

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Empirical Validation of Simulated Models through the GSL-div: an Illustrative Application

Francesco Lamperti [°]

[°] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy, and
FEEM, Milan, Italy

2016/18

November 2017

ISSN(ONLINE) 2284-0400

Empirical Validation of Simulated Models through the *GSL-div*: an Illustrative Application

Francesco Lamperti

Abstract

A major concern about the use of simulation models regards their relationship with the empirical data. The identification of a suitable indicator quantifying the distance between the model and the data would help and guide model selection and output validation. This paper proposes the use of a new criterion, called *GSL-div* and developed in Lamperti (2017), to assess the degree of similarity between the dynamics observed in the data and those generated by the numerical simulation of models. As an illustrative application, this approach is used to distinguish between different versions of the well known asset pricing model with heterogeneous beliefs proposed in Brock and Hommes (1998). Once the discrimination ability of the *GSL-div* is proved, model's dynamics are directly compared with actual data coming from two major stock market indexes (EuroSTOXX 50 for Europe and CSI 300 for China). Results show that the model, once calibrated, is fairly able to track the evolution of both the two indexes, even though a better fit is reported for the Chinese stock market. However, I also find that many different combinations of traders' behavioural rules are compatible with the same observed dynamics. Within this heterogeneity, an emerging common trait is found: to be empirically valid, the model has to account for a strong trend following component, which might either come from a unique trend type that heavily extrapolates information from past observations or the combinations of different types with milder, or even opposite, attitudes towards the trend.

Keywords Simulated Models · Empirical Validation · Model Selection · *GSL-div*

1 Introduction

Empirical validation is crucial for all modelling frameworks providing support to policy decisions, independently of their theoretical background. Even though Agent Based Models (ABMs) have often been advocated as promising alternatives to neoclassical models rooted in the dogmatic paradigms of rational expectations and representative agents, there are still some concerns about how to bring them *down* to the data (Windrum et al, 2007; Gallegati and Richiardi, 2009; Grazzini and Richiardi, 2015). In macroeconomics, for example, Giannone et al (2006), Canova and Sala (2009) and Paccagnini (2009) provide details about how to estimate and validate Dynamic Stochastic General Equilibrium models. However,

F. Lamperti

Scuola Superiore Sant'Anna, Institute of Economics, Piazza Martiri della Libertà 33, 56127, Pisa (Italy)

Fondazione Eni Enrico Mattei, corso Magenta 63, 20123, Milan (Italy)

E-mail: f.lamperti@santannapisa.it

their approach cannot be extended to settings where an analytical solution of the model (or an equilibrium) does not exist, which are typical cases in ABMs, system dynamics and complex systems more in general. Broadly speaking, these numerical models are validated through a comparison of the statistical properties emerging from simulated and real data. In many cases, this amounts at replicating the largest possible number of stylized facts characterizing the phenomenon of interest (see Dosi et al, 2010, 2013, 2015 for business cycle properties, credit and interbank markets or Pellizzari and Forno, 2006; Jacob Leal et al, 2015 for financial markets). Recent attempts are trying to enrich empirical validation beyond the simple replication of empirical regularities, thereby requesting models to generate series that exhibit the same dynamics (Marks, 2013; Lamperti, 2017), conditional probabilistic structure (Barde, 2016b) and causal relations (Guerini and Moneta, 2016) as those observed in the real world data.¹ At least partially, such contributions have been motivated by the unsatisfactory results delivered by calibration. In general, it is difficult to justify the choice of one combination of model's parameters over another, and calibration can be thought exactly as the exercise of selecting the values of the parameter set that best fit the real data.

In this paper I present an application of the *GSL-div* developed in Lamperti (2017) to validate model's output against real word data and explore the behaviour of the model quantifying the distance between the dynamics observed in the data and those numerically simulated. *GSL-div* stands for Generalized Subtracted L-divergence and constitutes an information theoretic criterion that builds on the L-divergence (Lin, 1991) and measures the distance between distributions of patterns retrieved in time series data. Validation is achieved capturing the ability of a given model to reproduce the distributions of time changes (that is, changes in the process' values from one point in time to another) in the real-world series, without the need to resort to any likelihood function or to impose requirements of stationarity. The *GSL-div* adds something that seems missing in the literature: a precise quantification of the distance between the model and data with respect to their dynamics in the time domain. On this side, my work builds on Marks (2013) and extend it by capturing and emphasizing the dynamical nature of time series models, which is, for example, loosely represented by the longitudinal moments used in many calibration exercises. The *GSL-div* is tested on the series produced by the well known asset pricing model with heterogeneous traders developed in Brock and Hommes (1998).

The rest of the paper is organized as follows. Section 3 introduces the *GSL-div*, discusses its main properties and provides a simple example; section 4 summarizes the mathematical structure of the model that will be used throughout the paper and validated against historical data; section 5 constitutes the core of this contribution, it illustrates and discusses the results I obtained. Finally, section 6 concludes the paper and provides some insights into future research.

¹ See also Fagiolo et al (2017) for a survey.

2 Literature Review

In the last decade a variety of efforts have been carried out to address the issue of calibrating and validating simulation models and, more specifically, agent based models.²

In those cases where the model is sufficiently simple and well behaved, it would be possible to derive a closed form solution for the distributional properties of a specific output of the model and to estimate the parameters governing such distributions by standard statistical techniques (Alfarano et al, 2005, 2006; Boswijk et al, 2007). However, in the majority of cases, policy oriented models do not allow such procedures (see, for example, contributions in Dawid and Fagiolo, 2008 and LeBaron and Winker, 2008). When models' complexity prevents to obtain closed form solutions, more sophisticated techniques are required.

One of the possible ways forward has been identified in *indirect inference* (Gourieux and Monfort, 1997). Relying on this technique, Bianchi et al (2007, 2008) have targeted a specific medium-scale macroeconomic ABM and estimated a subset of its parameters. Starting from the same procedure, Gilli and Winker (2003) and Winker et al (2007) have introduced an algorithm and a set of statistics which led to the construction of an objective function subsequently applied to exchange rate models. Refining this framework, Franke and Westerhoff (2011, 2012, 2016) have produced a series of papers where various financial ABMs have been successfully estimated through what they call method of simulated moments (MSM), which has found application also in Fabretti (2012) and, more recently, in Franke (2016). Similarly, Grazzini and Richiardi (2015) have performed the estimation of a simple ergodic ABM both in the long run equilibrium and during transitional dynamics using what is labelled as simulated minimum distance. The fil-rouge linking together all these approaches is that they seek to identify numerical parameter values such that some summary statistics of interest — or “moments” — that are computed from the simulations of the model, come close to their empirical counterparts. An advantage of this approach, as well as the one proposed in this paper, consists in that it does not require the likelihood function.

According to Winker et al (2007), the moments and the statistics used in the objective function must be robust, reflect statistical properties of the real data and exhibit the potential to discriminate between alternative models or parameter values. Not all calibration procedures appear persuasive in this respect. For example, Amilon (2008) estimates a relatively simple model of financial markets with 15 parameters (but only 2 or 3 agents) by efficient method of moments and reports an high sensitivity of the model to the assumptions on the noise term and stochastic components, questioning the performance of calibration exercises more in general. In addition, it appears straightforward that even if calibration delivers one or more array of parameters that maximise model's fit with the data, it is not automatic that this fit is a reasonably good one. The *GSL-div* might contribute to such a strand of the literature. In particular, it

² In this paper, I use the terms *calibration* and *estimation* interchangeably. They both indicate an exercise where the problem of finding a vector of parameters minimizing some loss function expressing the fit between model's output and real data is solved.

might be included in the set of moments to be matched. However, as underlined in Barde (2016a), the main advantage of the *GSL-div* is that it prevents from the arbitrary selection of moments to match. This is because the models would be scored on the basis of their entire, simulated, conditional distributions. Such a feature is shared with the approach developed in Barde (2016b).

Beyond indirect inference, alternative techniques have been recently employed to estimate simulation models. In these exercises, the key choice seems to boil down to the function that measures models' fit with the data. Recchioni et al (2015) used a simple gradient-based calibration procedure to conveniently sample the parameter space minimizing a standard loss function based on the cumulative squared errors. Kukacka and Barunik (2016) have instead developed an approach that maximise a non-parametric version of the simulated likelihood function, while Grazzini et al (2017) have comparatively tested a variety of parametric and non-parametric loss functions.

In general, it appears difficult to single out a globally superior approach, which systematically outperforms the alternatives.³ In such a context, it gains importance the assessment of the the ex-post validity of simulation models, i.e. expressing the extent to which somehow calibrated models can reproduce the properties of the real data.⁴ Remarkably, these exercises should be performed by means of tools that are different from those employed for model estimation. Finally, in line with Grazzini et al (2017) and Lamperti et al (2017), I believe that validation should inform about the behaviour of the model within large regions of the parameter space.

3 Validation and the *GSL-div*

3.1 Validation

Validation is a complex task that encompasses diverse aspects of the overall modelling activity. In this paper I interpret validation as the exercise of assessing the fit of one or different models with empirical data. To provide a general context, Manson (2002) distinguishes between *output validation* and *structural validation*. The latter asks how well the simulation model represents the (prior) conceptual model of the real-world system, while the former asks how successfully the simulations' output exhibits the historical behaviours of the real-world target system. Output validation can be directly related to what Leombruni et al (2006) define as *empirical validity* of a model, i.e. validity of the empirically occurring true value relative to its indicator. Following Rosen (1985), it is useful to think of two parallel unfolding: the evolution of the system (an economy, a market, an industry) and the evolution of the model of the system. If the model is correct, properly calibrated and initial conditions have been fixed according to the initial status of the real system, the simulation should mirror the historical evolution of such system with respect to the variables, or statistics, of interest. It is relevant to notice that I do not interpret validation as a binary

³ The interest reader might want to look at Lux and Zwinkels (2017) for a review of validation and calibration approaches especially focused on models of financial markets.

⁴ As it will be briefly discussed below, this concept refers to what is called *output validation*.

test on model acceptability on the basis of its realism; rather, I interpret it as the process of assessing the relationship of similarity between empirical data and model's output. The paper provides and tests a statistical measure that allows such an exercise; it is then left to the modeller the task of establishing whether the similarity is satisfactory or not.⁵ This is also in line with (Windrum et al, 2007, section 2), and the exercises provided in Marks (2013) and Guerini and Moneta (2016).

In our context, a *model* is broadly defined as a representation of a system that is able to produce some synthetic output tracking the evolution of the system itself. Formally, the output of a model can be represented by the collection of all micro-states at time t , $\mathbf{X}_t \equiv \{x_{i,t}\}$ with $i = 1, \dots, N$ and $t = 1, \dots, T$ such that

$$x_{i,t} = f_i(\boldsymbol{\theta}, \mathbf{X}_{t-1}), \quad (1)$$

where f_i can be any (deterministic or stochastic) real valued function and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbf{R}^d$ is a vector of d parameters. I assume that, for each model, $\boldsymbol{\theta}$ and \mathbf{X}_0 are exogenously given. In other words, the model has already been calibrated: initial conditions and parameters are assigned precise values. In addition, *real world data* are defined as the empirically observable elements of the system.

