Comput Econ (2014) 44:507–536
DOI 10.1007/s10614-013-9406-7

Efficient Sampling and Meta-Modeling for Computational Economic Models

Isabelle Salle · Murat Yıldızoğlu

Accepted: 7 November 2013 / Published online: 22 November 2013

Abstract Extensive exploration of simulation models comes at a high computational cost, all the more when the model involves a lot of parameters. Economists usually rely on random explorations, such as Monte Carlo simulations, and basic econometric modeling to approximate the properties of computational models. This paper aims to provide guidelines for the use of a much more efficient method that combines a parsimonious sampling of the parameter space using a specific *design of experiments (DoE)*, with a well-suited metamodeling method first developed in geostatistics: *kriging*. We illustrate these guidelines by following them in the analysis of two simple and well known economic models: Nelson and Winter's industrial dynamics model, and Cournot oligopoly with learning firms. In each case, we show that our DoE experiments can catch the main effects of the parameters on the models' dynamics with a much lower number of simulations than the Monte-Carlo sampling (e.g. 85 simulations instead of 2,000 in the first case). In the analysis of the second model, we also introduce supplementary numerical tools that may be combined with this method, for characterizing configurations complying with a specific criterion (social optimal, replication of stylized facts, etc.). Our appendix gives an example of the R-project code that can be used to apply this method on other models, in order to encourage other researchers to quickly test this approach on their models.

I. Salle
CeNDEF, Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, Building J/K, 1018 XE Amsterdam, The Netherlands
e-mail: I.L.Salle@uva.nl

I. Salle
Tinbergen Institute,
Amsterdam, The Netherlands

M. Yıldızoğlu (✉)
GREThA (UMR CNRS 5113), Université de Bordeaux,
Avenue Léon Duguit , 33608 Pessac, France
e-mail: yildi@u-bordeaux4.fr

Keywords Computational economics · Exploration of agent-based models · Design of experiments · Meta-modeling

1 Introduction

Agent-based computational models (**ABM** hereafter) have become now a widely used tool in economic research. They have been notably applied to the investigation of markets, social dynamics, technological competition and learning dynamics, industrial dynamics and firms strategies, exchange or stock markets, *see* the surveys collected in Tesfatsion and Judd (2006) and Miller and Page (2007). Those models are highly non-linear, and generally do not allow for the derivation of analytical solutions. Intensive sensitivity analyses are required to investigate the behavior of those models, in order to understand their properties, to discriminate between key parameters and the others, to select optimal configurations regarding a predefined criterion, etc. However, as soon as the model involves many parameters, with wide variation domains, the computational cost of those analyses dramatically increases, and may quickly become prohibitive. For example, with ten parameters, each of them having five potential levels, we need almost 10 millions of simulation runs to cover all configurations, and even more if the model is not deterministic, and involves replications with different random number generator seeds. A very common method that is used in economics for this sensitivity analysis is the Monte-Carlo approach. This method consists in launching a high number of simulation runs (typically several thousands), with randomly drawn parameter values and initial conditions, to obtain a representative sample of the mapping between the parameter values and the model's behavior. Standard variance analysis and econometric methods are then applied to determine how parameter values influence the simulation results.

In this paper, we argue that the use of a specific *design of experiments* (hereafter, DoE), and an estimation model adapted to such a DoE for analysing the results is a much more efficient way of exploring computational models. By efficient, we mean that much less simulations than under a Monte Carlo procedure are needed to obtain a picture of the relationships between parameter values and model's outcomes, and this picture is more precise than the one coming out from standard econometric models. Basically, DoE aims to minimize the sample size of parameter configurations under the constraint of their representativeness. The DoE that we introduce here is called Nearly Orthogonal Latin Hypercube (**NOLH**) sampling. Based on the data collected from the model using this DoE, an appropriate meta-model is estimated, in order to approximate the true model, which connects the parameters to the variables of interest. The meta-model that we introduce is inspired by geostatistics and it is called **kriging**. The combination of this DoE and this meta-model is very interesting because we obtain a faithful vision of the model's behavior, and of its sensitivity to different parameter values using a considerably smaller set of simulation runs, as we will show below, in the example applications.

