropy

As noted above, model risk is reflected in the ambiguity with regard to the “correct” probability distribution to use for relative pricing or risk assessment. Following [Glasserman and Xu \(2014\)](#), we quantify this ambiguity while using the divergence between probability measures. In the present context, these can be classified as divergence measures that are defined as a function $D(\cdot||\cdot) : S \times S \rightarrow \mathbb{R}$, satisfying

$$D(Q||P) \geq 0 \quad \forall P, Q \in S \quad (1)$$

$$D(Q||P) = 0 \Leftrightarrow P = Q, \quad (2)$$

where S is a space of all probability measures with a common support. More specifically, most divergence measures belong to the class of f -divergence (see e.g., [Ali and Silvey \(1966\)](#); [Csiszár \(1967\)](#) or [Ahmadi-Javid \(2012\)](#)), which gives the divergence between two equivalent measures as:

$$D(Q||P) = \int_{\Omega} f\left(\frac{dQ}{dP}\right) dP, \quad (3)$$

where f is a convex function of the Radon–Nikodym derivative satisfying $f(1) = 0$. Kullback–Leibler divergence (a.k.a. relative entropy) is the most common f -divergence, which assigns $f(m) = m \ln m$. It is noted that the methodology of this paper applies to all types of statistical distances in principle, although, in the empirical study, the Kullback–Leibler divergence is adopted due to its simplicity and widespread use.

2.1. Quantifying Calibration Error

If we wish to quantify calibration error (Type 1 model risk) in this fashion, then, in Equations (1)–(3), the probability measure P corresponds to the calibrated model and thus is parametric in some form. The probability measure Q , on the other hand, serves as a reference measure exactly matching observed market option prices at a given point in time, unrestricted by the assumptions of the model under consideration. Note that, if

we only consider market prices for European options, then these only constrain the risk-neutral distributions Q for a finite number of (discrete) time horizons and, therefore, the requirement (because we are using f -divergence) that Q and P are equivalent measures is not a material restriction—for example, for a fixed time horizon T , the risk-neutral distribution of the underlying asset price implied by a [Black and Scholes \(1973\)](#) model is a probability measure equivalent to the risk-neutral distribution implied by a [Heston \(1993\)](#) stochastic volatility model, because both distributions have a strictly positive density on the same support.

On calibrating an option pricing model, we may regard the measure Q as some non-parametric risk-neutral measure that fits market option prices. However, in practice, the measure Q is not unique, as the market is usually incomplete. For example, based on the result of [Breedon and Litzenberger \(1978\)](#), in order to obtain Q for a single time horizon completely non-parametrically, one would need option prices for a continuum of strikes, which is clearly never available. Therefore, we define the space of all risk-neutral probability measures that fit market prices by $S_Q \subset S$.

We may further define the space of probability measures that are given by all possible choices of parameter values for the target model by $S_P \subset S$. The new calibration methodology that is proposed here aims to minimise the calibration error, as quantified by the divergence between the two measures P and Q , taken from their respective spaces, i.e.,

$$(Q^*, P^*) = \arg \min_{Q \in S_Q, P \in S_P} D(Q||P). \quad (4)$$

This is to say, the new approach attempts to calibrate a model measure P^* (i.e., a set of model parameters θ^*) and non-parametric perfect fit to the market (at a given point in time) Q^* , in a fashion that minimises the calibration error that is expressed by

$$\eta_1 = D(Q^*||P^*). \quad (5)$$

This is not an end in itself—it is required in order to compare model risk due to the calibration error and model risk due to recalibration (as specified in Section 2.2, below) in a unified framework. Here, it is worth reiterating the point that was made in the introduction, that using relative entropy to quantify model risk due to these two sources is motivated by the role of a “relative entropy budget”, as introduced by [Glasserman and Xu \(2014\)](#) in determining worst-case prices or risk measures—i.e., the greater the relative entropy budget, the worse the worst case outcome, with all other things being equal. Furthermore, choosing relative entropy as the divergence pre-metric means that solving (4) corresponds to what [Buchen and Kelly \(1996\)](#) call the “Principle of Minimum Cross-Entropy” for fitting a risk-neutral distribution to market option prices, and given P^* , Q^* can be obtained in closed form—this is exploited in the numerical implementation in Section 3, below.

The classical approaches of model calibration, such as minimising the mean-squared error between model and market prices for options, would be inappropriate in this context, as they would lead to unnecessarily high model risk quantities. It is the choice of divergence measure that informs the calibration procedure, which results in a pair of probability measures, (Q^*, P^*) , one of which corresponds to the calibrated model while the other provides a consistent reference measure fitting the market exactly.

2.2. Including Model Risk Due to Recalibration

In order to quantify the model risk due to recalibration, let us consider the more specific case where the model is Markovian in a vector of observable state variables X , the model is characterised by a vector of model parameters θ , and market prices are given for European option prices of a single maturity T . This last assumption of a single maturity T avoids the need to constrain the choice of Q in order to ensure the absence of calendar spread arbitrage between non-parametric risk-neutral measures for different time horizons—parametric models typically ensure this by construction. If we appropriately constrain Q , this assumption could be lifted.

Thus, a “model” in our context is a mapping from parameters θ and state variables X at a current time (call it t_i) to a risk-neutral distribution of an underlying asset (or assets) at time horizon T . The two models that are considered below, [Black and Scholes \(1973\)](#) and [Heston \(1993\)](#), both provide such a mapping. Of course, these models also provide more than that, in particular specifying the dynamics of their respective state variables, but that aspect would fall under “Type 3” model risk not considered here.

Suppose that we solved (4) yesterday (at time t_{i-1}) to obtain a P^* —to be as explicit as possible, denote this by

$$P_{t_{i-1}, X_{t_{i-1}}(\omega), \theta_{t_{i-1}}^*}(A) \quad \forall A \in \mathcal{F}_T. \quad (6)$$

i.e., this is a (conditional) probability measure that is defined on all \mathcal{F}_T -measurable events, where the conditioning is on the state variables at time t_{i-1} , $X_{t_{i-1}}$, and we write $X_{t_{i-1}}(\omega)$ in order to express that the time t_{i-1} realisations of the state variables are known at the time that these probabilities are evaluated. We write the subscript $\theta_{t_{i-1}}^*$ to express that these probabilities are evaluated in a model with parameters that are calibrated by solving (4) at time t_{i-1} . Furthermore, denote the non-parametric measure Q^* that results from solving (4) at time t_{i-1} by $Q_{t_{i-1}}$.

Now, if we recalibrate today (at time t_i) by solving (4), we obtain Q_{t_i} and

$$P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}(A) \quad \forall A \in \mathcal{F}_T. \quad (7)$$

We can then define the model risk quantity due to recalibration, as

$$\eta_2 = D(P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*} || P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}), \quad (8)$$

which is the divergence between the (conditional) probability measures that were evaluated at time t_i , where one measure is based on the recalibrated parameters $\theta_{t_i}^*$ and the other is based on the previously calibrated parameters $\theta_{t_{i-1}}^*$ (thus, expressing, in terms of divergence, the inconsistency with the model assumptions due to the fact that we are going “outside of the model” to change parameters in recalibration). The aggregate of calibration error and model risk due to recalibration is then

$$\eta_3 = D(Q_{t_i} || P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}), \quad (9)$$

i.e., the divergence between the non-parametric probability measure Q_{t_i} obtained by solving (4) at time t_i , and the non-recalibrated parametric probability measure, consisting of probabilities that are conditional on the state at time t_i , but based on model parameters that were obtained by solving (4) at time t_{i-1} . However, this approach minimises the divergence between the reference distribution and the recalibrated distribution, thus arguably overstating the divergence to the non-recalibrated (i.e. model-consistent) distribution and, therefore, overstating the aggregate model risk η_3 .