3.2 The *GSL-div*

The *GSL-div* is a measure, developed in Lamperti (2017), that determines the degree of similarity between the dynamics observed in real data and those produced by the numerical simulation of a model. The only input it requires are the real and simulated series. It should be noticed that such a comparison might involve objects having different dimensions. While real quantities can be observed once and only once, a (stochastic) simulation model provides different realizations of the same process each time the seed of the (pseudo) random number generator is changed. To the contrary, if the model is deterministic, a unique (or many identical) series will be obtained for each variable of interest. The *GSL-div* can be used in both the two cases, but since many economic decisions or events (e.g. innovation outcome) naturally entail a random component, it has been thought to treat stochastic models. Further, time series (real and simulated) that are comparable in terms of sampling frequency should be preferred. The *GSL-div* does not autonomously distinguish between different frequencies. For example, a macroeconomic model that is meant to produce quarterly data should be compared to quarterly empirical counterparts.

The approach proposed in this paper builds on a solid theoretical background. It uses the *L-div* (Lin, 1991) as building block to measure the distance between distributions of patterns retrieved in empirical

⁵ For example, a straightforward procedure could rely on the selection of a threshold distinguishing acceptable outcomes from those that are not. This would then allow to frame validation as in Marks (2007). However, the choice of the threshold is likely to depend dramatically on the specific model or phenomenon under study and, therefore, I think it should remain on the modeller's.

and simulated data.⁶ Further, when the L -div is available for patterns of different lengths, it provides a straightforward aggregation that accounts for the fact that multiple model runs should be compared with the same empirical data. This step ensures that similarity between model and data comes from a systematic evaluation of the simulation output and not just from “lucky” runs.

The estimation of the GSL -div follows a simple, four-steps procedure that is discussed below and exemplified in section 3.3. The algorithm (pseudo-code) to compute the GSL -div is also provided in the Appendix. Shortly, the procedure is as follows:

1. Time series (both real and simulated) are symbolized
2. Patterns of symbols (i.e. *words*) are observed through rolling windows of different length $l = 1, \dots, L$.
3. Distributions of patterns, f_l , are estimated for each windows' length
4. The distance between distributions from real and simulated data are evaluated and aggregated.

The first step consists in series' symbolization. This procedure is carried out to constrain them to take only a finite set of values. Let $\{x(t)\}_{t=1}^T$ be a time series of total length T where each $x(t)$ is a real number. To symbolize it, I could take the interval $[x_{min}; x_{max}]$ and partition it in $b \in \mathbf{N}_0$ subintervals, each of equal length. Then, subintervals would be numbered increasingly from 1 to b , with 1 assigned to $[x_{min}; x_{min} + \frac{(x_{max}-x_{min})}{b})$. However, this procedure would be very sensitive to outliers, which might distort the entire symbolization of the series. To cope with this issue we suggest to replace the use of the sample minimum and maximum with some extreme percentiles, say $x_{1\%}$ and $x_{99\%}$. Then, the same procedure as above applies and observations falling in the tails will be numbered using 1 if $x \leq x_{1\%}$ and b if $x \geq x_{99\%}$. Obviously, the choice about the percentiles is upon the researcher and might depend on the nature of the data. In the present paper I will use an arbitrary interval for the didactic examples presented below and in section 3.3 while I will move to $[x_{1\%}; x_{99\%}]$ for all the other exercises. The parameter b controls for the precision of the symbolization: if $b = 1$ the symbolized series takes one and only one value (namely 1), while $b \rightarrow \infty$ implies that we are back to the (scaled) real-valued process. The symbolization is simple and works as follows: each $\{x(t)\}_{t=1}^T$ is mapped into the natural number corresponding to the partition interval where it falls (see the Appendix for a graphical insight). As an example, consider the following time series $x(t)$ with $T = 3$: $\{0; 0.4; 1\}$. Choosing $b = 2$, the symbolized series will be $x^s(t) = \{1, 1, 2\}$, while choosing $b = 9$ the symbolized series becomes $x^s(t) = \{1, 4, 9\}$, where the apex s stands for *symbolized*. Obviously, the information loss about the behaviour of the stochastic process due to the symbolization becomes smaller and smaller as b increases. On the other side, low values of b would likely wash away processes' noise the modeller might not be interested in. In general, the GL -div is very precise in recognizing similarity in time series dynamics even for very small values of b (see section 5.1 and the Appendix).

Now, I iteratively construct *words* of symbols having lengths $l = 1, \dots, L$ with $L \leq T$ by pasting together successive symbolized observations. Each of the resulting words with $l > 1$ corresponds to a

⁶ The interest reader might want to know that the L -div is a symmetric generalization of the more widely known Kullback-Leibler divergence (Kullback and Leibler, 1951).

realized pattern of time to time changes in the process. For example, with $b = 9$ and $l = 1$ previous symbolized time series, $x^s(t)$, comprehends words $\{1, 4, 9\}$; with $l = 2$ it comprehends $\{14, 49\}$, indicating two increasing trends with the second being more pronounced than the first. The collection of all words of length l would be a vocabulary. In particular, we define $S_{l,b}$ as the vocabulary of words having length l that can be composed using the alphabet $A_b = \{1, 2, \dots, b\}$. Formally, it would be possible to use the following notation, $S_{l,b} := \binom{A_b}{l}$, which indicates that $S_{l,b}$ is the set of all the l -combinations of A_b . The cardinality of the **vocabulary** is defined as $a_{l,b} = 2^{S_{l,b}} = b^l$, $\forall l = 1, \dots, L$ and corresponds to the number of all possible l -words that can be composed once b is chosen. Once vocabularies have been established, time series are explored through rolling windows that determine which words are retrieved and which are not. It should be noticed that words in the same series overlap. This is a relevant feature allowing to capture each possible pattern independently of the initial (and final) observation in the sample. Obviously, if series are of length T , $T - l + 1$ words will be obtained for each value of l . L represents the maximum length of the windows used to compare the behaviour of real data with synthetic ones. It has to be chosen considering both (i) the nature of the phenomenon of interest and (ii) the size of the available real-world time series.⁷

Now, frequency of symbols in each series are estimated. Let $x^s(t)$ and $y^s(t)$ be two symbolized time series. In general, one can think of the first as the real series and the second as the synthetic output of a simulation. S_l is the vocabulary of symbols available at length $l = 1, \dots, L$ for any given b , and $f_{x,l}, f_{y,l}$ are vectors collecting the occurrence frequencies. Similarity in the behaviour of the time series is inferred by the systematic comparison of the frequency distributions of symbols. In particular, for each value of $l = 1, \dots, L$ the distance between $f_{x,l}$ and $f_{y,l}$ is evaluated through the a simple modification of the *L-div* (Lin, 1991). The *L-div* can be written as the difference between twice the entropy of $(f_{x,l} + f_{y,l})/2$ and the sum of the entropies of $f_{x,l}$ and $f_{y,l}$. Since the real series is observed once and only once, and provided it might be short and/or non-stationary, its entropy can be difficult to estimate. However, it is immediate to see that every multi-model comparison with the same set of empirical data would be independent from such a quantity. For these reasons, I “subtract” the entropy of $f_{x,l}$ from the *L-div* and focus on the term $2H(f_{x,l} + f_{y,l})/2 - H(f_{y,l})$, where $H(\cdot)$ indicates the Shannon’s entropy (Shannon, 1948).⁸ Finally, the aggregation of the subtracted *L-divs* gives the Generalized Subtracted L-divergence. A formal definition and additional details follow immediately.

Definition 1 (*GSL-div* between distributions)

Assume that b and L have been fixed. Let $x^s(t)$ and $y^s(t)$ be two symbolized time series and use two

⁷ For example, one should notice that the larger L the more likely that simple shifts in time between two processes would not be penalized. The choice of whether to penalize is likely to depend on the nature of the model and the investigation. If the modeller is unsure about the length of the transient period of his/her model, not to penalize such shifts might be the preferred option.

⁸ It should be noticed that the Shannon’s entropy expresses the amount of uncertainty associated with the behaviour of a random variable. In time series it is also used as a measure of complexity, which should be intended as regularity of behaviour. The more the behaviour is irregular, the more complex the series, the larger the entropy.

arrays, $\mathbf{f}_x = \{f_{x,1} \dots f_{x,L}\}$ and $\mathbf{f}_y = \{f_{y,1} \dots f_{y,L}\}$, to collect the frequency distributions of patterns for $l = 1, \dots, L$. Finally, let w_l be aggregation weights that sum up to 1. We define the Generalized Subtracted L-divergence between the two distributions as

$$\begin{aligned} D_{GSL}(\mathbf{f}_x || \mathbf{f}_y) &= \sum_{l=1}^L w_l \left(-2 \sum_{s \in S_l} m_l(s) \log_{a_l} m_l(s) + \sum_{s \in S_l} f_{y,l}(s) \log_{a_l} f_{y,l}(s) \right) \\ &= \sum_{l=1}^L w_l (2H^{S_l}(\mathbf{m}_l) - H^{S_l}(\mathbf{f}_{y,l})), \end{aligned} \quad (2)$$

where the symbol $H^{S_l}(\cdot)$ indicates the Shannon entropy of a distribution over the state space S_l , $m_l = (f_{x,l} + f_{y,l})/2$ is the mean distribution and a_l represents the cardinality of the vocabulary available at length $l = 1, \dots, L$ and is used as the base for the logarithms.

The second equality in equation (2) represents the *GSL-div* as a (weighted) difference between the Shannon entropy of the mean distribution, which is obtained collecting the average frequencies of each word in the real and synthetic series, and the distribution of patterns retrieved in the model's simulation (the synthetic series). Obviously, if the distributions of patterns in the two series (\mathbf{f}_x and \mathbf{f}_y) were equal, also the entropies would be identical, indicating that $x^s(t)$ and $y^s(t)$ exhibit the same regularity of behaviour or, using the glasses of information theory, the same information about the unknown data generating process. To the contrary, when the two distributions are different, the entropy of the mean distribution increases with respect to the one of the simulation (because novel patterns are introduced), pointing to the fact that empirical data exhibits regularities that the simulation is not able to match.⁹

When dealing with a deterministic model, the use of (2) might be satisfactory in the characterization of the distance between real and simulated dynamics, since all the available information has already been exploited. In the more interesting case of stochastic models, one might want to estimate the distance between data and model relying on the probabilistic structure of the latter. For example, one would like to feed the *GSL-div* with the *true* probability that the model assigns to each sequence of symbols rather than its frequency. However, if the model is not solvable analytically, the only information about the stochastic process underlying the aggregate behaviour of these models is available through the synthetic series they produce. Now, imagine to take an ensemble of independent runs of the same, previously calibrated model. In this context, the following proposition is proved and discussed in details in Lamperti (2017, appendix B).

