We provide an economic valuation of the riskiness of risk models by directly measuring the impact of model risks (specification and estimation risks) on VaR estimates. We find that integrating the model risk into the VaR computations implies a substantial minimum correction of the order of 10-40% of VaR levels. We also present results of a practical method – based on a backtesting framework – for incorporating the model risk into the VaR estimates.

Keywords: Model Risk; Quantile Estimation; VaR; Basel II Validation Test.
JEL Classification: C14, C50, G11, G32.
1. Introduction

The recent worldwide financial crisis has dramatically revealed that risk management pursued by financial institutions is far from optimal. This paper proposes an economic evaluation of the impact of model uncertainty on VaR estimates based on a backtesting framework.

The Basel III committee has recently further proposed that financial institutions assess the model risk (BCBS, 2009). However, the model risk, whilst well studied in the case of specific price processes (e.g. Cont, 2006), is not yet taken into account practically in the building of risk models by the industry\(^1\).

The outline of the paper is as follows. Section 2 defines and illustrates the model risk in VaR estimates. Section 3 presents our practical approach for calibrating adjusted Empirical VaRs that deal with the model risk. Section 4 concludes.

2. Model Risk and VaR Computations

We first illustrate the model risk of VaR estimates, which is here defined as the consequence of two types of errors due to a model misspecification and a

\(^1\) Only a few recent papers (e.g. Kerkhof et al., 2010; Gouriéroux and Zakoïan, 2010) still aim to take model risk into account in the computation of risk measures.
parameter estimation uncertainty. Various VaR computation methods do indeed exist in the literature, from non-parametric, semi-parametric and parametric approaches (e.g. Christoffersen, 2009). However, the Historical-simulated VaR computation is still the one most used by practitioners (Christoffersen and Gonçalves, 2005) and will serve as the reference throughout this article.

Table 1 presents the Estimated VaR as well as the mean, minimum and maximum errors on these Estimated VaR. Errors are defined by the differences between the “true” asymptotic VaR (based on simulated DGP) and the Imperfect Historical-simulated Estimated VaR (because the latter are approximately specified and estimated with a limited data sample). Three different rolling time-windows (from 250 to 750 days in Panels A, B and C) and several levels of probability confidence thresholds (three rows for each Panel) are considered. The results are presented for three Data Generating Processes for the underlying stock price with various intensities of jumps (Brownian, Lévy and Hawkes$^2$).

As expected, the Estimated VaR is an increasing function of the confidence level and the presence of jumps in the process (Lévy and Hawkes’ cases). For a large number of trials, the mean bias of the Historical-simulated

$^2$ See Applebaum (2004) and Aït-Sahalia et al. (2010) for Lévy and Hawkes process definitions.
Method is quite small (inferior to 1% in relative terms) in the Brownian case, and rather insensitive to the number of datapoints in the sample. By contrast, this mean bias is quite large when jumps are considered (with an amplitude of 10% to 27% in relative terms\(^3\)).

Moreover, the observed range of potential relative errors (the difference between the maximum and minimum estimated errors divided by the estimated VaR) is substantial in our experiments, representing between around 40% of the VaR levels in the best case (for the simple Brownian DGP over the longer sample) to as high as 290% in the worst case scenario (for the Lévy DGP over the shorter sample). Furthermore, the potential relative under-estimation of the “true” VaR (too aggressive estimated VaR) is, in the main, large (ranging from 13% to 40%, depending on the sample length and the quantile considered\(^4\)). These results suggest that the Historical-simulated VaR should be corrected when safely taking into account the riskiness of risk models.

\(^3\) The relative error of 27% corresponds to the probability 99.50% with a window of 250 days for the Lévy DGP (i.e. 14.90 out of -54.26).

