

QUANTIFYING THE MODEL RISK INHERENT IN THE CALIBRATION AND RECALIBRATION OF OPTION PRICING MODELS

YU FENG, RALPH RUDD, CHRISTOPHER BAKER, QAPHELA MASHALABA, MELUSI MAVUSO,
AND ERIK SCHLÖGL

ABSTRACT. The renowned statistician George E. P. Box wrote that “essentially, all models are wrong, but some are useful.”¹ 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.”

In this paper, 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 recalibration of model parameters (in contradiction to the model assumptions). In this context, we follow the approach of Glasserman and Xu (2014) to use relative entropy as a consistent 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 applied to the models of Black and Scholes (1973) and Heston (1993), using option data for Apple (AAPL) and Google (GOOG).

1. INTRODUCTION

As a paper by the Board of Governors of the Federal Reserve System put it in 2011,² “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.” However, there has been surprisingly little research to date on quantifying this risk, or putting the analysis of this risk on a more rigorous footing. One notable exception is a paper by Glasserman and Xu (2014), who propose relative entropy as a consistent pre-metric by which to measure model risk from different sources. This allows one to optimise the choice of a model and the choice of the way a model is implemented in day-to-day practice under a unifying criterion, while also quantifying the remaining model risk.

In broad terms, one could identify four broad classes of model risk 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’s call this “Type 0” model risk for short. If model parameters need to be statistically estimated, they will only be known up to some level of statistical confidence, and this parameter uncertainty induces uncertainty about the correctness of the model outputs.³
- 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. To the extent that a model cannot match observed prices on a given day, single-day (a.k.a. “cross-sectional”) market data already contradicts the model assumptions. The classical example of this is the Black/Scholes implied volatility smile.

Date. This version: July 27, 2018. Incomplete draft — comments welcome.

¹See Box and Draper (1987).

²See Board of Governors of the Federal Reserve System Office of the Comptroller of the Currency (2011).

³One (relatively rare) example of where this type of risk is considered explicitly in the literature is Löffler (2003).

- Change in parameters due to recalibration — let’s 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 don’t match model dynamics — let’s 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).

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 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.

What matters in the application of mathematical models in finance is the probability distributions which 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). Each type of model risk manifests itself as some form of ambiguity about the “true” probability measures which should be used for these purposes, and being able to quantify different types of model risk in a unified setting using a pre-metric for the divergence between distributions (like relative entropy) allows one to make an informed choice about the trade-offs between different sources of model risk.

In the present paper, we focus on the model risk 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 focuses solely on calibration error, Heston (1993) will always be preferred to Black and Scholes (1973), and more frequent recalibration preferred to less. 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 which 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...

2. CALIBRATION ERROR, MODEL RISK DUE TO RECALIBRATION, AND TREATMENT OF LATENT STATE VARIABLES

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 using the divergence between probability measures. In the present context, these can be classified as divergence measures defined as a function

$D(\cdot||\cdot) : S \times S \rightarrow \mathbb{R}$ satisfying

- (1) $D(Q||P) \geq 0 \forall P, Q \in S$
- (2) $D(Q||P) = 0 \Leftrightarrow P = Q$

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, which gives the divergence between two equivalent measures as:⁴

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

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, though in the empirical study the Kullback–Leibler divergence is adopted due to its simplicity and widespread use.

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 prices at a given point in time, unrestricted by the assumptions of the model under consideration. On calibrating an option pricing model, we may regard the measure Q as some non-parametric risk–neutral measure that explains the market in full assuming absence of arbitrage. In practice, however, the measure Q is not unique as the market is usually incomplete. We therefore define the space of all probability measures that explains the market in full by $S_Q \subset S$.

We may further define the space of probability measures given by all possible choices of parameter values for the target model by $S_P \subset S$. The new calibration methodology 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.

