Validating Market Risk Models: A Practical Approach

Doug Gardner
Wells Fargo

November 2010

The views expressed in this presentation are those of the author and do not necessarily reflect the position of Wells Fargo. The examples, methods, approaches, and explanations are intended to reflect perceived best practices and do not necessarily represent methods or approaches as implemented or used by Wells Fargo.
Outline

• Market risk models
• Components of market risk models
• Validation of market risk model components
• Risk factor identification and distributions
• Pricing models
• Statistical measures
• Validation tips
Market Risk Definitions

- **Market risk**
  - *The risk of an increase or decrease in the market price of a financial instrument or portfolio*
  - *May be due to changes in stock prices, interest rates, credit spreads, foreign exchange rates, commodity prices, implied volatilities*

- **Regulatory classification**
  - *General market risk: due to systemic or general market factors (e.g., FX rates, equity indices, benchmark prices)*
  - *Specific risk: due to idiosyncratic factors unique to a financial instrument*
  - *Incremental risk: market risk arising from credit-related factors (rating downgrades or defaults)*
Market Risk Measurement

• Predict the distribution of possible change in value of an instrument over a given time horizon

• Summary statistics
  – Moments: volatility (standard deviation), skewness, kurtosis

• Value at Risk (VaR)
  – the maximum amount that can be lost over a given time horizon with a specified degree of confidence
  – Banks calculate regulatory capital based on a 99% confidence level over a 10 day time horizon

• CVaR (aka Expected Shortfall, Tail VaR)
  – Expected loss in the distribution tail (e.g., in worst 5% of outcomes)
  – Unlike VaR, CVAR is a coherent risk measure
P&L Distribution Through Time
Market Risk Models

- Model components
  - Risk factor multivariate distribution (more generally, process)
  - Instrument price distribution (as function of risk factors)
- Risk factor distributions
  - Parametric (e.g., multivariate normal)
  - Nonparametric (e.g., historical simulation)
  - Semi-parametric (e.g., historical simulation with volatility updating)
- Instrument pricing
  - Taylor series (e.g., delta, delta/partial gamma, delta/full gamma)
  - Full revaluation
  - Pricing grids
Risk Factor Selection

- RF selection is portfolio dependent
  - Requires in-depth knowledge of portfolio and products
  - Basis risk - modeled/ignored?
- Bucketing for curves/surfaces
  - Risk reports may reveal portfolio concentrations (e.g., large OTM positions), and hence suggest the need for additional granularity
- Proxies
  - New products often lack relevant historical data
  - A particular problem for historical stress scenarios and stressed VaR
- General market risk versus specific risk
  - Which risk factors where?
Risk Factor Properties

• Distributional assumption: Normal/lognormal/other
  – an issue even with historical simulation
• Volatility clustering
• Fat tails – leptokurtosis
• Time-varying correlations (?)
• Other considerations
  – Frequency of risk factor updates (daily/monthly/quarterly)
  – Curve/surface interpolation and extrapolation
  – Dealing with poor/missing data (particular issue with specific risk)
• Overlapping intervals
  – Observations no longer independent
  – Sun et al. (2009) examine impact on price change percentiles
Pricing Model Validation

- Pricing models used for market risk purposes should be validated to similar standard as models used for marking
  - Theoretical review
  - Review of key assumptions
  - Comparison to alternative models (including models used for marking if different)
  - Implementation review
**Backtesting Pricing Models**

- Compare model predicted P&L (using actual risk factor changes) to actual P&L
- Clean (frozen portfolio, exclude fee income and ad-hoc adjustments) versus dirty (i.e., actual) P&L
- Significant differences can reveal a number of possible issues
  - Model implementation errors
  - Missing risk factors
  - Inadequate pricing model approximations (portfolio nonlinearity)
  - Lack of granularity in curve bucketing
  - For general market risk, some differences are to be expected for instruments subject to specific risk
Backtesting VaR: Unconditional Coverage