Alternatively, we may choose as the non-parametric reference distribution at time t_i

$$\hat{Q}_{t_i} = \arg \min_{Q \in S_Q} D(Q || P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}), \quad (10)$$

resulting in a lower aggregate model risk of

$$\eta_3 = D(\hat{Q}_{t_i} || P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}). \quad (11)$$

Note that $\theta_{t_i}^*$ is still obtained by solving (4), because both Q_{t_i} and \hat{Q}_{t_i} represent non-parametric probability measures that fit observed market prices exactly, so $\theta_{t_i}^*$ remains the best available parametric fit to the market at time t_i (\hat{Q}_{t_i} is only used to determine minimum divergence of the non-recalibrated model to a measure giving a perfect fit).

In the heuristic schematic shown in Figure 1a, point A represents $P_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}$, being the parametric probability measure “closest” to the set of non-parametric probability measures fitting the market exactly, where point C represents Q_{t_i} . If we do not recalibrate at time t_i , then we end up with the parametric probability measure $P_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}$ (point B), to which \hat{Q}_{t_i} (point D) is the “closest” non-parametric probability measure fitting the market exactly. Note that these graphs are for the purpose of heuristic illustration only—in particular, we are not requiring that the two sets of probability measures are convex.

In the case of Kullback–Leibler divergence, if Type 1 (calibration error) and 2 (recalibration) model risk involve independent Radon–Nikodym derivatives, then, in the first case considered above, the aggregate model risk equals the sum of the two components. In fact, the Radon–Nikodym derivatives, as random variables, take the key role in evaluating the two types of model risk. At the time the model is recalibrated, we again consider the optimisation (4), with S_Q now changed to reflect the change in observed market prices, so we have the following Radon–Nikodym derivatives:

$$\text{for calibration error: } m_1 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}} \tag{12}$$

$$\text{for model risk due to recalibration: } m_2 = \frac{dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}} \tag{13}$$

$$\text{for aggregate model risk: } m_1 m_2 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}} \tag{14}$$

Abbreviating $dP_{t_i, X_{t_i}(\omega), \theta_{t_{i-1}}^*}$ as P and $dP_{t_i, X_{t_i}(\omega), \theta_{t_i}^*}$ as P^* , the aggregate risk can be expressed in terms of m_1 and m_2 , as

$$\eta_3 = E^P [m_1 m_2 \ln(m_1 m_2)] \tag{15}$$

$$= E^{P^*} [m_1 \ln(m_1)] + E^P [m_1 m_2 \ln(m_2)] \tag{16}$$

$$= \eta_1 + [1 - \text{cov}^P(m_1, m_2)] \eta_2 + \text{cov}^P [m_1, m_2 \ln(m_2)]. \tag{17}$$

If m_1 and m_2 are independent, then $\text{cov}^P(m_2, m_1) = \text{cov}^P [m_2, m_1 \ln(m_1)] = 0$. The total model risk is equal to the sum of the calibration risk and recalibration risk. In our empirical exploration in Section 4, below, we found that this equality is followed quite well by the Black–Scholes model. However, it typically does not hold in the Heston model, suggesting substantial dependence (of Radon–Nikodym derivatives) between the calibration error and model risk due to recalibration. This already foreshadows the clear trade-off between calibration error and model risk, due to recalibration in the model that is better able to calibrate to option prices across different strikes (i.e., the Heston model) due to its larger number of parameters. If this greater freedom to fit the market reduces calibration error, but effectively only pushes the model risk from Type 1 to Type 2, this dependence between m_1 and m_2 is to be expected.

2.3. The Treatment of Latent State Variables

A model may also involve one or more latent state variables. An example of that is the class of stochastic volatility models where the volatility is taken as a latent state variable, rather than a model parameter (in the empirical examples in Section 4 below, we specifically consider the model of Heston (1993), which falls into this category). Under the framework of a single stochastic volatility state variable, a model that is specified by a given set of parameters forms a one-dimensional manifold (Figure 1b) for possible realisations of the state variable, rather than a point in the Black/Scholes world (Figure 1a).

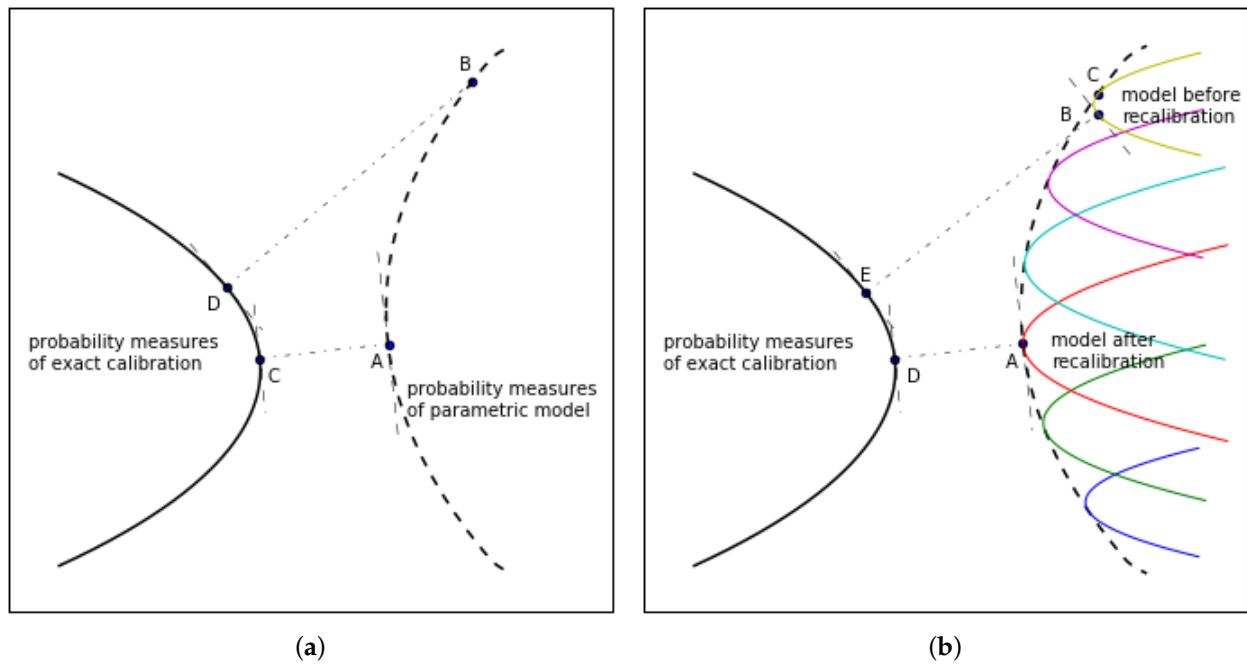


Figure 1. Graphic illustration of the mathematical definitions of model risks for (a) observable state variable, (b) latent state variable.

Thus, the model that we are now considering is Markovian in a vector of state variables (X, V) , where the state variables X are observable and the state variables V are latent (unobservable). Subsequently, the initial calibration problem (4) becomes

$$(Q^*, v^*, \theta^*) = \arg \min_{Q \in S_Q, v \in S_v, \theta \in S_\theta} D(Q || P(v; \theta)), \tag{18}$$

where S_v and S_θ are the sets of legitimate values of the state variables and the parameters, respectively. θ^* is the set of model parameters calibrated to the market, and v^* is the best estimate of the latent state variables under the calibrated model. This effectively treats the latent (unobserved) state variable as an additional parameter to be calibrated, but the recalibration of which does not contribute to (Type 2) model risk, due to recalibration, because it is consistent with the model assumptions for this latent state variable to evolve stochastically. This does shift Type 2 model risk to Type 3, the risk that the state variable dynamics are not (econometrically) consistent with the dynamics that are assumed in the model. However, in the present paper, we deliberately set aside Type 3 model risk for the purposes of our analysis, leaving the integration of all four types of model risk for future research.