Admittedly, other alternatives to Monte Carlo sampling with a fixed number of simulations have been contemplated in the literature to address execution time cost

of computer models. Sequential Monte Carlo procedure is one of the main ones:[1] this method chooses not to set the number of simulations beforehand, and runs the simulations until the results fulfill a given criterion. To that respect, Silva et al. (2009) provide a simple rule of thumb to stop as early as possible the simulations, and obtain results as reliable as those from traditional Monte Carlo procedures with preset number of simulations.[2] Sequential Monte Carlo sampling can with no doubt address the issue of statistical tests such as Student tests (of the null hypothesis that the value of the variable under interest is higher, equal or less that a given threshold) in computationally costly models, as rejection of the null hypothesis can be stated at each simulation step. However, it does not appear suited for conducting sensitivity analyses in multi-dimensional models. In that case, an analysis of variance (ANOVA) is typically performed to establish whether the factors significantly influence the variable under interest. This requires a *full sample* of runs in order to compute a F test (of the null hypothesis that the variable has the same value across all the possible values of the factor) for each factor, and possibly each two- or higher-order interaction.

Another alternative to traditional Monte Carlo procedures to reduce the number of simulations is importance sampling [see (Booth and Butler 1999)]. A main drawback of traditional Monte Carlo sampling is typically that many sample points are redundant by falling into regions of the parameter space containing very little informative value, because sampling is performed from a uniform distribution (i.e. each configuration of factors has an equal probability to be sampled). Importance sampling allows to address this drawback. This method performs sampling according to a specific distribution, which is selected to sample more frequently parameter values which are expected to have more impact on the variable under interest, while not excluding areas of the parameters' variation domain. Provided that some prior information is available regarding where the regions of high influence lie over the factors' variation domain, this sampling method allows to eliminate a large amount of the simulations. However, importance sampling becomes intractable when dealing with high-dimensional variation domains, while, as just mentioned, our proposed method can address sensitivity analysis of models containing up to 29 factors, and does not require any prior information on factor's influence.

We consider that the method we propose—the use of kriging meta-modeling based on an appropriate DoE—is better adapted to the analysis of computer experiments with ABMs. Indeed, this method both drastically reduces the number of simulations to run, and makes possible a high-dimensional sensitivity analysis: non-linear relations and interactions can be significantly identified with only 17 simulations up to 7 factors, 33 simulations up to 11 factors, 65 simulations up to 16 factors, 129 simulations up to

[1] See, notably, Besag and Clifford (1991). Applications in economics of this method mostly concern Bayesian estimation of models, see, for instance, Herbst and Schorfheide (2013).

[2] More precisely, consider that the null hypothesis is rejected if the observed values of the variable under interest are higher than a given value, at a frequency higher than a given confidence threshold $\alpha \leq 1$. Running simulations until one obtains $n \cdot \alpha \leq n$ observations leading to the rejection of the null hypothesis is shown to have the same power as running n simulations, and assessing only afterwards whether the null hypothesis has to be rejected or not. The number of simulations can hence be reduced by a factor up to $\frac{1}{\alpha}$. A small amount of power can further be traded with a larger decrease in the number of simulations, and Silva et al. (2009) provide related estimations of the power loss.

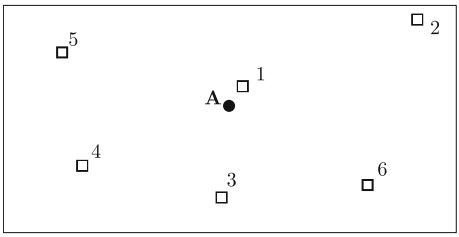


Fig. 1 Design of experiments and kriging meta-modeling: an illustrative introduction

22 factors and 257 up to 29 factors. Furthermore, this method provides the additional advantage of estimating the response surface over the whole variation domain of the factors. This estimation is of special interest when the purpose of the modeling exercise is to identify a factor configuration which minimizes or maximizes the response value. In this paper, we illustrate such a use of the meta-model within a Cournot oligopoly model.

The intuition behind our method, combining NOLH sampling and kriging meta-modeling, may be conveyed using an illustrative simple example. Suppose that a mining company aims to evaluate gold resources on a field.[3] A prospection of the entire area, or even a prospection of a high number of locations over the field would be with no doubt extremely long and costly. The company can only carry out a limited number of peripheral samples. The actual ore resources over the entire field are then estimated from the values measured at those sample points. Two issues arise here, one related to the sample selection, and one concerning the estimation procedure. First, how many sample points should the company evaluate, and where should the company arrange those points ? Second, what estimation procedure should the company use to estimate gold resources over the whole field from the limited number of sample data ? These two questions are related to each other. Indeed, the more accurate the estimation model, the less sample points needed to obtain a reliable picture of the gold coverage. The larger the sample data, the more precise the estimation of the resources. As the evaluation of resources is costly, it is in the company's interest to minimize the sample size, while maximizing the chances of extracting useful information from the collected data and, hence, obtaining an accurate estimation over the entire field. The company is clearly facing a trade-off between the sample size and the accuracy of the resulting estimation.