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

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 which minimises the calibration error expressed by

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

This is not an end in itself — it is required in order to compare model risk due to calibration error and model risk due to recalibration (as specified below) in a unified framework.

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, which informs the calibration procedure, resulting 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.

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 .⁵ Suppose we solved (4) yesterday (at time t_{i-1}) to obtain a P^* — to be as

⁴See e.g. Ali and Silvey (1966), Csiszár (1967) or Ahmadi-Javid (2012).

⁵This last assumption of a single maturity T avoids the need to constrain the choice of Q 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 can be lifted.

explicit as possible, denote this by

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

I.e., this is a (conditional) probability measure defined on all \mathcal{F}_T -measurable events, where the conditioning is on the state variables at time t_{i-1} , $X_{t_{i-1}}$, where we write $X_{t_{i-1}}(\omega)$ 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 calibrated by solving (4) at time t_{i-1} . Furthermore, denote the non-parametric measure Q^* resulting 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

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

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

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

which is the divergence between the (conditional) probability measures 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

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

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 conditional on the state at time t_i , but based on model parameters 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 :

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

resulting in a lower aggregate model risk of

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

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 fitting 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 of Figure 1(a), 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 , 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.

In the case of Kullback–Leibler divergence, note that if Type 1 (calibration error) and 2 (recalibration) model risk involve independent Radon–Nikodym derivatives, then, in the first case considered above, 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),

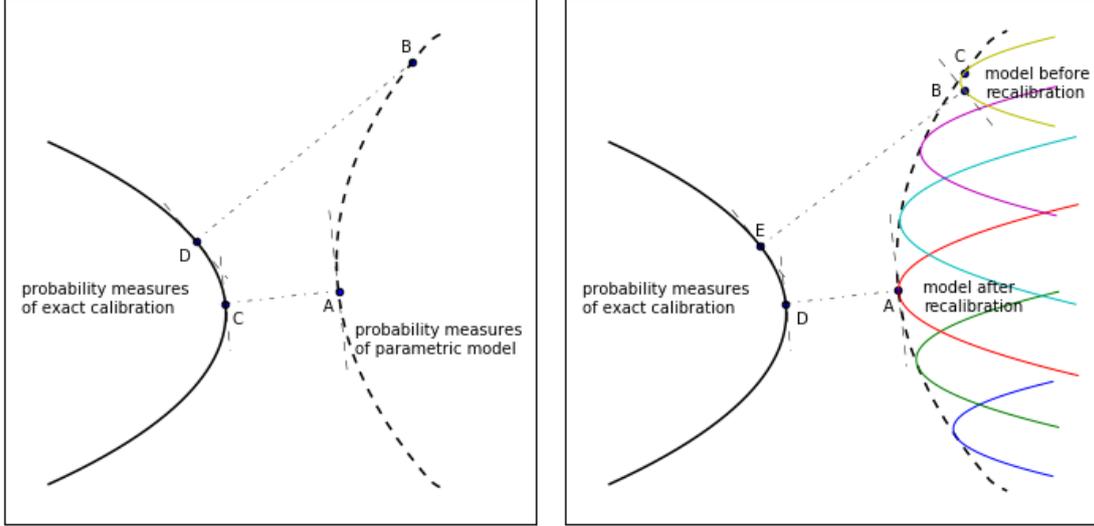


FIGURE 1. Graphic illustration of the mathematical definitions of model risks for (a) observable state variable, (b) latent state variable.

with S_Q now changed to reflect the change in observed market prices, so we have the following Radon–Nikodym derivatives:

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

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

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

So the aggregate risk is expressed in terms of m_1 and m_2 :

$$(15) \quad \eta_3 = E^P[m_1 m_2 \ln(m_1 m_2)]$$

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

$$(17) \quad = \eta_1 + [1 - cov^P(m_1, m_2)]\eta_2 + cov^f[m_1, m_2 \ln(m_2)]$$

If m_1 and m_2 are independent, $cov^P(m_2, m_1) = cov^P[m_2, m_1 \ln(m_1)] = 0$. The total model risk is equal to the sum of the calibration risk and the recalibration risk. Surprisingly, in our empirical analysis 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.