\(^4\) The relative error of 40% corresponds to the probability 99.50% with a window of 250 days for the Brownian DGP (i.e. -16.04 out of -39.95).
Table 1

Estimated annualized VaR and model-risk errors (%)

Three price processes of the asset returns are considered below, such as for \( t=1,\ldots,T \) and \( p=1, 2, 3 \):

\[
dS_t = S_t \left( \mu \, dt + \sigma \, dW_t + J^p_t \, dN_t \right)
\]

with:

\[
\begin{align*}
J^1_t &= 0 \\
J^2_t &= \lambda_2 \exp(-\lambda_2 t) \\
J^3_t &= \lambda_3 + \beta \exp(-\gamma(t-s))
\end{align*}
\]

where \( S_t \) is the price of the asset at time \( t \), \( W_t \) is a standard Brownian motion, independent from the Poisson process \( N_t \), governing the jumps of various intensities \( J^p_t \) (null, constant or time-varying according to the process \( p \)), defined by parameters, \( \lambda_2, \lambda_3, \beta \) and \( \gamma \), which are some positive constants, with \( s \) (in case 3) the date of the last observed jump.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VaR Error (in %):</td>
<td>Mean</td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>Panel A. One-year Rolling Window Calibration (( T=250 ))</td>
<td>-24.78</td>
<td>-0.06</td>
<td>-8.69</td>
</tr>
<tr>
<td>Panel B. Two-year Rolling Window Calibration (( T=500 ))</td>
<td>-35.74</td>
<td>-0.11</td>
<td>-14.21</td>
</tr>
</tbody>
</table>

Source: simulations by the authors. Errors are defined as the difference between the “true” asymptotic simulated VaR and the Estimated VaR. These statistics were computed with a series of 250,000 simulated daily returns with specific DGP (1. Brownian, 2. Lévy and 3. Hawkes), averaging the parameters estimated in Aït-Sahalia et al. (2010, Table 5, i.e. \( \beta=41.66\% \), \( \lambda_3=1.20\% \) and \( \gamma=22.22\% \)), and ex post recalibrated for sharing the same first two moments (i.e. \( \mu=0.12\% \) and \( \sigma=1.02\% \)) and the same mean jump intensity (for the two last processes such as \( J^2_t \) - which leads after rescaling here, for instance, to an intensity of the Levy such as: \( \lambda_2=1.06\%)\). Per convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative).

3. A Simple Procedure for adjusting Estimated VaR

We propose herein a simple procedure to calibrate a correction on VaR estimates to account for the impact of the model errors. This procedure is based on the “Traffic Light” test developed by the Basel Committee. The regulatory
backtesting process is carried out by comparing the last 250 daily 99% VaR estimates with corresponding daily trading outcomes.

The regulatory framework uses the proportion of a failure test based on the Unconditional Coverage test (Kupiec, 1995). This test is based on the so-called “hit variable” associated to the ex post observation of Estimated VaR violations at the threshold $\alpha$ and time $t$, denoted $I_{t}^{\text{VaR}}(\alpha)$, which is defined such:

$$I_{t}^{\text{VaR}}(\alpha) = \begin{cases} 1 & \text{if } r_{t} < -\text{EVaR}(P, \alpha)_{t-1} \\ 0 & \text{otherwise,} \end{cases}$$  \hspace{1cm} (1)$$

where $\text{EVaR}(\cdot)$ is the Estimated VaR on a portfolio $P$ at a threshold $\alpha$, and $r_{t}$ is the return on a portfolio $P$ at time $t$, with $t=[1,\ldots,T]$.

If we assume that the $I_{t}^{\text{VaR}}(\cdot)$ variables are Independently and Identically Distributed, then, under the unconditional coverage hypothesis (Kupiec, 1995), the total number of VaR exceptions (Cumulated Hits) follows a Binomial distribution (Christoffersen, 1998), denoted $B(T, \alpha)$, such as:

$$\text{Hit}_{T}^{\text{VaR}} = \sum_{t=1}^{T} I_{t}^{\text{VaR}}(\alpha) \xrightarrow{T \to +\infty} B(T, \alpha)$$  \hspace{1cm} (2)$$

A Perfect VaR (not too aggressive, but not too confident) in the sense of this test, is such that it provides a sequence of VaR denoted $\text{VaR}(\cdot)^{*}$ (i.e. all $\left\{\text{VaR}(P, \alpha)^{*}_{t}\right\}$ for $t=[0,T]$), that respects:
\[
\begin{align*}
T^{-1} \text{Hit}_T^{\text{VaR}(\cdot)^*} &< \alpha \\
T^{-1} \left[ \text{Hit}_T^{\text{VaR}(\cdot)^*} + 1 \right] &\geq \alpha,
\end{align*}
\tag{3}
\]

where \( \text{Hit}_T^{\text{VaR}(\cdot)^*} \) is the cumulated hit variable associated to the \( \text{VaR}(\cdot)^* \).

