



FEDERAL RESERVE BANK *of* NEW YORK

Performance testing

The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

October 2016

Anna Kovner

Performance testing

- Statistics offers many metrics to analyze model performance
- What complicates the use of standard econometric tools?
 - Data availability
 - Trade-offs between data availability and time series
 - Future and past may be different
- What matters in stress testing context?
 - Sensitivity to macro scenarios
 - Overfitting
 - Predictive content especially in bad times
 - Stability

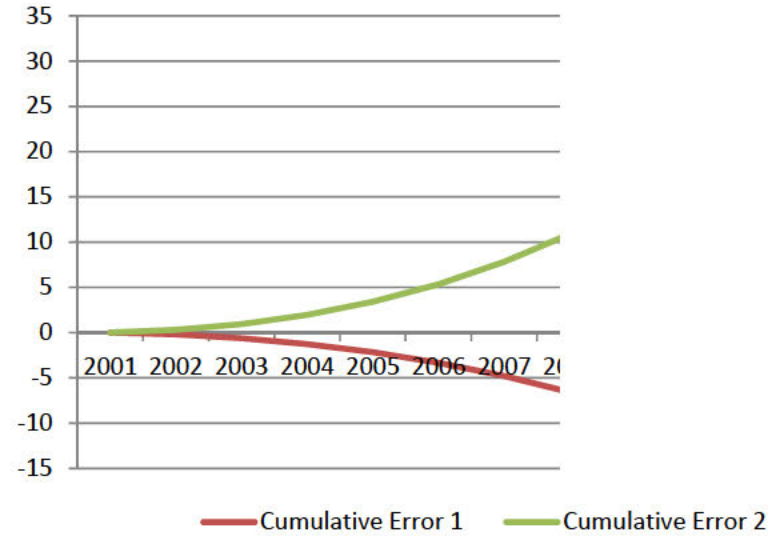
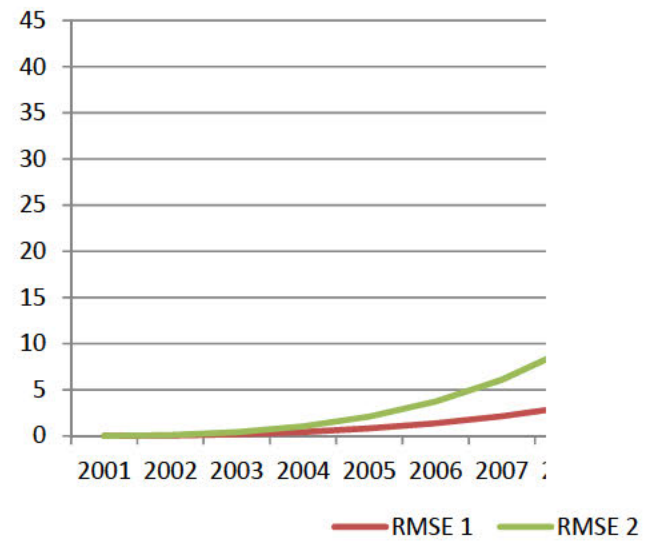
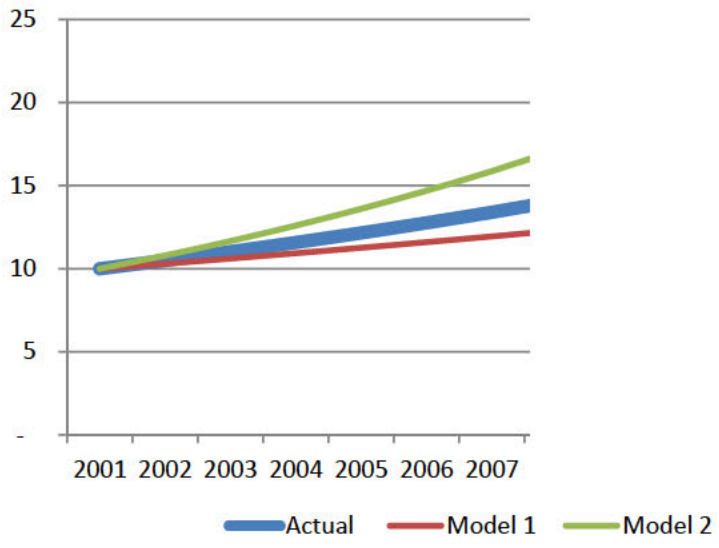


What goes wrong?

