

TIMS: Tests for building model confidence (Forrester and Senge, 1980)

- Model Testing: Comparing a model to the empirical reality to corroborate or refute the model.
- Model Validation: The process to establish confidence in a model's soundness and usefulness.
  - Soundness: Plausibility/Robustness of a model
    - ◆ The process to accumulate model builders' confidence
    - ◆ The ability to transfer confidence from model builders to model users
  - Usefulness:
    - ◆ The Scientific Usefulness: a useful model should be able to:
      - Generate insights into structures of reality
      - Make prediction
      - Stimulate meaningful research questions
    - ◆ Usefulness to practical leaders:
      - Explain causes of important problems
      - Provide basis for policy/improvement making
  - Confidence of a SD model accumulates through passing various tests
- Three dimensions of model testing: Model Structure (4), Model Behavior, and Policy Implications
- Tests of model structure: focus on structure and parameters
  - Structure-Verification Test
    - ◆ Model structure should exist in real world, and no contradiction to existing knowledge
    - ◆ Verify model assumptions through expert review, compare to literature, model builder's experience, comments from those with real world experiences
  - Parameter-Verification Test
    - ◆ Every parameter should have corresponding element in the real world
    - ◆ The numerical value of each parameter should be reasonable
    - ◆ A parameter may become a model variable when model purpose changed or time frame extended
  - Extreme-Conditions Test
    - ◆ A model should generate plausible output at extreme conditions
    - ◆ Extreme-Conditions Test could help to identify model structural flaws, discover omitted variables, and enhance the usefulness of model outside historical region
    - ◆ For each Rate Variable, trace back to its causal origin Level Variable, then evaluate the impact when that Level were in extreme consitions

- Boundary-Adequacy (Structure) Test:
  - ◆ Verify whether the model's aggregate level is appropriate
  - ◆ Make sure all relevant structures are included in model
  - ◆ The Model builder needs to clarify criticisms on model boundary from criticisms on model purpose
- Dimensional Consistency Test
  - ◆ The unit/dimension of all model variables should be consistent
  - ◆ This test usually operates jointly with parameter verification test
- Tests of model behavior
  - Behavior Reproduction Tests
    - ◆ Symptom generation: use known structure or policy to regenerate symptoms
    - ◆ Frequency/relative phase generation: able to regenerate frequency features
    - ◆ Multiple mode test: the ability to generate multiple mode of system behaviors
    - ◆ Behavior characteristics test: the ability to regenerate special behavior pattern or event without feeding historical time series data
  - Behavior Prediction Tests
    - ◆ Test model's ability to predict future behaviors
    - ◆ Pattern prediction test: generates qualitatively correct future behavior
    - ◆ Event prediction test:
      - identify the conditions for certain changes to occur
      - reveal the dynamic nature of an event
  - Behavior Anomaly Test
    - ◆ Trace the source of anomalous model behavior
      - Could help to discover model flaws
      - May be used as a support for some model assumptions
  - Family-Member Test
    - ◆ A good model should be able to represent other similar system (family members) through parameter changes (external validity)
  - Surprise Behavior Test
    - ◆ When encounter unexpected behavior
      1. Seek possible causes with the model, then
      2. compare to behaviors and causes in the real world, then
      3. identification of unrecognized real-world behavior
  - Extreme Policy Test
    - ◆ Test whether the model behaves as expected under extreme policy

- conditions/policy changes
- Boundary-Adequacy (behavior) Test
  - ◆ To make sure the model includes all necessary structures
  - ◆ Add/remove additional structure to analyze the behavior changes
  - ◆ Require model builders' ability of finding new concepts and analyzing the impact of the new concepts
- Behavior Sensitivity Test
  - ◆ Test whether plausible parameter changes will cause the model fail on previous tests
  - ◆ Typically SD models are insensitive to plausible parameter changes
  - ◆ Discovering sensitive parameters could be good to policy impacts
- Kalman filter
- Tests of policy implications
  - System improvement test
    - ◆ Whether effective policies to the model remain effective in the real world
    - ◆ It's difficult to prove the policy effectiveness, however confidence should be achieved through other tests
  - Changed-behavior prediction test
    - ◆ Whether a model can predict real world behavior change according to governing policy changes
    - ◆ Compare model responses with the historical real world responses to similar policy changes
    - ◆ Compare model responses with responses from similar real world systems
  - Boundary adequacy (policy) test
    - ◆ Test whether add/remove some structure will alter the effectiveness of certain policy
  - Policy sensitivity test
    - ◆ Test whether policy implications are insensitive to plausible parameter changes
- Core Tests
  - All structure tests
  - Core behavior tests
    - ◆ Behavior reproduction
    - ◆ Behavior anomaly
    - ◆ Behavior sensitivity
  - Core policy implication tests

- ◆ Change behavior prediction
- ◆ Policy sensitivity
- Why statistical tests are not appropriate to test structure
  - There are long debates on the using statistical correlation to test causal relation
  - Bias in hypothesis is one of the causes to statistical significance
    - ◆ Multi-collinearity among Xs
    - ◆ Measurement error
  - The ability of Point-by-point goodness of fit test to identify useful models is very low