The notation in (6) is amended to

$$P_{t_{i-1}, X_{t_{i-1}}(\omega), v_{t_{i-1}}^*, \theta_{t_{i-1}}^*}(A) \quad \forall A \in \mathcal{F}_T. \tag{19}$$

At time t_i , we have, for the calibration error,

$$\eta_1 = D(Q_{t_i} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}). \tag{20}$$

The model risk due to recalibration is

$$\eta_2 = D(P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_{i-1}}^*}). \tag{21}$$

The aggregate model risk while using Q_{t_i} from Q^* and $v_{t_i}^*$ from v^* in (18) is

$$\eta_3 = D(Q_{t_i} || P_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i-1}^*}) \quad (22)$$

or, alternatively, using \hat{Q}_{t_i} and \hat{v}_{t_i} determined analogously to (10), i.e.,

$$(\hat{Q}_{t_i}, \hat{v}_{t_i}) = \arg \min_{Q \in S_Q, v \in S_v} D(Q || P_{t_i, X_{t_i}(\omega), v, \theta_{t_i-1}^*}), \quad (23)$$

which results in

$$\hat{\eta}_3 = D(\hat{Q}_{t_i} || P_{t_i, X_{t_i}(\omega), \hat{v}_{t_i}, \theta_{t_i-1}^*}). \quad (24)$$

We then have the following Radon–Nikodym derivatives:

$$\text{for calibration error:} \quad m_1 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}} \quad (25)$$

$$\text{for model risk due to recalibration:} \quad m_2 = \frac{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i}^*}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i-1}^*}} \quad (26)$$

$$\text{for aggregate model risk:} \quad m_1 m_2 = \frac{dQ_{t_i}}{dP_{t_i, X_{t_i}(\omega), v_{t_i}^*, \theta_{t_i-1}^*}} \quad (27)$$

Note that the key difference between (12)–(14) and (25)–(27) is that the change in v , being permitted by the model assumptions, does not contribute to the model risk quantities. In (4) and (18), we are deliberately prioritising the minimisation of calibration error, as this is congruent to the (often exclusive) focus of practitioners on calibration error (with little or no regard to model risk due to recalibration). If desired, one could reformulate this approach in order to prioritise the minimisation of aggregate model risk or of model risk due to recalibration.

3. Numerical Implementation

In this section, we outline the numerical scheme for solving the minimisation problems that arise when taking into account calibration error and model risk, due to recalibration in the manner that is described in the previous section, including problems of the type (4) involving the optimal choice of two probability measures. In this case, an iterative process is required, optimising two probability measures Q and P , in turn, until convergence, in the following manner:

- (1) Produce $P^{(0)}$ from a parametric model that is based on an initial guess of the model parameters (and latent state variables, where required).
- (2) Solve, for $Q^{(0)}$, via Lagrange multipliers for the constrained problem that minimises $D(Q^{(0)} || P^{(0)})$.
- (3) Solve, for $P^{(1)}$, to obtain model parameters for the $P^{(1)}$ that minimises $D(Q^{(0)} || P^{(1)})$.
- (4) Iterate steps 2 and 3: $P^{(n)} \rightarrow Q^{(n)} \rightarrow P^{(n+1)}$ until convergence.

3.1. Step 1

In Step 1, the initial guess may be obtained in several different ways. A common way is to minimise the mean-squared error between model and market option prices at all available strikes (see e.g., Cui et al. (2017)). We opted for the Broyden/Fletcher/Goldfarb/Shanno (BFGS) algorithm for conducting this initial calibration of the model parameters and (where required) latent state variables. In practice, the particular choice of the initial guess in Step 1 appears to have little impact in our implementation, as Step 4 converged quite quickly.

3.2. Step 2

In Step 2, we solve the following constrained minimisation problem while using Lagrange multipliers:

$$Q^{(0)} = \arg \min_{Q \in \mathcal{S}_Q} D(Q||P^{(0)}) \tag{28}$$

$$s.t. \mathbf{B} \leq E^Q(\mathbf{Z}) \leq \mathbf{A}. \tag{29}$$

Note that, here, we specify the constraints in the form of expectations under the measure Q , where these expectation are the model prices for our calibration instruments for the model that is based on the non-parametric reference distribution Q . In general, \mathbf{B} , \mathbf{Z} , and \mathbf{A} are vectors; thus, (29) is a “stack” of inequality constraints representing observed market prices. Additionally, notice that, for generality, we “relax” each equality constraint into two inequality constraints. This is in order to account for the bid-ask spread of each option traded on the market. The vector \mathbf{B} denotes a list of bid prices, while the vector \mathbf{A} contains ask prices. In a simplified scenario, where exact option prices are given, we may set $\mathbf{B} = \mathbf{A}$. \mathbf{Z} denotes the vector of discounted option payoffs. By introducing vectors of Lagrange multipliers λ_B and λ_A , we convert the constrained problem to an unconstrained dual problem,

$$\max_{\lambda_B \geq 0, \lambda_A \leq 0} \min_{Q \in \mathcal{S}_Q} D(Q||P^{(0)}) - \lambda_B [E^Q(\mathbf{Z}) - \mathbf{B}] - \lambda_A [E^Q(\mathbf{Z}) - \mathbf{A}]. \tag{30}$$

In the case of Kullback–Leibler divergence, solving the inner problem corresponds to the “Principle of Minimum Cross-Entropy” in [Buchen and Kelly \(1996\)](#). By Equation (11) in that paper, the probability density function $q^{(0)}$ of $Q^{(0)}$ is given in terms of the density $p^{(0)}$ of $P^{(0)}$, by

$$q^{(0)} = \frac{p^{(0)} e^{\lambda \mathbf{Z}}}{E^{P^{(0)}}(e^{\lambda \mathbf{Z}})}. \tag{31}$$

Note that

$$D(Q^{(0)}||P^{(0)}) = \int_{\Omega} \frac{dQ^{(0)}}{dP^{(0)}} \ln \frac{dQ^{(0)}}{dP^{(0)}} dP^{(0)} = \int_{\Omega} \ln \frac{dQ^{(0)}}{dP^{(0)}} dQ^{(0)} = E^{Q^{(0)}} \left[\ln \frac{dQ^{(0)}}{dP^{(0)}} \right] \tag{32}$$

and (31) implies

$$\frac{dQ^{(0)}}{dP^{(0)}} = \frac{q^{(0)}}{p^{(0)}} = \frac{e^{\lambda \mathbf{Z}}}{E^{P^{(0)}}(e^{\lambda \mathbf{Z}})}. \tag{33}$$

Substituting (31) into (30), we obtain a maximisation problem,

$$\max_{\lambda_B \geq 0, \lambda_A \leq 0} E^{Q^{(0)}} \left[(\lambda_B + \lambda_A) \mathbf{Z} - \ln E^{P^{(0)}} \left[e^{(\lambda_B + \lambda_A) \mathbf{Z}} \right] \right] - \lambda_B [E^Q(\mathbf{Z}) - \mathbf{B}] - \lambda_A [E^Q(\mathbf{Z}) - \mathbf{A}] \tag{34}$$

$$= \max_{\lambda_B \geq 0, \lambda_A \leq 0} - \ln E^{P^{(0)}} \left[e^{(\lambda_B + \lambda_A) \mathbf{Z}} \right] + \lambda_B \mathbf{B} + \lambda_A \mathbf{A} \tag{35}$$

$$= \max_{\lambda_B \geq 0, \lambda_A \leq 0} - \ln E^{P^{(0)}} \left(e^{(\lambda_B + \lambda_A) \mathbf{Z}} \right) + (\lambda_B + \lambda_A) \frac{\mathbf{B} + \mathbf{A}}{2} - (\lambda_B - \lambda_A) \frac{\mathbf{A} - \mathbf{B}}{2}. \tag{36}$$