To illustrate that trade-off, and how DoE and kriging meta-models provide an interesting way to handle it, suppose that the rectangle in Fig. 1 represents the field, and points 1–6 stand for the six sample locations where gold resources have been evaluated.

First, it sounds intuitive to scatter the samples all over the field in order to obtain a rough idea of the gold resources on all areas of the field, rather than selecting sample points in a purely random manner. A DoE does allow to scatter a minimum number of

[3] Kriging models have been named after Danie G. Krige, a South African mining engineer who developed those models to improve ore evaluation techniques at the Witwatersrand reef complex in South Africa, pioneering the field of geostatistics, see Krige (1951). As for the statistical theory of DoE, it was developed in agriculture in the 1920's, for real, non-simulated experiments, see Fisher (1935).

sample points over a given space, so that the representativeness of the obtained sample regarding the whole space is ensured. Such a sampling is said to be *parsimonious* because it minimizes the sample size.

Second, in order to estimate the gold resources, for instance at point A, given the data collected at points 1–6, it sounds reasonable to adopt an estimation procedure which gives more importance to sample 1 than to sample 2, as sample 2 is further from point A than sample 1. Interpolation through kriging meta-models is based on such a principle: values at non-sampled locations are interpolated by setting the relative weight of each sample data according to the distance of the sample from the point to be estimated.

The same principle exactly applies to exploration of computer simulations. The field to be prospected is the parameter space, which has as many dimensions as parameters. The DoE allows to choose only a small subset of parameter values to be run and, based on the model's outcomes evaluated for those values, a kriging meta-model interpolates the model's behavior for unsampled parameter values. This method drastically reduces the number of simulations to run, while giving a precise picture of the impact of parameter values on the model. It is particularly appealing when performing sensitivity analyses of models which involve a high computational cost, in the sense that running one single simulation takes several minutes, or even several hours.

This method is very common in other scientific fields, such as industry, chemistry, electronic, biology, physics, computer science... [*see* for example (Goupy and Creighton 2007)], but is almost unknown in economics: to the best of our knowledge, the only applications are Oeffner (2008), Yıldızoğlu et al. (2012) and Salle et al. (2012). This paper aims at providing guidelines to apply this method to economic computational models. In order to do so, we use DoE and kriging meta-modeling to explore the properties of two standard economic models, namely the Nelson and Winter (1982) model and a Cournot oligopoly model, and contrast the results with the ones obtained through Monte Carlo simulations and basic econometric analysis. Our results can be stated as follows.

First, the kriging meta-model is able to account for the main effects of the parameters, as well as their interactions, on the concentration of the industry in the Nelson and Winter (1982) model with only 17 parameter configurations (replicated each five times), while a 2,000 simulations Monte Carlo sampling, at least, is necessary for capturing the direct effects, and the two-way interactions cannot be captured with less then 10,000 simulations. Importantly, the identified effects are in line with the ones discussed in Nelson and Winter (1982). A similar exercise within the Cournot oligopoly model with learning firms shows that running only 33 experiments (replicated each 20 times) is enough to catch the determinants of the convergence towards the Cournot equilibrium, while, again, at least 3,800 Monte-Carlo experiments are necessary to capture the direct effects, and 4,600 simulations for capturing all effects. These results clearly highlight the computational gain associated with the use of the NOLH DoE and kriging meta-modeling, and show that no significant informational loss results from the reduction of the number of simulations. Second, we show how a kriging meta-model predicts the behavior of the model as a function of the critical parameters in a much more accurate manner than a standard OLS regression model does. We further discuss a method to identity the parameter configuration which minimizes the distance to the Cournot equilibrium through the kriging meta-model, and

provide the corresponding R-project (R Development Core Team 2013) codes. This additional point turns out to be useful in many economic simulation analyses, in which optimizing a specific criterion is the goal of the modeling exercise.

The rest of the paper is organized as follows. Section 2 presents the mathematical foundations of the NOLH DoE and the kriging meta-modeling, Sect. 3 details the two applications of this method and Sect. 4 concludes.