Proposition 1 *Let $\bar{p}_\mu(s)$ be the average probability that model μ assign to each symbol in the interval $t \in [1, \dots, T]$ and $p(s)$ the frequency of the same symbol observed in the real data. The *GSL-div* between $\bar{p}_\mu(s)$ and $p(s)$ is approximately equal to*

⁹ This interpretation is in line with the one of divergences in information theory. For example, the usual understanding of the Kullback and Leibler (1951)'s divergence suggests that it is a measure of the inefficiency of assuming a given distribution when the true one is different.

$$\begin{aligned}
GSL(p(s) \parallel \bar{p}_\mu(s)) &\cong \sum_{l=1}^L w_l \mathbf{E} \left(-2 \sum_{s \in S_l} m_l(s) \log_{a_l} m_l(s) + \sum_{s \in S_l} f(s) \log_{a_l} f(s) \right) \\
&+ \sum_{l=1}^L w_l \left(\frac{B_l^m - 1}{4T_l} - \frac{B_l^{\bar{p}_\mu} - 1}{2T_l} \right). \tag{3}
\end{aligned}$$

where $\mathbf{E}(\cdot)$ is the expectation over an ensemble of independent runs, w_i are arbitrary positive weights such that $\sum_l w_l = 1$, $B^j \geq 1$ is the cardinality of the support of $j = \{m, \bar{p}_\mu\}$ and $T_l = T - l + 1$ is the number of observations of size l . The second line of (3) represents a correction term for the systematic bias arising from the use of frequency distributions in place of true probabilities.

The use of a sufficiently large ensemble of runs allows to capture the overall degree of similarity between models and the data, washing away run-specific effects. Since it is not infrequent that ABMs exhibit chaotic dynamics, stochastic shocks and/or tipping points, one run of the model might be completely different from the others.¹⁰ To adequately explore behaviour of the model, a relatively large number of runs have to be considered and, when it comes to validation (after having conveniently calibrated and/or estimated the model) different runs should ideally exhibit relatively similar dynamics.¹¹ Finally, it is relevant to underline that, the application of (3) in the case of a deterministic model boils down exactly to (2).

Now, the only element that remains to be determined is the vector of aggregation weights, w_l with $l = 1, \dots, L$. In general, they are chosen to increase with l for two reasons. On one side, such choice reflects the greater importance assigned to patterns of similar behaviour lasting over longer time-windows and, on the other, it compensates for the increasing value of the logarithms' base a_l . A detailed discussion about weights' selection is found in Lamperti (2017, section 2.3), together with different robustness exercises showing results' poor sensibility to the choice of weights. For the purpose of this paper we consider *additively progressive* weights, that is, weights such that their first difference is constant and collectively sum up to one. This choice is additionally justified by the fact that additively progressive weights are unique; once L is fixed there is a unique vector satisfying previous requirements:

$$w_{l+1} = w_l + \frac{2}{L(L-1)} \quad \text{with } l = 1, \dots, L.$$

The *GSL-div* exhibits a set of interesting properties, even though it is important to recall it does not satisfy triangular inequality and, therefore, it is not a metric. In particular,

1. The *GSL-div* is well defined for all p and \bar{p}_μ
2. $0 < GSL(p \parallel \bar{p}_\mu) < 2$
3. $GSL(p \parallel \bar{p}_\mu) = H(p) \iff p = \bar{p}_\mu \quad \forall l = 1, \dots, L \text{ and } s \in S_{l,b}$.

¹⁰ See the discussion in section 3.3 of Pyka and Fagiolo (2007) and, for a recent contribution on the issue of tipping points in macroeconomic agent based models, Gualdi et al (2015).

¹¹ The implicit assumption behind this reasoning is that the model is ergodic or, in case it is not, that all runs refers to the same statistical equilibrium. See Grazzini (2012) for details and tests.

The first property guarantees that for any couple of probability vectors, the *GSL-div* between the two exists and can be computed independently of their support, which might either be the same or not.¹² The second property indicates that the *GSL-div* is bounded both from above and below. This is interesting and desirable for validation purposes, as benchmarks for extreme cases are naturally provided. Finally, property 3. shows the effective lower bound for the *GSL-div*: it is equal to the entropy of the real time series if and only if for every word length model μ assigns each symbol with the same probability the observed time series does. This would be the case of perfect matching between the dynamics of the simulated series and the data, whose patterns are mirrored exactly in all model's runs. Conversely, the *GSL-div* moves towards its upper bound, indicating that two series exhibit completely different behaviours, as soon as they tend to constantly persist over different states (i.e. symbols). For example, if we were to compare the dynamics of the inflation rates of two countries, this approach would detect maximally divergent behaviours when prices are constant in one country (zero inflation) but constantly rising (or falling) in the other or, in the same way, when deflation affects the first and hyperinflation the second.¹³ However, it should be noticed that the upper bound of 2 constitutes a theoretical value, which is extremely unusual to reach in practice. Finally, it is relevant that for correction terms sufficiently close to zero, the *GSL-div* boils down to the mean over an ensemble of independent runs. Therefore, it would be possible to formally test for the difference between *GSL-divs* using a t-test for the means. Otherwise, the construction of a proper test is required but goes beyond the scope of the present paper.

3.3 A simple example

Here I propose a simple and brief example showing how the *GSL-div* estimation works in practice. I consider three time series of length $T = 10$, called x , y and z respectively, and omit dependence on time to ease notation. In particular,

$$\begin{aligned} x &= \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \\ y &= \{10, 9, 8, 7, 6, 5, 4, 3, 2, 1\} \\ z &= \{2, 2.5, 3, 3.5, 4, 4.5, 4, 4.5, 5, 5.5\}. \end{aligned}$$

These series might be thought as a real-world quantity (say, x) and the output of two deterministic competing models (y and z). Figure 1 graphically shows these three time series. By inspection, we notice that z behaves much more closely to x than y does; even though they have different slopes, they are both increasing over time, apart from one downward step in z . To the contrary, y is always decreasing, at a constant pace, which exactly corresponds to the opposite with respect to x 's one. Therefore, x and y touch the same states, each once and only once, but with reversed dynamics.

¹² This constitutes a direct advantage vis-à-vis, for example, the Kullback-Leibler divergence (Kullback and Leibler, 1951).

¹³ Note that in these example the observation units are inflation rates, and not price levels, which would deliver different results.

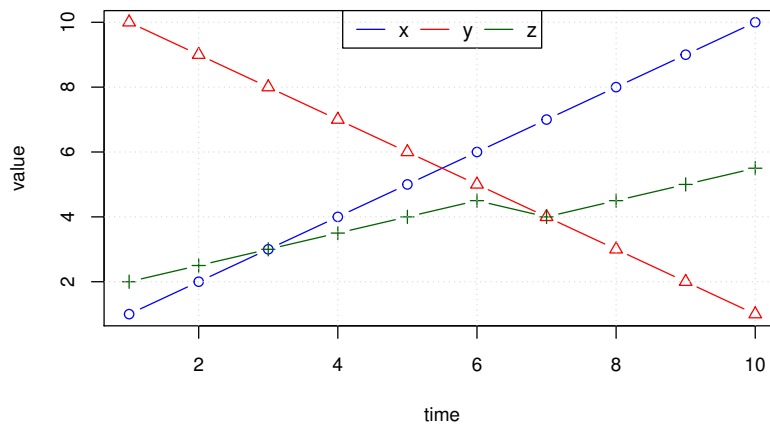


Fig. 1: Three simple time series.

Now we want to employ the *GSL-div* to detect similarities between x and both y and z . Obviously, we expect it to identify a much closer dynamics between x and z rather than between x and y . First, we proceed with the symbolization process using $b = L = 3$. These values are chosen for narrative reasons; on one hand they are quite low, thereby making it more difficult for the *GL-div* to capture similarities in series' behaviour, on the other they allow for a relatively short alphabet, which eases the representation of symbols. Table 1 reports symbols observed in the symbolized version of each series and their frequency. Each symbol represents a pattern that is observed in the data and its frequency measures the recursion of such symbol over time. The more similar frequency distributions over the different alphabets, the more time series exhibit analogous dynamics.

Some similarity of behaviour among x , y and z can be singled out just watching at the symbols reported in Table 1. Considering length equal to one ($l = 1$), the focus falls on the persistence of each series within each state. Moving to $l = 2$ patterns of length two are analysed. At this level, it starts emerging the key difference amongst our three series: two are (almost always) increasing while the other decreases over time. Hence, the supports of the distribution of patterns for x and z share a larger number of elements than it happens for x and y .¹⁴ Beyond such similarities, the present approach recognizes from the very beginning the presence of a downward sloping episode in z , which is absent in x . When words of length three are studied, such an evidence is confirmed and, in addition, we capture the fact that z 's downward phase lasts just one period.

The straightforward application of (2) leads then to the comparison of the *GSL-divs* for the couples (x, y) and (x, z) . As expected, $D_{GSL}(\mathbf{f}_x | \mathbf{f}_y) = 1.037 > 0.799 = D_{GSL}(\mathbf{f}_x | \mathbf{f}_z)$, which gives a preliminary

¹⁴ We notice that some symbol can be retrieved in all the three supports; this is due to the low precision of the symbolization process, which does not allow to readily capture the strictly monotonic nature of series x and y . With $b = 10$, to the contrary, x and y 's supports at $l = 2$ would be completely different.

Table 1: Observed symbols and their frequencies for time series x , y and z .

	x	y	z
length	observed symbols		
1	{1}, {2}, {3}	{1}, {2}, {3}	{1}, {2}, {3}
2	{11}, {12}, {22}, {23}, {33}	{11}, {21}, {22}, {32}, {33}	{11}, {12}, {22}, {23}, {32}, {33}
3	{111}, {112}, {122}, {222}, {223}, {233}, {333}	{111}, {211}, {221}, {222}, {223}, {233}, {333}	{111}, {112}, {122}, {223}, {232}, {233}, {323}, {333}
	frequencies		
1	0.4; 0.3; 0.3	0.4; 0.3; 0.3	0.3; 0.3; 0.4
2	0.33; 0.11; 0.22; 0.11; 0.22	0.33; 0.11; 0.22; 0.11; 0.23	0.22; 0.11; 0.11; 0.22; 0.11; 0.22
3	0.25; 0.125; 0.125; 0.125; 0.125; 0.125; 0.125;	0.25; 0.125; 0.125; 0.125; 0.125; 0.125; 0.125;	0.125; 0.125; 0.125; 0.125; 0.125; 0.125

yet positive insight into the present approach's ability to identify similarities and differences in time series' dynamics, even with low precisions (b and L) and time series that are extremely short.

4 The Brock and Hommes Asset Pricing Model

To show the approach proposed in this paper more extensively, I rely on the widely known asset pricing model with heterogeneous agents proposed in Brock and Hommes (1998). The model is ideal for illustrative purposes. It is relatively simple and costless to simulate but, on the other side, it offers variegated dynamics that are linked to a rich parameter space. Further, it has already been used as a test-model for different calibration and validation exercises (e.g. Boswijk et al, 2007; Recchioni et al, 2015).

