# Validating Simulations with Historical Data: The State Similarity Measure

Sydney Agents
UTS
19 August 2010

Robert Marks,
Economics,
UNSW, Sydney

bobm@agsm.edu.au

from:
http://www.agsm.edu.au/bobm/papers/asa.pdf

>

## Outline

1. **Sufficiency and Necessity**

2. **Validation, Measurement**

3. **The Issue — Heterogeneous Agents, Sets of Time-series of Prices**

4. **The Method — Measuring the Distance Between Sets of Time-series using the *State Similarity Measure* (SSM)**

5. **The Results**

6. **Conclusions**

## 1. Sufficiency and Necessity

**Simulation demonstrates existence, sufficiency, but not necessity.**

## 1. Sufficiency and Necessity

**Simulation demonstrates existence, sufficiency,**

**but not necessity.**

**Simulation can demonstrate the untruth of a proposition,**

**but not provide proofs or theorems,**

**simulations cannot provide generality.**

## 1. Sufficiency and Necessity

**Simulation demonstrates existence, sufficiency,**

**but not necessity.**

**Simulation can demonstrate the untruth of a proposition,**

**but not provide proofs or theorems,**

**simulations cannot provide generality.**

**What, never?**

## 1. Sufficiency and Necessity

Simulation demonstrates existence, sufficiency,

but not necessity.

Simulation can demonstrate the untruth of a proposition,

but not provide proofs or theorems,

simulations cannot provide generality.

What, never?

Does this matter?

## Formal Simulation

**Mathematical "model $A$" comprises the conjunction ($a_1 \wedge a_2 \wedge a_3 \cdots \wedge a_n$), where $\wedge$ means "AND", and the $a_i$ denote the elements (equations, parameters, initial conditions, etc) that constitute the model.**

## Formal Simulation

**Mathematical "model $A$" comprises the conjunction ($a_1 \wedge a_2 \wedge a_3 \cdots \wedge a_n$), where $\wedge$ means "AND", and the $a_i$ denote the elements (equations, parameters, initial conditions, etc) that constitute the model.**

***Sufficiency*: If model $A$ exhibits the desired target behaviour $B$, then model $A$ is sufficient to obtain exhibited behaviour $B$: $A \Rightarrow B$**

**Thus, any model that exhibits the desired behaviour is sufficient, and demonstrates one conjunction of conditions (or model) under which the behaviour can be simulated.**

## Formal Simulation

**Mathematical "model $A$" comprises the conjunction ($a_1 \wedge a_2 \wedge a_3 \cdots \wedge a_n$), where $\wedge$ means "AND", and the $a_i$ denote the elements (equations, parameters, initial conditions, etc) that constitute the model.**

***Sufficiency*: If model $A$ exhibits the desired target behaviour $B$, then model $A$ is sufficient to obtain exhibited behaviour $B$: $A \Rightarrow B$**

**Thus, any model that exhibits the desired behaviour is sufficient, and demonstrates one conjunction of conditions (or model) under which the behaviour can be simulated.**

**But if there are several such models, how can we choose among them? And what is the set of all such conjunctions (models)?**

## Necessity

***Necessity:*** **Only those models $A$ belonging to the set of necessary models $\mathcal{N}$ exhibit target behaviour $B$.**

**That is, $(A \in \mathcal{N}) \Rightarrow B$, and $(D \notin \mathcal{N}) \not\Rightarrow B$.**

## Necessity

***Necessity:*** **Only those models $A$ belonging to the set of necessary models $\mathcal{N}$ exhibit target behaviour $B$.**

**That is, $(A \in \mathcal{N}) \Rightarrow B$, and $(D \notin \mathcal{N}) \not\Rightarrow B$.**

**A difficult challenge: determine the set of necessary models, $\mathcal{N}$.**

## Necessity

***Necessity:*** **Only those models $A$ belonging to the set of necessary models $\mathcal{N}$ exhibit target behaviour $B$.**

**That is, $(A \in \mathcal{N}) \Rightarrow B$, and $(D \notin \mathcal{N}) \not\Rightarrow B$.**

**A difficult challenge: determine the set of necessary models, $\mathcal{N}$.**

**Since each model $A = (a_1 \wedge a_2 \wedge a_3 \cdots \wedge a_n)$, searching for the set $\mathcal{N}$ of necessary models means searching in a high-dimensional space, with no guarantee of continuity, and a possible large number of non-linear interactions among elements.**

## Lack of Necessity Means ...

For instance, if $D \not\Rightarrow B$, it does not mean that all elements $a_i$ of model $D$ are invalid or wrong, only their conjunction, that is, model $D$.

It might be only a single element that precludes model D exhibiting behaviour $B$.

## Lack of Necessity Means ...

For instance, if $D \not\Rightarrow B$, it does not mean that all elements $a_i$ of model $D$ are invalid or wrong, only their conjunction, that is, model $D$.

It might be only a single element that precludes model D exhibiting behaviour $B$.

But determining whether this is so and which is the offending element is a costly exercise, in general, for the simulator.

Therefore, without clear knowledge of the boundaries of the set of necessary models, it is difficult to generalise from simulations.

## Simulation Can Demonstrate Necessity . . .

**only when the set $\mathcal{N}$ of necessary models is known to be small (such as in the case of DNA structure by the time Watson & Crick were searching for it) is it relatively easy to use simulation to derive necessity.**

## Simulation Can Demonstrate Necessity . . .

only when the set $\mathcal{N}$ of necessary models is known to be small (such as in the case of DNA structure by the time Watson & Crick were searching for it) is it relatively easy to use simulation to derive necessity.

They had much information about the properties of DNA (from others):

when they hit on the simulation we know as the "double helix", they knew it was right.

## Simulation Can Demonstrate Necessity . . .

only when the set $\mathcal{N}$ of necessary models is known to be small (such as in the case of DNA structure by the time Watson & Crick were searching for it) is it relatively easy to use simulation to derive necessity.

They had much information about the properties of DNA (from others):

when they hit on the simulation we know as the "double helix", they knew it was right.

But still "*A* structure ...", not "*The* structure" in the title of their 1953 *Nature* paper.

## 2. Formalisation of Validation

**Let set $P$ be the possible range of observed (historical) outputs of the real-world system.**

## 2. Formalisation of Validation

**Let set $P$ be the possible range of observed (historical) outputs of the real-world system.**

**Let set $M$ be the exhibited outputs of the model in any week.**

## 2. Formalisation of Validation

**Let set $P$ be the possible range of observed (historical) outputs of the real-world system.**

**Let set $M$ be the exhibited outputs of the model in any week.**

**Let set $S \subset P$ be the specific, historical output of the real-world system in any week.**

## 2. Formalisation of Validation

**Let set $P$ be the possible range of observed (historical) outputs of the real-world system.**

**Let set $M$ be the exhibited outputs of the model in any week.**

**Let set $S \subset P$ be the specific, historical output of the real-world system in any week.**

**Let set $Q$ be the intersection, if any, between the set $M$ and the set $S$, $Q \equiv M \cap S$.**

**We can characterise the model output in several cases. (Mankin et al. 1977).**

## Five Cases for Validation

a.

## Five Cases for Validation

a. no intersection between $M$ and $S$ ($Q = \varnothing$), then the model is *useless.*

b.

## Five Cases for Validation

a. no intersection between $M$ and $S$ ($Q = \emptyset$), then the model is *useless.*

b. intersection $Q$ is not null, then the model is *useful,* to some degree: will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. Both incomplete and inaccurate.

c.

## Five Cases for Validation

a. no intersection between $M$ and $S$ ($Q = \emptyset$), then the model is *useless.*

b. intersection $Q$ is not null, then the model is *useful,* to some degree: will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. Both incomplete and inaccurate.

c. If $M$ is a proper subset of $S$ ($M \subset S$) then all the model's behaviours are correct (match historical behaviours), but the model doesn't exhibit all behaviour that historically occurs: accurate but *incomplete.*

d.

## Five Cases for Validation

a. no intersection between $M$ and $S$ ($Q = \emptyset$), then the model is *useless.*

b. intersection $Q$ is not null, then the model is *useful,* to some degree: will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. Both incomplete and inaccurate.

c. If $M$ is a proper subset of $S$ ($M \subset S$) then all the model's behaviours are correct (match historical behaviours), but the model doesn't exhibit all behaviour that historically occurs: accurate but *incomplete.*

d. If $S$ is a proper subset of $M$ ($S \subset M$) then all historical behaviour is exhibited, but will exhibit some behaviours that do not historically occur: complete but *inaccurate.*

e.