2 Method

This section presents the mathematical background of the kriging-based meta-modeling technique and the associated DoE, and highlights several potential pitfalls in the modeling choices as well as the main available applications of this approach. For the sake of clarity, bold mathematical expressions stand for multi-dimensional variables (i.e. vectors), while normal characters stand for one-dimensional variable (i.e. scalars), and $\mathbf{x}$ indexes indicate values observed at the sampled points using the DoE.

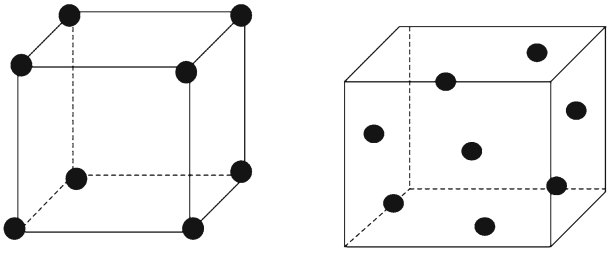
2.1 Preliminary Definitions

Let $x_1, \ldots, x_k$ be the $k \geq 1$ parameters of the model. The parameters are called *factors* (or inputs). The variation domain of each factor is the set of all possible values for this factor. Let $D \subset \mathbb{R}^k$ be the *experimental domain*, *i.e.* the k-dimensional space of the variation domains of the k factors. An *experimental point* (or point) $\mathbf{x}_i = (x_{1,i}, \ldots, x_{k,i})$ is a $1 \times k \in D$ vector, which is a point of the experimental domain (a particular configuration of the factor values in which $x_1 = x_{1,i}, \ldots, x_k = x_{k,i}$). The *DoE* is the $n \times k$ matrix of the n experimental points, which are selected for the sample. Each column represents a factor and each row represents an experimental point. The DoE is thus denoted by $\mathbf{X} \equiv (\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n)' \in M_{n,k}(\mathbb{R})$. An experiment (or simulation) is a run of the model, with a particular parameter configuration $\mathbf{x}_i$, $i = 1, \ldots, n$. Finally, let $y : \mathbf{x} = (x_1, \ldots, x_k) \in D \subset \mathbb{R}^k \rightarrow y(\mathbf{x})$ be the *response* variable, i.e. the variable under interest.

DoE aims at choosing a minimal number n of points $\mathbf{x}_i$, $i = 1, \ldots, n$ in order to approximate the true model y by a kriging meta-model Y of it. As mentioned in the introduction, where to arrange the sample points over the experimental domain is closely related to the choice of the meta-model, i.e. the choice of the form of Y [*see* (Wang and Shan 2007)]. Two main methods are available [*see* for example (Jourdan 2005)]: the first one relies on widely used Ordinary Least Squares (OLS) regressions and is outlined in Sect. 2.2; the second one is based on kriging interpolation and requires specific properties of the DoE. This is reviewed in Sect. 2.3.

2.2 The Classical Approach

One can combine a classical DoE with the OLS estimation of a second-order polynomial model (possibly including two-way interactions) in order to estimate the value of the response y at any unsampled location $\mathbf{x}$ of the experimental domain:



(a) A classical DoE **(b)** A uniform DoE

Fig. 2 Examples of DoE [3 factors, 8 points, source: (Goupy and Creighton 2007)]

$$Y(\mathbf{x}) = \beta_0 + \sum_{g=1}^{k} \beta_g x_g + \sum_{g=1}^{k} \beta_g x_g^2 + \sum_{g=1}^{k} \sum_{h>g} \beta_{g,h} x_g x_h + \varepsilon \tag{1}$$

where ε is a usual error term. Classical DoE are factorial type DoE [see for example (Box and Draper 1987)]. They are very simple to generate and optimal for estimating models of form (1). However, they put experimental points at the extremities of the experimental domain (see Fig. 2a). Consequently, they are not adapted if the response is irregular over the domain. We should use these DoE only if the response is expected to be smooth on the entire domain, or if we investigate only a restricted domain, on which we can locally approximate the response with a smooth function. Nevertheless, Iman and Helton (1988) find that this approach is useful for ranking the relative influence of the factors on the response, even if model (1) is not able to adequately represent the complex response surface.

2.3 The Kriging-Based Approach

A more accurate meta-model Y over the entire experimental domain can be obtained through the use of a spatial interpolation model, namely a kriging model.[4] It is particularly relevant for the analysis of computer simulations, which produce non-linear dynamics and can be highly sensitive to small changes in parameter values. In that case, the response is not smooth over the experimental domain, and a meta-model accounting for that irregularity should be preferred to the classical approach.