We also may consider a more complicated type of model which involves 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 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 specified by a given set of parameters forms a one-dimensional manifold (Fig. 1(b)) for possible realisations of the state variable, rather than a point in the Black-Scholes world (Fig. 1(a)).

Thus, the model which 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). Then, the initial calibration problem (4) becomes

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

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.⁶ The notation in (6) is amended to

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

At time t_i , we have for the calibration error

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

The model risk due to recalibration is

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

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

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

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

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

which results in

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

We then have the following Radon–Nikodym derivatives:

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

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

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

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 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 arising when taking into account calibration error and model risk due to recalibration in the manner 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 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 $D(Q^{(0)} \| P^{(0)})$.

⁶This 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 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.

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.

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

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

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

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 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. Also 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 instead. In a simplified scenario, where exact option prices are given, we may set $\mathbf{B} = \mathbf{A}$. By introducing vectors of Lagrange multipliers λ_B and λ_A , we convert the constrained problem to an unconstrained dual problem,

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

In the case of Kullback-Leibler divergence, solving the inner problem gives the probability density function $q^{(0)}$ of $Q^{(0)}$ in terms of the density $p^{(0)}$ of $P^{(0)}$,

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

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

$$(32) \quad \max_{\lambda_B \geq 0, \lambda_A \leq 0} -\ln E^{P^{(0)}}(e^{(\lambda_B + \lambda_A)\mathbf{Z}}) + (\lambda_B + \lambda_A) \frac{\mathbf{B} + \mathbf{A}}{2} - (\lambda_B - \lambda_A) \frac{\mathbf{A} - \mathbf{B}}{2}$$

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. We may therefore transform the Lagrange multipliers by

$$(33) \quad \lambda_+ = \lambda_B + \lambda_A$$

$$(34) \quad \lambda_- = \lambda_B - \lambda_A$$

and the objective function becomes

$$(35) \quad \max_{\lambda_+ \in \mathbb{R}^d, \lambda_- \geq |\lambda_+|} -\ln E^{P^{(0)}}(e^{\lambda_+ \mathbf{Z}}) + \lambda_+ \frac{\mathbf{B} + \mathbf{A}}{2} - \lambda_- \frac{\mathbf{A} - \mathbf{B}}{2}$$

$$(36) \quad = \max_{\lambda_+ \in \mathbb{R}^d} -\ln E^{P^{(0)}}(e^{\lambda_+ \mathbf{Z}}) + \lambda_+ \frac{\mathbf{B} + \mathbf{A}}{2} - |\lambda_+| \frac{\mathbf{A} - \mathbf{B}}{2}$$

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

$$(37) \quad E^{P^{(0)}}(\mathbf{Z} e^{\lambda_+ \mathbf{Z}}) = \left[\frac{\mathbf{B} + \mathbf{A}}{2} - \mathbf{sgn}(\lambda_+) \frac{\mathbf{A} - \mathbf{B}}{2} \right] E^{P^{(0)}}(e^{\lambda_+ \mathbf{Z}})$$

where the element-wise sign function $\mathbf{sgn}(\lambda_+)$ assigns 1, -1 or 0 to each element of λ_+ . However, due to the discontinuity of the sign function (37) cannot be solved directly in a stable

way. To bypass this problem, we approximate the sign function with a continuous step function:

$$(38) \quad \text{sgn}(x) = 1 - \frac{2}{1 + \exp\left(\frac{x}{\delta}\right)}$$

We use Powell's hybrid method to solve the multidimensional equations (37).

In Step 3, we use L-BFGS-B algorithm to minimise the divergence with respect to model parameters (or latent variables or both). Step 2 and Step 3 are repeated until convergence. The convergence criterion adopted here is that all the percentage changes of parameters after one iteration do not exceed a certain threshold, say 0.1%.