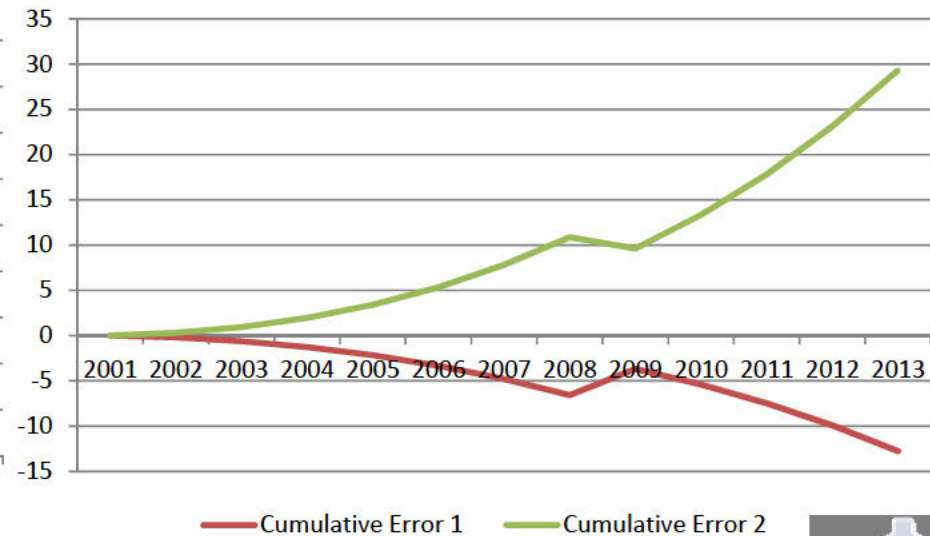
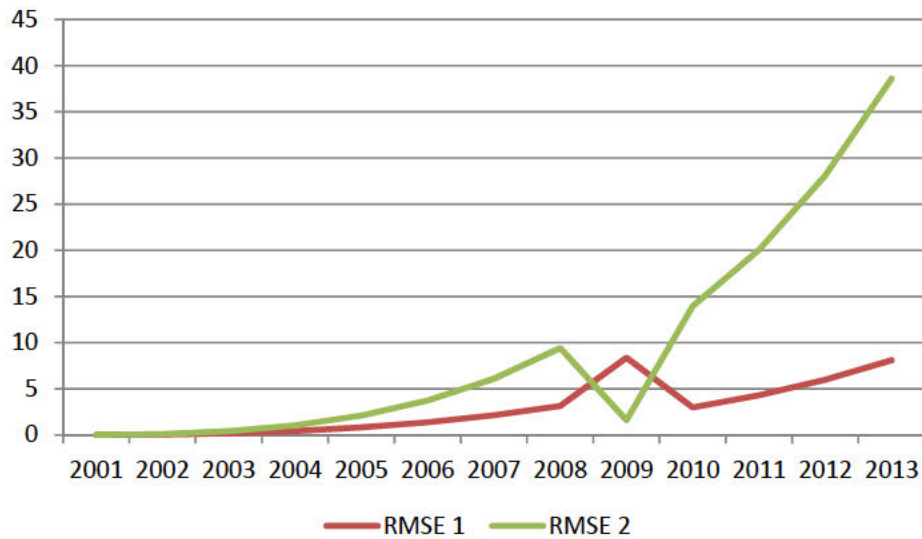
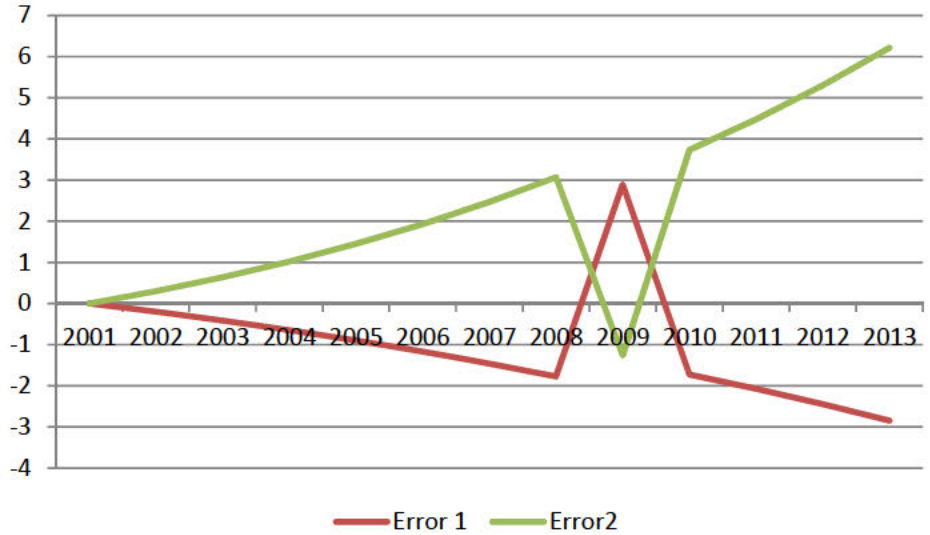
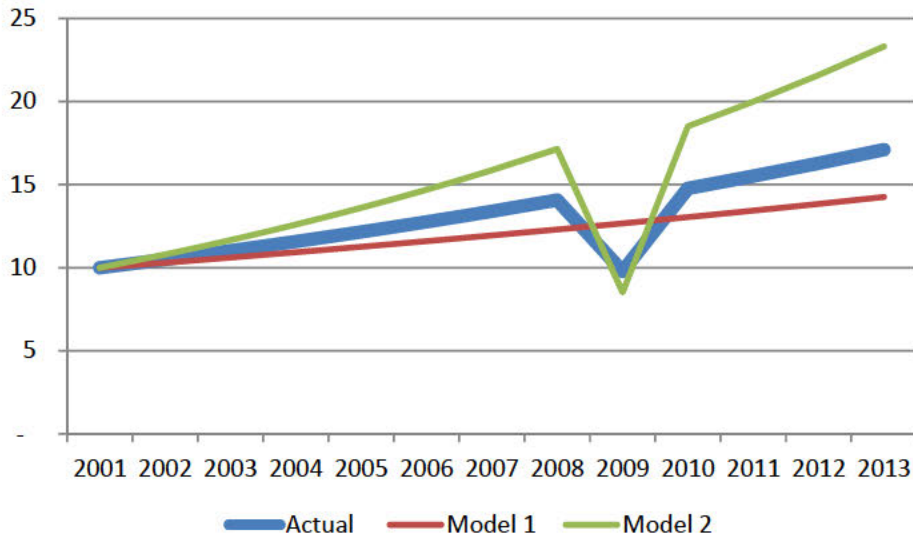
- Not enough data to test
- Data is not consistent
 - Reporting changes over time: fields are added, moved, recoded, and removed quarter-to-quarter
 - Firms change over time: businesses are acquired and divested
 - “One-time” items add noise (or are not really one-time)
- Underlying relationships are not stable
- Unobservable variables matter and are correlated with macro data



