Performance Testing Models within Systems

Nathaniel Hoover
Federal Reserve Bank of Boston
Presentation at Seventh Annual Stress Test Modeling Symposium
October 11, 2018

Background and Disclaimers

- Model Risk Management Group, an internal advisory team
  - Implementing industry standards and best practices for model risk management into the Federal Reserve's internal stress testing operations
- Today I will discuss issues and approaches to performance testing models that are a part of a large, inter-connected system of models
- This presentation is not intended to represent specific guidance for any firm relative to their CCAR, DFAST, or any other models

The views expressed in this presentation are my own and do not necessarily represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

Why move beyond individual model testing to system-level?

- Some (most?) models are primarily used as part of a larger system
- Ideally, measuring the performance of these models requires evaluating their impact on the performance of the system
- Today I will be discussing instances where individual-level model performance testing my not provide a full picture of the risk associated with a model when it interacts with other models
  - System features may make individual model test results difficult to interpret or insufficient
  - System information can be used to make individual model performance testing more informative

Background

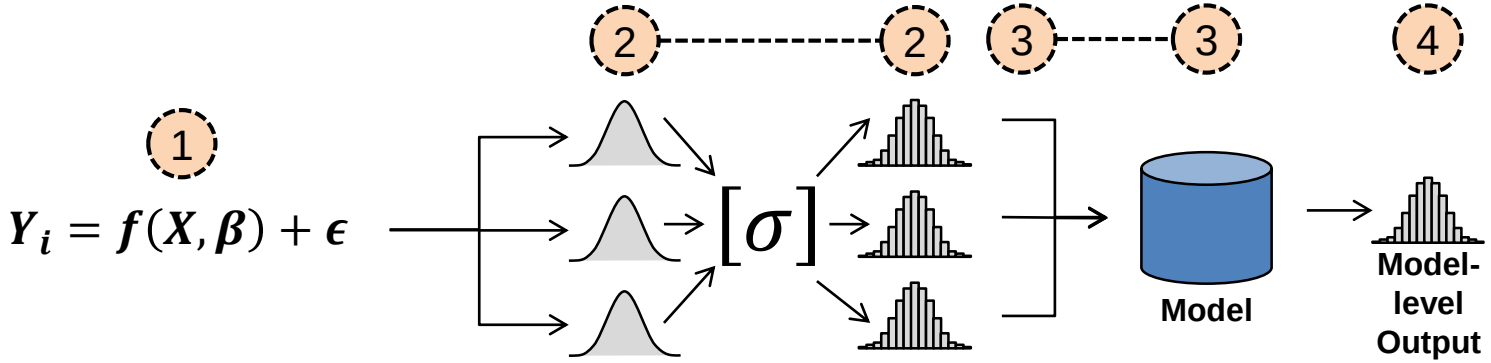
- Model performance testing definition
  - Any analysis that is primarily quantitative in nature that is intended to evaluate the ability of a model to produce accurate and stable estimates
- Objectives of performance testing
  - Support the model theory, specification, and variable selection
  - Assess the accuracy of model estimates and forecasts
  - Evaluate stability of the model
  - Identify limitations on appropriate usage

Direct Testing of System Outputs to Address Potential Issues

- Issue: Reasonable model outputs combine to produce unreasonable system-level results
  - Sub-model results sends a denominator to zero, cause negative rates, etc.
  - Sub-models that test well may combine to produce inaccurate final results
    - Incorrect formulation used to combine accurate intermediate outputs
    - Not properly accounting for correlation across intermediate outputs/errors
- Approach: Direct final output benchmarking, back testing, and limitations testing
  - Evaluate final system-level outputs by comparing directly to alternative models and observed results
  - Ensures models are properly integrated
  - Ensures sub-models are conceptually consistent with system design
  - Several peer reviewed, “top-down” models are available to benchmark “bottom-up” stress testing systems