We start by discussing the statistical background of kriging meta-models, as it is useful to understand the specific features of the associated DoE that these meta-models require.

2.3.1 Form of the Meta-Model

The response y can be predicted through the meta-model Y:

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}) \tag{2}$$

[4] See, notably, Matheron (1963), Sacks et al. (1989), van Beers and Kleijnen (2004) and Roustant et al. (2010).

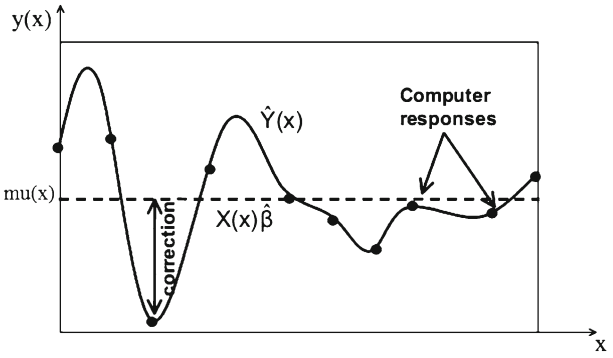


Fig. 3 Ordinary kriging meta-model: $\mu(\mathbf{x}) = \bar{\mu}$ [source: (Jourdan 2005)]

where $\mu : \mathbf{x} \in D \subset \mathbb{R}^k \rightarrow \mu(\mathbf{x}) \equiv \sum_{j=1}^{l} \beta_j f_j(\mathbf{x}) \in \mathbb{R}, l > 0$, is the global trend of the model, composed by predetermined functions f_j (possibly non-linear) and a vector $\boldsymbol{\beta} \equiv \{\beta_j\}_{1,\ldots,l}$ of coefficients, to be estimated. Z is a stochastic process, representing local deviations of the model from the global trend μ (see Fig. 3). The meta-model is said to be *global*, as it is defined over the whole experimental domain D.

Most of the time, Z is assumed to be second-order stationary, with zero mean, and a covariance given by $C \equiv \sigma^2 R$, with σ^2 a scale parameter called the process variance. The correlation function R is a $n \times n$ matrix, whose (i, j) element is $corr\left(Z(\mathbf{x}_i), Z(\mathbf{x}_j)\right)$. Kriging assumes that the closer the points $\mathbf{x}_i$ and $\mathbf{x}_j$, the higher the correlation between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$, and the higher the correlation between the responses $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$. That is why kriging is said to be a *spatial* estimator. This assumption translates into the form of the correlation R. In practice, an exponential function is often used and the (i, j) element of R is computed as:

$$corr\left(Z(\mathbf{x}_i), Z(\mathbf{x}_j)\right) = \exp\left(-\sum_{g=1}^{k} \theta_g \mid x_{g,i} - x_{g,j} \mid\right) \tag{3}$$

where $x_{g,i}$ denotes the value of factor x_g at the experimental point $\mathbf{x}_i$.

Let $\boldsymbol{\theta} \equiv \{\theta_1, \ldots, \theta_k\}$ be the $1 \times k$ vector of positive values quantifying the relative importance of the k factors on the response y. Element $\theta_g \geq 0, g = 1, \ldots, k$, measures the relative importance of factor g. The higher θ_g, the lower the correlation between the responses evaluated for close values of factor x_g, and the smaller the importance of factor g on the response y [see (van Beers and Kleijnen 2004), p. 145]. We note that the correlation between responses does decrease as the distance between points increases, for any given value of θ_g.

The absolute value in (3), representing the distance between two points, can also be replaced by the squared differences $(x_{g,i} - x_{g,j})^2$ to obtain a smoother process (Gaussian correlation).

2.3.2 Estimation of the Meta-Model

Parameters to be estimated are the l coefficients of vector $\boldsymbol{\beta}$, the k coefficients of vector $\boldsymbol{\theta}$ and σ^2, for a total of $l + k + 1$ parameters. The estimation of the kriging model involves two steps (Sacks et al. 1989).