4. EXAMINING THE TRADE-OFF BETWEEN CALIBRATION ERROR AND MODEL RISK DUE TO RECALIBRATION

As an application example of the method described in the previous two sections, we consider historical data consisting of daily market prices for call options on AAPL and 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 free of extraneous complications,⁷ while still covering reasonably liquid options and including a period of "interesting" market volatility (2007/8). From this data, we remove options 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 which had zero trading volume on a given day, to ensure that the prices we are using are not stale.

On this 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 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 risks under a nominal Black/Scholes model, i.e. where the underlying asset price is assumed to follow geometric Brownian motion, with dynamics under the risk-neutral measure given by

$$(39) \quad dS(t) = rS(t)dt + \sigma S(t)dW(t)$$

where r is the continuously compounded risk-free rate of interest and σ is a constant volatility parameter. We note that in the Black/Scholes model we obtain a simple closed form expression for the recalibration risk defined in (8):

$$(40) \quad \eta_2 = \left(\frac{\sigma^{*2}}{\sigma^2} - 1 \right) \left[\frac{1}{2} + \frac{T}{8}(\sigma^{*2} - \sigma^2) \right] + \ln \left(\frac{\sigma}{\sigma^*} \right)$$

where σ^* is the correctly recalibrated Black/Scholes volatility parameter and σ is the parameter value 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 total model risk as the sum of the calibration error, the recalibration risk and a residual. As noted in equation (15), the residual is zero if the likelihood ratios involved in the calibration and recalibration risks are two independent random variables. 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.⁸ Unsurprisingly (as it is well documented that the

⁷Although 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)).

⁸The vertical axis denotes the numerical value of the relative entropy.