If $\mathbf{B} = \mathbf{A}$, then the last term vanishes, representing the problem with exact market prices. If (component-wise) $\mathbf{B} > \mathbf{A}$, then the last term reflects a penalty on the objective function that is proportional to the difference of the two Lagrange multipliers. Therefore, we may transform the Lagrange multipliers by

$$\lambda_+ = \lambda_B + \lambda_A \quad (37)$$

$$\lambda_- = \lambda_B - \lambda_A \quad (38)$$

and the objective function becomes

$$\max_{\lambda_+ \in \mathbb{R}^d, \lambda_- \geq |\lambda_+|} -\ln E^{P^{(0)}}(e^{\lambda_+ Z}) + \lambda_+ \frac{B+A}{2} - \lambda_- \frac{A-B}{2}. \quad (39)$$

Because λ_- only appears in the final, negative term, maximising (39) implies that λ_- must be as small as possible, so we can set $\lambda_- = |\lambda_+|$, and the objective function becomes

$$\max_{\lambda_+ \in \mathbb{R}^d} -\ln E^{P^{(0)}}(e^{\lambda_+ Z}) + \lambda_+ \frac{B+A}{2} - |\lambda_+| \frac{A-B}{2}. \quad (40)$$

We can numerically solve the maximization problem by taking its gradient with respect to λ_+ ,

$$E^{P^{(0)}}(Z e^{\lambda_+ Z}) = \left[\frac{B+A}{2} - \text{sgn}(\lambda_+) \frac{A-B}{2} \right] E^{P^{(0)}}(e^{\lambda_+ Z}), \quad (41)$$

where the element-wise sign function $\text{sgn}(\lambda_+)$ assigns 1, -1 , or 0 to each element of λ_+ . However, due to the discontinuity of the sign, function (41) cannot be solved directly in a stable way. In order to bypass this problem, we approximate the sign function with a continuous step function:

$$\text{sgn}(x) = 1 - \frac{2}{1 + \exp\left(\frac{x}{\delta}\right)}. \quad (42)$$

We use Powell's hybrid method in order to solve the multidimensional Equations (41), where δ controls the steepness of the function and choosing this value is critical for a fast and stable convergence of the method, i.e., with a careful choice of δ , Step 2 converges quickly.

3.3. Steps 3 and 4

In Step 3, we use the L-BFGS-B algorithm to minimise the divergence with respect to model parameters (or latent variables or both). This step corresponds to the quite standard process of calibrating a parametric option pricing model, and it does not result in any difficulties in obtaining convergence. Step 2 and Step 3 are repeated until convergence. The convergence criterion that is adopted here is that all of the percentage changes of parameters after one iteration do not exceed a certain threshold, say 0.1%. On this "outer" convergence, in practice we did not encounter any difficulties either, which may be due to the convexity of the sets constraining P and Q .

4. Examining the Trade-Off between Calibration Error and Model Risk Due to Recalibration

As an application example of the method that is described in the previous two sections, we consider historical data that consist of daily market prices for call options on Apple (AAPL) and Google (GOOG) stock over a period from 6 January 2004 to 19 December 2008 for AAPL and 4 January 2005 to 19 December 2008 for GOOG. This gives us a reasonably straightforward application example that is free of extraneous complications, while still reasonably covering liquid options and including a period of "interesting" market volatility (August 2007): although these options are of the American type, i.e., permitting early exercise, AAPL and GOOG did not pay any dividends during this period. Thus, the possibility of early exercise may be ignored (see Merton (1973)).

From these data, we remove options that are very far away from the money, restricting the range of strikes from delta 2.5% to delta 97.5%. Furthermore, we remove prices of options that had zero trading volume on a given day, in order to avoid using prices that are likely to be stale.

On these data, we consider two parametric models, [Black and Scholes \(1973\)](#) and [Heston \(1993\)](#)—arguably the two most popular option pricing models available, where the latter introduces a latent variable for stochastic volatility. The unified methodology, quantifying calibration error, model risk due to recalibration, and the aggregate of the two allows for us to explore the trade-off between calibration error (which is, unsurprisingly, reduced by moving from [Black and Scholes \(1973\)](#) to [Heston \(1993\)](#)) and model risk due to recalibration (which has hitherto been largely ignored) when moving from one parametric model to another, as well as when changing the frequency with which the model is recalibrated.

We start by evaluating the calibration, recalibration, and aggregated model risk quantities under a Black/Scholes model, i.e., where the underlying asset price is assumed to follow geometric Brownian motion, with dynamics under the risk-neutral measure that is given by

$$dS(t) = rS(t)dt + \sigma S(t)dW(t), \quad (43)$$

where r is the continuously compounded risk-free rate of interest and σ is a constant volatility parameter. In [Appendix A](#), we note that, in the Black/Scholes model, we obtain a simple closed form expression for the recalibration model risk quantity that is defined in [\(8\)](#):

$$\eta_2 = \left(\frac{\sigma^{*2}}{\sigma^2} - 1 \right) \left[\frac{1}{2} + \frac{T}{8}(\sigma^{*2} - \sigma^2) \right] + \ln\left(\frac{\sigma}{\sigma^*}\right), \quad (44)$$

where σ^* is the correctly recalibrated Black/Scholes volatility parameter and σ is the parameter value that was obtained in a previous calibration. This formula is a consequence of the log-normal distribution of returns assumed in the Black/Scholes model.

We can express the aggregate model risk as the sum of the calibration error, the recalibration risk, and a residual. The residual is zero if the likelihood ratios that are involved in the calibration and recalibration risks are two independent random variables, as noted in [Equation \(15\)](#). In practice, the residual usually takes a small non-zero value. In [Figure 2](#), we demonstrate the decomposition of the total model risk into the three components. Unsurprisingly (as it is well documented that the Black/Scholes model cannot fit the implied volatility “smile” observed in most options markets), we see that calibration error typically predominates.

In the Heston model, the dynamics [\(43\)](#) are extended in order to allow for stochastic volatility, i.e.,

$$dS(t) = rS(t)dt + \sigma(t)S(t)dW_1(t) \quad (45)$$

$$d(\sigma^2(t)) = \kappa(\theta - \sigma^2(t))dt + \eta\sigma(t)dW_2(t). \quad (46)$$

This model involves two state variables, the underlying asset price $S(t)$ and the volatility $\sigma(t)$, and five model parameters: r , κ , θ , η , and ρ , where ρ is the correlation coefficient between the two Wiener processes:

$$d\langle W_1, W_2 \rangle_t = \rho dt.$$

r is the risk-free rate, and κ , θ and η relate to the volatility process, with κ being the rate of mean reversion, θ the long-run mean, and η the volatility of this process.

In our empirical application examples, we take the risk-free rate as one of the financial variables that were observed in the market, but we do not explicitly take the interest rate risk in our empirical analysis into account. For the short-dated options that were considered here, interest rate risk is known to be of relatively little importance—for a discussion of this issue, see e.g., [Cheng et al. \(2018\)](#) and the literature cited therein.