In other words, since the Estimated VaR and the bounding range are known, we now have to search, amongst all possibilities, for the minimum (unconditional) adjustment that allows us to recover a Corrected Estimated VaR that respects condition (3) over the whole sample, \textit{i.e.}:

\[
\text{adj}(P, \alpha) = q^* = \text{ArgMax}_{q \in \mathbb{R}} \left\{ \text{VaR}(P, \alpha)^* \right\}
\]

\text{s.t.}: \begin{align*}
T^{-1} \text{Hit}_T^{\text{VaR}(\cdot)^*} &< \alpha \\
T^{-1} \left[ \text{Hit}_T^{\text{VaR}(\cdot)^*} + 1 \right] &\geq \alpha,
\end{align*}
\tag{4}

with:
\[
\text{VaR}(P, \alpha)^* = \text{EVaR}(P, \alpha) + q^*.
\]

Figure 1 represents the minimum adjustments (absolute errors) to be applied to estimated VaR, denoted \( q^* \) as solutions of the optimization program (4), for one-year (two-year and three-year) Historical-simulated VaR computed on the DJIA over more than one century (from the 1\textsuperscript{st} of January, 1900 to the 15\textsuperscript{th} of October, 2010). In other words, it represents the minimal global constants that we should have added to the quantile estimations for having reached a VaR sequence that would have passed the Hit test on the full sample for the considered levels of
confidence. We observe that the Historical-simulated error is quite significant for all quantiles (between -0.5% and -7% in absolute terms, i.e. 15% or so in relative terms) and significantly increases with the confidence level.

**Figure 1**
Minimum Model Risk Adjustments associated to Historical-simulated VaR

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January 1900 to the 15th October 2010; computations by the authors. The first plot (on the left hand side) represents the non-adjusted average annualized VaR level. The minimal adjustment is represented in the second plot and is expressed in absolute value (on the right hand side). The minimal adjustment necessary to respect the hit ratio criterion is here considered as a proxy of the economic value of the model risk. The historical VaR is computed on a daily horizon as an annualized empirical quantile using respectively 1 year, 2 years and 3 years of past returns. Without any adjustment, the imperfect Estimated VaR is underestimated (too permissive) in each of these cases.

Besides, the smaller the estimation period, the more important the adjustment (both in absolute and relative terms). This phenomenon can be explained by the fact that using larger estimation periods leads more likely to take into consideration extreme realizations and crisis episodes.

4. Conclusion
This paper proposes a practical method to incorporate model risk into risk measure estimates by adjusting the estimated VaR according to the frequency of past exceptions. This VaR adjustment allows a joint treatment of theoretical and estimation risks, taking into account their possible dependence.

We find that the model risk can represent a significant part of the risk measure in simulations and, secondly, that the required correction may be substantial (of the order of 10-40% of VaR levels in extensive simulations, and in the range of 1-15% for a real risky stock market index). Recently, this kind of bias estimation and correction of risk measures has been proposed by Kerkhof et al. (2010) with a similar range of estimated corrections based on different DGP.

References


Appendix. Complementary Results at the Referees’ Attention.

Figure A1 presents simple illustrations on a limited series of 250 returns of the behaviours of the three underlying processes (Brownian, Levy and Hawkes) used in Table 1.

**Figure A1**

An Illustration of Simulated Processes used in Table 1 for a Series of 250 Returns