## Five Cases for Validation

a. no intersection between $M$ and $S$ ($Q = \emptyset$), then the model is *useless.*

b. intersection $Q$ is not null, then the model is *useful,* to some degree: will correctly exhibit some real-world system behaviours, will not exhibit other behaviours, and will exhibit some behaviours that do not historically occur. Both incomplete and inaccurate.

c. If $M$ is a proper subset of $S$ ($M \subset S$) then all the model's behaviours are correct (match historical behaviours), but the model doesn't exhibit all behaviour that historically occurs: accurate but *incomplete.*

d. If $S$ is a proper subset of $M$ ($S \subset M$) then all historical behaviour is exhibited, but will exhibit some behaviours that do not historically occur: complete but *inaccurate.*

e. If the set $M$ is equivalent to the set $S$ ($M \Leftrightarrow S$), then (in your dreams!) the model is complete and accurate.

# Or Graphically ...

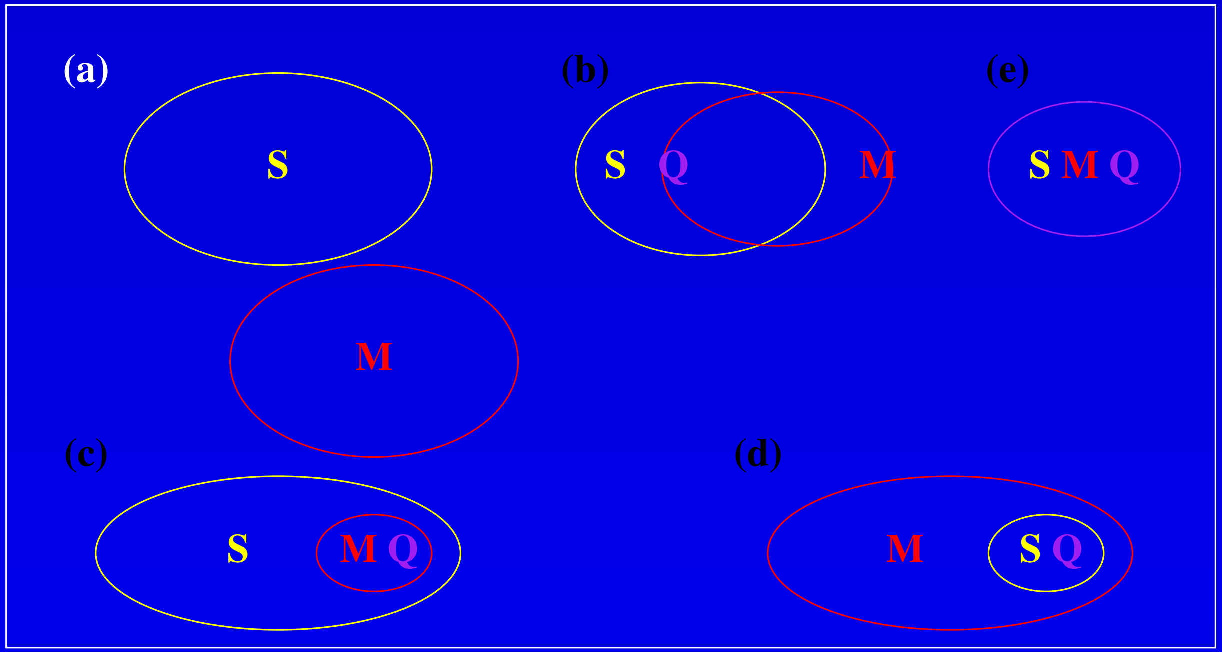



**Figure 1:** Validity relationships (after Haefner (2005)).

- a. useless
- b. useful, but incomplete and inaccurate
- c. accurate but incomplete
- d. complete but inaccurate ← possibly the best to aim for
- e. complete and accurate

## Modelling Goals

**One goal: to construct and calibrate the model so that**

**$M \approx Q \approx S$: there are very few historically observed behaviours that the model does not exhibit,**

**and there are very few exhibited behaviours that do not occur historically.**

**The model is close to being both complete and accurate.**

## Modelling Goals

One goal: to construct and calibrate the model so that

$M \approx Q \approx S$: there are very few historically observed behaviours that the model does not exhibit,

and there are very few exhibited behaviours that do not occur historically.

The model is close to being both complete and accurate.

In practice, a modeller might be happier to achieve case d., where the model is complete (and hence provides sufficiency for all observed historical phenomena), but not accurate.

## Modelling Goals

**One goal: to construct and calibrate the model so that**

**$M \approx Q \approx S$: there are very few historically observed behaviours that the model does not exhibit,**

**and there are very few exhibited behaviours that do not occur historically.**

**The model is close to being both complete and accurate.**

**In practice, a modeller might be happier to achieve case d., where the model is complete (and hence provides sufficiency for all observed historical phenomena), but not accurate.**

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

## Four Levels of Validation (Axtell & Epstein 1994)

**Level 0: Qualitatively similar at the micro level of individuals (agents)**

## Four Levels of Validation (Axtell & Epstein 1994)

**Level 0: Qualitatively similar at the micro level of individuals (agents)**

**Level 1: Qualitatively similar at a higher, macro, level**

## Four Levels of Validation (Axtell & Epstein 1994)

**Level 0: Qualitatively similar at the micro level of individuals (agents)**

**Level 1: Qualitatively similar at a higher, macro, level**

**Level 2: Quantitative agreement of macro structures eg. means, moments, distributions, statistical tests**

## Four Levels of Validation (Axtell & Epstein 1994)

**Level 0: Qualitatively similar at the micro level of individuals (agents)**

**Level 1: Qualitatively similar at a higher, macro, level**

**Level 2: Quantitative agreement of macro structures**
**eg. means, moments, distributions, statistical tests**

**Level 3: Quantitative agreement at the micro level**
**eg. agents behave exactly the same.**

## Four Levels of Validation (Axtell & Epstein 1994)

**Level 0: Qualitatively similar at the micro level of individuals (agents)**

**Level 1: Qualitatively similar at a higher, macro, level**

**Level 2: Quantitative agreement of macro structures**
**eg. means, moments, distributions, statistical tests**

**Level 3: Quantitative agreement at the micro level**
**eg. agents behave exactly the same.**

**Here we address Level 2, with a new moment, the SSM.**

## Measurement

**Q: how can we measure the degree of similarity of two sets of time-series?**

## Measurement

**Q: how can we measure the degree of similarity of two sets of time-series?**

**One: the historical record of the rivalrous dance among the sellers in an oligopoly, while**

## Measurement

**Q: how can we measure the degree of similarity of two sets of time-series?**

**One: the historical record of the rivalrous dance among the sellers in an oligopoly, while**

**The other: the output from a (agent-based) simulation model of the market, where each seller agent prices this week as a function of the state of the market last week (or earlier).**

## Measurement

**Q: how can we measure the degree of similarity of two sets of time-series?**

**One: the historical record of the rivalrous dance among the sellers in an oligopoly, while**

**The other: the output from a (agent-based) simulation model of the market, where each seller agent prices this week as a function of the state of the market last week (or earlier).**

**Q: how can we *output validate* our model against history?**

## Measurement

**Q: how can we measure the degree of similarity of two sets of time-series?**

**One: the historical record of the rivalrous dance among the sellers in an oligopoly, while**

**The other: the output from a (agent-based) simulation model of the market, where each seller agent prices this week as a function of the state of the market last week (or earlier).**

**Q: how can we *output validate* our model against history?**

**Or: how can we derive a *degree of confidence* in the model output?**

## 3. The Issue: Heterogenous Agents and Time-series Price

**Two reasons to compare such model output against history:**

1.

## 3. The Issue: Heterogenous Agents and Time-series Price

**Two reasons to compare such model output against history:**

1. **To choose better parameter values, to "calibrate" or (more formally) "estimate" the model against the historical record.**
2.

## 3. The Issue: Heterogenous Agents and Time-series Price

**Two reasons to compare such model output against history:**

1. **To choose better parameter values, to "calibrate" or (more formally) "estimate" the model against the historical record.**
2. **To measure how closely the output reflects history, to *validate* the model.**

## 3. The Issue: Heterogenous Agents and Time-series Price

**Two reasons to compare such model output against history:**

1. **To choose better parameter values, to "calibrate" or (more formally) "estimate" the model against the historical record.**
2. **To measure how closely the output reflects history, to *validate* the model.**

**We are interested in the second, having used machine learning (the GA) to derive the model parameters in order to improve each agent's weekly profits (instead of fitting to history) in our agent-based model.**

## 3. The Issue: Heterogenous Agents and Time-series Price

**Two reasons to compare such model output against history:**

1. **To choose better parameter values, to "calibrate" or (more formally) "estimate" the model against the historical record.**
2. **To measure how closely the output reflects history, to *validate* the model.**

**We are interested in the second, having used machine learning (the GA) to derive the model parameters in order to improve each agent's weekly profits (instead of fitting to history) in our agent-based model.**

**Figure 2 shows historical data from a U.S. supermarket chain's sales of (heterogeneous) brands of sealed, ground coffee, by week in one city (Midgley et al. 1997).**

## Historical Data: Prices and Volumes in Chain 1


*Figure 2:* Weekly Sales and Prices (Source: Midgley et al. 1997)

## Stylised facts of the historical data:

1.