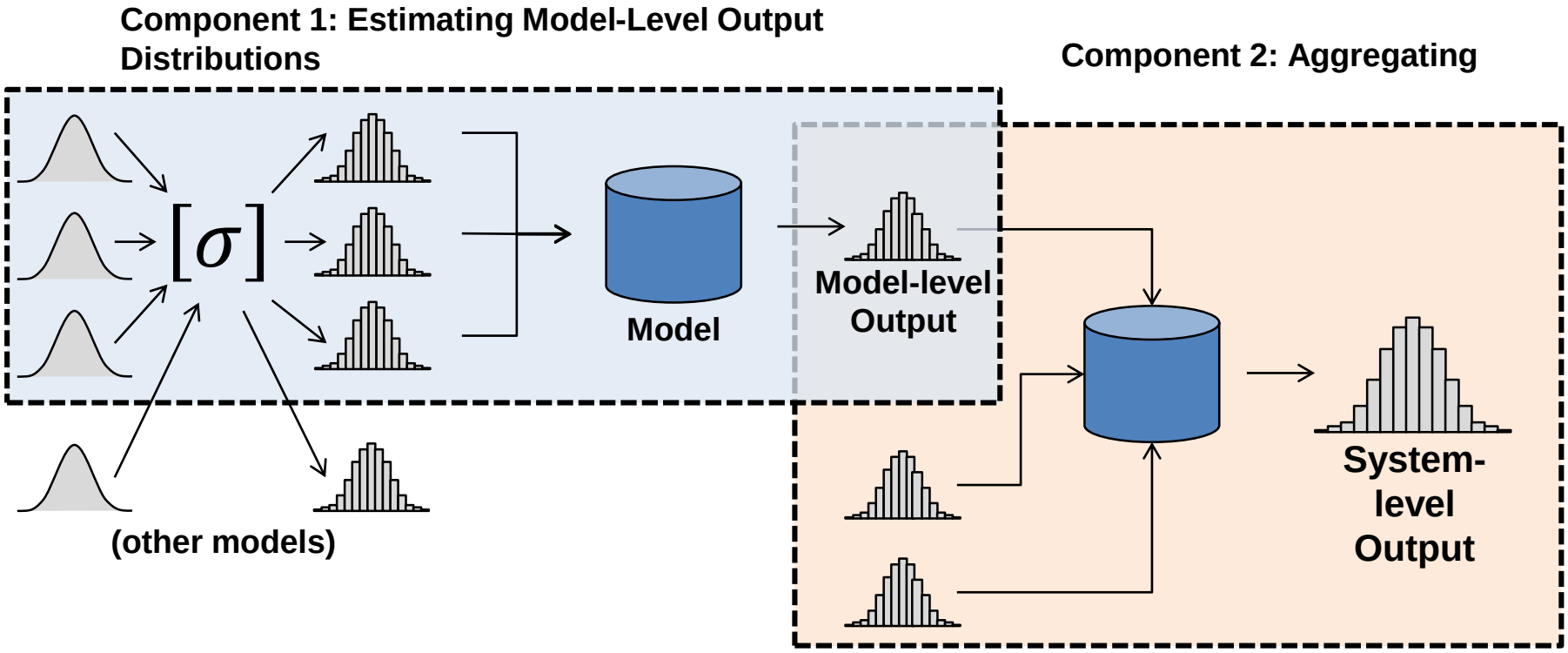
Direct Estimation of System Uncertainty and Sensitivities

- Issue: Final output uncertainty and sensitivities are difficult to measure
  - Goal: estimate aggregate uncertainty surrounding final model outputs and identify sources of risk
  - Impact of model-level parameter estimation error and prediction uncertainty may be difficult to assess as it propagates through models
- Approach: Bootstrapping system uncertainty using model-level measures.
  1. Assess individual models to identify estimated parameters and error terms
  2. Estimate joint distribution of parameter estimates and joint distribution of forecast errors
  3. Repeatedly sample from distributions
  4. Produce model-level outputs for each sample to produce a distribution of outcomes



Direct Estimation of System Uncertainty and Sensitivities

5. For each draw, aggregate model-level results and calculate distribution of final estimates



- Difficulties: Estimating a very large correlation matrix (computational burden), estimating distribution around "non-modeled" parameters, and logistics of simultaneously producing and combining results from all models.

Enhancing Individual Model Performance Testing Using System Data

- Issue: Correlation in model-level forecast errors may result in unexpected uncertainty at system level
- Approach: Joint evaluation of model-level forecast errors
  - Can be statistical or visual test
  - Correlation in individual model forecast errors can indicate that system forecast error will be larger (smaller) than expected
  - May argue for joint estimation of individual models
- Issue: Sensitivity testing using deviations may not utilize all information
  - Perturbing inputs by a uniform amount (percent) does not account for different volatility of individual inputs
- Approach: Evaluating sensitivity of downstream models using measures of uncertainty of outputs from upstream models
  - Sensitivity analyses may benefit from empirically determining the amount to perturb inputs when conducting sensitivity tests

Performance Tests of Individual Models Using System Data

- Issue: Difficult to interpretation individual model performance metrics
  - It may be difficult to assess some metrics in isolation ("good" performance is always relative)
- Approach: Evaluate individual models that are known to be economically related jointly to inform performance evaluation
    - E.g. Delinquency rates and income from account late fees estimated separately. Sensitivity to macro factors across both models likely should be similar.

Concluding Thoughts

- Increasingly attempting applying model-level principles to the system of models wherever possible
- This can result in improved identification and assessment of model risk
- This process is challenging both technically and logistically
- This is a process that is a work in progress within our group (far from completed and integrated into decision making processes).