An Academic Perspective on Backtesting and Stress-Testing

Presentation for

"Credit Risk Models and the Future of Capital Management"

Federal Reserve Bank of San Francisco

August 28-29, 2000

Jose A. Lopez

Financial & Regional Studies

Economic Research Department

Federal Reserve Bank of San Francisco

jose.a.lopez@sf.frb.org

Introduction

- Over the past decade or so, financial firms have greatly increased use of financial models to measure and manage their risk exposures.

- Financial regulators have focused their attention on the use of such models for two reasons:
  - supervision of banks’ risk management systems
  - the possibility of setting regulatory capital requirements that more accurately reflect banks’ risk exposures

Model Validation

A key component to the implementation of model-based risk management is **model validation**.

That is, determining whether the model chosen is accurate and performing consistently?

Important both to firms and their regulators.

Q: How should we conduct model validation?

Fortunately, we have available a large number of tools from the statistical and econometric literature.

The industry has used “backtesting” and “stress-testing”

Backtesting:	compare observed outcomes with model’s expected outcomes	• forecast evaluation established empirical issue with a large academic literature
Stress-testing:	examine model’s expected outcomes under extreme conditions	• projection analysis • outlier analysis • scenario analysis/ case studies

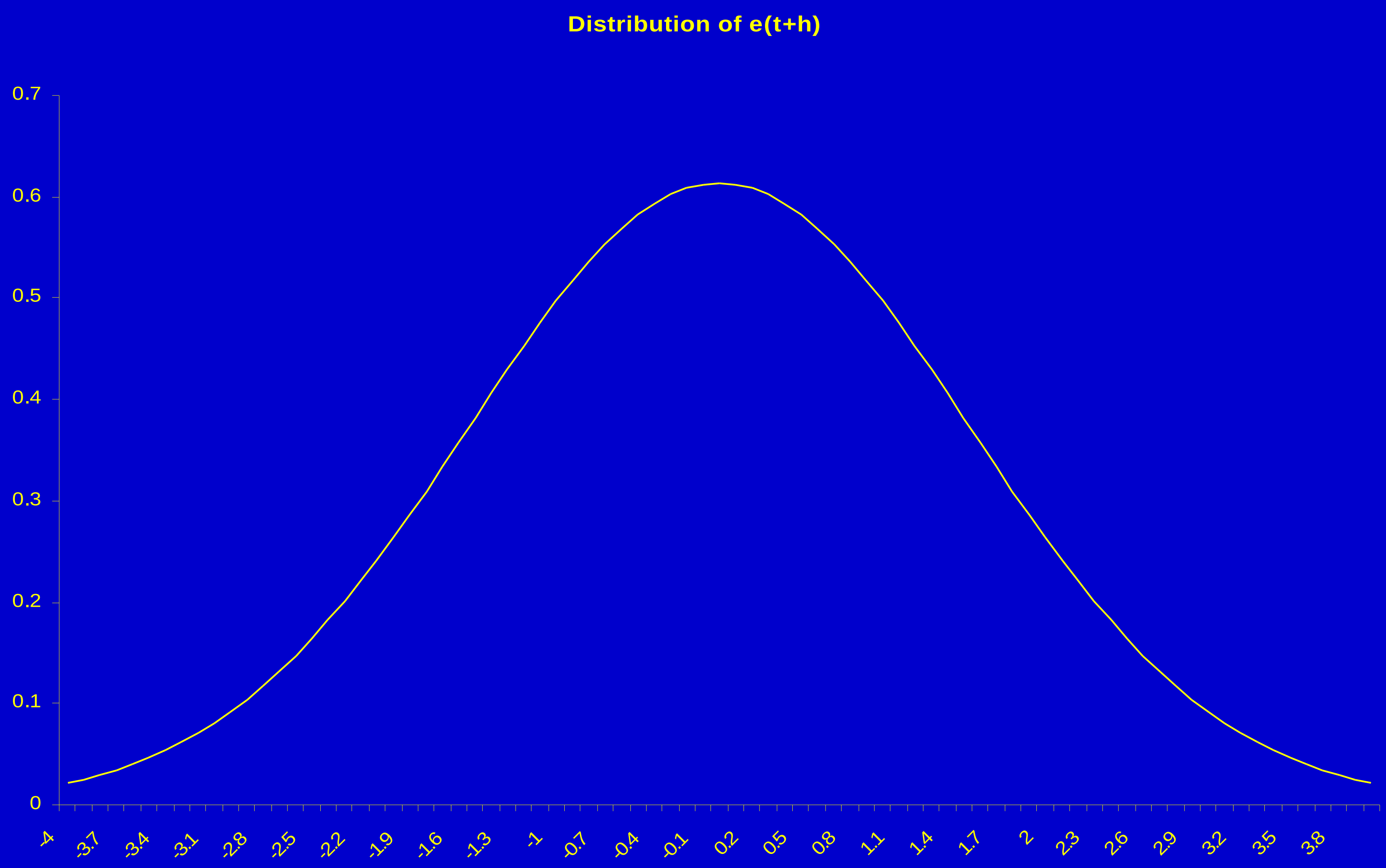
How can we backtest market risk models?