## Stylised facts of the historical data:

1. Much movement in the prices and volumes of four strategic brands — a rivalrous dance.

2.

## Stylised facts of the historical data:

1. Much movement in the prices and volumes of four strategic brands — a rivalrous dance.

2. For these four (coloured) brands, high prices (and low volumes) are punctuated by a low price (and a high volume).

3.

## Stylised facts of the historical data:

1. Much movement in the prices and volumes of four strategic brands — a rivalrous dance.

2. For these four (coloured) brands, high prices (and low volumes) are punctuated by a low price (and a high volume).

3. The remaining five brands exhibit stable prices and volumes, by and large. For this reason we abstract away from these five brands, and focus solely on the first four.

## Stylised facts of the historical data:

1. **Much movement in the prices and volumes of four strategic brands — a rivalrous dance.**

2. **For these four (coloured) brands, high prices (and low volumes) are punctuated by a low price (and a high volume).**

3. **The remaining five brands exhibit stable prices and volumes, by and large. For this reason we abstract away from these five brands, and focus solely on the first four.**

**In addition, the competition is not open slather: the supermarket chain imposes some restrictions on the timing and identity of the discounting brands.**

## A Model of Strategic Interaction

**We assume that the price $P_{b,w}$ of brand $b$ in week $w$ is a function of the state of the market $M_w$ at week $w$, where $M_w$ in turn is the product of the weekly prices $S_w$ of all brands over several weeks:**

$$P_{b,w} = f_b(M_w) = f_b(S_{w-1} \times S_{w-2} \times S_{w-3} \cdots)$$

## A Model of Strategic Interaction

We assume that the price $P_{b,w}$ of brand $b$ in week $w$ is a function of the state of the market $M_w$ at week $w$, where $M_w$ in turn is the product of the weekly prices $S_w$ of all brands over several weeks:

$$P_{b,w} = f_b(M_w) = f_b(S_{w-1} \times S_{w-2} \times S_{w-3} \cdots)$$

Earlier in the research program undertaken with David Midgley et al., we used the Genetic Algorithm to search for "better" (i.e. more profitable) brand-specific mappings, $f_b$, from market state to pricing action.

And derived the parameters of the model, and derived its simulated behaviour, as time-series patterns (below).

## 4. The Method — Measuring the Distance Between Sets of Time-series using the *State Similarity Measure*

**The SSM method introduced here reduces the dimensionality of the historical behaviour (and sometimes the model output too) by partitioning the price line in order to derive a measure of similarity or distance beween two sets.**

## 4. The Method — Measuring the Distance Between Sets of Time-series using the *State Similarity Measure*

**The SSM method introduced here reduces the dimensionality of the historical behaviour (and sometimes the model output too) by partitioning the price line in order to derive a measure of similarity or distance beween two sets.**

**Marks (1998) explores partitioning while maximising information (using an entropy measure); maximising profits would be a better criterion. Finds that dichotomous partition is sufficient.**

## 4. The Method — Measuring the Distance Between Sets of Time-series using the *State Similarity Measure*

**The SSM method introduced here reduces the dimensionality of the historical behaviour (and sometimes the model output too) by partitioning the price line in order to derive a measure of similarity or distance beween two sets.**

**Marks (1998) explores partitioning while maximising information (using an entropy measure); maximising profits would be a better criterion. Finds that dichotomous partition is sufficient.**

**Here: use symmetric dichotomous partitioning: a brand's price is labelled 0 if above its midpoint, else 1 below.**

## 4. The Method — Measuring the Distance Between Sets of Time-series using the *State Similarity Measure*

**The SSM method introduced here reduces the dimensionality of the historical behaviour (and sometimes the model output too) by partitioning the price line in order to derive a measure of similarity or distance beween two sets.**

**Marks (1998) explores partitioning while maximising information (using an entropy measure); maximising profits would be a better criterion. Finds that dichotomous partition is sufficient.**

**Here: use symmetric dichotomous partitioning: a brand's price is labelled 0 if above its midpoint, else 1 below.**

**Then defining market states first by week $S_w$ and then by multi-week window $M_w$, counting the frequency of each state, subtracting the two sets' frequencies, and summing the absolute difference.**

## Dichotomous Symmetric Price Partitioning of Chain 1

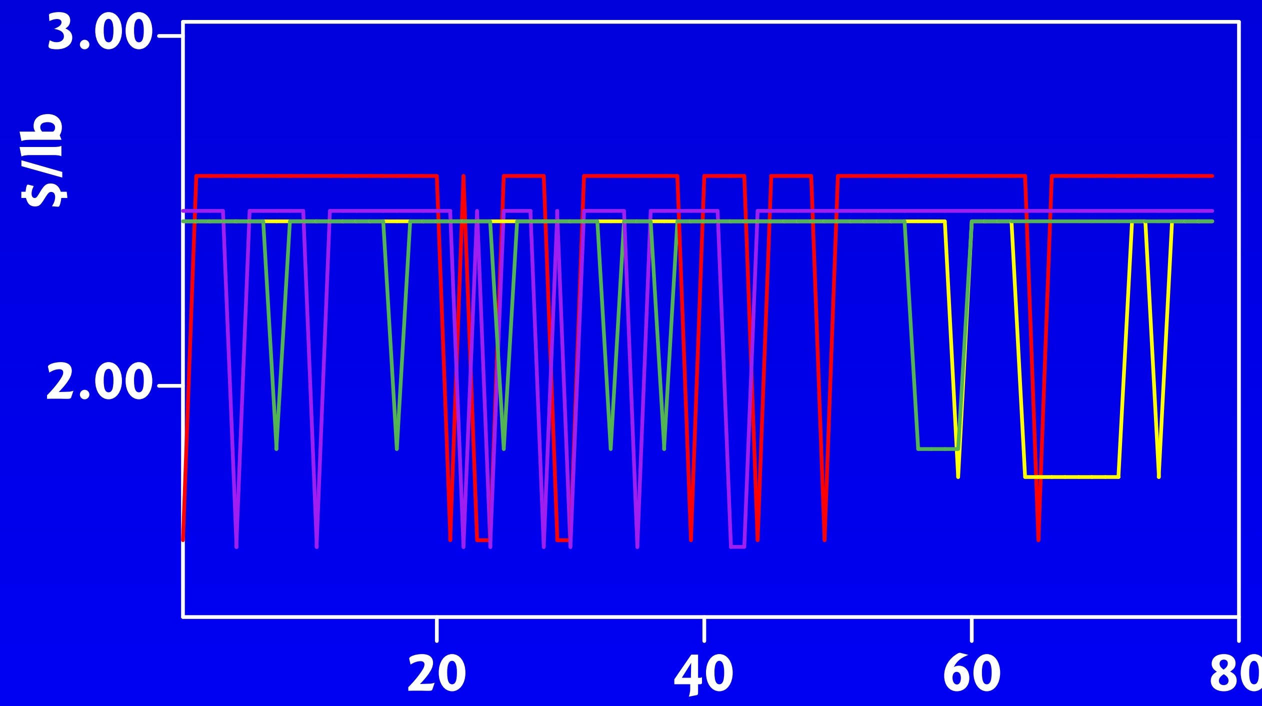



***Figure 3:*** **Partitioned Weekly Prices of the Four Chain-One Brands**

# Calculating the Weekly $S_w$ and Window $M_w$ States

| Week | Red | Purple | Green | $\therefore S_w$ | $\therefore M_w$ |
|---|---|---|---|---|---|
| 18 | 0 | 0 | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | |
| 20 | 0 | 0 | 0 | 0 | 0 |

## Calculating the Weekly $S_w$ and Window $M_w$ States

| Week | Red | Purple | Green | $\therefore S_w$ | $\therefore M_w$ |
|---|---|---|---|---|---|
| 18 | 0 | 0 | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | |
| 20 | 0 | 0 | 0 | 0 | 0 |
| 21 | 1 | 0 | 0 | 4 | 256 |

## Calculating the Weekly $S_w$ and Window $M_w$ States

| Week | Red | Purple | Green | $\therefore S_w$ | $\therefore M_w$ |
|---|---|---|---|---|---|
| 18 | 0 | 0 | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | |
| 20 | 0 | 0 | 0 | 0 | 0 |
| 21 | 1 | 0 | 0 | 4 | 256 |
| 22 | 0 | 1 | 0 | 2 | 160 |