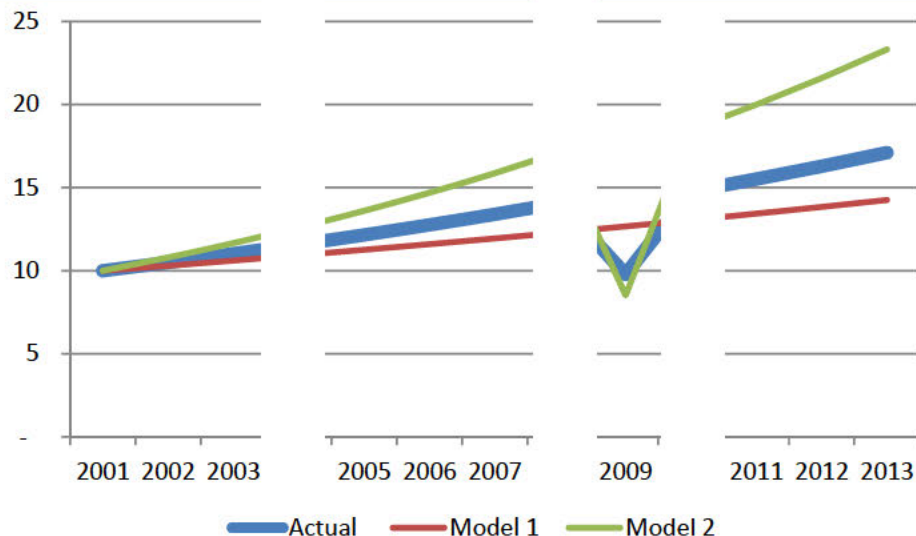
What's the best model?