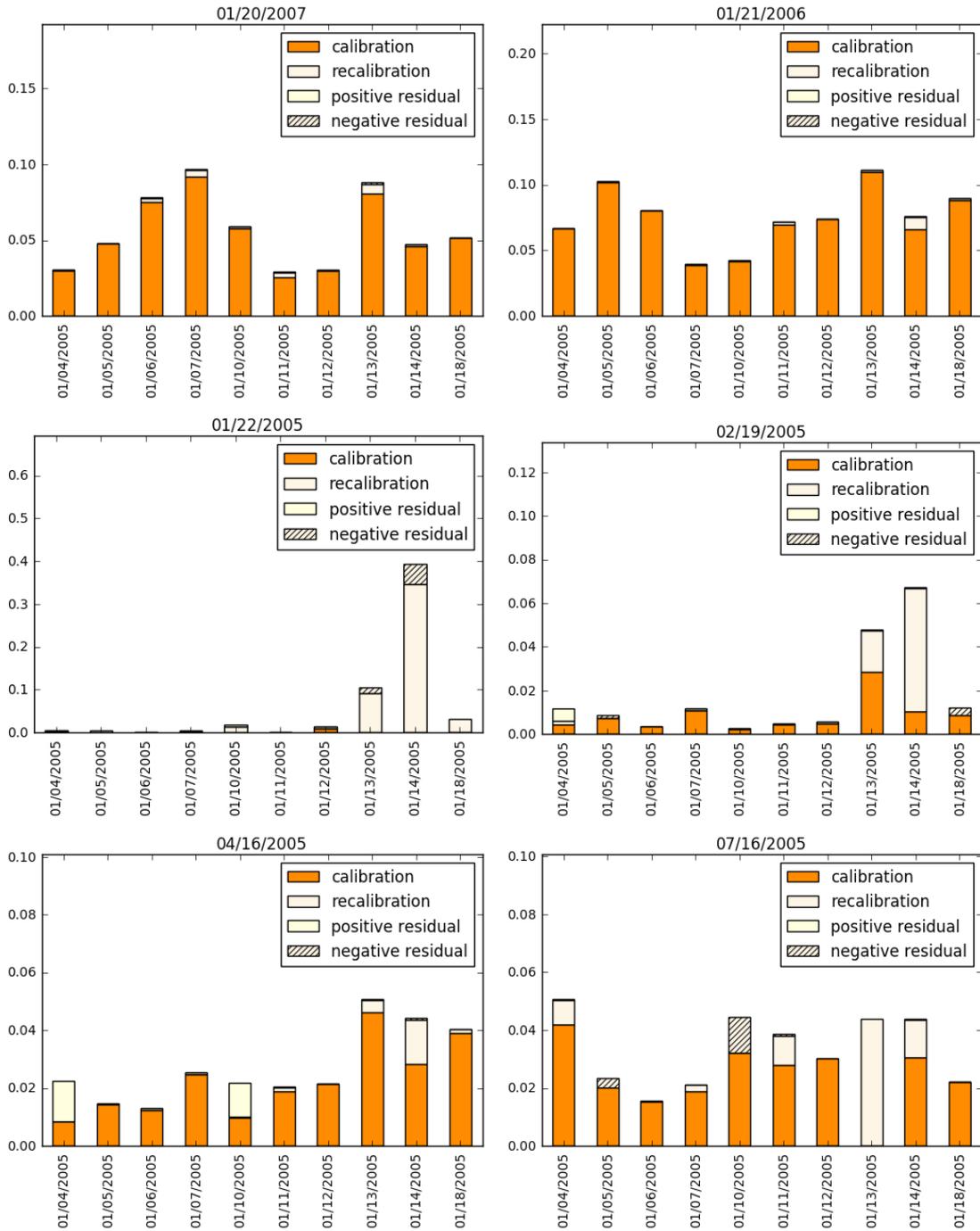


FIGURE 2. Decomposition of model risks of AAPL options using the Black/Scholes model as the nominal model, for a selection of option maturity dates (given as the title of each graph).

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 (39) are extended to allow for stochastic volatility, i.e.

$$(41) \quad dS(t) = rS(t)dt + \sigma S(t)dW_1(t)$$

$$(42) \quad d(\sigma^2(t)) = \kappa(\theta - \sigma^2(t))dt + \eta\sigma(t)dW_2(t)$$

This model involves two state variables, the underlying asset price $S(t)$ and the volatility $\sigma(t)$, and five model parameters: $r, \kappa, \theta, \eta, \rho$ where ρ is the correlation coefficient between the two