First, we define $F_{\mathbf{x}} \equiv (\mathbf{f}(\mathbf{x}_1), \ldots, \mathbf{f}(\mathbf{x}_n))'$ as the so-called $n \times l$ *experimental matrix*, $\mathbf{f}(\mathbf{x})$ is the $(1 \times l)$ vector of the trend values at some point $\mathbf{x}$, and $\mathbf{y}_{\mathbf{x}}$ is the $(n \times 1)$ vector of the observed values of the response at the n points of the DoE. The l trend coefficients β_j $(j = 1, \ldots, l)$ are estimated using generalized least squares (GLS):[5]

$$\hat{\boldsymbol{\beta}} = (F_{\mathbf{x}}' C^{-1} F_{\mathbf{x}})^{-1} F_{\mathbf{x}}' C^{-1} \mathbf{y}_{\mathbf{x}}$$

Let $\Upsilon(\mathbf{x}) \equiv Y(\mathbf{x}) - \mathbf{f}(\mathbf{x})\hat{\boldsymbol{\beta}}$ be the detrended process defined at any given point $\mathbf{x}$ of the experimental domain, and $\boldsymbol{\Upsilon}_{\mathbf{x}}$ the $(n \times 1)$ vector of the values of the detrended process at the n points of the DoE, i.e. $\Upsilon_{\mathbf{x}} = \left(Y(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_1)\hat{\boldsymbol{\beta}}, \ldots, Y(\mathbf{x}_n) - \mathbf{f}(\mathbf{x}_n)\hat{\boldsymbol{\beta}}\right)'$.

Second, residuals are interpolated, making abstraction of the trend. The best linear unbiased predictor (BLUP) $\hat{\Upsilon}$ of Υ is obtained as a linear combination of the n observations of Υ at the n points of the DoE, with weights $\boldsymbol{\lambda}_{\mathbf{x}} = (\lambda(\mathbf{x}_1), \ldots, \lambda(\mathbf{x}_n))'$. For any point $\mathbf{x}$ of the experimental domain, the value of the detrended process Υ is therefore interpolated as follows:

$$\hat{\Upsilon}(\mathbf{x}) = \sum_{i=1}^{n} \lambda(\mathbf{x}_i)\Upsilon(\mathbf{x}_i) = \boldsymbol{\lambda}_{\mathbf{x}}^{'} \boldsymbol{\Upsilon}_{\mathbf{x}}$$

The mean squared error of the estimation given the weights $\boldsymbol{\lambda}_{\mathbf{x}}$ equals:

$$MSE \equiv E\left[\left(\Upsilon(\mathbf{x}) - \hat{\Upsilon}(\mathbf{x})\right)^2\right] = E\left[\left(Y(\mathbf{x}) - \mathbf{f}(\mathbf{x})\hat{\boldsymbol{\beta}} - \boldsymbol{\lambda}_{\mathbf{x}}^{'} \boldsymbol{\Upsilon}_{\mathbf{x}}\right)^2\right] \tag{4}$$

As the MSE is convex, minimizing it implies that the solution exists, is unique and is given by:

$$\boldsymbol{\lambda}_{\mathbf{x}}^{*} = R^{-1}\mathbf{r}(\mathbf{x})$$

where R is the correlation matrix previously introduced, and $\mathbf{r}(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}_i)_{1 \leqslant i \leqslant n}$ is the $(n \times 1)$ vector of the correlations between the interpolation at any point $\mathbf{x}$ of the experimental domain, $Y(\mathbf{x})$, and the values of Y at the sampled points $(Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))$.

In expression (4), weights $\boldsymbol{\lambda}_{\mathbf{x}}$ are now replaced by their optimal values $\boldsymbol{\lambda}_{\mathbf{x}}^{*}$, and the values of $\boldsymbol{\Upsilon}_{\mathbf{x}}$ by the actual (observed) values $y(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_i)\hat{\boldsymbol{\beta}}$ at the n points of the DoE. The so-called mean prediction of y at any point $\mathbf{x} \in D$ is hence given by:

[5] The estimation of the meta-model is actually done through feasible GLS as the covariance matrix C is unknown and its parameters have to be estimated, see below.

$$E(Y(\mathbf{x})) = \mathbf{f}(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}(\mathbf{x})' R^{-1} \left(\mathbf{y}_\mathbf{x} - F_\mathbf{x}\hat{\boldsymbol{\beta}}\right) \tag{5}$$