## Calculating the Weekly $S_w$ and Window $M_w$ States

| Week | Red | Purple | Green | $\therefore S_w$ | $\therefore M_w$ |
|---|---|---|---|---|---|
| 18 | 0 | 0 | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | |
| 20 | 0 | 0 | 0 | 0 | 0 |
| 21 | 1 | 0 | 0 | 4 | 256 |
| 22 | 0 | 1 | 0 | 2 | 160 |
| 23 | 1 | 0 | 0 | 4 | 276 |

## Calculating the Weekly $S_w$ and Window $M_w$ States

| Week | Red | Purple | Green | $\therefore S_w$ | $\therefore M_w$ |
|---|---|---|---|---|---|
| 18 | 0 | 0 | 0 | 0 | |
| 19 | 0 | 0 | 0 | 0 | |
| 20 | 0 | 0 | 0 | 0 | 0 |
| 21 | 1 | 0 | 0 | 4 | 256 |
| 22 | 0 | 1 | 0 | 2 | 160 |
| 23 | 1 | 0 | 0 | 4 | 276 |
| 24 | 1 | 1 | 0 | 6 | 418 |
| 25 | 0 | 0 | 1 | 1 | 116 |
| 26 | 0 | 0 | 0 | 0 | 14 |
| 27 | 0 | 0 | 0 | 0 | 1 |
| 28 | 0 | 1 | 0 | 2 | 128 |
| 29 | 1 | 0 | 0 | 4 | 272 |
| 30 | 1 | 1 | 0 | 6 | 418 |

Three Brands, 3-Week Window

## Calculating the SSM Between to Two Sets of Time Series

**More formally:**

1.

## Calculating the SSM Between to Two Sets of Time Series

**More formally:**

1. **For each set, partition the time-series $\{P_{b,w}\}$ of price $P_{b,w}$ of brand $b$ in week $w$ into {0,1}, where 0 corresponds to "high" price and 1 corresponds to "low" price to obtain time-series $\{P_{b,w}'\}$;**

2.

## Calculating the SSM Between to Two Sets of Time Series

**More formally:**

1. **For each set, partition the time-series $\{P_{b,w}\}$ of price $P_{b,w}$ of brand $b$ in week $w$ into {0,1}, where 0 corresponds to "high" price and 1 corresponds to "low" price to obtain time-series $\{P_{b,w}'\}$;**

2. **For the set of 3- or 4-brand time-series of brands' partitioned prices $\{P_{b,w}'\}$, calculate the time-series of the state of the market each week $\{S_w\}$;**

3.

## Calculating the SSM Between to Two Sets of Time Series

**More formally:**

1. **For each set, partition the time-series $\{P_{b,w}\}$ of price $P_{b,w}$ of brand $b$ in week $w$ into {0,1}, where 0 corresponds to "high" price and 1 corresponds to "low" price to obtain time-series $\{P_{b,w}'\}$;**

2. **For the set of 3- or 4-brand time-series of brands' partitioned prices $\{P_{b,w}'\}$, calculate the time-series of the state of the market each week $\{S_w\}$;**

3. **For each set, calculate the time-series of the state of the 3- or 4-week moving window of partitioned prices $\{M_w\}$, from the per-week states $\{S_w\}$;**

4.

4. **Count the numbers of each state observed for the set of time-series over the given time period; convey this by an $n \times 1$ vector $c$, where $c[s]$ = the number of observations of window state $s$ over the period;**

5.

4. **Count the numbers of each state observed for the set of time-series over the given time period; convey this by an $n \times 1$ vector $c$, where $c[s]$ = the number of observations of window state $s$ over the period;**

5. **Subtract the number of observations in set $A$ of time-series from the number observed in set B, across all $n$ possible states; $d^{AB} = c^{A} - c^{B}$;**

6.

4. Count the numbers of each state observed for the set of time-series over the given time period; convey this by an $n \times 1$ vector $\boldsymbol{c}$, where $\boldsymbol{c}[s]$ = the number of observations of window state $s$ over the period;

5. Subtract the number of observations in set $A$ of time-series from the number observed in set B, across all $n$ possible states; $\boldsymbol{d}^{AB} = \boldsymbol{c}^{A} - \boldsymbol{c}^{B}$;

6. Sum the absolute values of the differences across all possible states:

$$D^{AB} = \mathbf{1}' \times |\boldsymbol{d}^{AB}| \tag{1}$$

4. **Count the numbers of each state observed for the set of time-series over the given time period; convey this by an $n \times 1$ vector $c$, where $c[s]$ = the number of observations of window state $s$ over the period;**

5. **Subtract the number of observations in set $\mathcal{A}$ of time-series from the number observed in set $\mathcal{B}$, across all $n$ possible states; $d^{AB} = c^{A} - c^{B}$;**

6. **Sum the absolute values of the differences across all possible states:**

$$D^{AB} = 1' \times |d^{AB}| \tag{1}$$

**This number $D^{AB}$ is *the distance between two time-series sets $\mathcal{A}$ and $\mathcal{B}$.***

**This method is called *the State Similarity Measure.***

## 5. The Results

## 5. The Results

**Having derived the distance between two sets of time-series using the *State Similarity Measure*, by calculating the sum of absolute differences in observed window states between the two set, so what?**

## 5. The Results

**Having derived the distance between two sets of time-series using the *State Similarity Measure*, by calculating the sum of absolute differences in observed window states between the two set, so what?**

**First, the greater the sum, the more distant the two sets of time-series.**

## 5. The Results

**Having derived the distance between two sets of time-series using the *State Similarity Measure*, by calculating the sum of absolute differences in observed window states between the two set, so what?**

**First, the greater the sum, the more distant the two sets of time-series.**

**Second, we can calculate the maximum size of the summed difference: zero intersection between the two sets (no states in common) implies a measure of $2 \times S$ where $S$ is the number of possible window states, from the data.**

## 5. The Results

**Having derived the distance between two sets of time-series using the *State Similarity Measure*, by calculating the sum of absolute differences in observed window states between the two set, so what?**

**First, the greater the sum, the more distant the two sets of time-series.**

**Second, we can calculate the maximum size of the summed difference: zero intersection between the two sets (no states in common) implies a measure of $2 \times S$ where $S$ is the number of possible window states, from the data.**

**Third, we can derive some statistics to show that any pair of sets in not likely to include random series (below).**

## The Historical Data: A Diversity of Brands in the Chains

**There are seven chains, containing a variety of brands, some (1, 2, 4, 5) active rivals, the rest non-strategic.**

| | Brands | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Chain 1 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | |
| Chain 2 | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | | |
| Chain 3 | ✓ | ✓ | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | |
| Chain 4 | ✓ | ✓ | ✓ | | | | | ✓ | | | | ✓ |
| Chain 5 | ✓ | ✓ | ✓ | | | ✓ | | ✓ | | | ✓ | ✓ |
| Chain 6 | ✓ | ✓ | ✓ | | | | | | | | | ✓ |
| Chain 7 | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | | | ✓ |

*Table 1*: The Historical Data: The Seven Chains and the Twelve Brands

**(Brand 1=Folgers, 2=Maxwell House, 3=Master Blend, 4=Hills Bros, 5=Chock Full O Nuts, 6=Yuban, 7=Chase & Sanbourne, etc.)**

## SSMs Between Four Chains (with Brands 1, 2, 4, 5)

| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 0 | 128 | 112 | 110 |
| Chain 2 | 128 | 0 | 132 | 138 |
| Chain 3 | 112 | 132 | 0 | 124 |
| Chain 7 | 110 | 138 | 124 | 0 |
| Random | 150 | 150 | 150 | 150 |

*Table 2:* SSMs Between Four Chains (with Brands 1, 2, 4, 5)

## SSMs Between Four Chains (with Brands 1, 2, 4, 5)

| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 0 | 128 | 112 | 110 |
| Chain 2 | 128 | 0 | 132 | 138 |
| Chain 3 | 112 | 132 | 0 | 124 |
| Chain 7 | 110 | 138 | 124 | 0 |
| Random | 150 | 150 | 150 | 150 |

*Table 2:* SSMs Between Four Chains (with Brands 1, 2, 4, 5)

With two possible states per week per brand and four brands: $2^4$ possible weekly states; with a four-week window, there are $16^4 = 65,536$ possible window states.