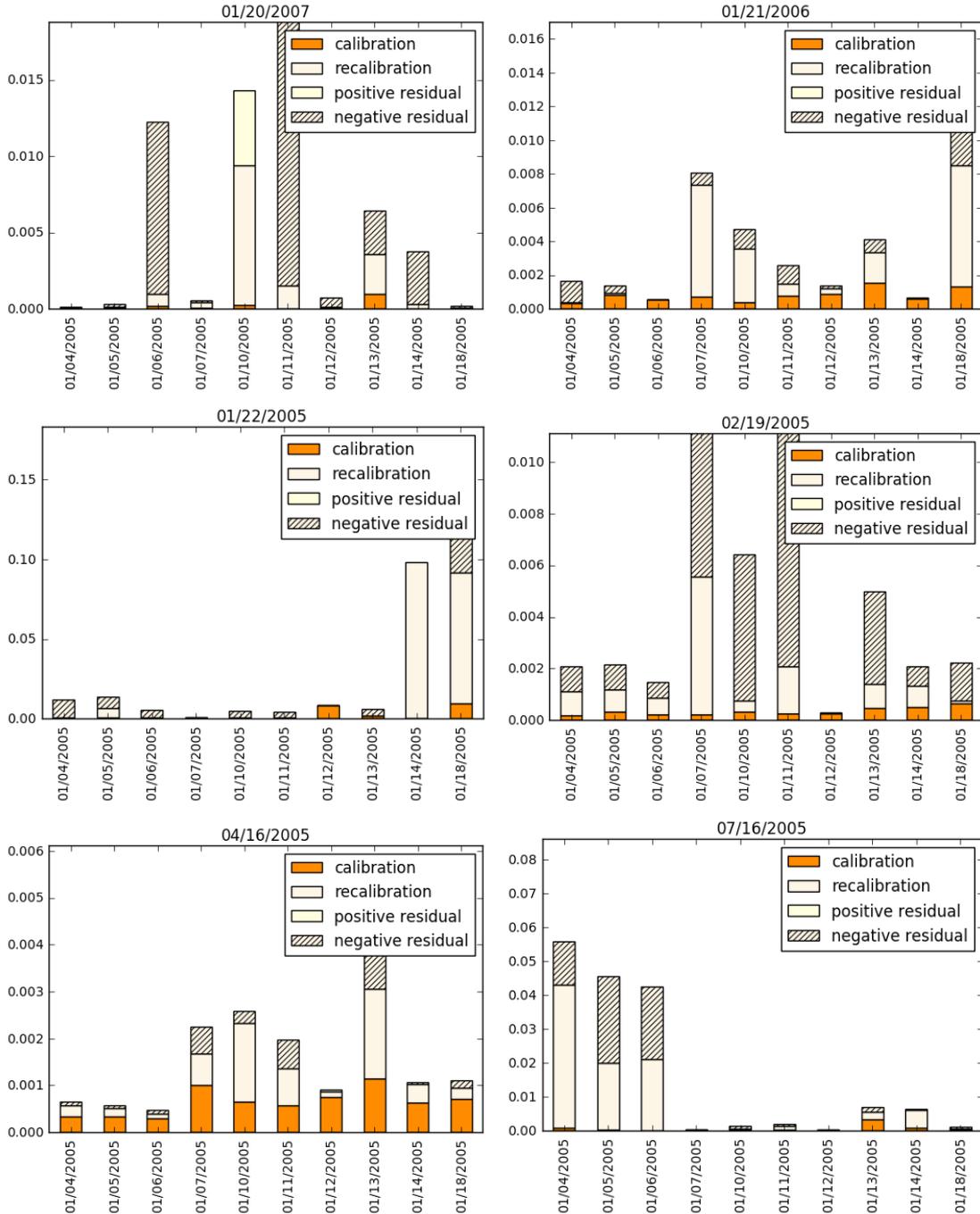


FIGURE 3. Decomposition of model risks of GOOG options using the Heston model as the nominal model, for a selection of option maturity dates (given as the title of each graph).

Wiener processes:

$$d \langle W_1, W_2 \rangle = \rho dt$$

r is the risk-free rate,⁹ and κ , θ and η relate to the volatility process, κ being the rate of mean reversion, θ the long-run mean and η the volatility of this process.

⁹In our empirical application examples, we take the risk-free rate as one of the financial variables observed in the market, but we do not explicitly take into account interest rate risk in our empirical analysis. For the short-dated options considered here, interest rate risk is known to be of relatively little importance — for a discussion of this issue, see e.g. Cheng, Nikitopoulos and Schlögl (2017) and the literature cited therein.

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

$$(43) \quad P_0(x, v, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{\exp\{C(u, \tau)\theta + D(u, \tau)v + iux\}}{iu} \right] du,$$

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,

$$(44) \quad C(u, \tau) = \kappa \left[r_- \tau - \frac{2}{\eta^2} \ln \left(\frac{r_+ - r_- e^{-d\tau}}{r_+ - r_-} \right) \right]$$

$$(45) \quad D(u, \tau) = r_- \frac{r_+ - r_+ e^{-d\tau}}{r_+ - r_- e^{-d\tau}}$$

$$(46) \quad r_\pm = \frac{\beta \pm d}{\eta^2} \quad d = \sqrt{\beta^2 - 2\alpha\eta^2}$$