Similarly, by plugging $\boldsymbol{\lambda}_\mathbf{x}^*$ in the expression of the MSE, the mean squared error of the predictor (the so-called kriging variance) is obtained:

$$\begin{aligned} s_Y(\mathbf{x}) = \sigma^2 \Bigg(1 - \mathbf{r}(\mathbf{x})' R^{-1} \mathbf{r}(\mathbf{x}) (\mathbf{f}(\mathbf{x}) - r(\mathbf{x})' R^{-1} F_\mathbf{x}) \\ + \left(F_\mathbf{x}' R^{-1} F_\mathbf{x}\right)^{-1} (\mathbf{f}(\mathbf{x}) - r(\mathbf{x})' R^{-1} F_\mathbf{x})' \Bigg) \end{aligned} \tag{6}$$

σ^2 is estimated as follows:

$$\hat{\sigma}^2 = \frac{1}{n} \Upsilon' R^{-1} \Upsilon \tag{7}$$

and parameters $\boldsymbol{\theta}$ are mostly estimated using the maximum of likelihood, under Gaussian assumptions [see (Welch et al. 1992)]. Consequently, the following expression measures the quality of the model, where smaller values represent a better fit [see, e.g., (Jeong et al. 2005)]:

$$\begin{aligned} -2 \ln \left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\boldsymbol{\theta}}\right) = n \ln(2\pi) + n \ln \left(\hat{\sigma}^2\right) + \ln (\mid R \mid) \\ + \frac{1}{\hat{\sigma}^2} \left(\mathbf{y}_\mathbf{x} - F_\mathbf{x}\hat{\boldsymbol{\beta}}\right)' R^{-1} \left(\mathbf{y}_\mathbf{x} - F_\mathbf{x}\hat{\boldsymbol{\beta}}\right) \end{aligned} \tag{8}$$

Variance $s_Y(\mathbf{x})$ tends towards zero when $\mathbf{x}$ gets close to experimental points $\mathbf{x}_i$. In other words, error is null at samples and increases with distance: kriging is an *exact* interpolator, i.e. $Y(\mathbf{x}_i) = y(\mathbf{x}_i), \forall i = 1, \ldots, n$. However, this property can be released in case of non-deterministic responses (see below). Contrary to the OLS regression where all observations $\mathbf{x}_i$ are given an equal weight in the estimation, kriging estimation adjusts the weights $\boldsymbol{\lambda}_\mathbf{x}$, depending on the point $\mathbf{x} \in D$ where the response y is interpolated. More precisely, experimental points $\mathbf{x}_i$ closer to the point $\mathbf{x}$ are given a stronger weight in the estimation of $Y(\mathbf{x})$ than further ones. This property makes kriging estimations more flexible, and results in more precise estimations than with OLS (van Beers and Kleijnen 2004). This feature also requires particular properties of the DoE.

2.3.3 Properties of the DoE for the Kriging Approach

In classical DoE, points are set on the extremities of the domain, as illustrated in Fig. 2a. This characteristic prevents the analysis from accurately estimating the parameters of $\boldsymbol{\theta}$ in kriging models because the response is only measured at very distant points. Consequently, kriging estimation requires a DoE with good space-filling properties. Points have to be uniformly distributed across the domain. Designs that fulfil this requirement are called *uniform designs* (see Fig. 2b). This criteria is essential if the modeller's aim is a wide exploration of the model, without a precise prior knowledge of the relations between the factors and the response (Fang et al. 2000). *Latin hypercubes*

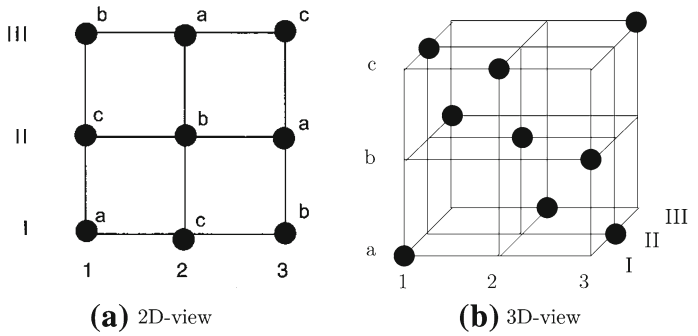


Fig. 4 A latin cube, with three factors and three levels [source: (Goupy and Creighton 2007)]

are often used to that end [see (Goupy and Creighton 2007)]. Figure 4 illustrates the way such a DoE is constructed in a case with three factors—A, B and C, each of them taking 3 values. There are then $3^3 = 27$ possible combinations. Let us start with a square, representing the values of factor A (denoted by Arabic numbers) and B (in Roman numbers). We then attribute the three Latin letters a, b, and c, representing the 3 values of factor C, so that a, b and c are uniquely displayed in all rows and all columns. Only 12 configurations are possible,[6] and we randomly take one. The DoE involves 9 points, over the 27 initially to be tested. The DoE can be represented in three dimensions, as a cube. When more than 3 dimensions are involved, the DoE is an hypercube. Hypercubes ensure that the non-collapsing criteria is fulfilled: each combination of three factors is tested only once. More precisely, if one of the three factors turns out to be unimportant and is eliminated, no points become identical in the 2D space constituted by the two remaining factors [see (van Beers and Kleijnen 2004), p. 166].