General notation for risk models:

- y_t = value of market risk portfolio at time t
- $y_{t+h|t}$ = future value of portfolio at time t+h at time t
- $y_{t+h|t} = M(y_t , e_{t+h})$; based on chosen risk model M
- e_{t+h} = random portfolio shock with distribution f (0, S)

Output of risk model M:

- $F_M(y_{t+h|t})$ = forecasted distribution of $y_{t+h|t}$



Distribution of e(t+h)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3.7
-3.4
-3.1
-2.8
-2.5
-2.2
-1.9
-1.6
-1.3
-1
-0.7
-0.4
-0.1
0.2
0.5
0.8
1.1
1.4
1.7
2
2.3
2.6
2.9
3.2
3.5
3.8

Object of interest:
VaR estimates / tail percentiles of $F_M(y_{t+h|t})$

Evaluation of VaR estimates is based on three numbers:

- the horizon of the forecast (denoted h)
- tail coverage required (the lower x% of $F_M(y_{t+h|t})$)
- number of observations to be evaluated

The backtesting component of the MRA regulation specifies:

- one-day horizon
- the lower 1% tail
- 250 observations (effectively 1 year)

Object of interest:
VaR estimates / tail percentiles of $F_M(y_{t+h|t})$

- The MRA uses the binomial method for VaR evaluation

- Statistically tests the hypothesis that the observed frequency of VaR exceptions equals the expected frequency.

- For example, for 250 days / observations, we expect to see 2.5 exceptions; do we see that?

- Simple, straightforward, well understood test

Object of interest: VaR estimates / tail percentiles of $F_M(y_{t+h|t})$

- However, the binomial test has poor power characteristics.
- That is, the probability of the test indicating that a set of VaR estimates is accurate when they are not is very high; limits its usefulness
- Intuition: we are discarding alot of information regarding the forecast; we only create an indicator variable
- How can we take greater advantage of the information within the VaR estimate framework?

Object of interest:
VaR estimates / tail percentiles of $F_M(y_{t+h|t})$

Multinomial tests:

- requires multiple VaR estimates; ex., lower 1% & 5%
- research ongoing to determine power
- requires reporting further VaR estimates

Magnitude tests:

- proposed by Berkowitz (2000, FRBOG)
- compares magnitude of exceptions with model's value
- shown to have reasonably good power properties
- requires no further reporting, but more complicated test

Object of interest: Probability forecasts

- Analyzed by Lopez (1999, Journal of Risk)
- Idea: $F_M(y_{t+h|t})$ is transformed into probability forecasts of what might happen to y_{t+h}; ex. what is the probability that y_{t+h} is above some fixed threshold value?
- Probability scoring rules and statistical tests can be used to compare probability forecasts to what actually occurs
- Intuition: does it actually rain 10% of the times that you forecast a 10% chance of rain?

Object of interest: Distribution forecasts

- Proposed by Berkowitz (2000, FRBOG)

- Transform forecasted distribution into an observed quantile using observed change in portfolio value and integration

- The empirical quantiles should be independent and uniformly distributed if accurate

- Benefit: good power properties; uses all the information
- Drawback: requires much more reporting

- For extensive example, see Lopez and Walter (2000)

Actual experience under the MRA:

Study by the BIS Committee on Banking Supervision (1999):

- 40 banks from 9 countries over 1998 Q3 and Q4
    - No losses exceeding regulatory capital
    - Half of sample had no one-day VaR exceptions!

Study by Berkowitz and O'Brien (2000, FRBOG):

- U.S. banks over the two-year implementation period
- Preliminary analysis finds that nearly all VaR exceptions occur during the third quarter of 1998.

How can we stress-test market risk models?

What are stress-tests?

- internally consistent possible states of the world
- possible, but unlikely, future events
- storylines that test risk managers' skills

No set criteria; practice varies widely and rightly so

Some formalization has been attempted in the literature:

- Kupiec (1998, Journal of Derivatives):
  - How to shock specific risk factors and not others in an internally consistent way?

Stress-testing within a backtesting framework

- Developed by Berkowitz (1999, FRBOG)

- Stress-testing is typically scenario analysis outside of the market risk model; the scenarios don't have probability attached to them as do the more standard outcomes

- Fold stress-test scenarios into backtesting framework:

  $e_{t+h} \sim f\,(\,0, S\,)$ with probability (1-a)

  $e_{t+h} \sim f_{stress}\,(\,0, S_{stress}\,)$ with probability a

- Thus, calculate $F_{M,stress}(y_{t+h|t})$ and examine specific stress scenarios using VaR calculations

Validating Credit Risk Models

Much more difficult for various reasons:

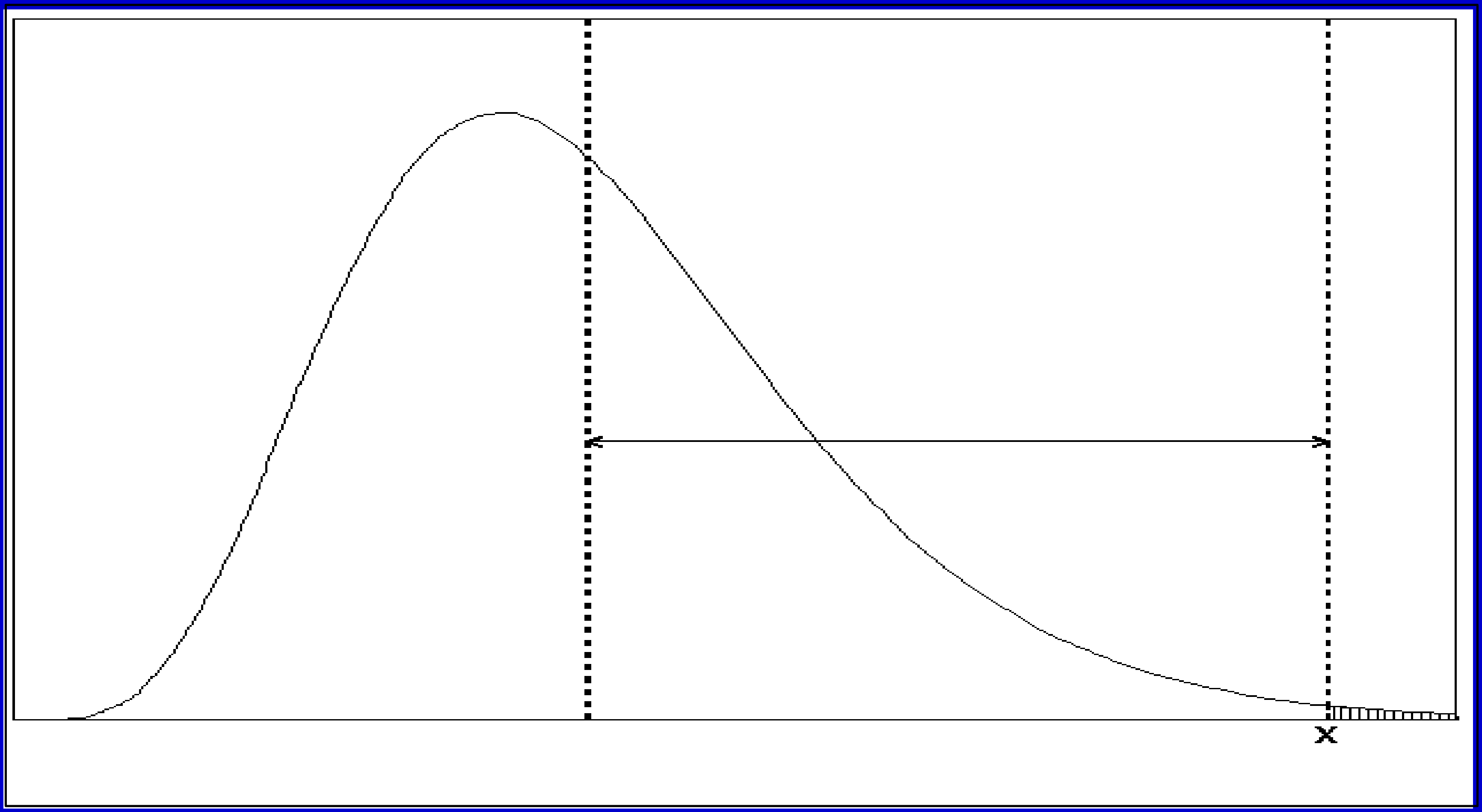
- Definition of loss: default mode vs. MTM mode
- Planning horizon: longer than market risk; greater dependence on the business cycle
- Asymmetric loss distribution with long tail:

  power properties not known, but should be fine
- Portfolio issues, such as default correlations, are harder

How can we backtest credit risk models?

- Basically, the same tools as before
- $F_M(y_{t+h|t})$ is still a probability distribution function, just asymmetric credit loss distribution

Probability Density Function of Portfolio Credit Losses (PDF)



Credit Losses Over Planning Horizon

How can we backtest credit risk models?

The main challenge is the lack of credit risk observations due to the longer forecast horizons; typically, one year.

- 250 observations is 250 years!
- If move to monthly, still need to wait more than 20 years!

<u>What can we do?</u>

Lopez & Saidenberg (2000, J. Banking & Finance)

- Mainly work in cross-section using simulation techniques
- Create many sample portfolios and run through model M
- Simulation results can be used for some model validation
- Note: still dependent on when in the business cycle

How can we stress-test credit risk models?

- Same as before:

  generate stories and scenarios to see what the model says.

- Again, scenarios are harder to construct for credit risk because of a lack of data and the greater degree of complexity present in the within portfolio correlations.

Conclusion

- Backtesting and stress-testing are key components to model-based risk measurement and management systems.

- Many tools are available for these purposes. No common criteria has been set as of yet; more work using different tools is needed to assist in this choice.