There is a population of N traders that can either invest in a risk free asset, which is perfectly elastically supplied at a gross return $R = (1 + r) > 1$, or in a risky one, which pays an uncertain dividend y and has a price denoted by p . Wealth dynamics is given by

$$W_{t+1} = RW_t + (p_{t+1} + y_{t+1} - Rp_t)z_t, \quad (4)$$

where p_{t+1} and y_{t+1} are random variables whose behaviour will be clarified in few lines and z_t is the number of the risky asset shares purchased at time t . Traders are heterogeneous in terms of their expectations about future prices and dividends and are assumed to be myopic mean-variance maximizers. In particular, each agent demands a number of shares that solves

$$\max_{z_{h,t}} \left\{ E_{h,t}(W_{t+1}) - \frac{\alpha}{2} V_{h,t}(W_{t+1}) \right\}, \quad (5)$$

which implies

$$Z_{h,t} = E_{h,t}(p_{t+1} + y_{t+1} - Rp_t)/(\alpha\sigma^2), \quad (6)$$

where h denotes a trader-specific quantity, α controls for the agents' risk aversion and σ indicates the conditional volatility, which is assumed to be equal across traders and constant in time. In the case of zero supply of outside shares and of different trader types, the market equilibrium equation can be written as

$$Rp_t = \sum n_{h,t} E_{h,t}(p_{t+1} + y_{t+1}), \quad (7)$$

where $n_{h,t}$ denotes the share of type h traders at time t . In presence of homogeneous traders, perfect information and rational expectations it is possible to derive the following no-arbitrage market equilibrium condition:

$$Rp_t^* = E_t(p_{t+1}^* + y_{t+1}), \quad (8)$$

where the expectation is conditional on all histories of prices and dividends up to time t and where p^* indicates the fundamental price. In case the process of dividends is independent and identically distributed with time unvarying mean, equation (8) has a unique solution where the fundamental price is constant and such that $p^* = E(y_t)/(R - 1)$. In what follows it is convenient to express prices as their deviations from the fundamental, $x_t = p_t - p_t^*$.

Trading happens over a number of periods, denoted by $t = \{1, 2, \dots, T\}$. At the beginning of each trading period t , agents make expectations about future prices and dividends. I assume that agents are heterogeneous in that they have different forecasts of p_{t+1} and y_{t+1} . Beliefs about future are assumed to take the following form:

$$E_{h,t}(p_{t+1} + y_{t+1}) = E_t(p_{t+1}^*) + f_h(x_{t-1}, \dots, x_{t-L}) \quad (9)$$

for all t and h . In words, investors believe that, in a heterogeneous world, prices may deviate from the fundamental value by some function $f_h(\cdot)$ depending upon past deviations from the fundamental price. Many forecasting strategies have been implemented in the economic literature, specifying different trading behaviours and attitudes (Banerjee, 1992; Brock and Hommes, 1997; Lux and Marchesi, 2000; Chiarella et al, 2009). I follow Brock and Hommes (1998) in using a simple linear representation of beliefs:

$$f_{h,t} = g_h x_{t-1} + b_h \quad (10)$$

where g_h is said to be the trend component and b_h the bias of trader type h . If $b_h \neq 0$, we call agent h a pure trend chaser if $g_h > 0$ (strong trend chaser if $g > R$) and a contrarian if $g < 0$ (strong contrarian if $g < R$). If $g_h \neq 0$, type h is said to be purely biased (upward resp. downward biased if $b_h > 0$ resp. $b_h < 0$). In the special case $g_h = b_h = 0$, we obtain (pure) fundamentalists, who believes that prices return to their fundamental value. It is also possible to include a prototype of rational agent, who is characterized

by $f_{rational,t} = x_{t+1}$. Rational agents have perfect foresight but, to obtain such a good prediction they are subjected to the payment of a cost C .¹⁵

To the purposes of the present application, I use a simple model with only two types of agents, whose behaviours vary according to the choice of trend components, biases and perfect forecasting costs. Combining equations (7), (9) and (10) it is possible to derive the following equilibrium condition:

$$Rx_t = n_{1,t}f_{1,t} + n_{2,t}f_{2,t}, \quad (11)$$

which allows to compute the price of the risky asset (in deviation from the fundamental) at time t . Traders' strategy is updated over time on the basis of accumulated wealth, which evolves according to equation (4). In particular the model allows for a switching behaviour that is governed by a parameter β in the following way. Each type h is associated with a fitness measure of the form:

$$U_{h,t} = (p_t + y_t - Rp_{t-1})z_{h,t} - C_h + \omega U_{h,t-1} \quad (12)$$

where $\omega \in [0, 1]$ is a weight attributed to past profits. As time goes by, a strategy may become more profitable than the other one in term of fitness. All agent starts with their own (pre-specified) strategy, however at the beginning of each successive period they reassess the profitability of their own type relatively to others. The probability that a trader chooses the strategy h is given by ‘‘Gibbs’’ probability:

$$n_{h,t} = \frac{\exp(\beta U_{h,t})}{\sum_h \exp(\beta U_{h,t})}. \quad (13)$$

The rewind algorithm is designed so that the successful strategy gains a higher number of followers. In addition, algorithm introduces a certain amount of randomness, and more profitable strategies has a finite probability not to be preferred over less successful ones. In this way, the model capture imperfect information and bounded rationality of agents. This randomness also helps unlocking the system from the situation where all traders ends up with the same strategy h . The parameter $\beta \in [0, +\infty)$ controls for the intensity of choice of the traders: the higher its values, the larger the likelihood of switching.

5 Model Selection and Validation

In this section I will illustrate and discuss the results obtained applying the *GSL-div* to the Brock and Hommes model described above, with two trader types denoted as 1 and 2. In particular, two main exercises are presented. First, I will show that the *GSL-div* is an adequate measure to distinguish between different versions and parameter configurations of the model. To the purposes of validation, this is an explicit requirement in Winker et al (2007). Secondly, I will move to the comparison of simulated dynamics,

¹⁵ More in general, one could allow for the possibility that a positive a cost might be by paid also by non-rational traders; this is to mirror the fact that some trader might want to buy additional information which, however, might not be able to use (e.g. because of computational mistakes).

obtained through a calibrated version of the model, with real data from two major stock market indexes, namely the EuroSTOXX 50 and the CSI 300 (which represent the main European and Chinese markets respectively).

5.1 Discriminating among different models

To illustrate the ability of the *GSL-div* in distinguishing amongst different models, a known Data Generating Process (DGP hereafter) is needed, as it will be used as a benchmark against alternative and competing model configurations. Given the relatively high number of parameters in the model, a nearly infinite number of choices, delivering a wide and variegated array of dynamics (see Brock and Hommes, 1998) were available at this stage. In that, I tried to balance — on one side — the need to be clear and concise and — on the other — the need of discriminating among objects that are inherently different but produce a relatively similar behaviour. Five models, in addition to the DGP, have been chosen and their configurations summarized in table 2. Despite being few, they account for a variety of traders' behaviours. The DGP is characterized by two trend-follower trader types, with the second extrapolating much stronger than the first, who is also upward biased in contrast to the other. The switching parameter (β), which might take any positive value, is relatively low and in line with the numerical exercises carried out in Brock and Hommes (1998). The other models, M1-M5, are obtained using the DGP as reference and modifying one or more characteristics defining the attitude of traders, while leaving unchanged those parameters that represent broader context conditions (e.g. the risk-free interest rate, r , or the volatility of the asset σ) and summarized in the lower part of table 2. M1 and M2 modify the DGP in the intensity of choice of the two types, implying a much higher (M2) and lower (M1) likelihood of switching towards the strategy delivering higher payoffs. M3 simply differs in the initial share of traders of first type, which is exactly balanced in the DGP while exhibits a strong dominance of second type in this model configuration. M4 maintains the bias of type 1 traders but assumes they are trend contrarians rather than followers, while type 2 keep their strategy but extract significantly more information from previous prices. Finally, M5 considers a fundamentalist trader type vis-à-vis a trend follower, with the same switching attitude as modelled in the DGP. In addition to these features, it is relevant to point out that an element of randomness is included to enrich the framework and show the performance of the *GSL-div* in presence of noise and stochastic models. As in Brock and Hommes (1998), the dividend process, $\{y_t\}_{t=1}^T$, follows a stochastic process such that $y_t = \bar{y} + \epsilon_t$ where the noise term ϵ_t is i.i.d. uniformly distributed between -0.5 and $+0.5$ and, in our case, $\bar{y} = 0$. Such a formalization of the dividend process is kept unchanged for all the models considered.

Figure 2 collects plots of a randomly chosen realization of the price process (in deviation from the fundamental value) produced both by the DGP (in red) and the various competing models (in blue). Direct inspection shows that notwithstanding each model accounts for different trading attitudes of the two types, the dynamics are quite similar, at least for an unaided eye. The relevant exception is provided

Table 2: Parameters' value for Data Generating Process (DGP) and different models

Parameter	Brief description	DGP	M1	M2	M3	M4	M5
β	intensity of choice	4	2	40	4	4	4
n_1	share of type 1 traders	0.5	0.5	0.5	0.1	0.5	0.5
b_1	bias of type 1 traders	0.2	0.2	0.2	0.2	0.2	0
b_2	bias of type 2 traders	0	0	0	0	0	0
g_1	trend component of type 1 traders	0.2	0.2	0.2	0.2	-0.2	0
g_2	trend component of type 2 traders	1.2	1.2	1.2	1.2	1.8	1.2
C	cost of obtaining type 1 forecasts	0	0	0	0	0	0
ω	weight to past profits	0.5	0.5	0.5	0.5	0.5	0.5
σ	asset volatility	0.1	0.1	0.1	0.1	0.1	0.1
α	attitude towards risk	10	10	10	10	10	10
r	risk-free return	0.1	0.1	0.1	0.1	0.1	0.1

Note: The symbol n_1 is used to indicate the initial (at $t = 0$) share of type 1 traders.

by model M5, which robustly generates a continuously falling price until it gets constant. Whatever device, tool or methodology that aims at validating models, should be able to distinguish between those configurations yielding a truly different dynamics and, on the other side, to pull together those producing reasonably alike ones.

In our case, which treats an asset pricing model, dynamics are observed and compared with reference to two different quantities, that is, *prices* and *normalized returns*. More formally, if p_t is the price of a given asset at time t and τ is the sampling frequency, the logarithmic difference of prices gives the returns:

$$r_t = \log(p_t) - \log(p_{t-\tau}) \approx \frac{p_t - p_{t-\tau}}{p_{t-\tau}}, \quad (14)$$

which can be normalized subtracting the longitudinal mean over the sample of interest and dividing by the standard deviation,

$$nr_t = \frac{r_t - \langle r \rangle}{\sigma_r}, \quad (15)$$

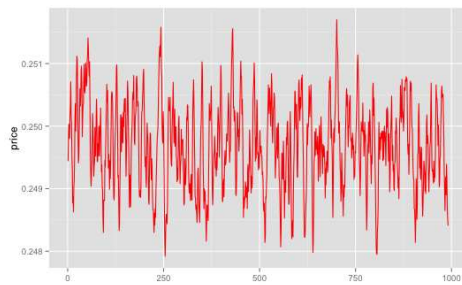
where nr_t indicates normalized returns at time t , σ_r is the standard deviation and $\langle \cdot \rangle$ is the time average over the considered period. In many applications assets' returns are normalized in order symmetrize their distribution and to wash away the effects of their long run volatility; the same convention is used here.