• Kupiec (1995)
  – Coverage test: compare actual VaR exceptions to predicted
  – Example: P&L is expected to exceed a 95% VaR 5 days out of 100 (on average)
  – Under the null hypothesis (model is correctly specified), the number of exceptions follows a binomial distribution (which can be approximated by a normal distribution for large number of observations)
  – Basis for regulatory backtesting (99% 1-day VaR)
Hypothesis Testing

<table>
<thead>
<tr>
<th>Decision</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>Incorrect</td>
</tr>
<tr>
<td>Accept</td>
<td>OK</td>
</tr>
<tr>
<td>Reject</td>
<td>Type 1 error</td>
</tr>
<tr>
<td></td>
<td>OK</td>
</tr>
</tbody>
</table>

- Type 1 error: reject a correct model
- Type 2 error: accept an incorrect model
### Regulatory Exception Zones

<table>
<thead>
<tr>
<th>Zone</th>
<th>Exceptions</th>
<th>Penalty K</th>
<th>Type 1 Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>0-4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Yellow</td>
<td>5</td>
<td>0.4</td>
<td>10.8%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.5</td>
<td>4.1%</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.65</td>
<td>1.4%</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.75</td>
<td>0.4%</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.85</td>
<td>0.1%</td>
</tr>
<tr>
<td>Red</td>
<td>10+</td>
<td>1</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

- Exceptions based on 99% 1-day VaR over past year
- Within the yellow zone, regulators apply discretion
Type 1/2 Error Tradeoffs

- Ideally, a test would have both low type 1 (reject a correct model) and type 2 (accept an incorrect model) errors
- In practice, need to balance these through choice of exception threshold: to reduce type 2 errors, must increase type 1 errors
- Regulators are most concerned about type 2 errors
- 99% confidence level produces an unreliable test due to the relative paucity of exceptions
- Using a lower confidence level (say 95%) produces a more reliable test (lower type 2 error rate for a given type 1 error rate), and thus can be a useful supplement
- A number of institutions entered the red zone in 2008 (RISK, Jan-2010)
Conditional Coverage Testing

• Exceptions should be serially independent (i.e., not “bunch”)
• The model should have the correct conditional coverage as well as the correct unconditional coverage
• Christoffersen (1998) proposes a test that tests individually for unconditional coverage and independence
• Lack of independence may suggest a model that fails to properly account for volatility clustering, or a poorly modeled portfolio with time-varying risk profile
Testing the Entire Distribution

• Rosenblatt transformation: Let $X(t) = F(Y(t))$ where $F$ is the model-predicted P&L cdf, $Y(t)$ is the P&L on day $t$
  – $X(t)$ should be uniformly distributed and iid
  – May examine this graphically as well as formally test statistically

• Berkowitz (2001) suggests using $Z(t) = N(X(t))$ where $N$ is the inverse of the standard normal
  – $Z(t)$ should then be distributed normally and iid
  – May then apply standard tests of normality
  – May also construct a test that focuses only on the tail of the distribution rather than the entire distribution
Multivariate Coverage Testing

• Perignon and Smith (2008) suggest a multivariate generalization of the unconditional coverage test of Kupiec
  – First select a set of coverage probabilities and associated K buckets (e.g., 0-1%, 1-5%, 5-10%, 10-100%)
  – Compare the expected frequency of observations X(t) falling in each bucket (1%, 4%, 5%, 90% in this example) to the observed frequency
  – The associated test statistic is chi-square distributed with K-1 degrees of freedom
  – This procedure is commonly applied in other areas including the evaluation of the calibration of credit scoring models (Hosmer-Lemeshow 2000)
Validation Tips

• Model validation, like model building, must start with a strong understanding of the phenomenon being modeled
  – Build a strong relationship with trading, finance and risk oversight

• Validate based on model usage
  – For example, evaluate model performance in stress scenarios if this is a possible use

• Use multiple methods
  – Examine each key model component individually, as well as together
  – Use a full range of tests
  – Evaluate model performance at multiple levels (product, desk, trading area, firm)