## SSMs Between Four Chains (with Brands 1, 2, 4, 5)

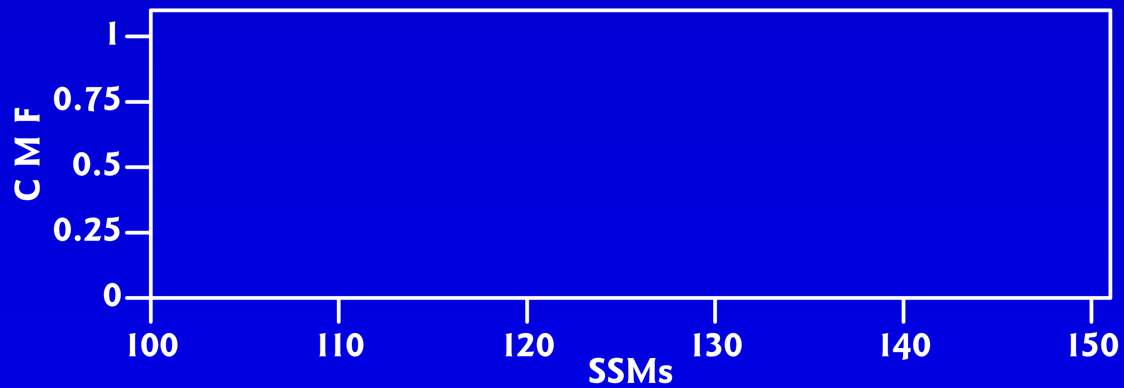
| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 0 | 128 | 112 | 110 |
| Chain 2 | 128 | 0 | 132 | 138 |
| Chain 3 | 112 | 132 | 0 | 124 |
| Chain 7 | 110 | 138 | 124 | 0 |
| Random | 150 | 150 | 150 | 150 |

*Table 2:* SSMs Between Four Chains (with Brands 1, 2, 4, 5)

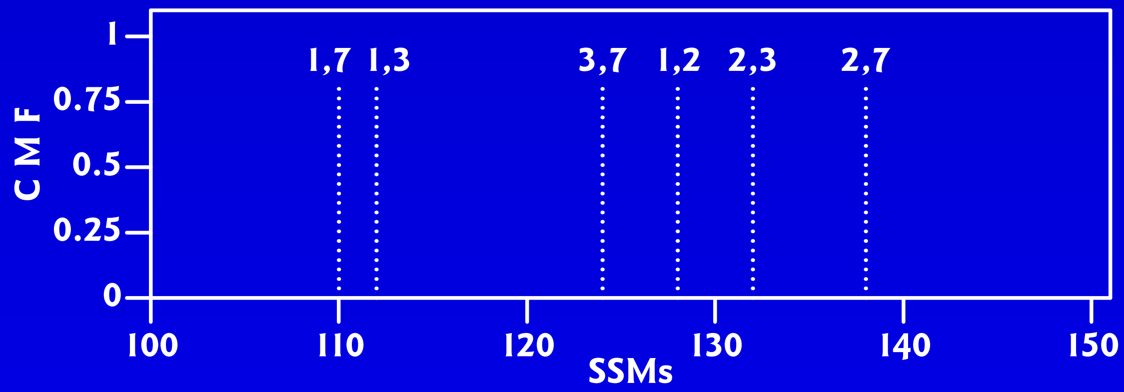
With two possible states per week per brand and four brands: $2^4$ possible weekly states; with a four-week window, there are $16^4 = 65,536$ possible window states.

With 75 overlapping four-week windows, $S = 75$, and the maximum measure (distance) is 150.

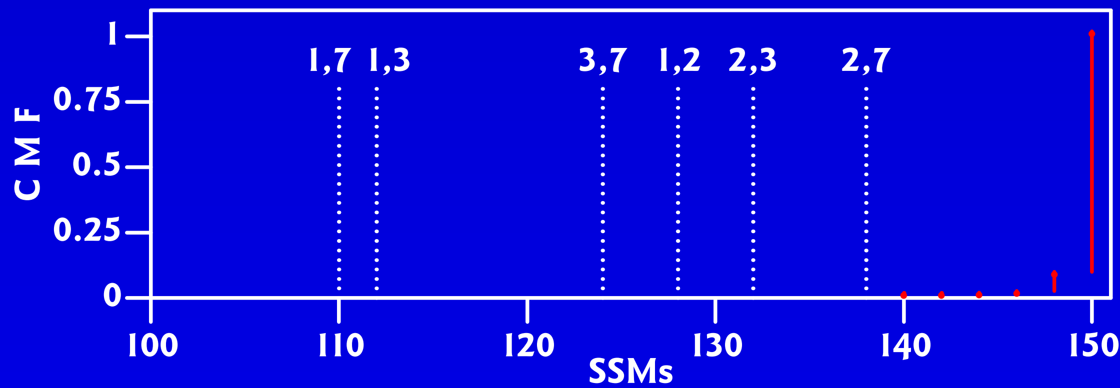
# Testing for Randomness Figure 4



## Testing for Randomness Figure 4

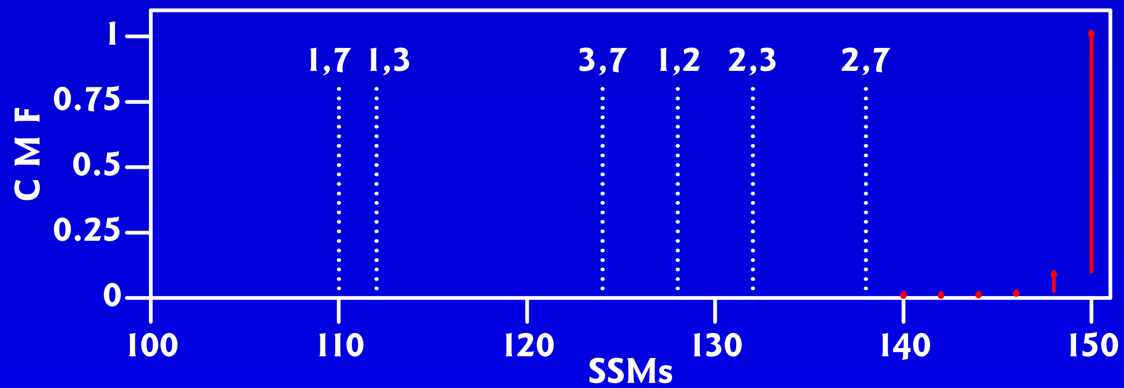


## Testing for Randomness Figure 4




**The red lines are the CMF of pairs of sets of random series (4 series, 75 observations) from 100,000 Monte Carlo parameter bootstraps.**

## Testing for Randomness Figure 4




The red lines are the CMF of pairs of sets of random series (4 series, 75 observations) from 100,000 Monte Carlo parameter bootstraps.

All six measured SSMs are significantly not random.

The one-sided c.i. at 1% corresponds to a SSM of 148, much exceeding the greatest distance (between Chains 2 and 7) of 138.

## Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 100 | 14.67 | 25.33 | 26.67 |
| Chain 2 | 14.67 | 100 | 12.0 | 8.0 |
| Chain 3 | 25.33 | 12.0 | 100 | 17.33 |
| Chain 7 | 26.67 | 8.0 | 17.33 | 100 |
| Random | 0 | 0 | 0 | 0 |

*Table 3:* Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

## Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 100 | 14.67 | 25.33 | 26.67 |
| Chain 2 | 14.67 | 100 | 12.0 | 8.0 |
| Chain 3 | 25.33 | 12.0 | 100 | 17.33 |
| Chain 7 | 26.67 | 8.0 | 17.33 | 100 |
| Random | 0 | 0 | 0 | 0 |

*Table 3*: Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

Table 3 is derived from Table 2, with 150 the maximum possible distance between sets.

## Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

| | Chain 1 | Chain 2 | Chain 3 | Chain 7 |
|---|---|---|---|---|
| Chain 1 | 100 | 14.67 | 25.33 | 26.67 |
| Chain 2 | 14.67 | 100 | 12.0 | 8.0 |
| Chain 3 | 25.33 | 12.0 | 100 | 17.33 |
| Chain 7 | 26.67 | 8.0 | 17.33 | 100 |
| Random | 0 | 0 | 0 | 0 |

*Table 3*: Percentage Matches Between Four Chains (with Brands 1, 2, 4, 5)

Table 3 is derived from Table 2, with 150 the maximum possible distance between sets.

Note that there is a 100% own match, and that there is zero match between the Random pricing process and any of the historical chains.