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

$$(47) \quad \alpha(u) = -\frac{u^2}{2} - i\frac{u}{2}$$

$$(48) \quad \beta(u) = \kappa - \rho\eta iu$$

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

$$(49) \quad p(S_T) = - \left. \frac{\partial P_0}{\partial K} \right|_{K=S_T} = - \left. \frac{\partial P_0}{\partial x} \frac{\partial x}{\partial K} \right|_{K=S_T}$$

$$(50) \quad = \frac{1}{2\pi S_T} \int_{-\infty}^{\infty} \exp \left[C(u, \tau)\theta + D(u, \tau)v + iu \ln \left(\frac{S_t}{B(t, T)S_T} \right) \right] du$$

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 :

$$(51) \quad p(y) = p(S_T) \left| \frac{\partial y}{\partial S_T} \right|^{-1}$$

$$(52) \quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp [C(u, \tau)\theta + D(u, \tau)v - iuy] du$$

$p(y)$ can be calculated by fast Fourier transform (FFT).

The decomposition of the total model risk into the three components (the components due 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, since 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 this comes at a price of increased model risk due to recalibration.

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.

Aggregated Risk										
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
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
	95%	0.212	0.221	0.221	0.212	0.251	0.173	0.169	0.177	0.185
	90%	0.158	0.160	0.165	0.160	0.192	0.096	0.097	0.098	0.105
	75%	0.092	0.094	0.095	0.099	0.112	0.027	0.027	0.029	0.031
Calibration Risk										
mean	0.055	0.055	0.055	0.055	0.054	0.008	0.007	0.007	0.006	0.006
median	0.038	0.038	0.038	0.038	0.037	0.001	0.001	0.001	0.001	0.001
Quantile	99%	0.239	0.250	0.256	0.246	0.239	0.163	0.129	0.161	0.115
	95%	0.171	0.175	0.176	0.174	0.173	0.015	0.016	0.017	0.015
	90%	0.135	0.135	0.135	0.134	0.135	0.008	0.008	0.008	0.008
	75%	0.075	0.075	0.075	0.074	0.074	0.003	0.003	0.003	0.003
Recalibration Risk										
mean	0.024	0.024	0.025	0.026	0.038	0.057	0.059	0.061	0.061	0.073
median	0.002	0.002	0.003	0.004	0.007	0.004	0.005	0.006	0.007	0.014
Quantile	99%	0.458	0.453	0.358	0.373	0.437	0.630	0.679	0.653	0.665
	95%	0.103	0.103	0.112	0.110	0.189	0.315	0.347	0.358	0.360
	90%	0.056	0.055	0.059	0.062	0.099	0.164	0.161	0.172	0.171
	75%	0.013	0.016	0.019	0.022	0.036	0.055	0.050	0.055	0.057

TABLE 1. Varying models and recalibration frequency for AAPL

Aggregated Risk										
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
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
	95%	0.519	0.530	0.521	0.496	0.549	0.475	0.438	0.455	0.467
	90%	0.393	0.391	0.387	0.385	0.408	0.330	0.287	0.320	0.324
	75%	0.219	0.223	0.214	0.215	0.256	0.145	0.137	0.142	0.153
Calibration Risk										
mean	0.125	0.122	0.122	0.121	0.124	0.055	0.051	0.053	0.054	0.056
median	0.082	0.077	0.080	0.078	0.082	0.006	0.006	0.006	0.006	0.007
Quantile	99%	0.655	0.631	0.644	0.606	0.636	0.626	0.551	0.548	0.586
	95%	0.400	0.395	0.391	0.385	0.384	0.329	0.311	0.308	0.316
	90%	0.295	0.289	0.282	0.285	0.285	0.172	0.166	0.167	0.178
	75%	0.171	0.168	0.163	0.164	0.167	0.034	0.028	0.036	0.031
Recalibration Risk										
mean	0.036	0.040	0.040	0.044	0.060	0.119	0.124	0.089	0.129	0.135
median	0.004	0.006	0.007	0.009	0.018	0.043	0.046	0.039	0.048	0.055
Quantile	99%	0.405	0.467	0.455	0.502	0.527	0.757	0.757	0.489	0.760
	95%	0.173	0.191	0.179	0.196	0.250	0.543	0.559	0.384	0.578
	90%	0.109	0.114	0.115	0.119	0.176	0.411	0.440	0.288	0.438
	75%	0.036	0.039	0.042	0.046	0.068	0.142	0.141	0.111	0.152