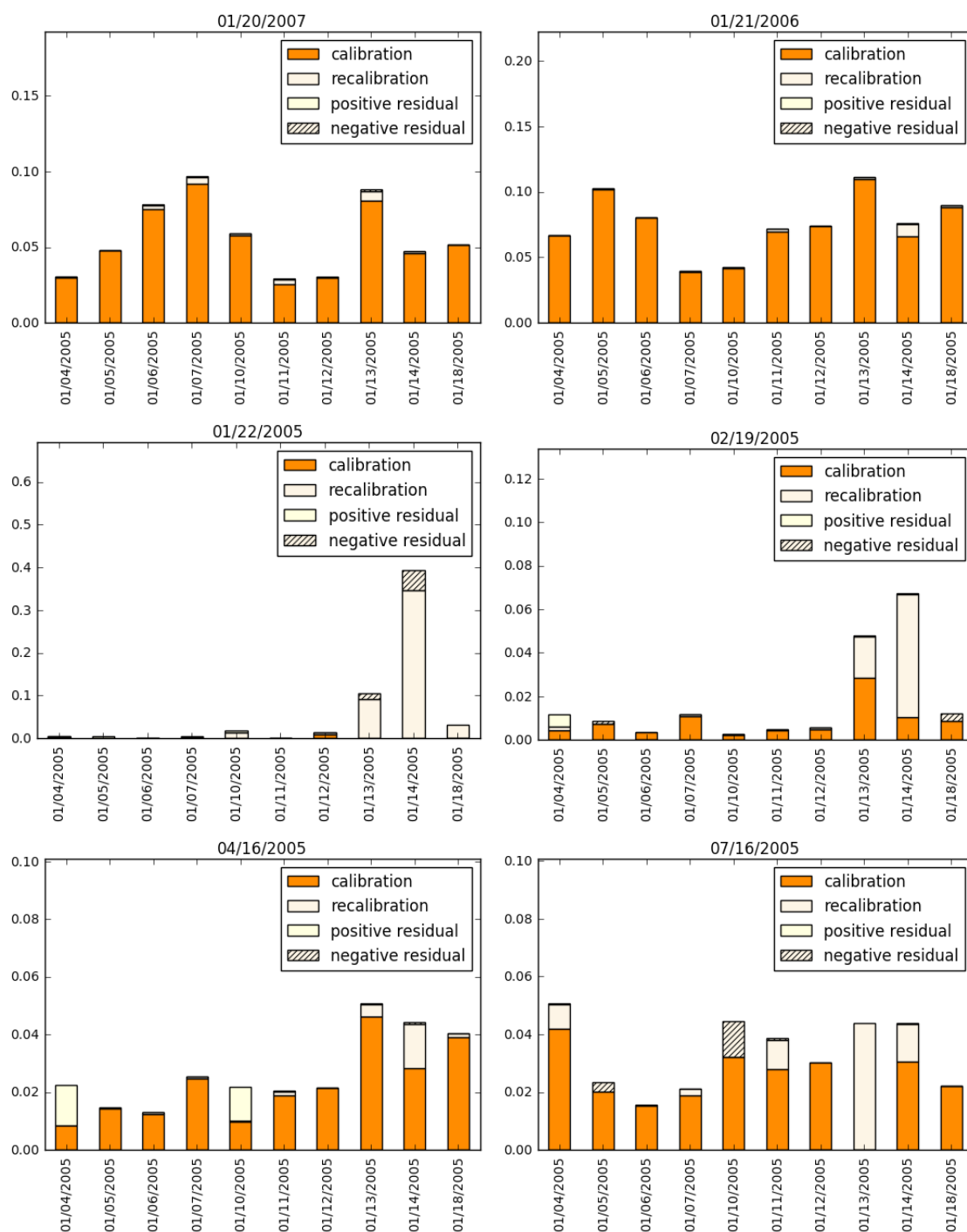


Figure 2. Decomposition of model risks of Apple (AAPL) options while using the Black/Scholes model as the nominal model, for a selection of option maturity dates (given as the title of each graph). The vertical axis denotes the numerical value of the relative entropy.

Following Gatheral (2006), Equation (2.13), the risk-neutral probability of exercise of a European call option with strike K in the Heston model is given by

$$P_0(x, v, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{\exp\{C(u, \tau)\theta + D(u, \tau)v + iux\}}{iu} \right] du, \quad (47)$$

where v is the current value of the volatility state variable $\sigma(t)$, $\tau = T - t$ is the time to maturity, and x is the logarithmic forward moneyness of the option, i.e.,

$$x : = \ln\left(\frac{F_{t,T}}{K}\right)$$

$$F_{t,T} = \frac{S(t)}{B(t,T)}$$

with $B(t, T)$ the time t price of a zero bond maturing in T . Furthermore,

$$C(u, \tau) = \kappa \left[r_- \tau - \frac{2}{\eta^2} \ln\left(\frac{r_+ - r_- e^{-d\tau}}{r_+ - r_-}\right) \right] \tag{48}$$

$$D(u, \tau) = r_- \frac{r_+ - r_+ e^{-d\tau}}{r_+ - r_- e^{-d\tau}} \tag{49}$$

$$r_{\pm} = \frac{\beta \pm d}{\eta^2} \quad d = \sqrt{\beta^2 - 2\alpha\eta^2}. \tag{50}$$

Parameters α and β are functions of u (Fourier transform variable of x):

$$\alpha(u) = -\frac{u^2}{2} - i\frac{u}{2} \tag{51}$$

$$\beta(u) = \kappa - \rho\eta iu. \tag{52}$$

It is noted that $P_0 = E^{\mathbb{Q}}(\mathbb{1}_{S_T > K})$, since, by definition, P_0 is the probability of exercise. Therefore, the probability density function of the risk-neutral measure is obtained:

$$p(S_T) = -\frac{\partial P_0}{\partial K} \Big|_{K=S_T} = -\frac{\partial P_0}{\partial x} \frac{\partial x}{\partial K} \Big|_{K=S_T} \tag{53}$$

$$= \frac{1}{2\pi S_T} \int_{-\infty}^{\infty} \Re \left[\exp \left[C(u, \tau)\theta + D(u, \tau)v + iu \ln\left(\frac{S_t}{B(t, T)S_T}\right) \right] \right] du. \tag{54}$$

For simplicity, e^y denotes the ratio of the forward price at t to its spot price at maturity T , i.e., $e^y = B(t, T)S_T/S_t$, we derive the risk-neutral probability with respect to y :

$$p(y) = p(S_T) \left| \frac{\partial y}{\partial S_T} \right|^{-1} \tag{55}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Re[\exp[C(u, \tau)\theta + D(u, \tau)v - iuy]] du. \tag{56}$$

$p(y)$ can be calculated by fast Fourier transform (FFT). We use this in each step of our iterative procedure, where a risk-neutral probability measure P as a function of the model parameters and state variables is required for the Heston model, i.e., when fitting an initial guess of the model parameters to market option prices in Step 1, when using $P^{(0)}$ in Step 2, when solving for $P^{(1)}$ in Step 3, and so on.

The decomposition of the total model risk into the three components (the components due to calibration and recalibration, and the positive or negative residual measuring the departure from independence between the first two components) when using the Heston model as the baseline is given in Figure 3. Again, because it is well documented that the Heston model can fit observed option prices better than Black/Scholes, it is unsurprising that, in this case, the relative entropy measuring calibration error is much lower—however, already in this set of example days, it is evident that this comes at the price of increased model risk due to recalibration.