## SSMs Between All Seven Chains (with Brands 1, 2, 3)

| | Chain 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Chain 1 | 0 | 70 | 82 | 76 | 102 | 132* | 74 |
| Chain 2 | 70 | 0 | 82 | 98 | 90 | 120† | 98 |
| Chain 3 | 82 | 82 | 0 | 100 | 96 | 122† | 102 |
| Chain 4 | 76 | 98 | 100 | 0 | 80 | 128* | 58 |
| Chain 5 | 102 | 90 | 96 | 80 | 0 | 114 | 92 |
| Chain 6 | 132* | 120† | 122† | 128* | 114 | 0 | 130* |
| Chain 7 | 74 | 98 | 102 | 58 | 92 | 130* | 0 |
| Random | 144 | 136 | 148 | 144 | 140 | 146 | 144 |

*Table 4:* SSMs Between All Chains (with Brands 1, 2, 3)

## SSMs Between All Seven Chains (with Brands 1, 2, 3)

| | Chain 1 | Chain 2 | Chain 3 | Chain 4 | Chain 5 | Chain 6 | Chain 7 |
|---|---|---|---|---|---|---|---|
| Chain 1 | 0 | 70 | 82 | 76 | 102 | 132* | 74 |
| Chain 2 | 70 | 0 | 82 | 98 | 90 | 120† | 98 |
| Chain 3 | 82 | 82 | 0 | 100 | 96 | 122† | 102 |
| Chain 4 | 76 | 98 | 100 | 0 | 80 | 128* | 58 |
| Chain 5 | 102 | 90 | 96 | 80 | 0 | 114 | 92 |
| Chain 6 | 132* | 120† | 122† | 128* | 114 | 0 | 130* |
| Chain 7 | 74 | 98 | 102 | 58 | 92 | 130* | 0 |
| Random | 144 | 136 | 148 | 144 | 140 | 146 | 144 |

*Table 4:* SSMs Between All Chains (with Brands 1, 2, 3)
(* : cannot reject the null of random at the 5% level)
(† : cannot reject the null of random at the 1% level)

# Table 4 — Historical Sets Compared using the SSM

## Table 4 — Historical Sets Compared using the SSM

**With three brands (1, 2, 3) treated as strategic, and three-week windowing, there are $8^3 = 512$ possible window states.**

## Table 4 — Historical Sets Compared using the SSM

With three brands (1, 2, 3) treated as strategic, and three-week windowing, there are $8^3 = 512$ possible window states.

The historical data include $S = 76$ overlapping three-week windows, so the maximum distance between any two chains is 152.

## Table 4 — Historical Sets Compared using the SSM

With three brands (1, 2, 3) treated as strategic, and three-week windowing, there are $8^3 = 512$ possible window states.

The historical data include $S = 76$ overlapping three-week windows, so the maximum distance between any two chains is 152.

The Random results are almost the maximum possible distance from the chains.

## Table 4 — Historical Sets Compared using the SSM

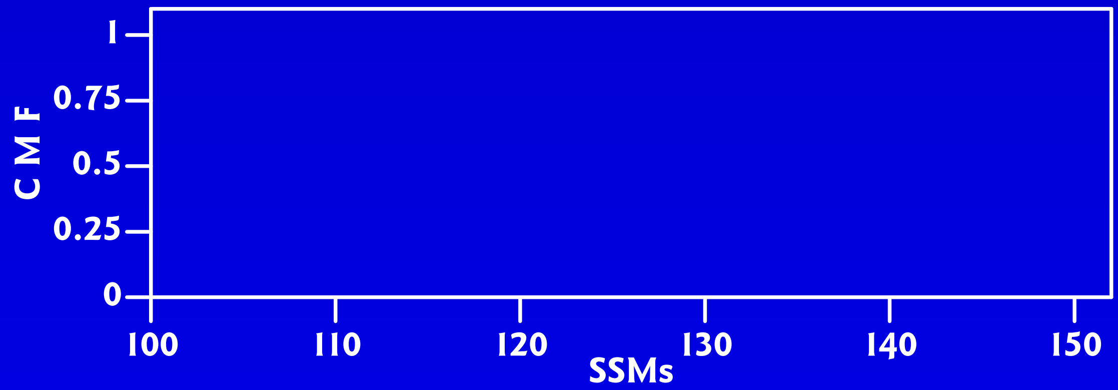
With three brands (1, 2, 3) treated as strategic, and three-week windowing, there are $8^3 = 512$ possible window states.

The historical data include $S = 76$ overlapping three-week windows, so the maximum distance between any two chains is 152.

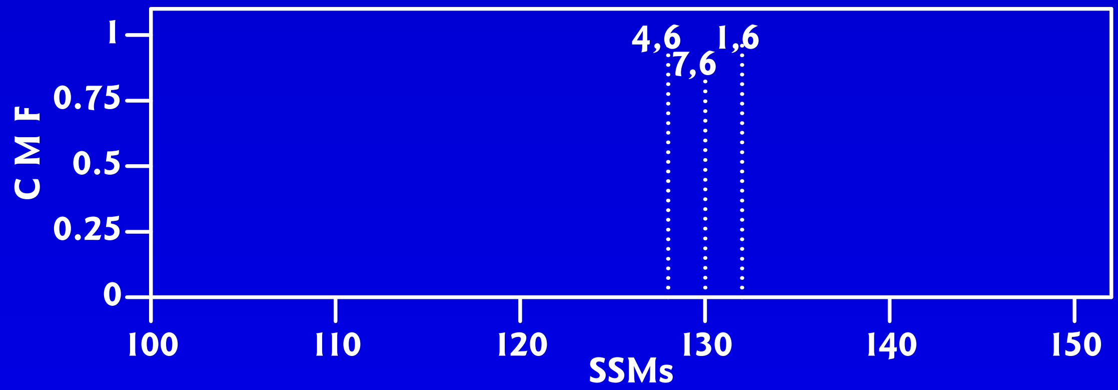
The Random results are almost the maximum possible distance from the chains.

The closest chains are Chain 4 and 7, with 152 − 58 = 94 states in common, or 61.84%.