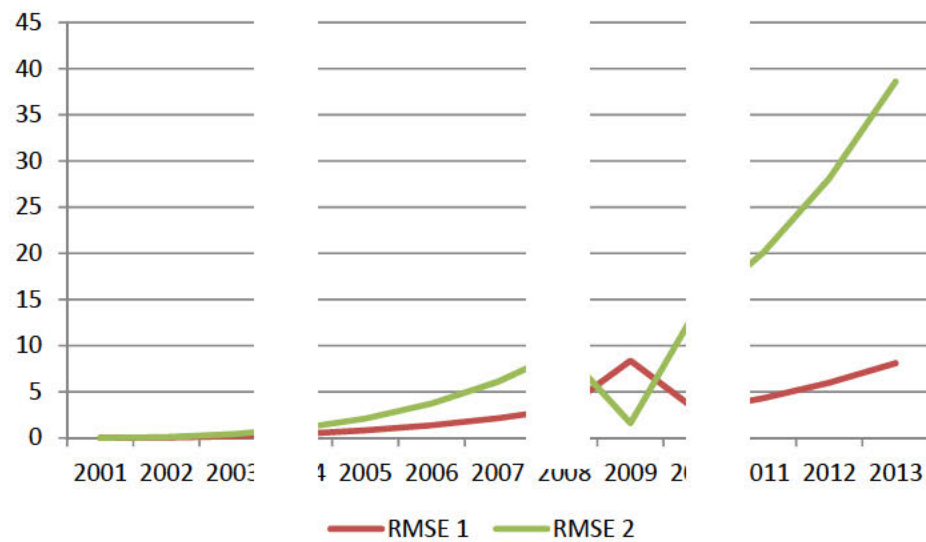
What's the best model?



What does this problem usually look like?



- Missing data
- Missing data is correlated with macroeconomic conditions
 - Acquisitions
 - Business model changes (fees, accounts etc)
- Not enough variation in macro conditions since business changes



Metrics

- So are you done if you look at enough metrics?
 - Errors
 - Root mean squared errors
 - Cumulative errors
 - Macro sensitivity
 - Forecasts in different scenarios
- Performance in different macro scenarios
 - Downturns
 - Interest income and expense → quarters with changing interest rates, other countries with negative rates
 - Non interest income → stock market drops and VIX increases
 - Changing competition

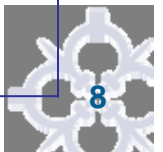
**OF
COURSE
NOT**



Performance evaluation

Method	Approaches	Pitfalls	Opinions*
Back testing	<ul style="list-style-type: none"> • Errors <ul style="list-style-type: none"> ➤ Simple, cumulative, RMSE ➤ In sample ➤ Out-of-sample • Different data <ul style="list-style-type: none"> ➤ Industry data ➤ Similar firms data ➤ Time periods 	<ul style="list-style-type: none"> • Data is not available • Business has “changed” 	<i>Given data issues, best to make smart assumptions and try to incorporate as much data from crisis as possible</i>
Benchmark models	<ul style="list-style-type: none"> • Aggregating vs. disaggregating • Different approaches <ul style="list-style-type: none"> ➤ Bayesian shrinkage ➤ Top down 	<ul style="list-style-type: none"> • Different data may be available for different approaches 	<i>The more different the benchmark, the more you can learn</i>
Stability	<ul style="list-style-type: none"> • How much do coefficients change in different time periods? • Weighted by actual scenarios to get a single number 	<ul style="list-style-type: none"> • Over-fitting and excessive series breaks 	<i>Models that are unstable may mean that the underlying relationships do not exist</i>

* Emphasizing that these are my opinions and not necessarily those of the Federal Reserve



Choosing among models

- Easy to choose if a model has better performance across all dimensions
- Weighting of different metrics can spark the right conversations about the objective of the forecasts
 - “intended to capture how ... each BHC would be affected by the macroeconomic and financial conditions described in the supervisory scenarios.... appropriately conservative and consistent with the purpose of a stress testing exercise”
(<http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20160623a1.pdf>)
- Other considerations such as complexity of implementation (model risk)
- Out-of-sample metrics and performance in severely adverse macro economic conditions



Understanding model(s)

- Consider putting errors (or RMSE) on the left hand side → what explains the model performance?
 - Macro conditions
 - Observable characteristics of portfolio
 - Breaks in performance that correspond to known business issues (regulatory changes, accounting)
- What change to the macro environment made the model unstable?
- Why are shorter and longer horizon errors giving different results?
- If the same model is estimated in different time periods how different are the results?

Interpretation of the performance statistics is key



“Why does it always have to represent something?”
The New Yorker, September 12, 2016