TABLE 2. Varying models and recalibration frequency for GOOG

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. 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 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.

5. CONCLUSION

Under our approach, less relative entropy implies less model risk, and we are able to consider 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. By “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, one can better assess the added value (if any) of more complicated stochastic models of financial markets.

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).¹⁰ Even in this case, we see that by including recalibration as one of the sources of aggregate model risk, recalibrating a model frequently 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.

REFERENCES

- Ahmadi-Javid, A.: 2012, Entropic value-at-risk: A new coherent risk measure, *Journal of Optimization Theory and Applications* **155**(3), 1105–1123.
- Ali, S. M. and Silvey, S. D.: 1966, A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 131–142.
- Black, F. and Scholes, M.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**(3), 637–654.
- 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*, Federal Reserve.
- Box, G. E. P. and Draper, N. R.: 1987, *Empirical Model-Building and Response Surfaces*, Wiley.
- Cheng, B., Nikitopoulos, C. S. and Schlögl, E.: 2017, Pricing of long-dated commodity derivatives: Do stochastic interest rates matter?, *Journal of Banking & Finance* .
- Csiszár, I.: 1967, Information-type measures of difference of probability distributions and indirect observations, *Studia Scientiarum Mathematicarum Hungarica* **2**, 299–318.
- Gatheral, J.: 2006, *The Volatility Surface: A Practitioner’s Guide*, John Wiley & Sons.
- Glasserman, P. and Xu, X.: 2014, Robust risk measurement and model risk, *Quantitative Finance* **14**(1), 29–58.
- Heston, S. L.: 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *The Review of Financial Studies* **6**, 327–343.
- Löffler, G.: 2003, The effects of estimation error on measures of portfolio credit risk, *Journal of Banking & Finance* **27**, 1427–1453.
- Merton, R. C.: 1973, Theory of rational option pricing, *The Bell Journal of Economics and Management Science* **4**(1), 141–183.

¹⁰If desired, one could reformulate this approach to prioritise the minimisation of aggregate model risk, or of model risk due to recalibration.

(Yu Feng) UNIVERSITY OF TECHNOLOGY SYDNEY, PO BOX 123, BROADWAY, NSW 2007, AUSTRALIA
E-mail address, Yu Feng: Yu.Feng-5@student.uts.edu.au

(Ralph Rudd) DEPARTMENT OF ACTUARIAL SCIENCE AND THE AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH, UNIVERSITY OF CAPE TOWN

(Christopher Baker) DEPARTMENT OF ACTUARIAL SCIENCE AND THE AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH, UNIVERSITY OF CAPE TOWN

(Qaphela Mashalaba) DEPARTMENT OF ACTUARIAL SCIENCE AND THE AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH, UNIVERSITY OF CAPE TOWN

(Melusi Mavuso) DEPARTMENT OF ACTUARIAL SCIENCE AND THE AFRICAN COLLABORATION FOR QUANTITATIVE FINANCE AND RISK RESEARCH, UNIVERSITY OF CAPE TOWN

(Erik Schlögl) UNIVERSITY OF TECHNOLOGY SYDNEY, PO BOX 123, BROADWAY, NSW 2007, AUSTRALIA
E-mail address, Corresponding author: Erik.Schlogl@uts.edu.au