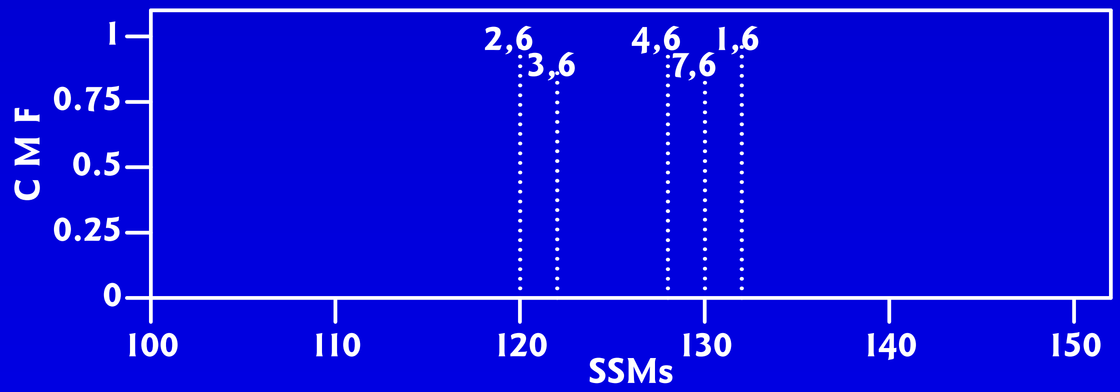
## Testing for Randomness Figure 5



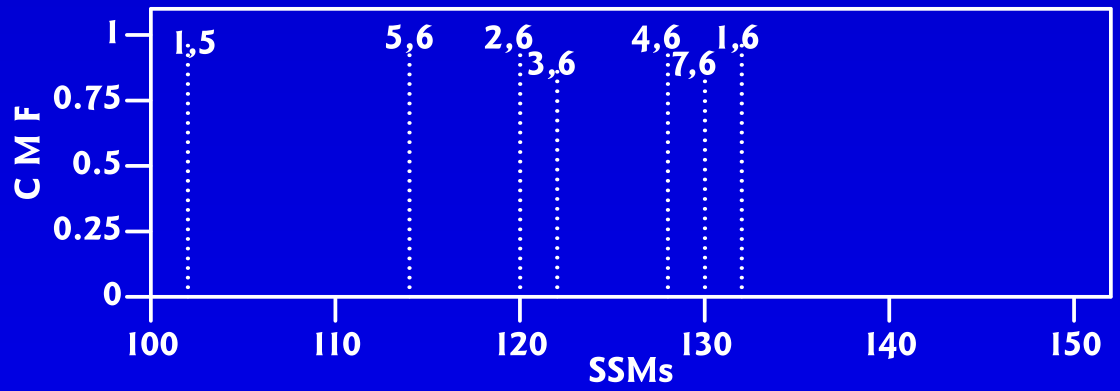
# Testing for Randomness Figure 5



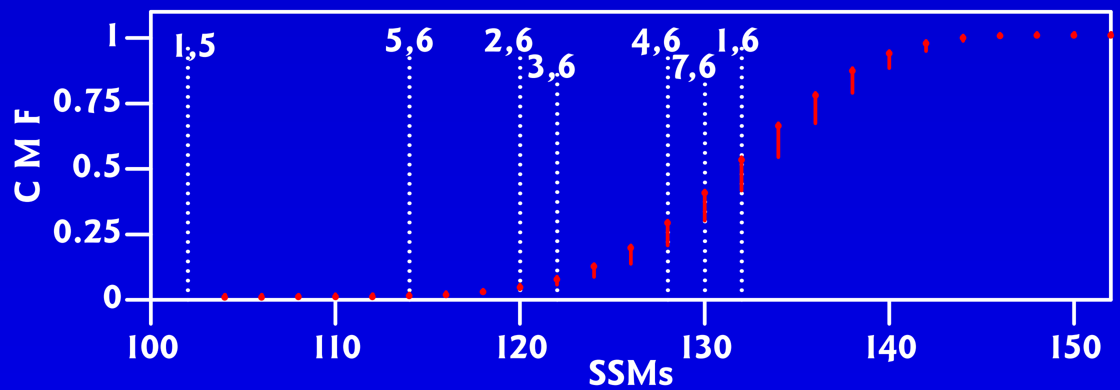
## Testing for Randomness Figure 5



# Testing for Randomness Figure 5

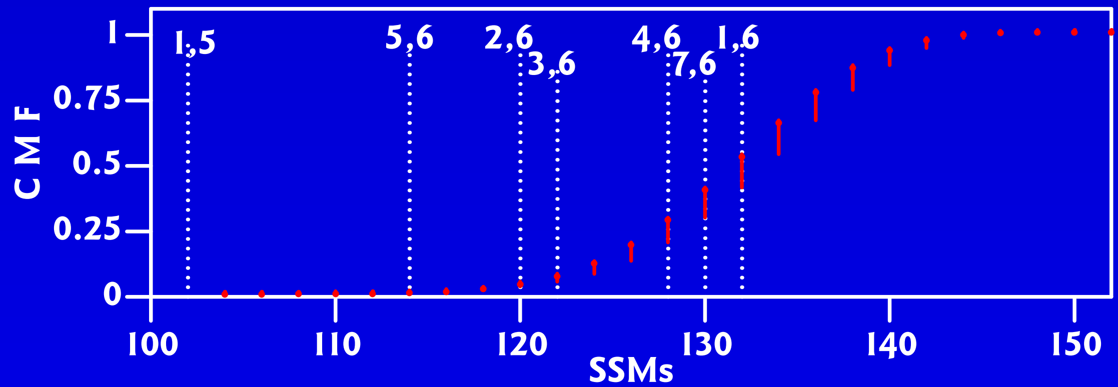


## Testing for Randomness Figure 5




**The red lines are the CMF of pairs of sets of random series (3 series, 76 observations) from 100,000 Monte Carlo parameter bootstraps.**

## Testing for Randomness Figure 5




The red lines are the CMF of pairs of sets of random series (3 series, 76 observations) from 100,000 Monte Carlo parameter bootstraps.

The one-sided c.i. at 1% corresponds to a SSM of 118, and at 5% 122.

Cannot reject the null hypothesis (random sets) for Chain 6 and Chains 1, 4, or 7 (5%) or for Chain 6 and Chains 2 or 3 (1%). The null is rejected for all other pairs.

## Example of a Simulated Oligopoly (Marks et al. 1995)

**Simulating rivalry between the three asymmetric brands: 1, 2, and 5, Folgers, Maxwell House, and Chock Full O Nuts.**

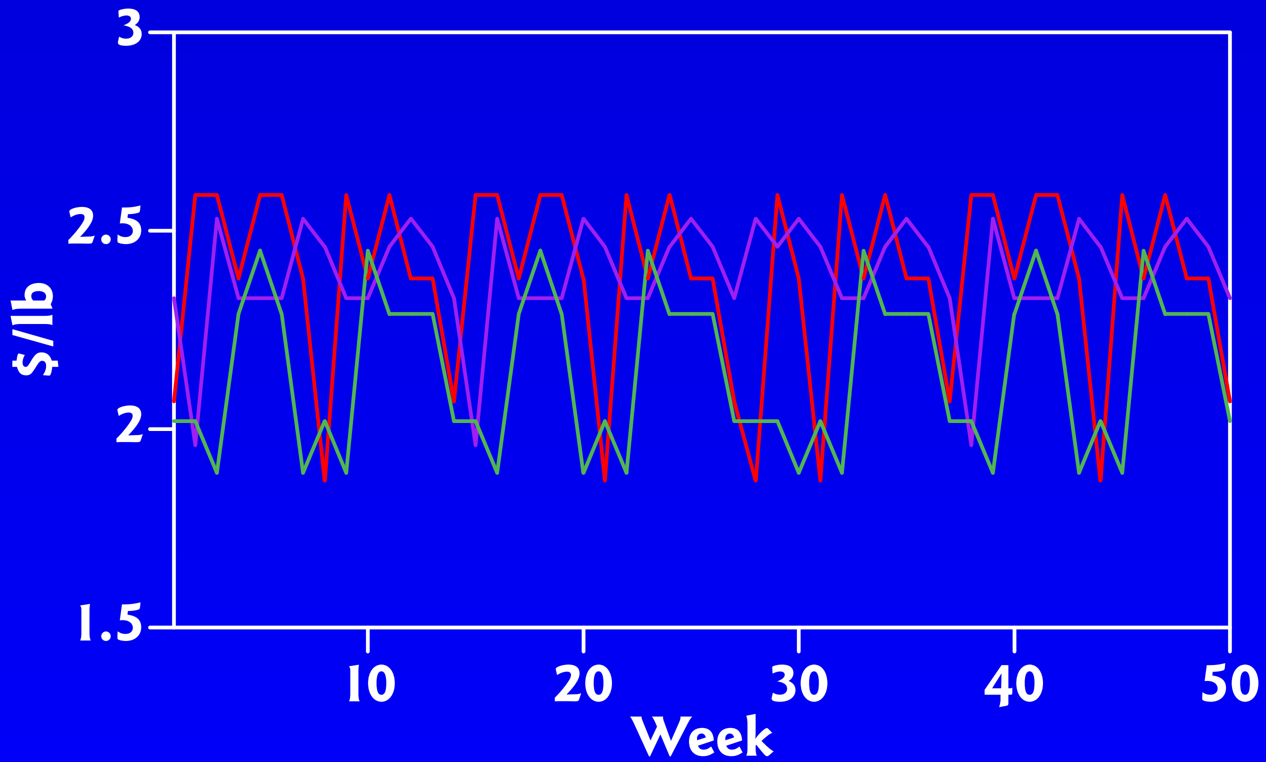



***Figure 6:*** **Example of a Simulated Oligopoly (Marks et al. 1995)**

## SSMs Between Chain 1 and Three Runs (Brands 1, 2, 5)

| | Chain 1 | Run 11 | Run 26a | Run 26b |
|---|---|---|---|---|
| Chain 1 | 0 | 82* | 68 | 68 |
| Run 11 | 82* | 0 | 66 | 60 |
| Run 26a | 68 | 66 | 0 | 30 |
| Run 26b | 68 | 60 | 30 | 0 |

*Table 5:* SSMs Between Chain 1 and Three Runs (Brands 1, 2, 5)
(* : cannot reject the null at the 5% level)

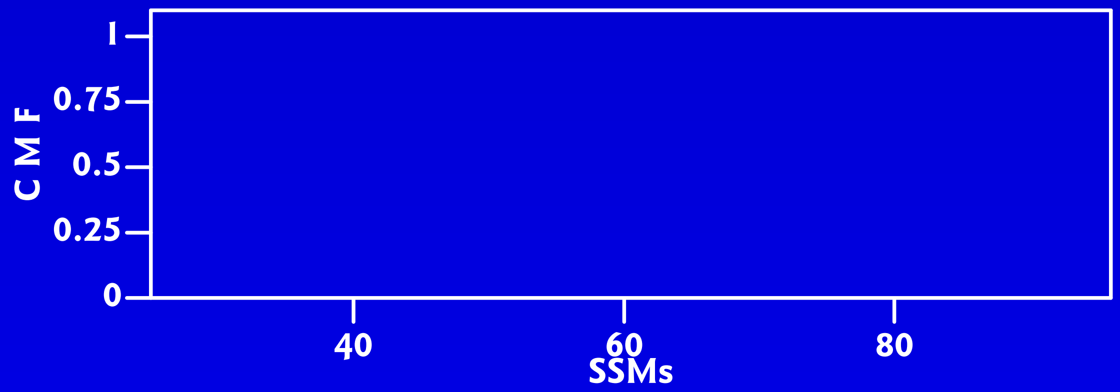
## SSMs Between Chain 1 and Three Runs (Brands 1, 2, 5)

| | Chain 1 | Run 11 | Run 26a | Run 26b |
|---|---|---|---|---|
| Chain 1 | 0 | 82* | 68 | 68 |
| Run 11 | 82* | 0 | 66 | 60 |
| Run 26a | 68 | 66 | 0 | 30 |
| Run 26b | 68 | 60 | 30 | 0 |