Source: simulations by the authors. The first two plots illustrate, respectively, the behaviour of the jump components within the Lévy and Hawkes processes. The three last plots present simulations of return series using Brownian, Lévy (Brownian plus Simple Jump Component) and Hawkes processes (Brownian + Autoregressive Jump Component). These five figures are representative and are computed from one sample of 250 simulated daily returns. They are generated using specific DGP (1. Brownian, 2. Lévy and 3. Hawkes), averaging the parameters estimated for major stock markets in Aït-Sahalia et al. (2010, Table 5, i.e. $\beta=41.66\%$, $\lambda=1.20\%$ and $\gamma=22.22\%$), and ex post recalibrated for sharing the same first two moments (i.e. $\mu=0.12\%$ and $\sigma=1.02\%$) and the same mean jump intensity (for the last two processes such as $I_t^r - I_t$ - which leads after rescaling here, for instance, to an intensity parameter of the Lévy process such as $\lambda_2=1.06\%$). According to the three DGP realizations, the annualized estimated 95.00%, 99.00% and 99.50% VaR (based on daily computations) are here respectively: -23.81%/-24.42%/-26.33%; -38.54%/-40.74%/-47.05% and -42.01%/-44.93%/-56.00%.
The Table A1 confirms the convergence of the results presented in the *corpus* of the text (see Table 1). We independently reproduced Table 1 with another set of 250,000 simulations and show the resulting differences for each cell. Except in three cases, most of the relative differences are below 1.00%, and none are above 5.00%.

**Table A1**

Differences between Results in Two Sets of 250,000 Times-series of Computations reported in Table 1

Three price processes of the asset returns\(^5\) are considered below, such as for \(t=1, ..., T\) and \(p=1, 2, 3\):

\[
\begin{align*}
    dS_t &= S_t \left( \mu dt + \sigma dW_t + J^P_t dN_t \right) \\
    J^1_t &= 0 \quad (1. \text{Brownian}) \\
    J^2_t &= \lambda_2 \exp(-\lambda_2 t) \quad (2. \text{Lévy}) \\
    J^3_t &= \lambda_3 + \beta \exp[-\gamma(t-s)] \quad (3. \text{Hawkes})
\end{align*}
\]

where \(S_t\) is the price of the asset at time \(t\), \(W_t\) is a standard Brownian motion, independent from the Poisson process \(N_t\), governing the jumps of various intensities \(J^p_t\) (null, constant or time-varying according to the process \(p\)), defined by parameters, \(\lambda_2, \lambda_3, \beta, \gamma\) which are some positive constants, with \(s\) (in case 3) the date of the last observed jump.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VaR and Error (in %):</td>
<td>Mean</td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>Estimated VaR Error Error Error Error</td>
<td>VaR Error Error Error Error</td>
<td>VaR Error Error Error Error</td>
<td>VaR Error Error Error Error</td>
</tr>
<tr>
<td>Probability</td>
<td>Panel A. One-year Rolling Window Calibration ((T=250))</td>
<td>Panel B. Two-year Rolling Window Calibration ((T=500))</td>
<td>Panel C. Three-year Rolling Window Calibration ((T=750))</td>
</tr>
<tr>
<td>95.00%</td>
<td>-0.26</td>
<td>-0.01</td>
<td>-0.10</td>
</tr>
<tr>
<td>99.00%</td>
<td>-0.37</td>
<td>-0.02</td>
<td>-0.15</td>
</tr>
<tr>
<td>99.50%</td>
<td>-0.41</td>
<td>-0.04</td>
<td>-0.17</td>
</tr>
<tr>
<td>95.00%</td>
<td>-0.26</td>
<td>0.00</td>
<td>-0.07</td>
</tr>
<tr>
<td>99.00%</td>
<td>-0.37</td>
<td>-0.18</td>
<td>-0.11</td>
</tr>
<tr>
<td>99.50%</td>
<td>-0.41</td>
<td>-0.04</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Panel B. Two-year Rolling Window Calibration \((T=500)\)

| Probability | Panel C. Three-year Rolling Window Calibration \((T=750)\) |
| 95.00% | -0.26 | 0.01 | -0.07 | 0.05 | -0.26 | 0.01 | -0.04 | 0.08 | -0.28 | 0.02 | -0.04 | 0.11 |
| 99.00% | -0.37 | 0.03 | -0.10 | 0.10 | -0.39 | 0.03 | -0.07 | 0.65 | -0.44 | 0.04 | -0.09 | 0.38 |
| 99.50% | -0.41 | 0.04 | -0.12 | 0.13 | -0.49 | 0.08 | -0.10 | 0.93 | -0.53 | 0.06 | -0.15 | 0.45 |

Source: simulations by the authors. This table presents the differences (in relative terms expressed in %) between results in two Sets of 250,000 Times-series as a robustness check of the convergence of computations of Imperfect Historical-simulated Estimated VaR and Model-risk Errors as reported in Table 1, using the very same methodology explained in the source of Table 1 in the *corpus* of the text. Except in three cases, most of the relative differences are below 1.00%, and none are above 5.00%.