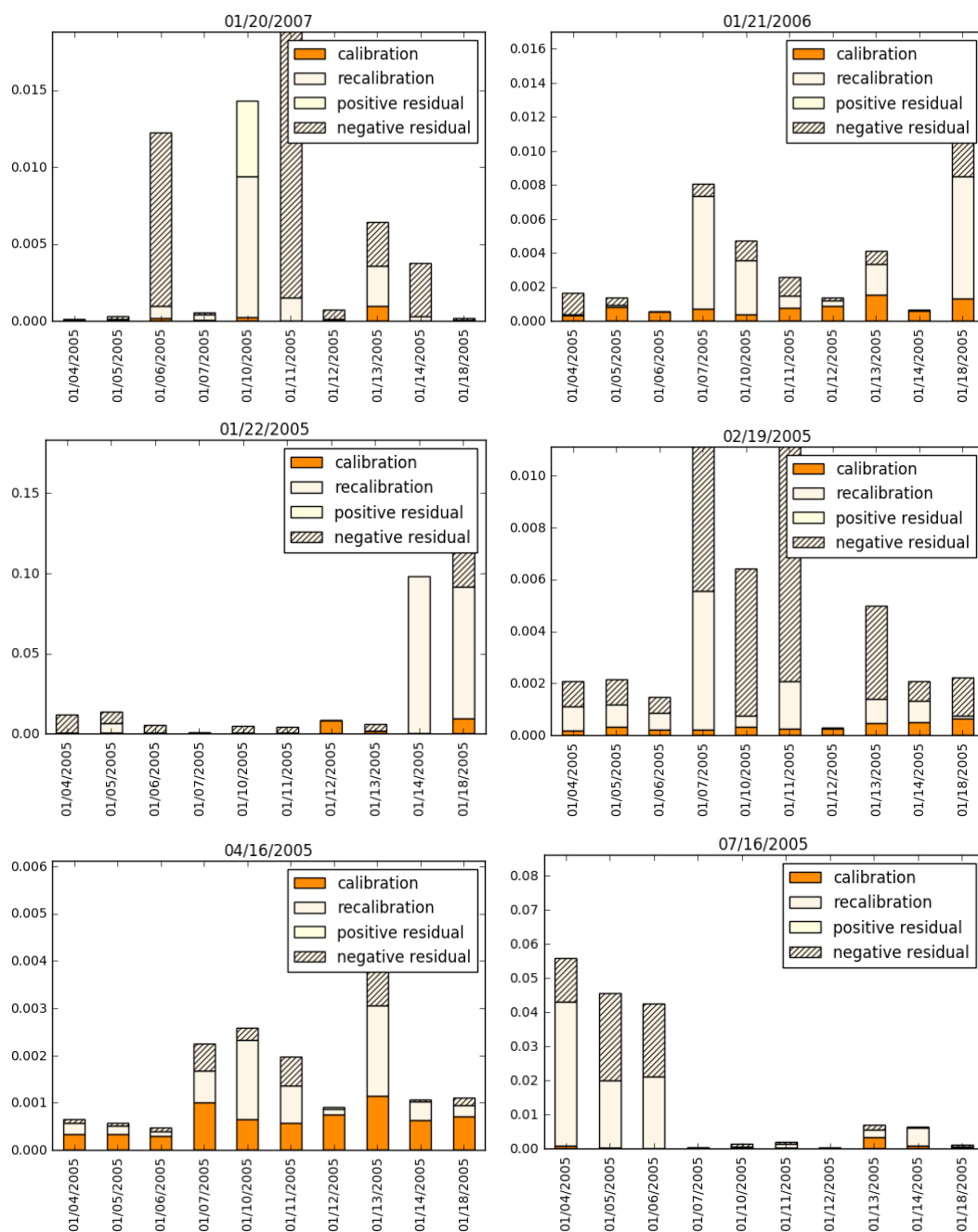


Figure 3. Decomposition of model risks of Google (GOOG) options while using the Heston model as the nominal model, for a selection of option maturity dates (given as the title of each graph). The vertical axis denotes the numerical value of the relative entropy.

These observations are reinforced when we consider aggregate model risk, calibration error, and model risk due to recalibration over the entire sample period, as presented in Tables 1 and 2. Note that the absolute numbers refer to relative entropy and, thus, lack direct financial interpretation—what matters are the relative values when comparing the model across models and different recalibration frequencies, in particular when considering the aggregate model risk. Here, we consider recalibrating the Black/Scholes and Heston models either daily, every three days, every week, every two weeks, or every quarter year. Thus, for each model and each recalibration frequency, we get a (for example, daily) sequence of model risk quantities for each of the three types (calibration, recalibration, and aggregate). In each column, we report summary

statistics of such a sequence, i.e., its mean, median, and the 99%, 95%, 90%, and 75% quantiles. Note that on days on which we do not recalibrate, the model risk, due to recalibration being zero, because (consistent with the model assumptions) we are keeping previously calibrated parameters unchanged—so, on those days, aggregate model risk is entirely due to calibration error (which increases because the fit to market prices deteriorates when we do not recalibrate).

We see that recalibrating more frequently has little effect on the aggregate model risk, neither when using the Black/Scholes model nor when using the Heston model. Essentially, recalibrating more frequently simply shifts calibration error into model risk, due to recalibration, highlighting the dangers in the common practice of focusing solely on the calibration of derivative pricing models, at the expense of all other sources of model risk.

In addition, we observe that if we are interested in “robustness” at a high level of confidence (looking at, say, the 99% quantile of aggregate model risk), moving from Black/Scholes to Heston also does not appear to deliver any advantage (it does yield some improvement at lower quantiles, or average or median, aggregate model risk). This means that, when high levels of confidence are required, any gain in calibration accuracy that is delivered by the Heston model is offset by higher model risk due to recalibration. One should note that this last point holds, even before considering Type 3 model risk, which may well be worse when additional state variables are introduced (as in the Heston model). For these results, in Tables 1 and 2, we calculated the means, medians, and quantiles across all available option maturities. If we only consider particular maturity “buckets” (i.e., considering only subsets of options, with maturities in a particular range), the same qualitative conclusions are evident—Tables 3 and 4 illustrate this in the case of daily recalibration.

Table 1. Model risks (in terms of relative entropy) under different models and recalibration frequencies for AAPL.

Risk Measure	Black-Scholes					Heston					
	1 Day	3 Days	1 Week	2 Weeks	1 Quarter	1 Day	3 Days	1 Week	2 Weeks	1 Quarter	
Aggregate Model Risk											
mean	0.070	0.071	0.073	0.073	0.085	0.037	0.035	0.038	0.039	0.046	
median	0.045	0.047	0.047	0.051	0.057	0.005	0.005	0.005	0.006	0.010	
Quantile	99%	0.474	0.427	0.471	0.455	0.462	0.508	0.462	0.512	0.503	0.649
	95%	0.212	0.221	0.221	0.212	0.251	0.173	0.169	0.177	0.177	0.185
	90%	0.158	0.160	0.165	0.160	0.192	0.096	0.097	0.098	0.105	0.121
	75%	0.092	0.094	0.095	0.099	0.112	0.027	0.027	0.029	0.031	0.034
Calibration Error											
mean	0.055	0.066	0.069	0.072	0.084	0.008	0.026	0.032	0.036	0.045	
median	0.038	0.043	0.045	0.049	0.057	0.001	0.003	0.004	0.005	0.009	
Quantile	99%	0.239	0.397	0.433	0.443	0.455	0.163	0.416	0.496	0.495	0.648
	95%	0.171	0.207	0.212	0.210	0.249	0.015	0.130	0.158	0.168	0.182
	90%	0.135	0.151	0.158	0.158	0.191	0.008	0.065	0.083	0.097	0.119
	75%	0.075	0.087	0.091	0.096	0.111	0.003	0.012	0.019	0.026	0.034
Model Risk due to Recalibration											
mean	0.024	0.009	0.005	0.002	0.001	0.057	0.019	0.011	0.006	0.001	
median	0.002	0.001	0.001	0.000	0.000	0.004	0.001	0.001	0.001	0.000	
Quantile	99%	0.458	0.196	0.057	0.030	0.012	0.630	0.226	0.117	0.067	0.011
	95%	0.103	0.034	0.021	0.010	0.005	0.315	0.108	0.062	0.032	0.006
	90%	0.056	0.020	0.011	0.006	0.002	0.164	0.052	0.031	0.016	0.004
	75%	0.013	0.006	0.004	0.002	0.001	0.055	0.016	0.011	0.005	0.001

Table 2. Model risks (in terms of relative entropy) under different models and recalibration frequencies for GOOG.