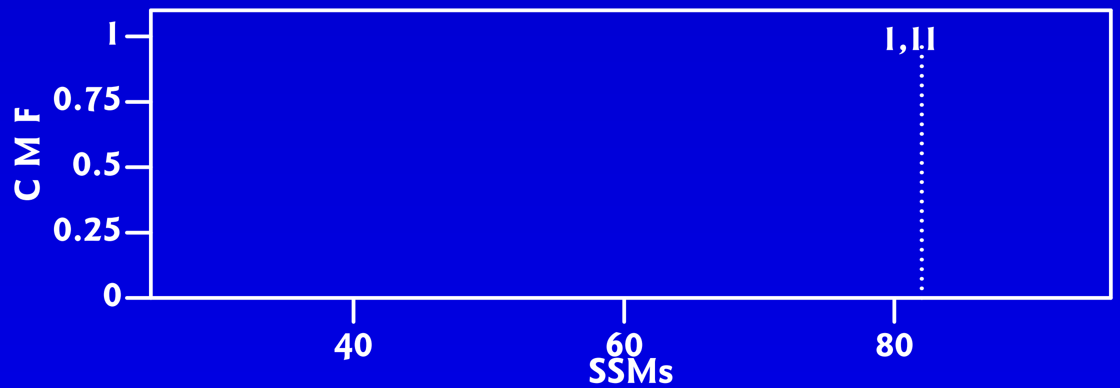
*Table 5:* SSMs Between Chain 1 and Three Runs (Brands 1, 2, 5)
(* : cannot reject the null at the 5% level)

Here, $S$, the maximum number of states = 48, so the maximum distance apart is 96. The three Runs are closer to each other than to historical Chain 1; Runs 26a and 26b are very close, only 30/96 = 31.25% apart.

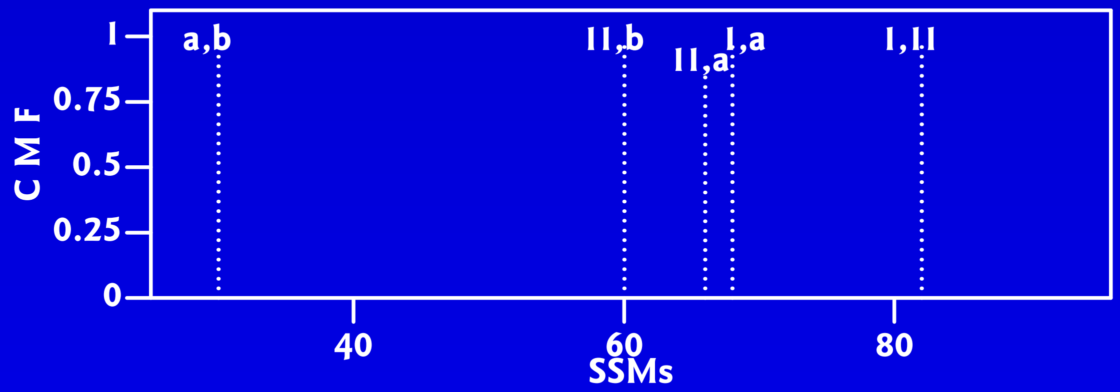
## Testing for Randomness Figure 7



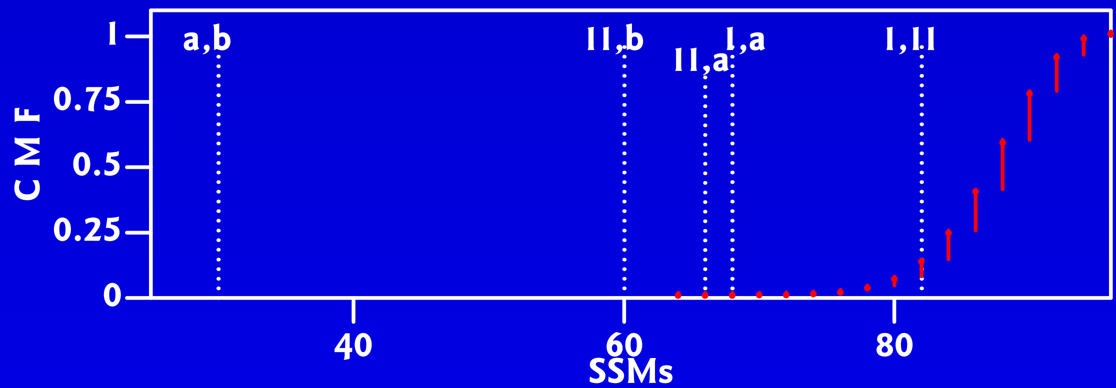
## Testing for Randomness Figure 7



## Testing for Randomness Figure 7

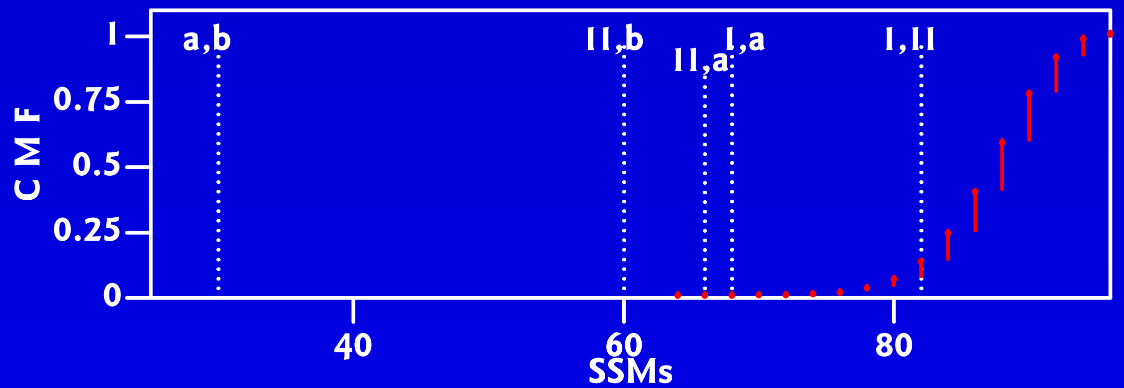


## Testing for Randomness Figure 7




The red lines are the CMF of pairs of sets of random series (3 series, 48 observations) from 100,000 Monte Carlo parameter bootstraps.

## Testing for Randomness Figure 7




The red lines are the CMF of pairs of sets of random series (3 series, 48 observations) from 100,000 Monte Carlo parameter bootstraps.

The one-sided c.i. at 1% corresponds to a SSM of 76, and at 5% 80.

Cannot reject the null hypothesis (random sets) for Chain 1 and Run 11; reject the null (random) hypothesis for all other pairs.

## 6. Conclusions — the State Similarity Measure

**This measure, *the State Similarity Measure* (SSM), is sufficient to allow us to put a number on the degree of similarity between two sets of time-series which embody dynamic responses. There is no limit to the number of time-series in each set, although the two sets must contain an equal number of series.**

<

## 6. Conclusions — the State Similarity Measure

This measure, *the State Similarity Measure* (SSM), is sufficient to allow us to put a number on the degree of similarity between two sets of time-series which embody dynamic responses. There is no limit to the number of time-series in each set, although the two sets must contain an equal number of series.

Such a metric is necessary for scoring the distance between any two such sets, which previously was unavailable.

## 6. Conclusions — the State Similarity Measure

**This measure, *the State Similarity Measure* (SSM), is sufficient to allow us to put a number on the degree of similarity between two sets of time-series which embody dynamic responses. There is no limit to the number of time-series in each set, although the two sets must contain an equal number of series.**

**Such a metric is necessary for scoring the distance between any two such sets, which previously was unavailable.**

**Here, the SSM has been developed to allow us to measure the extent to which a simulation model that has been chosen on some other criterion (e.g. weekly profitability) is similar to historical sets of time-series.**

## 6. Conclusions — the State Similarity Measure

**This measure, *the State Similarity Measure* (SSM), is sufficient to allow us to put a number on the degree of similarity between two sets of time-series which embody dynamic responses. There is no limit to the number of time-series in each set, although the two sets must contain an equal number of series.**

**Such a metric is necessary for scoring the distance between any two such sets, which previously was unavailable.**

**Here, the SSM has been developed to allow us to measure the extent to which a simulation model that has been chosen on some other criterion (e.g. weekly profitability) is similar to historical sets of time-series.**

**The SSM will also allow us to measure the distance between any two sets of time-series and so to estimate the parameters, or to help calibrate a model against history, or to compare any two such sets.**