---

\(^5\) See Applebaum (2004) and Aït-Sahalia et al. (2010) for Lévy and Hawkes process definitions.
The following Table A2 (similar to Table 1, but with a weekly frequency of observations) checks the influence of the jump autocorrelation structure in the Hawkes Process onto the VaR Levels. Since shocks may accumulate during a week, the VaR get worse when a Hawkes process is considered.

Table A2

Estimated VaR and Model-risk Errors based on Weekly Returns (%)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VaR and Error (in %):</td>
<td>Mean</td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>VaR</td>
<td>Estimated VaR</td>
<td>VaR</td>
<td>VaR</td>
</tr>
<tr>
<td>VaR Error</td>
<td>Error</td>
<td>Error</td>
<td>Error</td>
</tr>
</tbody>
</table>

Panel A. One-year Rolling Window Calibration (250 overlapping weekly returns)

<table>
<thead>
<tr>
<th>Probability</th>
<th>95.00%</th>
<th>99.00%</th>
<th>99.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Brownian</td>
<td>-20.56</td>
<td>-30.00</td>
<td>-34.22</td>
</tr>
<tr>
<td>2. Lévy</td>
<td>-22.13</td>
<td>-37.70</td>
<td>-42.18</td>
</tr>
</tbody>
</table>

Panel B. Two-year Rolling Window Calibration (500 overlapping weekly returns)

<table>
<thead>
<tr>
<th>Probability</th>
<th>95.00%</th>
<th>99.00%</th>
<th>99.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Brownian</td>
<td>-20.05</td>
<td>-30.29</td>
<td>-32.48</td>
</tr>
<tr>
<td>2. Lévy</td>
<td>-22.08</td>
<td>-37.35</td>
<td>-45.02</td>
</tr>
<tr>
<td>3. Hawkes</td>
<td>-22.47</td>
<td>-36.43</td>
<td>-43.57</td>
</tr>
</tbody>
</table>

Panel C. Three-year Rolling Window Calibration (750 overlapping weekly returns)

<table>
<thead>
<tr>
<th>Probability</th>
<th>95.00%</th>
<th>99.00%</th>
<th>99.50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Brownian</td>
<td>-20.69</td>
<td>-30.36</td>
<td>-33.84</td>
</tr>
<tr>
<td>2. Lévy</td>
<td>-22.09</td>
<td>-37.41</td>
<td>-45.41</td>
</tr>
<tr>
<td>3. Hawkes</td>
<td>-22.41</td>
<td>-36.04</td>
<td>-43.29</td>
</tr>
</tbody>
</table>

Source: simulations by the authors. Errors are defined as the difference between the “true” asymptotic simulated VaR and the Estimated VaR. These statistics (in absolute terms, expressed in %, such as returns) were computed with series of 250,000 simulated daily returns with specific DGP (1. Brownian, 2. Lévy and 3. Hawkes), averaging the parameters estimated in Aït-Sahalia et al. (2010, Table 5, i.e. $\beta=41.66\%$, $\lambda_2=1.20\%$ and $\gamma=22.22\%$), and ex post recalibrated for sharing the same first two moments (i.e. $\mu=12\%$ and $\sigma=102\%$) and the same mean jump intensity (for the two last processes such as $\bar{\lambda}_2=\bar{\lambda}_3$ - which leads after rescaling here, for instance, to an intensity of the Levy, such as: $\lambda_3=1.06\%$). Using several rolling windows (250 weekly returns for Panel A, 500 weekly returns for Panel B and 750 weekly returns for Panel C), annualized Estimated VaR at the 95.00%, 99.00% and 99.50% confidence levels are presented in this table. The first column in each block related to a process represents the Mean Estimated VaR with specification and estimation errors, whilst the following cells indicate the mean-minimum-maximum of the adjustment term corresponding to the observed differences between the Imperfect Historical-simulated Estimated VaR, empirically recovered in 250,000 draws of limited samples of 250, 500 or 750 weekly returns (Panel A, B and C), and the asymptotic (true) VaR (computed with the 250,000 data points of the full original sample for each process). Per convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative).