ISSUES IN VALIDATING AGENT-BASED MODELS

Robert Marks, University of New South Wales

Daniel Klapper, Goethe University

David Midgley, INSEAD

2009 Computing in Economics and Finance
University of Technology, Sydney
July 15–17

Objective

Building on various literatures, Midgley, Marks and Kunchamwar (2007) proposed a procedure for “assuring” ABMs:
Assuring = Verifying & Validating

- Does the code implement the model?
- Do the model outputs behave reasonably?
- Do these outputs fit empirical data or stylized facts?

Our objective here is to illustrate what we have learnt since 2007, particularly about the practical hurdles in assuring ABMs, & to raise some unresolved issues.

Objective

Building on various literatures, Midgley, Marks and Kunchamwar (2007) proposed a procedure for “assuring” ABMs:
Assuring = Verifying & Validating

- Does the code implement the model?
- Do the model outputs behave reasonably?
- Do these outputs fit empirical data or stylized facts?

Our objective here is to illustrate what we have learnt since 2007, particularly about the practical hurdles in assuring ABMs, & to raise some unresolved issues.

We use the same ABM, which has moved on from Version 1 (2007) to Version 3 (2009).

Objective

Building on various literatures, Midgley, Marks and Kunchamwar (2007) proposed a procedure for "assuring" ABMs:
Assuring = Verifying & Validating

- Does the code implement the model?
- Do the model outputs behave reasonably?
- Do these outputs fit empirical data or stylized facts?

Our objective here is to illustrate what we have learnt since 2007, particularly about the practical hurdles in assuring ABMs, & to raise some unresolved issues.

We use the same ABM, which has moved on from Version 1 (2007) to Version 3 (2009).

Our main focus is on validation, although we first need to outline the model and the verification results.

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The essential paradigm is *learning by doing:*

- **Brands & stores retain a detailed memory of recent actions & results; their goals are to improve their profits.**
-

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The essential paradigm is *learning by doing:*

- **Brands & stores retain a detailed memory of recent actions & results; their goals are to improve their profits.**
- **Consumers learn about brands from the environment & their experience; they vary according to the importance they place on price, quality, or affect in their decisions.**
-

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The essential paradigm is *learning by doing:*

- **Brands & stores retain a detailed memory of recent actions & results; their goals are to improve their profits.**
- **Consumers learn about brands from the environment & their experience; they vary according to the importance they place on price, quality, or affect in their decisions.**
- **Consumers do not have detailed memories, just a "choice set" of preferred brands.**
-

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The essential paradigm is *learning by doing:*

- **Brands & stores retain a detailed memory of recent actions & results; their goals are to improve their profits.**
- **Consumers learn about brands from the environment & their experience; they vary according to the importance they place on price, quality, or affect in their decisions.**
- **Consumers do not have detailed memories, just a "choice set" of preferred brands.**
- **They react to the stimuli they receive shortly before or during their shopping trip; their goals are to maximize their satisfaction.**

The ABM is of the manufacturers & retailers who provide a frequently purchased product & the end-customers who buy and consume it.

The implementation here has 5 manufacturer agents (*brands*), 2 retail store agents (*stores*) & 600 end-customer agents (*consumers*).

The marketing environment includes product quality, retail & wholesale price, store promotions, manufacturer advertising, & promotional discounts to retailers (detergent brands in a German city).

The essential paradigm is *learning by doing:*

- **Brands & stores retain a detailed memory of recent actions & results; their goals are to improve their profits.**
- **Consumers learn about brands from the environment & their experience; they vary according to the importance they place on price, quality, or affect in their decisions.**
- **Consumers do not have detailed memories, just a "choice set" of preferred brands.**
- **They react to the stimuli they receive shortly before or during their shopping trip; their goals are to maximize their satisfaction.**

The ABM has a 14-page specification and, while complex, is a simplification of reality.

Three important aspects of this model:

1.

Three important aspects of this model:

1. Three classes of agents with conflicting goals: brands, stores, consumers.
2.

Three important aspects of this model:

1. Three classes of agents with conflicting goals:
   brands, stores, consumers.
2. High-involvement, data-driven decision making:
   brands maximize their profits, and
   stores maximize their profits,
3.

Three important aspects of this model:

1. Three classes of agents with conflicting goals:
   brands, stores, consumers.
2. High-involvement, data-driven decision making:
   brands maximize their profits, and
   stores maximize their profits,
3. Low-involvement decision-making:
   consumers maximize their satisfaction.

Seven issues addressed here:

1. We need *norms* for verification
2. When and how often to verify?
3. Degrees of freedom: Our views on validation have changed
4. *Pre-validation*, incomplete data and scaling
5. Computing power needed
6. Which data to fit and how to fit them?
7. How to test the “reasonableness” of the ABM?

Issue 1: We need *norms* for verification

Why is verification this necessary?

-

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
-

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

The ABM field needs norms for verification; these could include:

1.

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

The ABM field needs norms for verification; these could include:

1. **A written specification of the model as an online technical appendix,**
2.

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

The ABM field needs norms for verification; these could include:

1. **A written specification of the model as an online technical appendix,**
2. **An agreed process for checking that the code matches this specification, with metrics for showing acceptable matching,**
3.

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

The ABM field needs norms for verification; these could include:

1. **A written specification of the model as an online technical appendix,**
2. **An agreed process for checking that the code matches this specification, with metrics for showing acceptable matching,**
3. **The code being made available to others (online).**

Issue 1: We need *norms* for verification

Why is verification this necessary?

- **Journal reviewers might review the specification, but it is difficult to imagine them checking the code!**
- **Yet, without these checks, how can one be sure the code implements the proposed model?**

The ABM field needs norms for verification; these could include:

1. **A written specification of the model as an online technical appendix,**
2. **An agreed process for checking that the code matches this specification, with metrics for showing acceptable matching,**
3. **The code being made available to others (online).**

Here, we had *two independent coders* check the most important procedures against our specification.

Issue 1: We need *norms* for verification

Why is verification this necessary?

- Journal reviewers might review the specification, but it is difficult to imagine them checking the code!
- Yet, without these checks, how can one be sure the code implements the proposed model?

The ABM field needs norms for verification; these could include:

1. A written specification of the model as an online technical appendix,
2. An agreed process for checking that the code matches this specification, with metrics for showing acceptable matching,
3. The code being made available to others (online).

Here, we had *two independent coders* check the most important procedures against our specification.

This is a common approach in developing commercial software, although other approaches exist, including:

— Tool-based or automated code analysis, deriving automata from the program to check theorems, and finite state verification.

Issue 2: When and how often to verify?

Our ABM contains 1900 lines of Java, although most are input/output, housekeeping, or standard library functions (& hence not verified).

Issue 2: When and how often to verify?

Our ABM contains 1900 lines of Java, although most are input/output, housekeeping, or standard library functions (& hence not verified).

The two independent coders raised issues about 7% of the lines in the important procedures (the core of the model):

- 4% coding errors
- 2% where the specification was not followed but the code yielded the desired result
- 1% code that did not do anything

Issue 2: When and how often to verify?

Our ABM contains 1900 lines of Java, although most are input/output, housekeeping, or standard library functions (& hence not verified).

The two independent coders raised issues about 7% of the lines in the important procedures (the core of the model):

- 4% coding errors
- 2% where the specification was not followed but the code yielded the desired result
- 1% code that did not do anything

This was followed by a "review of the reviewers" & code modification

Issue 2: When and how often to verify?

Our ABM contains 1900 lines of Java, although most are input/output, housekeeping, or standard library functions (& hence not verified).

The two independent coders raised issues about 7% of the lines in the important procedures (the core of the model):

- 4% coding errors
- 2% where the specification was not followed but the code yielded the desired result
- 1% code that did not do anything

This was followed by a "review of the reviewers" & code modification

But of course ABMs evolve:

— we are now on Version 3, while verification was of Version 2.

This raises the issues of when and how often is verification needed?

— both to support progress in the research project and for peer review?

Issue 2: When and how often to verify?

Our ABM contains 1900 lines of Java, although most are input/output, housekeeping, or standard library functions (& hence not verified).

The two independent coders raised issues about 7% of the lines in the important procedures (the core of the model):

- 4% coding errors
- 2% where the specification was not followed but the code yielded the desired result
- 1% code that did not do anything

This was followed by a “review of the reviewers” & code modification

But of course ABMs evolve:
— we are now on Version 3, while verification was of Version 2.

This raises the issues of when and how often is verification needed?
— both to support progress in the research project and for peer review?

Also how much of the code should be verified, given the expense of doing this?

Issue 3: Our views on validation have changed

Issue 3: Our views on validation have changed

Our initial view of validation had three steps:

1. Check that the model behaves “reasonably”
2. Revise if necessary, and
3. Either fit it to data or compare its outputs with stylized facts.

Issue 3: Our views on validation have changed

Our initial view of validation had three steps:

1. Check that the model behaves "reasonably"
2. Revise if necessary, and
3. Either fit it to data or compare its outputs with stylized facts.

& our view of the first step was heavily influenced by Miller's (1998) ANTS (Automated Non-linear Testing Systems).

ANTS uses optimization methods to test the sensitivity of the model to perturbations in parameter values

e.g. find the combinations of small changes in values which produce wildly different outcomes (Club of Rome World3 model)

Issue 3: Our views on validation have changed

Our initial view of validation had three steps:

1. **Check that the model behaves "reasonably"**
2. **Revise if necessary, and**
3. **Either fit it to data or compare its outputs with stylized facts.**

& our view of the first step was heavily influenced by Miller's (1998) ANTS (Automated Non-linear Testing Systems).

ANTS uses optimization methods to test the sensitivity of the model to perturbations in parameter values

e.g. find the combinations of small changes in values which produce wildly different outcomes (Club of Rome World3 model)

But, in a high-dimensional parameter space (LeBaron & Tesfatsion 2009)—Version 2 had 37 d.f.—and with *arbitrary starting values* for these parameters, it is not clear what such results mean:

e.g. we could be testing "reasonableness" around values far from those that would fit empirical data.

Issue 3: Our views on validation have changed

Our initial view of validation had three steps:

1. **Check that the model behaves "reasonably"**
2. **Revise if necessary, and**
3. **Either fit it to data or compare its outputs with stylized facts.**

& our view of the first step was heavily influenced by Miller's (1998) ANTS (Automated Non-linear Testing Systems).

ANTS uses optimization methods to test the sensitivity of the model to perturbations in parameter values

e.g. find the combinations of small changes in values which produce wildly different outcomes (Club of Rome World3 model)

But, in a high-dimensional parameter space (LeBaron & Tesfatsion 2009)—Version 2 had 37 d.f.—and with *arbitrary starting values* for these parameters, it is not clear what such results mean:

e.g. we could be testing "reasonableness" around values far from those that would fit empirical data.

So we now think of "pre-validation"—getting a rough fit to empirical data before applying the ANTS perturbations.

Issue 4: *Pre-validation*, incomplete data and scaling

It is difficult to search high-dimensional spaces for even a "rough" fit to empirical data.

Especially when the ABM is non-linear & has stochastic elements which make the optimization objective function "noisy".

Issue 4: *Pre-validation*, incomplete data and scaling

It is difficult to search high-dimensional spaces for even a "rough" fit to empirical data.

Especially when the ABM is non-linear & has stochastic elements which make the optimization objective function "noisy".

One solution proposed in the literature is to use whatever external data to "micro-calibrate" as many parameters as possible, leaving only a smaller number to be fitted.

Issue 4: *Pre-validation*, incomplete data and scaling

It is difficult to search high-dimensional spaces for even a "rough" fit to empirical data.

Especially when the ABM is non-linear & has stochastic elements which make the optimization objective function "noisy".

One solution proposed in the literature is to use whatever external data to "micro-calibrate" as many parameters as possible, leaving only a smaller number to be fitted.

Here, our ultimate objective is to reproduce brand and store sales over 53 weeks, and we focus on the sales of the 5 main detergent brands in 2 stores in a German city.

These 5 brands and 2 stores represent about 75% of the market.

But we also obtained consumer panel data to micro-calibrate consumption & purchase amounts, starting brand shares, and probabilities of buying on promotion.

Issue 4, continued—Incomplete data and scaling

From Panel Data:

Agents who buy in:	Store 1 only	Store 2 only	Both stores
Light buyers	100 agents	100 agents	100 agents
Heavy buyers	100 agents	100 agents	100 agents

Need to scale these store-level data.

Issue 4, continued—Incomplete data and scaling

After exploratory analysis of the panel data, we decided to represent the consumer agents as 6 types, and:

- we decided that each agent would represent 10 real consumers, i.e. 6000 total

Issue 4, continued—Incomplete data and scaling

After exploratory analysis of the panel data, we decided to represent the consumer agents as 6 types, and:

- we decided that each agent would represent 10 real consumers, i.e. 6000 total

But we do not know how the panel sample relates to the 3 stores populations

- Anecdotally, the panel is thought to be heavily skewed.

Issue 4, continued—Incomplete data and scaling

After exploratory analysis of the panel data, we decided to represent the consumer agents as 6 types, and:

- we decided that each agent would represent 10 real consumers, i.e. 6000 total

But we do not know how the panel sample relates to the 3 stores populations

- Anecdotally, the panel is thought to be heavily skewed.

So we introduced scaling factors as parameters to estimate.

Issue 4, continued—Incomplete data and scaling

After exploratory analysis of the panel data, we decided to represent the consumer agents as 6 types, and:

- we decided that each agent would represent 10 real consumers, i.e. 6000 total

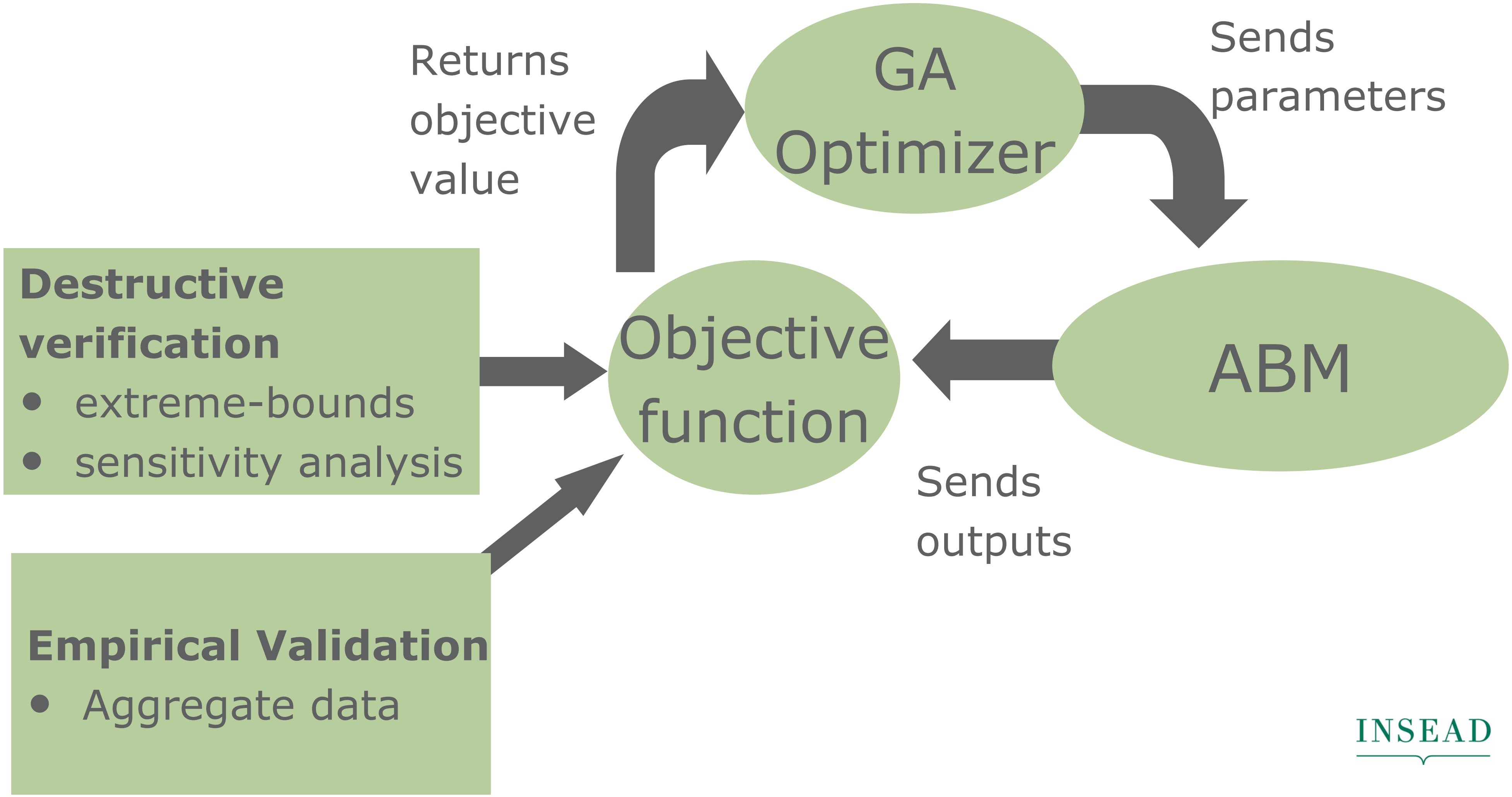
But we do not know how the panel sample relates to the 3 stores populations

- Anecdotally, the panel is thought to be heavily skewed.

So we introduced scaling factors as parameters to estimate.

Issue: On the one hand the panel reduces the number of parameters to estimate, but on the other it introduces scaling complications.

Work at the macro-level: Embed the ABM in an Automated Nonlinear Testing System



INSEAD

Issue 5: Computing power needed

As well as micro-calibrating some aspects of the consumer agents, we also:

- Fix some parameters arbitrarily (e.g. the size of a brand or store's memory of previous results)
- and for rough fitting we focus on those remaining parameters that previous testing reveals outputs are sensitive to.

Currently, this leaves 17 brand, store and consumer parameters to estimate (17 d.f.) — see below.

Issue 5: Computing power needed

As well as micro-calibrating some aspects of the consumer agents, we also:

- Fix some parameters arbitrarily (e.g. the size of a brand or store's memory of previous results)
- and for rough fitting we focus on those remaining parameters that previous testing reveals outputs are sensitive to.

Currently, this leaves 17 brand, store and consumer parameters to estimate (17 d.f.) — see below.

The ABM also has transitory state values, so we need to run it for a number of periods before we can extract valid outputs for estimation (the "burn-in").

Issue 5: Computing power needed

As well as micro-calibrating some aspects of the consumer agents, we also:

- Fix some parameters arbitrarily (e.g. the size of a brand or store's memory of previous results)
- and for rough fitting we focus on those remaining parameters that previous testing reveals outputs are sensitive to.

Currently, this leaves 17 brand, store and consumer parameters to estimate (17 d.f.) — see below.

The ABM also has transitory state values, so we need to run it for a number of periods before we can extract valid outputs for estimation (the "burn-in").

On a PC, one run of the ABM with a fixed set of parameters takes a few seconds,

But, once the ABM is embedded within a Genetic Algorithm (GA) optimizer for estimation, this takes days and fries laptops!

We have therefore ported the code to a 300-node supercomputer.

Issue 5, continued: The 17 variables

Consumer threshold on satisfaction with brand experience
Parameters used to generate different rankings on affect, quality and price
Parameters used to generate different rankings on affect, quality and price
Parameters used to generate different rankings on affect, quality and price
Percentage markup on wholesale price to get retail price, Retailer 1
Percentage markup on wholesale price to get retail price, Retailer 2
Slotting fees - Retailer 1
Slotting fees - Retailer 2
Quality of brand 1
Probability price and advertising will be changed for brand 1
Quality of brand 2
Probability price and advertising will be changed for brand 2
Factors to scale consumer types up from panel to universe
Factors to scale consumer types up from panel to universe
Factors to scale consumer types up from panel to universe
Intercept on the unit cost of production equation
Slope on the unit cost of production equation

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every "tick" (or week).

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every "tick" (or week).

The answer to the question is determined by the empirical data one has.

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every "tick" (or week).

The answer to the question is determined by the empirical data one has.

But, as in all modeling, real data contain phenomena that one does not model!

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every “tick” (or week).

The answer to the question is determined by the empirical data one has.

But, as in all modeling, real data contain phenomena that one does not model!

Here our main issue is that our brand and retail agents decide to use store promotions at times different from those in the historical data.

•

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every "tick" (or week).

The answer to the question is determined by the empirical data one has.

But, as in all modeling, real data contain phenomena that one does not model!

Here our main issue is that our brand and retail agents decide to use store promotions at times different from those in the historical data.

- Which means there is inherently no match between the output & real time series for brand and store sales.

Issue 6: Which data to fit and how to fit them?

ABMs can potentially produce many output time series—it is easy to observe both the internal states and behavior of all agents at every "tick" (or week).

The answer to the question is determined by the empirical data one has.

But, as in all modeling, real data contain phenomena that one does not model!

Here our main issue is that our brand and retail agents decide to use store promotions at times different from those in the historical data.

- Which means there is inherently no match between the output & real time series for brand and store sales.

Our solution has been to sort both ABM outputs and actual data on both magnitudes and first differences, and seek the best fit between these.

This is a version of Operational Validity testing (Sargent 2005), although other statistical methods are possible.

Issue 6, continued: What and how to fit?

Issue 6, continued: What and how to fit?

Thus there are 14 time series fitted, with implicit constraints.

Issue 6, continued: What and how to fit?

Thus there are 14 time series fitted, with implicit constraints.

We give each series *equal* weight in the objective function, although other schemes are possible, including estimated weights.

And: we could fit other data, e.g. prices.

Issue 6, continued: What and how to fit?

Thus there are 14 time series fitted, with implicit constraints.

We give each series *equal* weight in the objective function, although other schemes are possible, including estimated weights.

And: we could fit other data, e.g. prices.

Or we could test the hypothesis that the simulated model output and the historical data are generated by the "same" process (up to a level of specificity).

Issue 7: How to test the "reasonableness" of the ABM?

Issue 7: How to test the "reasonableness" of the ABM?

We use the rough fit as the starting point for stress-testing the ABM.

Issue 7: How to test the "reasonableness" of the ABM?

We use the rough fit as the starting point for stress-testing the ABM.

Here, the GA starts from the rough-fit parameter values achieved in *pre-validation* and tries to perturb these values to achieve "unreasonable" behavior

Issue 7: How to test the "reasonableness" of the ABM?

We use the rough fit as the starting point for stress-testing the ABM.

Here, the GA starts from the rough-fit parameter values achieved in *pre-validation* and tries to perturb these values to achieve "unreasonable" behavior

The 5 objectives we use for the GA here are different, namely:

- **Maximize:**
  - **The total profits of the five brands**
  - **The total profits of the two stores**
  - **The market share of one brand across both stores**
  - **The sum of customer satisfaction**
- **Equalize market shares across the five brands (minimize standard deviation of the share distribution)**

Issue 7: How to test the “reasonableness” of the ABM?

We use the rough fit as the starting point for stress-testing the ABM.

Here, the GA starts from the rough-fit parameter values achieved in *pre-validation* and tries to perturb these values to achieve “unreasonable” behavior

The 5 objectives we use for the GA here are different, namely:

- Maximize:
  - The total profits of the five brands
  - The total profits of the two stores
  - The market share of one brand across both stores
  - The sum of customer satisfaction
- Equalize market shares across the five brands (minimize standard deviation of the share distribution)

And we then observe whether:

- The model breaks down in any sense
- Unrealistic parameter values (or combinations) appear
- Mutually inconsistent time series emerge
- The competing objectives of the brand, store and consumer agents are not being balanced.

Issue 7, continued: Testing the ABM

Results of our ANTS tests:

•

Issue 7, continued: Testing the ABM

Results of our ANTS tests:

- **Choosing the 17 variables to Equalize the brands’ market shares explodes the model.**
-

Issue 7, continued: Testing the ABM

Results of our ANTS tests:

- **Choosing the 17 variables to Equalize the brands' market shares explodes the model.**
- **Maximizing the retailers' profits, and**
  Maximizing the brands' profits
  — both appear to lead to convergence to some local optima, and have parameter values very different from the best fit, and
-

Issue 7, continued: Testing the ABM

Results of our ANTS tests:

- **Choosing the 17 variables to Equalize the brands' market shares explodes the model.**
- **Maximizing the retailers' profits, and**
  Maximizing the brands' profits
  — both appear to lead to convergence to some local optima, and have parameter values very different from the best fit, and
- **the revenue and profit figures seem to follow what might seem logical, given the two objectives.**

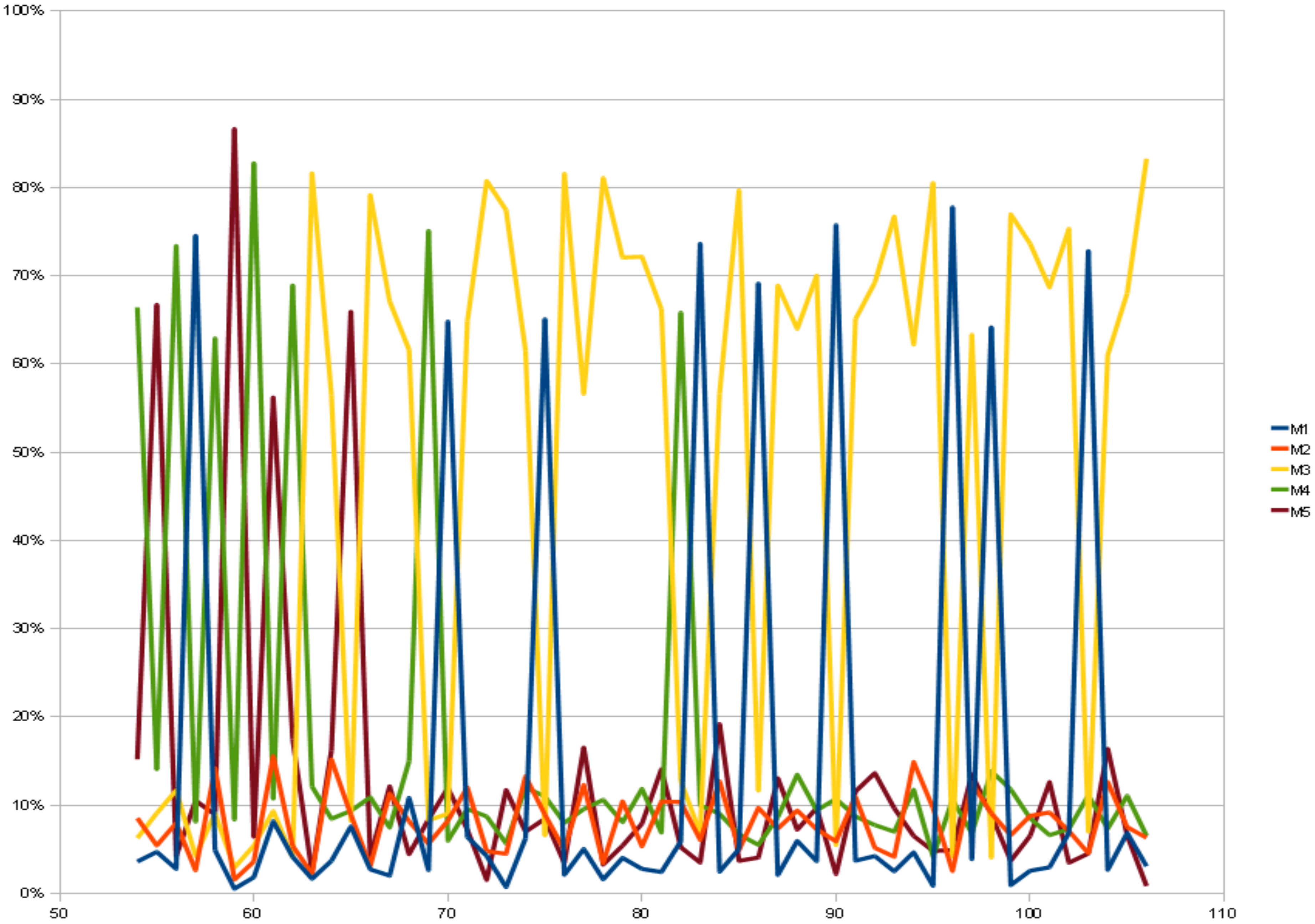
Issue 7, continued: Testing the ABM

Results of our ANTS tests:

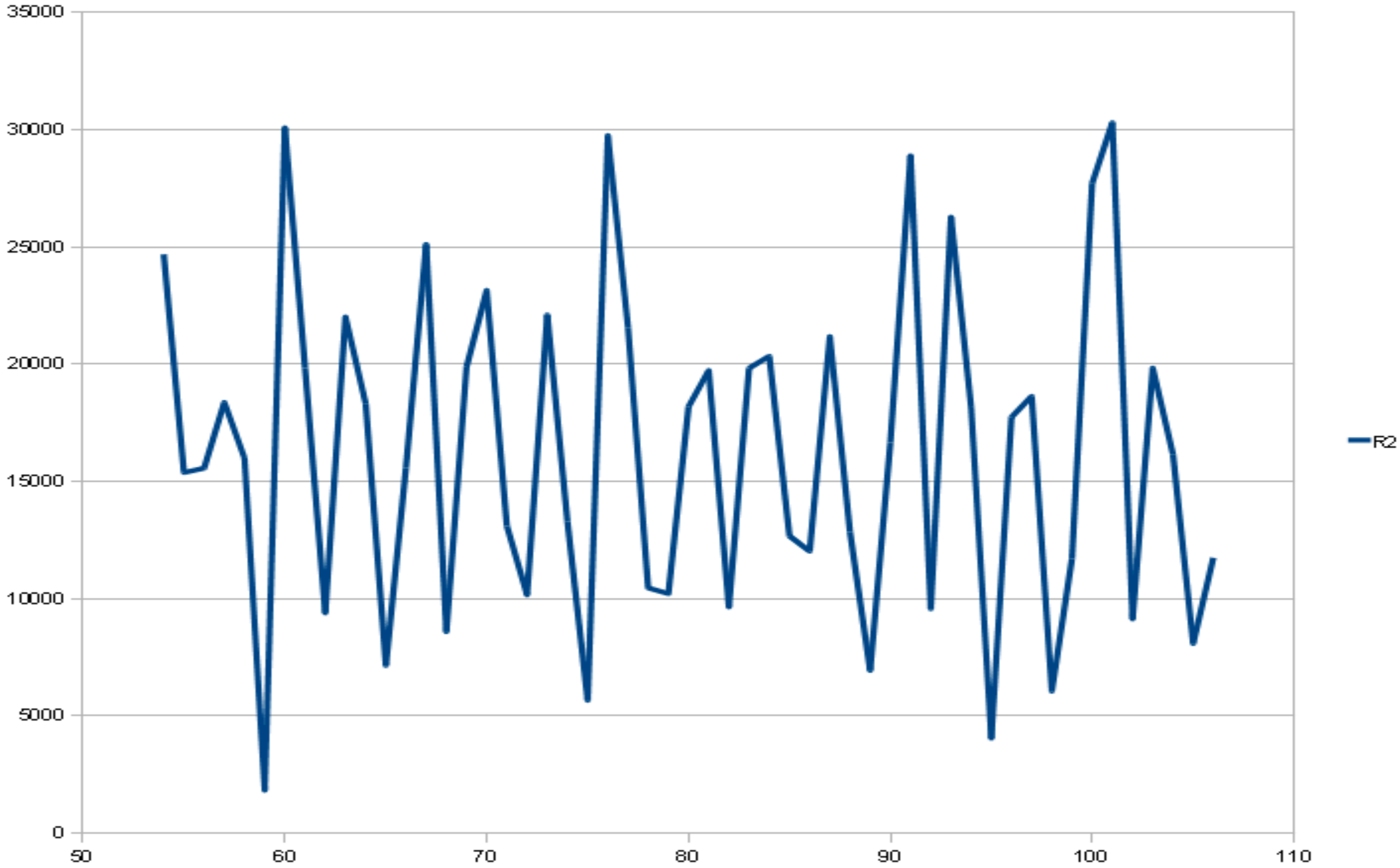
- **Choosing the 17 variables to Equalize the brands' market shares explodes the model.**
- **Maximizing the retailers' profits, and**
  Maximizing the brands' profits
  — both appear to lead to convergence to some local optima, and have parameter values very different from the best fit, and
- **the revenue and profit figures seem to follow what might seem logical, given the two objectives.**