Risk Measure	Black-Scholes					Heston					
	1 Day	3 Days	1 Week	2 Weeks	1 Quarter	1 Day	3 Days	1 Week	2 Weeks	1 Quarter	
Aggregate Model Risk											
mean	0.165	0.168	0.165	0.165	0.188	0.115	0.105	0.113	0.114	0.127	
median	0.109	0.111	0.112	0.115	0.138	0.053	0.050	0.055	0.053	0.059	
Quantile	99%	0.705	0.722	0.722	0.699	0.728	0.726	0.668	0.711	0.699	0.744
	95%	0.519	0.530	0.521	0.496	0.549	0.475	0.438	0.455	0.467	0.560
	90%	0.393	0.391	0.387	0.385	0.408	0.330	0.287	0.320	0.324	0.380
	75%	0.219	0.223	0.214	0.215	0.256	0.145	0.137	0.142	0.153	0.157
Calibration Error											
mean	0.125	0.155	0.158	0.162	0.187	0.055	0.083	0.103	0.107	0.126	
median	0.082	0.102	0.106	0.111	0.137	0.006	0.022	0.043	0.046	0.058	
Quantile	99%	0.655	0.701	0.709	0.696	0.728	0.626	0.651	0.706	0.688	0.744
	95%	0.400	0.482	0.507	0.494	0.548	0.329	0.381	0.438	0.445	0.557
	90%	0.295	0.359	0.362	0.379	0.408	0.172	0.250	0.298	0.306	0.375
	75%	0.171	0.208	0.207	0.212	0.255	0.034	0.103	0.127	0.141	0.155
Model Risk due to Recalibration											
mean	0.036	0.014	0.008	0.004	0.001	0.119	0.044	0.012	0.012	0.003	
median	0.004	0.002	0.001	0.001	0.000	0.043	0.016	0.003	0.003	0.001	
Quantile	99%	0.405	0.172	0.091	0.050	0.010	0.757	0.255	0.087	0.077	0.012
	95%	0.173	0.066	0.035	0.018	0.004	0.543	0.187	0.056	0.064	0.011
	90%	0.109	0.038	0.022	0.012	0.003	0.411	0.151	0.034	0.045	0.009
	75%	0.036	0.013	0.008	0.004	0.001	0.142	0.052	0.016	0.013	0.003

Table 3. Model risks (in terms of relative entropy) under different models, by maturity buckets, for AAPL (daily recalibration frequency).

Risk Measure	Black-Scholes				Heston				
	All	0–0.2 Year	0.2–0.7 Year	>0.7 Year	All	0–0.2 Year	0.2–0.7 Year	>0.7 Year	
Aggregate Model Risk									
mean	0.070	0.041	0.066	0.109	0.037	0.039	0.024	0.046	
median	0.045	0.021	0.052	0.081	0.005	0.004	0.004	0.008	
Quantile	99%	0.474	0.471	0.266	0.567	0.508	0.590	0.302	0.472
	95%	0.212	0.120	0.185	0.282	0.173	0.243	0.091	0.191
	90%	0.158	0.081	0.143	0.213	0.096	0.074	0.066	0.142
	75%	0.092	0.047	0.090	0.141	0.027	0.021	0.015	0.054
Calibration Error									
mean	0.055	0.026	0.057	0.087	0.008	0.012	0.004	0.007	
median	0.038	0.014	0.048	0.068	0.001	0.001	0.001	0.001	
Quantile	99%	0.239	0.157	0.230	0.289	0.163	0.072	0.052	0.116
	95%	0.171	0.091	0.157	0.212	0.015	0.021	0.011	0.017
	90%	0.135	0.062	0.126	0.181	0.008	0.006	0.007	0.010
	75%	0.075	0.036	0.077	0.128	0.003	0.002	0.003	0.003
Model Risk due to Recalibration									
mean	0.024	0.020	0.013	0.038	0.057	0.062	0.048	0.060	
median	0.002	0.002	0.001	0.003	0.004	0.004	0.003	0.009	
Quantile	99%	0.458	0.506	0.116	0.662	0.630	0.702	0.536	0.512
	95%	0.103	0.066	0.066	0.169	0.315	0.443	0.274	0.247
	90%	0.056	0.030	0.042	0.102	0.164	0.223	0.127	0.160
	75%	0.013	0.009	0.010	0.031	0.055	0.030	0.048	0.091

Table 4. Model risks (in terms of relative entropy) under different models, by maturity buckets, for GOOG (daily recalibration frequency).

Risk Measure	Black-Scholes				Heston				
	All	0–0.2 Year	0.2–0.7 Year	>0.7 Year	All	0–0.2 Year	0.2–0.7 Year	>0.7 Year	
Aggregate Model Risk									
mean	0.165	0.167	0.162	0.164	0.115	0.115	0.120	0.109	
median	0.109	0.104	0.099	0.120	0.053	0.052	0.057	0.048	
Quantile	99%	0.705	0.736	0.709	0.661	0.726	0.714	0.707	0.736
	95%	0.519	0.527	0.551	0.459	0.475	0.468	0.500	0.476
	90%	0.393	0.417	0.382	0.371	0.330	0.351	0.322	0.304
	75%	0.219	0.221	0.216	0.218	0.145	0.149	0.155	0.136
Calibration Error									
mean	0.125	0.121	0.123	0.132	0.055	0.055	0.052	0.057	
median	0.082	0.074	0.075	0.099	0.006	0.005	0.006	0.006	
Quantile	99%	0.655	0.662	0.650	0.639	0.626	0.658	0.531	0.665
	95%	0.400	0.403	0.423	0.391	0.329	0.349	0.311	0.330
	90%	0.295	0.298	0.301	0.285	0.172	0.155	0.185	0.171
	75%	0.171	0.162	0.160	0.176	0.034	0.031	0.034	0.034
Model Risk due to Recalibration									
mean	0.036	0.038	0.039	0.033	0.119	0.117	0.117	0.124	
median	0.004	0.005	0.003	0.005	0.043	0.043	0.037	0.049	
Quantile	99%	0.405	0.419	0.479	0.284	0.757	0.746	0.767	0.754
	95%	0.173	0.176	0.188	0.155	0.543	0.510	0.558	0.554
	90%	0.109	0.111	0.115	0.098	0.411	0.390	0.418	0.418
	75%	0.036	0.038	0.032	0.036	0.142	0.144	0.137	0.146

5. Conclusions

Under our approach, less relative entropy implies less model risk, and we are able to evaluate two hitherto disparate sources of model risk (calibration error and model risk due to recalibration) in a unified fashion, and examine the potential trade-off between the two. We have considered a simple choice between two models, and between different recalibration frequencies. “Putting a number on model risk” by calculating quantiles for the maximum model risk (quantified by relative entropy) over a time series of market data allows for one to assess the added value (if any) of more complicated stochastic models of financial markets.

One should note that, instead of using a relative entropy pre-metric, one could approach quantifying model risk in terms of optimal-transport distance, while using for example Wasserstein distance, which most recently has become popular for this purpose (see [Bartl et al. \(2020\)](#); [Blanchet et al. \(2018\)](#) and [Feng and Schlögl \(2018\)](#)). In the present paper, we have followed the more established approach while using relative entropy, which, in economics, has its roots in the seminal work of Hansen and Sargent (see e.g., [Hansen and Sargent \(2006\)](#)), as well as a long history of application in other fields, such as thermodynamics and information theory.

In our application, we are deliberately prioritising the minimisation of calibration error, as this is congruent to the (often exclusive) focus of practitioners on calibration error (with little or no regard to model risk due to recalibration). Even in this case, we see that, by including recalibration as one of the sources of aggregate model risk, frequently recalibrating a model to a changing market simply interchanges one source of model risk for another, and more complicated stochastic models may well underperform when aggregate model risk is taken into account.

Author Contributions: Conceptualisation, Y.F. and E.S.; Formal analysis, Y.F.; Investigation, Y.F., R.R., C.B., Q.M. and M.M.; Methodology, Y.F. and E.S.; Resources, E.S.; Software, Y.F.; Supervision, E.S.; Visualisation, Y.F.; Writing-original draft, Y.F. and E.S.; Writing-review & editing, E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data was obtained from IVolatility and are available via https://www.ivolatility.com/data/historical_data2.html.

Acknowledgments: Part of the initial research for this paper was conducted at the Financial Mathematics Team Challenge 2016 at the University of Cape Town. We would like to thank Sam Cohen and Hardy Hulley for helpful comments on earlier drafts of this paper. The usual disclaimers apply.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds ... are available from the authors.

Appendix A. Derivation of the Recalibration Model Risk Quantity in the Black/Scholes Model

Denote by $g_\sigma(S_T|S_0)$ the risk-neutral probability density of the stock price at time T given the stock price at time 0, in a Black and Scholes (1973) model with volatility σ . This is given by

$$g_\sigma(S_T|S_0) = \frac{1}{S_T\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^2T)^2}{2\sigma^2T}\right\}. \quad (\text{A1})$$

Furthermore, denote by σ^* the correctly recalibrated Black/Scholes volatility parameter and by σ the parameter value obtained in a previous calibration. The model risk quantity due to recalibration defined in (8) then becomes

$$\eta_2 = \int_0^\infty \frac{g_{\sigma^*}(S_T|S_0)}{g_\sigma(S_T|S_0)} \left(\ln \frac{g_{\sigma^*}(S_T|S_0)}{g_\sigma(S_T|S_0)} \right) g_\sigma(S_T|S_0) dS_T \quad (\text{A2})$$

$$= \int_0^\infty \left(\ln \frac{g_{\sigma^*}(S_T|S_0)}{g_\sigma(S_T|S_0)} \right) g_{\sigma^*}(S_T|S_0) dS_T. \quad (\text{A3})$$

Therefore,

$$\eta_2 = \int_0^\infty \left(\frac{(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^2T)^2}{2\sigma^2T} - \frac{(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^{*2}T)^2}{2\sigma^{*2}T} + \ln(S_T\sigma\sqrt{2\pi}) - \ln(S_T\sigma^*\sqrt{2\pi}) \right) g_{\sigma^*}(S_T|S_0) dS_T. \quad (\text{A4})$$

Consider the different parts (A4) in turn:

$$\int_0^\infty (\ln(S_T\sigma\sqrt{2\pi}) - \ln(S_T\sigma^*\sqrt{2\pi})) g_{\sigma^*}(S_T|S_0) dS_T = \ln \frac{\sigma}{\sigma^*} \quad (\text{A5})$$

$$\int_0^\infty \frac{(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^{*2}T)^2}{2\sigma^{*2}T} g_{\sigma^*}(S_T|S_0) dS_T = \frac{\text{Var}[\ln S_T]}{2\sigma^{*2}T} = \frac{1}{2} \quad (\text{A6})$$

and

$$\int_0^{\infty} \frac{(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^2 T)^2}{2\sigma^2 T} g_{\sigma^*}(S_T|S_0) dS_T \quad (\text{A7})$$

$$= \frac{1}{2\sigma^2 T} \int_0^{\infty} \left(\left(\ln S_T - \ln S_0 - rT + \frac{1}{2}\sigma^{*2} T \right)^2 + 2(\ln S_T - \ln S_0 - rT) \left(-\frac{1}{2}\sigma^{*2} T \right) - 2(\ln S_T - \ln S_0 - rT) \left(-\frac{1}{2}\sigma^2 T \right) - \frac{1}{4}\sigma^{*4} T^2 + \frac{1}{4}\sigma^4 T^2 \right) g_{\sigma^*}(S_T|S_0) dS_T \quad (\text{A8})$$

$$= \frac{1}{2\sigma^2 T} \left(\sigma^{*2} T - \sigma^2 T \left(-\frac{1}{2}\sigma^{*2} T \right) + \sigma^2 T \left(-\frac{1}{2}\sigma^{*2} T \right) - \frac{1}{4}\sigma^{*4} T^2 + \frac{1}{4}\sigma^4 T^2 \right) \quad (\text{A9})$$

$$= \frac{\sigma^{*2}}{2\sigma^2} + \frac{T}{8} \left(\frac{\sigma^{*4}}{\sigma^2} - 2\sigma^{*2} + \sigma^2 \right). \quad (\text{A10})$$

Combining (A5), (A6) and (A10) gives

$$\eta_2 = \left(\frac{\sigma^{*2}}{\sigma^2} - 1 \right) \left[\frac{1}{2} + \frac{T}{8} (\sigma^{*2} - \sigma^2) \right] + \ln \left(\frac{\sigma}{\sigma^*} \right). \quad (\text{A11})$$

References

- Ahmadi-Javid, Amir. 2012. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications* 155: 1105–23. [\[CrossRef\]](#)
- Ali, Syed Mumtaz, and Samuel D. Silvey. 1966. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B* 28: 131–42. [\[CrossRef\]](#)
- Bannör, Karl F., and Matthias Scherer. 2013. Capturing parameter risk with convex risk measures. *European Actuarial Journal* 3: 97–132. [\[CrossRef\]](#)
- Bartl, Daniel, Samuel Drapeau, and Ludovic Tangpi. 2020. Computational aspects of robust optimized certainty equivalents and option pricing. *Mathematical Finance* 30: 287–309. [\[CrossRef\]](#)
- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–54. [\[CrossRef\]](#)
- Blanchet, Jose, Lin Chen, and Xun Yu Zhou. 2018. Distributionally robust mean–variance portfolio selection with Wasserstein distances. *arXiv arXiv:1802.04885*.
- Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency. 2011. *Supervisory Guidance on Model Risk Management*. Technical Report OCC 2011-12 Attachment. Washington D.C.: Office of the Comptroller of the Currency.
- Boucher, Christophe M., Jón Danielsson, Patrick S. Kouontchou, and Bertrand B. Maillet. 2014. Risk models—at-risk. *Journal of Banking & Finance* 44: 72–92.
- Box, George E. P., and Norman R. Draper. 1987. *Empirical Model-Building and Response Surfaces*. Hoboken: Wiley.
- Breeden, Douglas T., and Robert H. Litzenberger. 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business* 51: 621–51. [\[CrossRef\]](#)
- Breuer, Thomas, and Imre Csiszár. 2016. Measuring distribution model risk. *Mathematical Finance* 26: 395–411. [\[CrossRef\]](#)
- Buchen, Peter W., and Michael Kelly. 1996. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis* 31: 143–59. [\[CrossRef\]](#)
- Cheng, Benjamin, Christina Sklibosios Nikitopoulos, and Erik Schlögl. 2018. Pricing of long-dated commodity derivatives: Do stochastic interest rates matter? *Journal of Banking & Finance* 95: 148–66.
- Csiszár, Imre. 1967. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2: 299–318.
- Cui, Yiran, Sebastian del Baño Rollin, and Guido Germano. 2017. Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research* 263: 625–38. [\[CrossRef\]](#)
- Detering, Nils, and Natalie Packham. 2016. Model risk of contingent claims. *Quantitative Finance* 16: 1357–74. [\[CrossRef\]](#)
- Feng, Yu, and Erik Schlögl. 2018. Model risk measurement under Wasserstein distance. *arXiv arXiv:1809.03641*. [\[CrossRef\]](#)
- Gatheral, Jim. 2006. *The Volatility Surface: A Practitioner's Guide*. Hoboken: John Wiley & Sons.
- Glasserman, Paul, and Xingbo Xu. 2014. Robust risk measurement and model risk. *Quantitative Finance* 14: 29–58. [\[CrossRef\]](#)
- Hansen, Lars Peter, and Thomas J. Sargent. 2006. *Robustness*. Princeton: Princeton University Press.
- Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6: 327–43. [\[CrossRef\]](#)
- Kerkhof, Jeroen, Bertrand Melenberg, and Hans Schumacher. 2010. Model risk and capital reserves. *Journal of Banking & Finance* 34: 267–79.

-
- Löffler, Gunter. 2003. The effects of estimation error on measures of portfolio credit risk. *Journal of Banking & Finance* 27: 1427–53.
- Merton, Robert C. 1973. Theory of rational option pricing. *The Bell Journal of Economics and Management Science* 4: 141–83. [[CrossRef](#)]