The simulation setup is simple and constructed to mirror a real problem. The DGP is used to obtain a single realization, which is then labelled as the real world data. As it happens in practice, this series will be the unique term of comparison for all the five competing models and the DGP itself. I will test the ability of different models to replicate the dynamics observed in the data and I will rely only on the output of the simulations. Therefore, each of the configurations included in table 2 is initialized with the

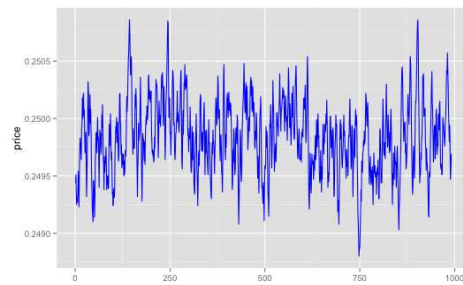
same conditions (but n_1 in the case of M3) and used to generate an ensemble of $R = 500$ independent runs each of length $T = 1000$.¹⁶ Then, the *GSL-div*, as expressed in equation (3), is employed to assess the similarity between what we have called real world data and models output. It is relevant to remark that the two free parameters in the procedure leading to the estimation of the *GSL-div*, that is, the precision

¹⁶ The unique exception is the DGP, which is run $R + 1$ times. Then one of these realizations is randomly selected as the real data while the others R are used to compare the DGP with these data.

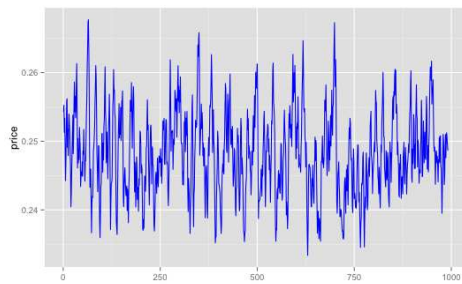
Fig. 2: Randomly chosen realizations of Data Generating Process and other models.



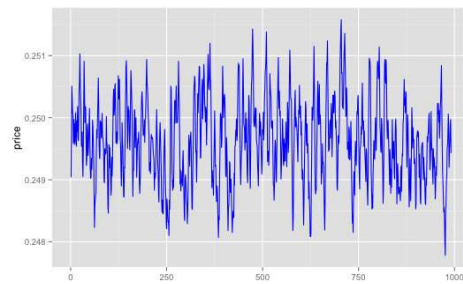
(a) Data Generating Process



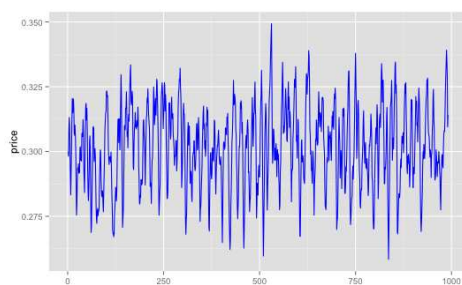
(b) Model 1



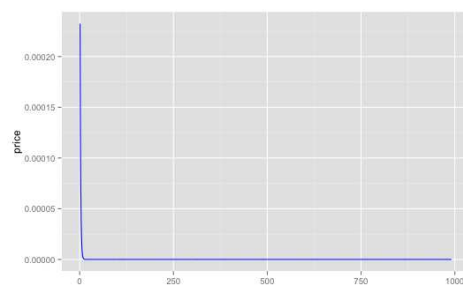
(c) Model 2



(d) Model 3



(e) Model 4



(f) Model 5

of the symbolization (b) and the maximum words' length (L) are set accordingly to Lamperti (2017).¹⁷ Table 3 reports both the distances estimated for different lengths of symbols and the *GSL-div*.¹⁸

Table 3: *GSL-div* between data and models for price and normalized returns.

Prices							
word length	weights	DGP	M1	M2	M3	M4	M5
1	0.05	0.739642	1.000337	0.859059	0.739737	0.741820	1.565000
2	0.10	0.610181	0.764265	0.681810	0.610095	0.673060	1.031641
3	0.14	0.565686	0.683832	0.622172	0.565780	0.590421	0.854372
4	0.19	0.542421	0.646387	0.592035	0.542452	0.558735	0.774317
5	0.24	0.528242	0.634293	0.573851	0.528184	0.539088	0.747398
6	0.29	0.516634	0.631059	0.559818	0.516774	0.526845	0.750572
GSL-div		0.55084	0.67256	0.60407	0.55089	0.56908	0.83472
MC s.d.		0.00830	0.01378	0.01401	0.00800	0.00648	0.00000
Normalized Returns							
word length	weights	DGP	M1	M2	M3	M4	M5
1	0.05	0.934610	1.042379	0.971981	0.934126	0.980298	1.544846
2	0.10	0.921513	1.011695	0.951396	0.921090	0.966484	1.294409
3	0.14	0.916101	0.996016	0.944652	0.915775	0.959991	1.178729
4	0.19	0.908013	0.969535	0.932854	0.907876	0.954993	1.173795
5	0.24	0.879334	0.925126	0.902375	0.879560	0.939478	1.262933
6	0.29	0.810997	0.887264	0.849656	0.812324	0.903308	1.267835
GSL-div		0.87717	0.94672	0.90714	0.87747	0.93955	1.25175
MC s.d.		0.00996	0.02703	0.00917	0.00929	0.01023	0.00667

As a first observation, one can notice that the *GSL-div* distinguishes clearly among the majority of proposed models, both for what concerns the dynamics of prices and that of returns, delivering consistent results. In particular, the DGP is correctly identified as the closest model to the real world data; further, the low Monte Carlo standard deviation indicate the distance from the data to be statistically different

¹⁷ Therefore, $b = 5$ and $L = 6$. In Lamperti (2017) the *GSL-div* is proved to be robust to changes in these two parameters. Their choice is problem-specific and depends on the degree of precision requested by the modeller, the number of competing models or configurations, the time scale of interest and the available computational power. In practice, as a rule of thumb, one can select the combination of b and L starting by $b = L = 2$ and then increasing their value one at the time stopping when any further increase in one and the other parameter do no change the order of models provided by the *GSL-div*. That will be the minimum order or complexity which is needed to discriminate robustly among a set of models. In my experience, $b = 5$ and $L = 6$ are generally sufficient to the scope. Robustness analysis are reported also in the Appendix.

¹⁸ All the experiments in this paper have been performed using a 2.8 GHz Intel Core i7 processor, with dual independent processor cores and 8 GB of RAM. The average time to compute the *GSL-div* between data and the output of the model for a single parameter vector and using an ensemble of 1000 independent runs of size 1000 periods is 22mins and 34secs, which corresponds to an evaluation time of about 1.3 secs per run.

from the one reported for any other models but one (M3).¹⁹ This result is extremely reasonable and helps prove the good performance of the *GSL-div*. In particular, M3 exhibits the same parametric structure of the DGP and differs only in the initial number of type 1 traders. However, being $n_1 = 0.1$ a disequilibrium starting point, $\beta = 4$ a reasonably high intensity of choice and having M3 the same stable steady state as DGP (see Brock and Hommes, 1998, for details on steady states of the model with two types), the share of type 1 traders suddenly converges to the equilibrium value and then fluctuates around it because of the effects introduced by the random noise. From such a point on, M3 and DGP can be seen as completely identical models and, therefore, I believe it is a good signal that the *GSL-div* cannot distinguish significantly between the two. Finally, as a consistency check, it is worth remarking that M5, which exhibits a dynamics of prices clearly at odds with the real data, is successfully found as the most distant model.

5.2 Validation against real data

Once the *GSL-div* has been proven to successfully discriminate amongst different models, it can be used to validate them against actual data. To this purpose I draw on the results obtained in Recchioni et al (2015), where a particular version of the Brock and Hommes model described in section 4 has been estimated using real stock market data. Specifically, this simplified model does not include dividends and has been constrained to incorporate two particular types of agents, namely a pure fundamentalist ($g_1 = b_1 = 0$) and an unbiased trend follower ($g_2 > 0$ and $b_2 = 0$). Further, it has been calibrated using daily information covering the period ranging from February 25, 2011 to December 16, 2011, for a total of 200 observations. In what follows, I refer exactly to this interval and target two major stock market indexes, namely the EuroSTOXX 50 (which is an index composed by main corporations in the Euro area) and the CSI 300 (which is one of the most important Asian indexes, designed to replicate the performance of 300 stocks traded in the Shanghai and Shenzhen stock exchanges). Table 4 collects the values of parameters set or calibrated in Recchioni et al (2015). In the exercises that follows some of these parameters will be used as benchmark. Figure 3 shows the behaviour produced by the model calibrated on the EuroSTOXX

Table 4: Parameters of the calibrated model for EuroSTOXX 50 and CSI 300.

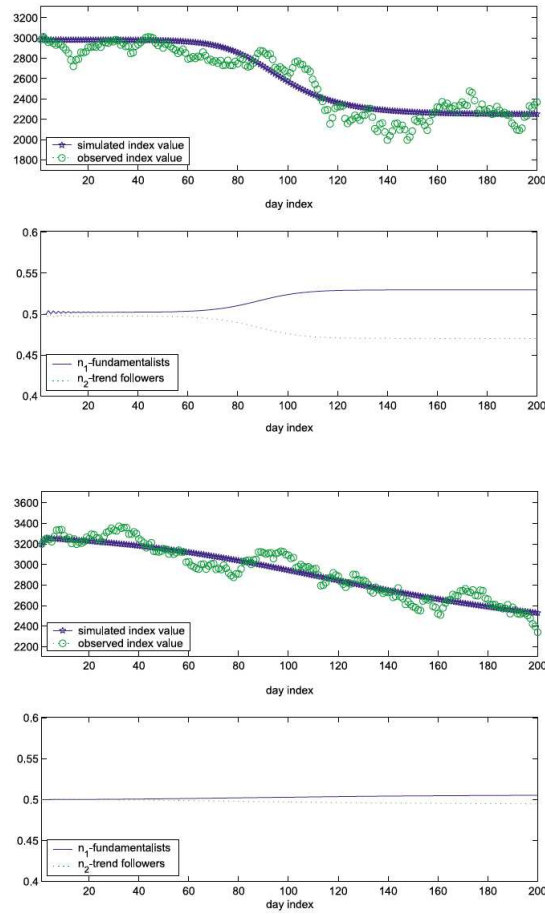
	β	g_1	g_2	α	p^*	C	ω	σ
EuroSTOXX 50	0.642	0	2.0	18.207	0.746	0	1	0.1
CSI 300	0.078	0	1.996	13.999	0.682	0	1	0.1

(top two panels) and on the CSI (bottom two panels), together with the real data and the share of

¹⁹ The Monte Carlo standard deviation is simply the standard deviation of the *GSL-div* computed in (2) across the ensemble. Being the latter composed by 500 independent runs of the same model, it is possible to interpret it as a Monte Carlo exercise on the seed of the random number generator.

traders following one or the other strategy along the simulation. It is evident that, despite being a simple model, it provides a reasonably good performance in tracking the real indexes. The additional value brought by the use of ABMs should be the possibility to analyse the micro-determinants of these macro-behaviours and, in this case for example, to single out differences in the attitude of traders operating in different geographical areas. In what follows I perform two different exercises by means of the *GSL-div*.

Fig. 3: Actual prices and model simulation in the calibration interval. Source: Recchioni et al (2015).



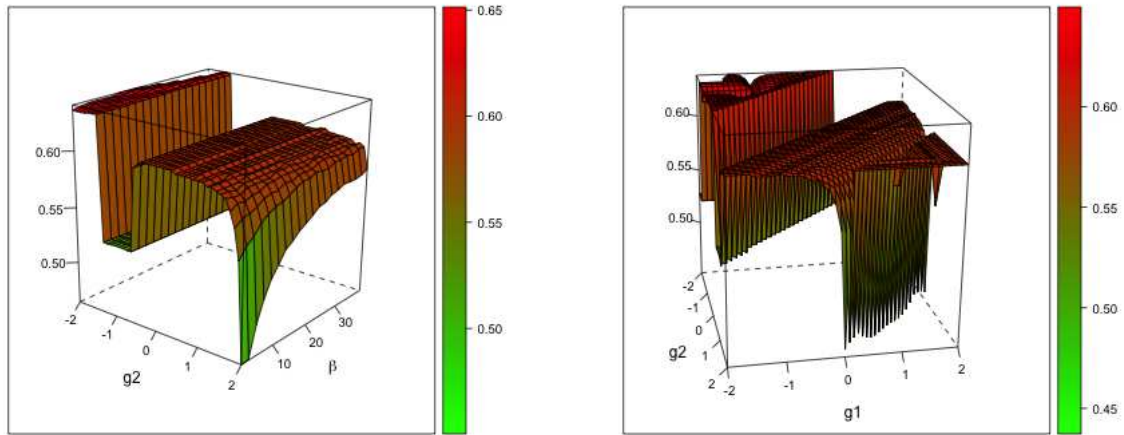
Note: This figure is composed by four panels. The first two from the top refers to the EuroSTOXX index and the corresponding model, while the bottom two to the CSI. For each index, the first (top) panel plots the behaviour of the price (observed and simulated) in the time interval used for calibration, while the second (bottom) the share of the two types of traders in the same time interval.

First, I explore the similarity between real data and model configurations obtained maintaining the same structure (fundamentalists vs. trend followers) and parameter values as in table 4, but I let vary the trend-following component, g_2 , and the switching parameter, β . Second, I allow the model to account for richer combinations of traders' attitudes (e.g. two trend followers, one trend-follower and one contrarian, one contrarian and one fundamentalist) and I check whether some of them are able to provide a better

account for the dynamics observed in the data. Obviously enough, if the calibrated model turns out to be the closest to the data both for EuroSTOXX and CSI indexes and, keeping fixed its structure, no combinations of parameters produce significantly better results, I will conclude in favor of the empirical validity of such a model, at least for what concerns its ability to track the historical behaviour of the targeted system.

To start with, I need to build a convenient subspace of parameters whose points will be used to construct model configurations that, in turns, will serve to produce simulated output to feed the *GSL-div*. In the case of the first exercise, I consider all possible combinations of parameters' values found in a two-dimensional grid obtained using the following intervals $1 \leq \beta \leq 40$ and $-2 \leq g_2 \leq 2$, where the former is discretized in 21 equally spaced segments, while the latter in 41. For the second exercise, instead, the same interval used for g_2 is allowed to characterize the trend component of the first trader type, g_1 .²⁰ This procedure leads to the inclusion of 861 model configurations for the first exercise and 1681 for the second. All these models are run starting from the same initial conditions and for $T = 200$ periods. The *GSL-div* is used to measure their distance with respect to the real data (both EuroSTOXX and CSI). Figure 4 shows the results obtained for the first (a) and second (b) exercise using EuroSTOXX data, while figure 5 does the same for the CSI index.

Fig. 4: *GSL-div* between model and EuroSTOXX50 for different portions of the parameters space.



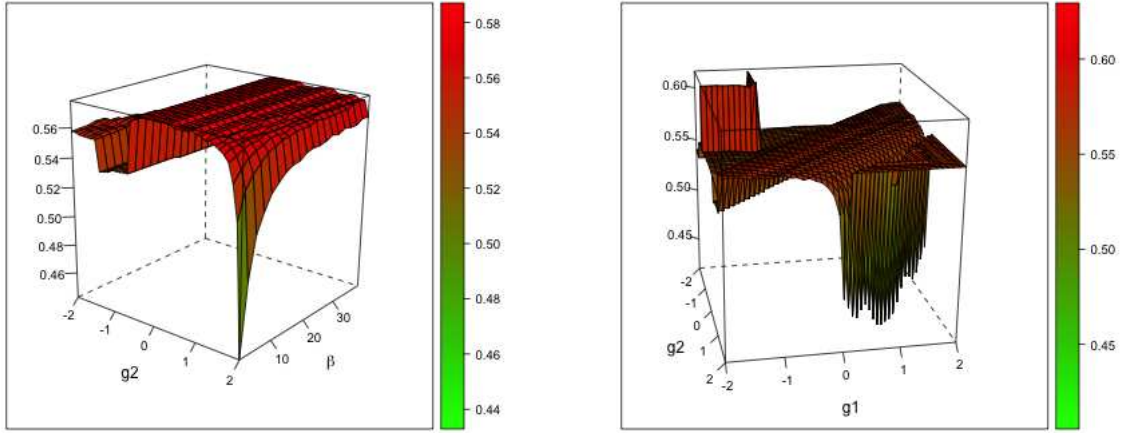
(a) Space generated by β and g_2 ($g_1 = 0$ as in Table 4). (b) Space generated by g_1 and g_2 ($\beta = 0.642$ as in Table 4).

Different remarks apply. To begin with, subfigures 4a and 5a contains two insights. On one hand, they show that in large parts of the explored parameters' subspace the dynamics of the Brock and Hommes model with a trend follower and a fundamentalist traders' types are guided by g_2 , that is, by the strength at which the trend followers extrapolate information on the basis of past observations. This finding is

²⁰ These intervals are in line with the literature and include many values attached to the relevant parameters in, for example, Brock and Hommes (1998) and Boswijk et al (2007).

partially in line with Teräsvirta (1994), Boswijk et al (2007) and Recchioni et al (2015), where it is shown that the intensity of choice has little significance in switching models vis--vis trend attitudes. However, a notable difference emerges: β seems not to affect the distance between models and data for the majority of values, but in those cases where it is much higher than data would admit, having the correct specification of the trend-following attitude does not suffice to obtain aggregate (price) dynamics consistent with the empirical data. Moreover, taking the best parameters' values as reference, the Brock and Hommes model is found equally sensitive to changes in the switching parameter and to g_2 , as confirmed by the steepness of the *GSL-div* surface represented in subfigures 4a and 5a. I also notice that such a sensitivity appears stronger when the model is calibrated to CSI data. On the basis of these results, it seems that the model exhibit tipping points in the space of parameters, i.e. areas where the behaviour of the model changes suddenly and dramatically in response to small variations in conditions (see also Gualdi et al, 2015). For example, in the case of the EuroSTOXX calibration, subfigure 4a shows a not-so-small area where g_2 is around -1.5 and the model shifts immediately its behaviour in response to a small change in the strength of the trend chasing attitude, irrespectively of β ; then it keeps constant for a while and shifts again (as g_2 approaches -1). A second remark concerns the general behaviour of the model calibrated in Recchioni et al (2015). Surfaces constructed through the *GSL-div* confirm that a switching parameter below 3 coupled with a trend-following intensity really close to 2 deliver, by far, the most similar dynamics with respect to the data in both the two analysed markets. Finally, comparing the EuroSTOXX and CSI cases one can notice that the role of the switching parameter is much more important in the Asian market rather than in the European one, meaning that the latter is more compatible, in relative terms, with a larger set of values for β . This is rapidly explained by the different behaviour of the data, where it is possible to see that, in the considered period, CSI exhibits an approximately linear negative trend while the EuroSTOXX does not. Since the switching parameters controls, other things being equal, for the steepness of the downward dynamics (larger β corresponds to an *S*-shaped behaviour with increasing steep of the *S* due to the higher facility of moving to the most profitable strategy), it is natural that a linear trend is compatible with a much more restricted portion of the parameter space.

Even though the calibrated Brock and Hommes model with a trend-following and a pure fundamentalist types is decently able to replicate the dynamics observed in the real data, much more caution is suggested by subfigures 4b and 5b. They indicate that, keeping fixed other parameters, there are many combinations of traders attitudes that guarantee approximately the same behaviour with respect to the data. The finding is robust across the different stock markets considered in this paper and, in a comparative perspective, it is more evident for the EuroSTOXX case, where there are two separate areas of the parameter space delivering a good matching with the data. Such a simple problem of multiple minima in the subspace of parameters spanned by g_1 and g_2 might be particularly harmful for the empirical validity of the model, because it supports the claim that many different combinations of traders' attitudes are compatible with the same dynamics observed in the data. In addition, since the estimated *GSL-divs* for these model configurations are pretty much close one to the other, it is hardly arguable why one (e.g.

Fig. 5: $GSL-div$ between model and CSI300 for different portions of the parameters space.(a) Space generated by β and g_2 ($g_1 = 0$ as in Table 4). (b) Space generated by g_1 and g_2 ($\beta = 0.078$ as in Table 4).

involving trend followers and fundamentalists) should be preferred to another (e.g. involving two trend follower types with different extrapolation strengths) in the attempt to explain some dynamics of the price.

To better investigate this issue I enlarged the possible combinations of traders' attitudes by considering a new grid where g_1 and g_2 are allowed to take values ranging from -4 (which models a really strong trend contrarian type) and $+4$ (which models a really strong trend follower type). As a result, 861 additional model configurations are tested against the EuroSTOXX and CSI by means of the $GSL-div$. Table 5 reports the configurations yielding the lowest ten values of the $GSL-div$.

Table 5: Model configurations yielding the 10 lowest $GSL-div$ values.

rank	EuroSTOXX 50			CSI 300		
	g_1	g_2	$GSL-div$	g_1	g_2	$GSL-div$
1	1.2	0.8	0.450562	1.2	0.8	0.394443
2	1.4	0.6	0.450133	1	1	0.395813
3	1	1	0.451551	1.4	0.6	0.400043
4	0	2	0.457403	1.6	0.4	0.414630
5	1.6	0.4	0.457498	0	2	0.418374
6	1.8	0.2	0.461870	1.8	1.2	0.424121
7	2.2	-0.2	0.462571	2.2	-0.2	0.425062
8	2.4	-0.4	0.476505	2.4	-0.4	0.440579
9	2.6	-0.6	0.491549	2.6	-0.6	0.456101
10	2.8	-0.8	0.504678	2.8	-0.8	0.470627

Variegated combinations of the behavioural strategies adopted by the two trader types are found within the ten cases that are closest to the data, and the configuration assuming a pure fundamentalist and a trend follower does not provide the best result in any of the two markets analysed in this paper. What emerges as a general trait is that the sum of the trend components for the two agents has been close to 2 in order for the model to be consistent with the dynamics in the data. While the same happens considering parameters estimated in Boswijk et al (2007) and Recchioni et al (2015) (confirming again the *GSL-div* provides results consistent with other methodologies), it is now shown that what matters is not the strict presence of mean-reverting and trend follower types (as in Boswijk et al, 2007). Rather, the model has to account for a strong trend following component, which might either come from a unique type that heavily extrapolates information from past observations or the combinations of different types with milder, or even opposite, attitudes towards the trend. However, results suggest that if one of the trader types follows the trend too strongly, a compensating trend contrarian type reduces the bullish pressure created in the market. Taken as an aggregate, these conclusions are not surprising. From equation (7) and under the assumption that trader types do not have bias towards particular price levels ($b_1 = b_2 = 0$), it would be possible to express the dynamics of price as $Rx_t = n_1 g_1 x_{t-1} + (1 - n_1) g_2 x_{t-1}$. Therefore, in those cases where traders are equally divided among strategies ($n_1 = 0.5$), $g_1 + g_2 = 2$ implies that the process is very close to a random walk, which is reasonable for a variety of asset classes. A correct identification of g_1 and g_2 is then fundamental to provide intuitions on what behavioural attitudes are consistent with the dynamics of asset prices and why. Our results, in contrast to Recchioni et al (2015), suggest that the combination of two mild trend follower types consists in the representation of agents' trading attitudes that best fit both the EuroSTOXX and CSI markets. However, having many different combinations of g_1 and g_2 yielding similar results, reduces the amount of information that the model provides beyond a simple random walk and, at the very end, it dampens the validity of the model.

Summing up, this paper uses the *GSL-div* to explore and validate the Brock and Hommes asset pricing model reporting a reasonably good ability to resemble dynamics observed in actual stock markets. The main condition for such a result to emerge is the inclusion of a balanced trend following attitude, which destabilizes the asset market from the fundamentals, but not strongly enough to create large bubbles. Relevantly, such an aggregate attitude can emerge from different mixtures of individual types' trading behaviour.

6 Conclusions

Validation of simulated models is still an open issue. One way of tackling this problem is via the identification of a measure quantifying the distance between simulated and real-world data with respect to the observed dynamics. This paper presents an illustrative application on the use of the *GSL-div* developed in Lamperti (2017) to the validation of simulated models. In particular, different versions of the asset pricing model presented in Brock and Hommes (1998) are analysed. The proposed approach is found to

successfully discriminate amongst alternative, competing models with reasonable precision, both when price or return dynamics are at stake. This suggests the *GSL-div* might be a good indicator to measure similarity of behaviours and co-movements in financial data.²¹ The Brock and Hommes model is then validated against data from two major stock market indexes, namely the EuroSTOXX 50 and the CSI 300. What emerges is that when the model is constrained to include two specific types of traders (a fundamentalist and a trend follower) it seems to achieve a reasonable similarity with the real data; however, when different combinations of traders' types are allowed, it is difficult to argue what configuration should be preferred over another, since many different attitudes are compatible with the same dynamics as those observed in actual data. What is found as a general trait, is that empirical validity of the model requires to account for a strong trend following component, which might come both from a unique trend follower type that heavily extrapolates information from past observations or the combinations of different, milder, or even opposite, trend follower types. Finally, it is worth to recall that even though in this paper the *GSL-div* is used as a validation tool applied to already calibrated models, an interesting application will be the development of a calibration procedure including it in the objective function. While the present approach relies on the comparison of univariate series, extensions to multivariate settings are possible. Different routes might be taken; on one side it would be possible to treat each series (i.e. output variable) separately, identify the values of b and L that are adequate to conveniently capture similarities and finally combine the *GSL-div* for each series in a composite indicator. On the other side, one could define the dynamics of the model in a multidimensional space and use the joint distributions of patterns followed by multiple variables. The second approach would mirror exactly the one described in the present paper from a methodological perspective, but computational costs might increase if many output variables would be considered. The first approach would allow a better characterization of the different series (through different values of b and L) but introduces the additional issue of aggregation. Future research will be devoted to address these issues.

Acknowledgements The author would like to thank Mattia Guerini, Mauro Napoletano and Andrea Roventini for valuable comments and suggestions. All the shortcomings are the author's.

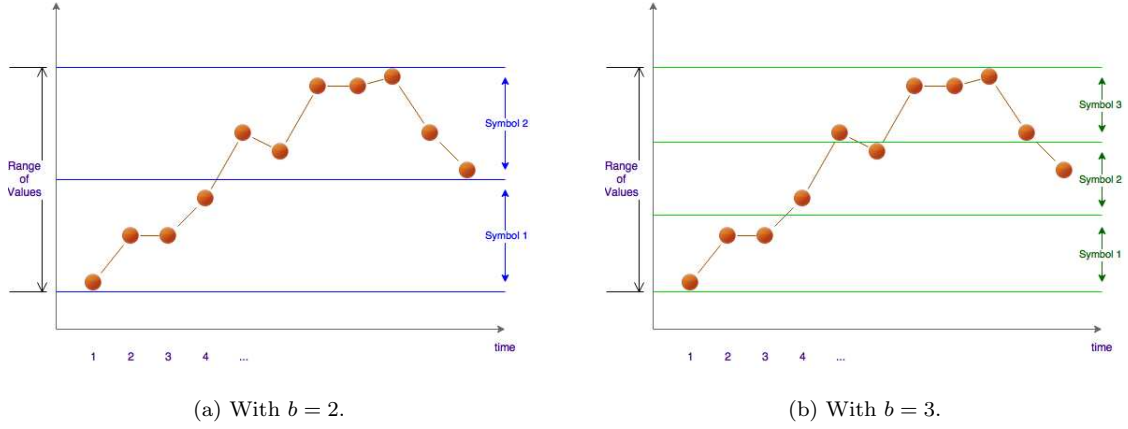
7 Appendix

7.1 Symbolization

Figure 6 provides a graphical insight into the process of series' symbolization described in section 3. Once a convenient support for the series is identified (in this figure it coincides with $[x_{min}; x_{max}]$, it is partitioned in b intervals of equal size. Then, symbols are assigned to each interval and observations are labelled using the symbol of the interval they fall in.

²¹ The exploration of the latter issue is left to future research.

Fig. 6: Symbolization



7.2 Algorithm to compute the $GSL-div$

The following algorithm, illustrated in figure 7, specifies how the $GSL-div$ as expressed in equation 3 can be computed through a simple iterative procedure.

Fig. 7: Algorithm for the $GSL-div$.

Data: load N runs of the model $(\{y_{i,t}\}_{t=1}^T$ for $i = 1, \dots, N$).

Data: load empirical data $(\{x_t\}_{t=1}^T)$.

Data: choose b, L and the aggregation weights.

Result: computation of the $GSL-div$.

begin

symbolize empirical data;

observe words of length l in the real data using a rolling window of the same length;

compute the frequency distribution of words $(f_{y,l})$;

for l in $1 : L$ **do**

for i in $1 : N$ **do**

symbolize simulation run;

observe words of length l in the simulation data using rolling windows of the same length;

compute the frequency distribution of words $(f_{x,l})$;

obtain the mean distribution $m_l = \frac{f_{x,l} + f_{y,l}}{2}$;

compute the subtracted L-divergence, $SL-div = 2H(m_l) - H(f_{y,l})$;

end

compute the average subtracted L-divergence over the N runs;

end

aggregate the average subtracted L-divergences using aggregation weights;

correct for the systematic bias.

end

7.3 Robustness

Here I present the results from different robustness exercises that test whether the discriminatory ability of the *GSL-div* is affected by small changes in the values of b (precision of the symbolization), L (maximum word's length) or by the choice of the aggregation weights, w_l . In particular, I repeated the same procedure of section 5.1 using uniform weights such that $w_l = w_k$ for all l and k and geometrically progressive weights such that $w_l/w_{l-1} = 1.2$. Further, I also changed the setting of b and L letting them vary by one unit (either in addition or subtraction) with respect to the baseline ($b = 5$ and $L = 6$). Table 6 reports the values of the *GSL-div* between the DGP and each model in the various cases. Results largely confirm the robustness of the proposed approach.

Table 6: Robustness exercises for the rankings reported in table 3. The *GSL-div* under different settings is reported. The position in the ranking is in parenthesis and the baseline ranking, resulting from parameters used in the main text, in bold.

Prices						
Setting	DGP	M1	M2	M3	M4	M5
geometrically progressive weights	0.56442 (1)	0.69534 (5)	0.62227 (4)	0.56446 (2)	0.58349 (3)	0.88508 (6)
uniform weights	0.58380 (1)	0.72669 (5)	0.64812 (4)	0.58384 (2)	0.60500 (3)	0.95388 (6)
b=4; L=5	0.54011 (2)	0.66196 (5)	0.58919 (4)	0.53989 (1)	0.54846 (3)	0.82155 (6)
b=4; L=6	0.53283 (2)	0.65311 (5)	0.58097 (4)	0.53279 (1)	0.54081 (3)	0.81762 (6)
b=4; L=7	0.52891 (2)	0.64901 (5)	0.57663 (4)	0.52702 (1)	0.53592 (3)	0.81197 (6)
b=5; L=5	0.55991 (1)	0.67805 (5)	0.60802 (4)	0.56000 (2)	0.57498 (3)	0.84102 (6)
b=5; L=6	0.55084 (1)	0.67256 (5)	0.60407 (4)	0.55089 (2)	0.56908 (3)	0.83472 (6)
b=5; L=7	0.54712 (1)	0.66670 (5)	0.59890 (4)	0.54734 (2)	0.56207 (3)	0.83101 (6)
b=6; L=5	0.57282 (1)	0.70164 (5)	0.62611 (4)	0.57348 (2)	0.58922 (3)	0.86499 (6)
b=6; L=6	0.56881 (1)	0.66588 (5)	0.61793 (4)	0.56923 (2)	0.58293 (3)	0.86108 (6)
b=6; L=7	0.56307 (1)	0.65918 (5)	0.60948 (4)	0.56194 (2)	0.57677 (3)	0.85429 (6)
Normalized Returns						
Setting	DGP	M1	M2	M3	M4	M5
geometrically progressive weights	0.88320 (1)	0.95592 (5)	0.91353 (3)	0.88341 (2)	0.94346 (4)	1.26863 (6)
uniform weights	0.89509 (1)	0.97200 (5)	0.92549 (3)	0.89513 (2)	0.95076 (4)	1.28709 (6)
b=4; L=5	0.85212 (1)	0.93072 (5)	0.88696 (3)	0.85209 (2)	0.91821 (4)	1.23795 (6)
b=4; L=6	0.84997 (1)	0.92855 (5)	0.88082 (3)	0.84508 (2)	0.91022 (4)	1.22916 (6)
b=4; L=7	0.84314 (1)	0.92076 (5)	0.87218 (3)	0.84326 (2)	0.90610 (4)	1.22349 (6)
b=5; L=5	0.88121 (1)	0.95282 (5)	0.91522 (3)	0.88130 (2)	0.94233 (4)	1.26019 (6)
b=5; L=6	0.87717 (1)	0.94672 (5)	0.90714 (3)	0.87747 (2)	0.93955 (4)	1.25175 (6)
b=5; L=7	0.87087 (1)	0.93911 (5)	0.89570 (3)	0.87141 (2)	0.93481 (4)	1.24489 (6)
b=6; L=5	0.90186 (1)	0.97870 (5)	0.92933 (3)	0.90298 (2)	0.95934 (4)	1.28170 (6)
b=6; L=6	0.89221 (1)	0.97218 (5)	0.92037 (3)	0.89308 (2)	0.95210 (4)	1.28002 (6)
b=6; L=7	0.88875 (1)	0.97031 (5)	0.91998 (3)	0.89269 (2)	0.94779 (4)	1.27809 (6)

References

- Alfarano S, Lux T, Wagner F (2005) Estimation of agent-based models: the case of an asymmetric herding model. *Computational Economics* 26(1):19–49
- Alfarano S, Lux T, Wagner F (2006) Estimation of a simple agent-based model of financial markets: An application to australian stock and foreign exchange data. *Physica A: Statistical Mechanics and its Applications* 370(1):38–42
- Amilon H (2008) Estimation of an adaptive stock market model with heterogeneous agents. *Journal of Empirical Finance* 15(2):342 – 362, DOI <http://dx.doi.org/10.1016/j.jempfin.2006.06.007>, URL <http://www.sciencedirect.com/science/article/pii/S0927539807000606>
- Banerjee AV (1992) A simple model of herd behavior. *The Quarterly Journal of Economics* 107(3):797–817
- Barde S (2016a) Direct comparison of agent-based models of herding in financial markets. *Journal of Economic Dynamics and Control* 73:329 – 353, DOI <http://dx.doi.org/10.1016/j.jedc.2016.10.005>, URL <http://www.sciencedirect.com/science/article/pii/S0165188916301622>
- Barde S (2016b) A practical, accurate, information criterion for Nth order markov processes. *Computational Economics* pp 1–44
- Bianchi C, Cirillo P, Gallegati M, Vagliasindi PA (2007) Validating and calibrating agent-based models: a case study. *Computational Economics* 30(3):245–264
- Bianchi C, Cirillo P, Gallegati M, Vagliasindi PA (2008) Validation in agent-based models: An investigation on the cats model. *Journal of Economic Behavior & Organization* 67(3):947–964
- Boswijk HP, Hommes CH, Manzan S (2007) Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control* 31(6):1938 – 1970, DOI <http://dx.doi.org/10.1016/j.jedc.2007.01.001>, URL <http://www.sciencedirect.com/science/article/pii/S0165188907000061>, tenth Workshop on Economic Heterogeneous Interacting Agents WEHIA 2005
- Brock WA, Hommes CH (1997) A rational route to randomness. *Econometrica* 65(5):1059–1095, URL <http://www.jstor.org/stable/2171879>
- Brock WA, Hommes CH (1998) Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control* 22(89):1235 – 1274, DOI [http://dx.doi.org/10.1016/S0165-1889\(98\)00011-6](http://dx.doi.org/10.1016/S0165-1889(98)00011-6), URL <http://www.sciencedirect.com/science/article/pii/S0165188998000116>
- Canova F, Sala L (2009) Back to square one: identification issues in dsge models. *Journal of Monetary Economics* 56(4):431–449
- Chiarella C, Iori G, Perelló J (2009) The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control* 33(3):525–537
- Dawid H, Fagiolo G (2008) Agent-based models for economic policy design: Introduction to the special issue. *Journal of Economic Behavior & Organization* 67(2):351–354

- Dosi G, Fagiolo G, Roventini A (2010) Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics and Control* 34(9):1748–1767
- Dosi G, Fagiolo G, Napoletano M, Roventini A (2013) Income distribution, credit and fiscal policies in an agent-based Keynesian model. *Journal of Economic Dynamics and Control* 37(8):1598–1625
- Dosi G, Fagiolo G, Napoletano M, Roventini A, Treibich T (2015) Fiscal and monetary policies in complex evolving economies. *Journal of Economic Dynamics and Control* 52(0):166 – 189, DOI <http://dx.doi.org/10.1016/j.jedc.2014.11.014>, URL <http://www.sciencedirect.com/science/article/pii/S016518891400311X>
- Fabretti A (2012) On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination* 8(2):277–293, DOI 10.1007/s11403-012-0096-3, URL <http://dx.doi.org/10.1007/s11403-012-0096-3>
- Fagiolo G, Guerini M, Lamperti F, Moneta A, Roventini A (2017) Validation of agent-based models in economics and finance. LEM Papers Series 2017/23, Laboratory of Economics and Management (LEM), Sant’Anna School of Advanced Studies, Pisa, Italy
- Franke R (2016) Competitive moment matching of a new-Keynesian and an old-Keynesian model. *Journal of Economic Interaction and Coordination* pp 1–39, DOI 10.1007/s11403-016-0181-0, URL <http://dx.doi.org/10.1007/s11403-016-0181-0>
- Franke R, Westerhoff F (2011) Estimation of a structural stochastic volatility model of asset pricing. *Computational Economics* 38(1):53–83, DOI 10.1007/s10614-010-9238-7, URL <http://dx.doi.org/10.1007/s10614-010-9238-7>
- Franke R, Westerhoff F (2012) Structural stochastic volatility in asset pricing dynamics: Estimation and model contest. *Journal of Economic Dynamics and Control* 36(8):1193 – 1211, DOI <http://dx.doi.org/10.1016/j.jedc.2011.10.004>, URL <http://www.sciencedirect.com/science/article/pii/S0165188912000802>, quantifying and Understanding Dysfunctions in Financial Markets
- Franke R, Westerhoff F (2016) Why a simple herding model may generate the stylized facts of daily returns: explanation and estimation. *Journal of Economic Interaction and Coordination* 11(1):1–34, DOI 10.1007/s11403-014-0140-6, URL <http://dx.doi.org/10.1007/s11403-014-0140-6>
- Gallegati M, Richiardi M (2009) Agent based modelling in economics and complexity, vol In: Meyer, R.A. (Ed.), *Encyclopedia of Complexity and System Science*. Springer, New York, USA
- Giannone D, Reichlin L, Sala L (2006) Vars, common factors and the empirical validation of equilibrium business cycle models. *Journal of Econometrics* 132(1):257–279
- Gilli M, Winker P (2003) A global optimization heuristic for estimating agent based models. *Computational Statistics & Data Analysis* 42(3):299–312
- Gourieux C, Monfort A (1997) *Simulation based econometric methods*. Oxford University Press
- Grazzini J (2012) Analysis of the emergent properties: Stationarity and ergodicity. *Journal of Artificial Societies and Social Simulation* 15(2):7, URL <http://jasss.soc.surrey.ac.uk/15/2/7.html>

- Grazzini J, Richiardi M (2015) Estimation of ergodic agent-based models by simulated minimum distance. *Journal of Economic Dynamics and Control* 51:148 – 165, DOI <http://dx.doi.org/10.1016/j.jedc.2014.10.006>, URL <http://www.sciencedirect.com/science/article/pii/S0165188914002814>
- Grazzini J, Richiardi MG, Tsionas M (2017) Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control* 77:26 – 47, DOI <http://dx.doi.org/10.1016/j.jedc.2017.01.014>, URL <http://www.sciencedirect.com/science/article/pii/S0165188917300222>
- Gualdi S, Tarzia M, Zamponi F, Bouchaud JP (2015) Tipping points in macroeconomic agent-based models. *Journal of Economic Dynamics and Control* 50:29 – 61, DOI <http://dx.doi.org/10.1016/j.jedc.2014.08.003>, URL <http://www.sciencedirect.com/science/article/pii/S0165188914001924>, crises and Complexity Complexity Research Initiative for Systemic Instabilities (CRISIS) Workshop 2013
- Guerini M, Moneta A (2016) A method for agent-based model validation. LEM Papers Series 2016/16, Laboratory of Economics and Management (LEM), Sant’Anna School of Advanced Studies, Pisa, Italy
- Jacob Leal S, Napoletano M, Roventini A, Fagiolo G (2015) Rock around the clock: An agent-based model of low- and high-frequency trading. *Journal of Evolutionary Economics* 26(1):49–76, DOI 10.1007/s00191-015-0418-4, URL <http://dx.doi.org/10.1007/s00191-015-0418-4>
- Kukacka J, Barunik J (2016) Simulated ML Estimation of Financial Agent-Based Models. Working Papers IES 2016/07, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies, URL https://ideas.repec.org/p/fau/wpaper/wp2016_07.html
- Kullback S, Leibler RA (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22:49–86
- Lamperti F (2017) An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics* pp –, DOI <http://dx.doi.org/10.1016/j.ecosta.2017.01.006>, URL <http://www.sciencedirect.com/science/article/pii/S2452306217300084>
- Lamperti F, Roventini A, Sani A (2017) Agent based model calibration using machine learning surrogates. LEM Papers Series forthcoming, Laboratory of Economics and Management (LEM), Sant’Anna School of Advanced Studies, Pisa, Italy
- LeBaron B, Winker P (2008) Introduction to the Special Issue on Agent-Based Models for Economic Policy Advice. *Journal of Economics and Statistics* 228(2+3):141–148
- Leombruni R, Richiardi M, Saam NJ, Sonnessa M (2006) A common protocol for agent-based social simulation. *Journal of Artificial Societies and Social Simulation* 9(1):15, URL <http://jasss.soc.surrey.ac.uk/9/1/15.html>
- Lin J (1991) Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory* 37:145–151
- Lux T, Marchesi M (2000) Volatility clustering in financial markets: a microsimulation of interacting agents. *International journal of theoretical and applied finance* 3(04):675–702
- Lux T, Zwinkels RC (2017) Empirical validation of agent-based models
- Manson S (2002) Validation and verification of multi-agent systems, in *Complexity and Ecosystem Management*, edited by m.a. janssen edn. Cheltenham: Edward Elgar

- Marks R (2007) Validating simulation models: A general framework and four applied examples. *Computational Economics* 30(3):265–290, DOI 10.1007/s10614-007-9101-7, URL <http://dx.doi.org/10.1007/s10614-007-9101-7>
- Marks R (2013) Validation and model selection: Three similarity measures compared. *Complexity Economics* 2(1):41–61
- Paccagnini A (2009) Model validation in the dsge approach: A survey, working Paper, mimeo
- Pellizzari P, Forno AD (2006) A comparison of different trading protocols in an agent-based market. *Journal of Economic Interaction and Coordination* 2(1):27–43, DOI 10.1007/s11403-006-0016-5, URL <http://dx.doi.org/10.1007/s11403-006-0016-5>
- Pyka A, Fagiolo G (2007) Agent-based modelling: a methodology for neo-schumpeterian economics. In: Pyka, Horst (eds) *Elgar companion to neo-schumpeterian economics*, Edward Elgar Publishing
- Recchioni MC, Tedeschi G, Gallegati M (2015) A calibration procedure for analyzing stock price dynamics in an agent-based framework. *Journal of Economic Dynamics and Control* 60:1 – 25, DOI <http://dx.doi.org/10.1016/j.jedc.2015.08.003>, URL <http://www.sciencedirect.com/science/article/pii/S0165188915001517>
- Rosen R (1985) *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations*. Oxford: Pergamon
- Shannon CE (1948) A mathematical theory of communication. *Bell system technical journal* 27
- Teräsvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89(425):208–218, DOI 10.1080/01621459.1994.10476462, URL <http://dx.doi.org/10.1080/01621459.1994.10476462>, <http://dx.doi.org/10.1080/01621459.1994.10476462>
- Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10(2):8, URL <http://jasss.soc.surrey.ac.uk/10/2/8.html>
- Winker P, Gilli M, Jeleskovic V (2007) An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination* 2(2):125–145, DOI 10.1007/s11403-007-0020-4, URL <http://dx.doi.org/10.1007/s11403-007-0020-4>