We need to explore which of the 17 values offend, in these cases.

Market Shares of the Five Manufacturers in Store 2

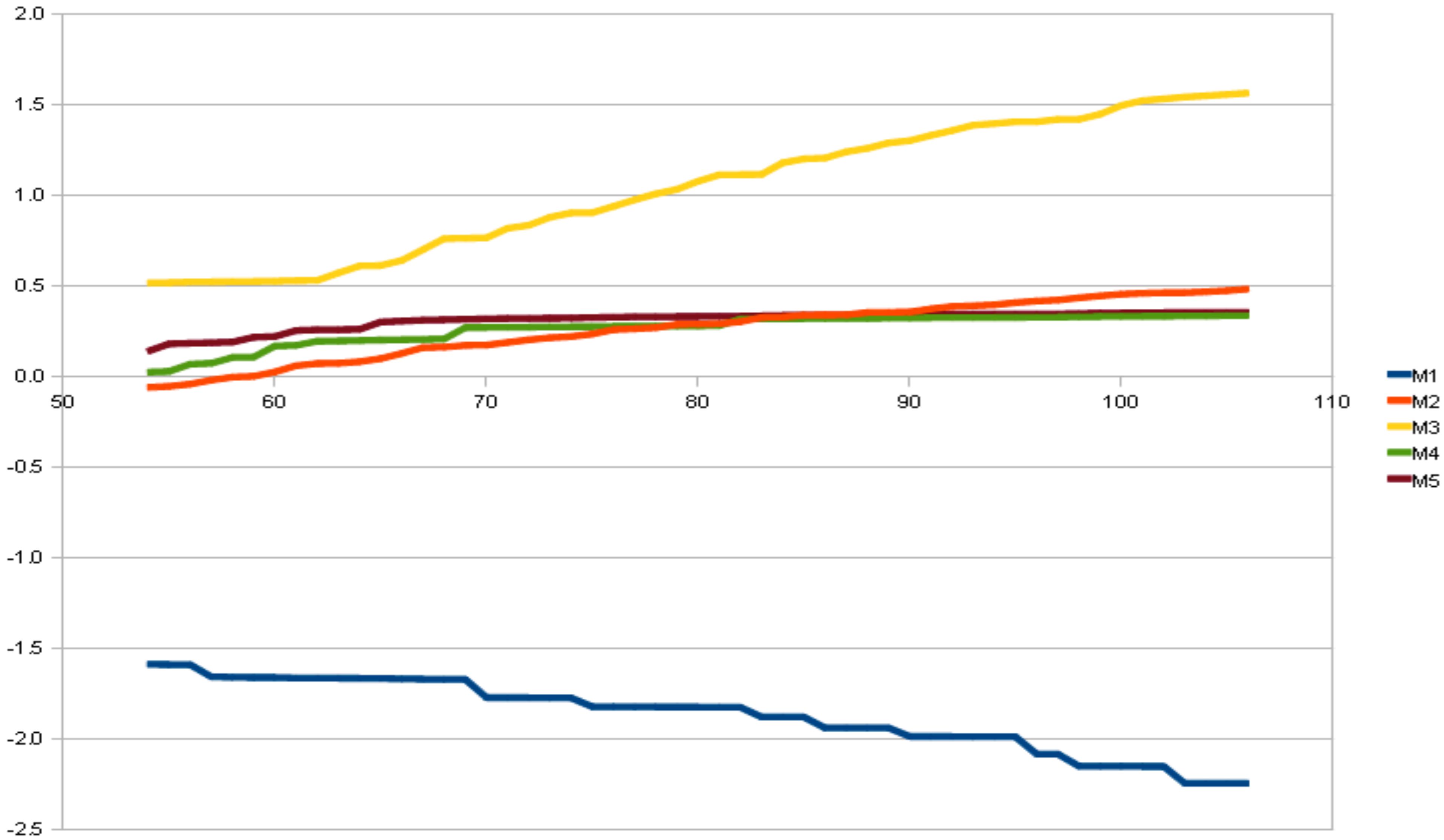


Retail Revenue for Store 2 (disguised)



Customer Satisfaction with the Five Manufacturers

(standard deviations from overall mean)



Issue 7, continued: Testing the ABM

Moreover, using the best pre-validation chromosome obtained to date, we find that the model behaves much as it should in a one-off run:

1. **The graph shows the spikes of the 5 Brands' market shares.**
2. **The graph shows the spikes in one Retailer's revenues.**
3. **The graph shows the slow evolution of higher Consumers' satisfaction.**

Where have we reached?

Where have we reached?

A rough fit which implies our output time series are within ±40% of the real data.

Where have we reached?

A rough fit which implies our output time series are within ±40% of the real data.

We are trying to find additional data to place constraints on, or externally estimate, the scaling factors, as the optimization is sensitive to these.

Where have we reached?

A rough fit which implies our output time series are within ±40% of the real data.

We are trying to find additional data to place constraints on, or externally estimate, the scaling factors, as the optimization is sensitive to these.

Some stress-testing completed:

- **This is gradually revealing the role of each of the focal parameters and where the model might need refining,**
- **although stress-testing also suffers from problems of high-dimensional spaces**
  E.g. easy enough to push one parameter until the model breaks, more difficult to uncover combinations that break it or produce unreasonable behavior.

Where have we reached?

A rough fit which implies our output time series are within ±40% of the real data.

We are trying to find additional data to place constraints on, or externally estimate, the scaling factors, as the optimization is sensitive to these.

Some stress-testing completed:

- **This is gradually revealing the role of each of the focal parameters and where the model might need refining,**
- **although stress-testing also suffers from problems of high-dimensional spaces**
  E.g. easy enough to push one parameter until the model breaks, more difficult to uncover combinations that break it or produce unreasonable behavior.

We also need to look at the other parameters which were not used to get a rough fit.

Where have we reached?

A rough fit which implies our output time series are within ±40% of the real data.

We are trying to find additional data to place constraints on, or externally estimate, the scaling factors, as the optimization is sensitive to these.

Some stress-testing completed:

- This is gradually revealing the role of each of the focal parameters and where the model might need refining,
- although stress-testing also suffers from problems of high-dimensional spaces
  E.g. easy enough to push one parameter until the model breaks, more difficult to uncover combinations that break it or produce unreasonable behavior.

We also need to look at the other parameters which were not used to get a rough fit.

Debating whether to fit store price as well as sales (this may also simplify some of the model).

At which point we can refine the model & move to a final close fit.

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

-

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

- **Verification: norms, metrics and tools for verifying code**
-

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

- **Verification: norms, metrics and tools for verifying code**
- **Pre-validation: quick and dirty methods for finding rough fits in high-dimensional spaces**
-

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

- **Verification: norms, metrics and tools for verifying code**
- **Pre-validation: quick and dirty methods for finding rough fits in high-dimensional spaces**
- **Validation: more debate on:**
  - **what is “reasonable” or “unreasonable” ABM behavior**
  - **what to fit, especially given growing sources of data, and**
  - **how to weight multiple outputs in objective functions**
-

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

- **Verification: norms, metrics and tools for verifying code**
- **Pre-validation: quick and dirty methods for finding rough fits in high-dimensional spaces**
- **Validation: more debate on:**
  - **what is "reasonable" or "unreasonable" ABM behavior**
  - **what to fit, especially given growing sources of data, and**
  - **how to weight multiple outputs in objective functions**
- **Better ways of automatically visualizing output data**
  - **Especially "latent" states and agent interactions**
-

Conclusions

If ABMs are to fulfill their promise, then we need better ways of assuring them, including:

- **Verification: norms, metrics and tools for verifying code**
- **Pre-validation: quick and dirty methods for finding rough fits in high-dimensional spaces**
- **Validation: more debate on:**
  - **what is “reasonable” or “unreasonable” ABM behavior**
  - **what to fit, especially given growing sources of data, and**
  - **how to weight multiple outputs in objective functions**
- **Better ways of automatically visualizing output data**
  - **Especially “latent” states and agent interactions**
- **More flexible ABM development environments, that allow easier refining and re-verification of models.**

On the one hand this is a daunting challenge, but on the other it is also a rich research agenda.

References

[1] LeBaron, Blake and Leigh Tesfatsion. "Modeling macroeconomies as open-ended systems of interacting agents." *American Economic Review: Papers and Proceedings* 98. 2 (2008): 246–50.

[2] Marks, R.E. (2007), Validating Simulation Models: A General Framework and Four Applied Examples, *Computational Economics*, 30(3): 265–290, October.

[3] Robert E. Marks and David F. Midgley (2008), The retailer needs to be paid for sophisticated decisions: modeling promotional interactions among consumers, retailers, and brand managers, http://www.agsm.edu.au/bobm/papers/MarksMidgley2008-V6.pdf

[4] Midgley D.F., Marks R.E., and Kunchamwar D. (2007), The Building and Assurance of Agent-Based Models: An Example and Challenge to the Field, *Journal of Business Research*, Special Issue: Complexities in Markets, 60(8): 884–893, August.

[5] Robert G. Sargent (2005), Verification and validation of simulation models, *Proceedings of the 2005 Winter Simulation Conference*, M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines, eds. IEEE.