We propose to use the Latin hypercube of Cioppa (2002), based on previous works of Ye (1998), because it provides interesting properties, besides having good space filling properties. Latin hypercubes are not generally orthogonal, some pairs of columns can be correlated. This feature can create multicolinearity issues in the estimation of the meta-model. Ye (1998) develops a way to obtain orthogonal hypercubes, and he even retains a more restrictive condition: not only each pair of columns has to be uncorrelated, but also the squares as well as the cross-products of each column. However, orthogonality is obtained at the expense of the space-filling properties of the DoE. Cioppa (2002) obtains an interesting trade-off between the two properties, while limiting the size of the DoE. The orthogonality criterion is released and the author defines a *near-orthogonality* criterion, according to which absolute values of correlations cannot exceed 0.03. Results are impressive: Cioppa demonstrates that non-linear relations and interactions can be significantly identified with only 17 experiments up to 7 factors, 33 experiments up to 11 factors, 65 experiments up to 16 factors, 129 experiments up to 22 factors and 257 up to 29 factors. Moreover, those corresponding

[6] There are two possible triplets of rows (bac, cba, acb) and (abc, cab, bca), each can be permuted in $3! = 6$ different ways, so that one obtains $2 \times 6 = 12$ possible configurations.

DoE are easy to generate: a spreadsheet file helping their computation is available in Sanchez (2005). The DoE are also constructed with a minimum of *a priori* restrictions on the relations between the factors and the response, as they allow for the estimation of a polynomial model of the form (1) as well.

2.3.4 Additional Issues in the Meta-Model Choice

More complex forms of the correlation function R can be chosen but a more sophisticated correlation function requires more observations to accurately estimate its parameters.

In practice, the trend μ is often reduced to a constant $\bar{\mu}$, which is thus interpreted as the mean of the process Y (see Fig. 3).[7] In that case, Eq. (5) is reduced to:

$$E(Y(\mathbf{x})) = \bar{\mu} + \mathbf{r}(\mathbf{x})' R^{-1} (\mathbf{y}_\mathbf{x} - \bar{\mu}) \tag{9}$$

However, if the trend is constant, the model is more sensitive to the specification of the correlation function R and to the estimations of the parameters $\boldsymbol{\theta}$ (Jourdan 2005). The trend or the correlation function may also be chosen according to an optimality criterion. Either one can use cross-validation, or external validation.

Cross validation consists in removing one or several points of the DoE, reestimating the model, and comparing the error between the estimations and the observed values at the removed points. The assessment of the estimation quality is then based on the Q^2 *predictivity coefficient* [see (Durrande et al. 2012)], which is a proxy of the R^2 of standard linear regressions, and is computed as:[8]

$$Q^2 \equiv 1 - \frac{\sum_{i=1}^{n} \left(y(\mathbf{x}_i) - \hat{Y}(\mathbf{x}_i) \right)^2}{\sum_{i=1}^{n} \left(y(\mathbf{x}_i) - \bar{y}_\mathbf{x} \right)^2} \tag{10}$$

where $\bar{y}_\mathbf{x}$ is the mean of y over the n observations of the DoE, $\hat{Y}(\mathbf{x}_i)$ is the predicted value of y at the sampled point $\mathbf{x}_i$, when the estimation is performed by using the $n-1$ points $\mathbf{X}_{-i}$ (i.e. removing point $\mathbf{x}_i$ from the DoE). Values close to zero denote a weak predictive power, whereas values close to one indicate a better fit.

External validation requires the estimation of the response at additional points, which are randomly chosen outside the DoE. The observed values at those points are then compared with the kriging predictions based on the original n points of the DoE, and the kriging meta-model which minimizes the root mean square error (RMSE) is chosen. Both criteria are broadly consistent with each other but relying on external validation is preferable when the DoE involves a small number of points because cross-validation may lead to imprecise estimations.

In the case of non-deterministic responses, different runs with the same parameter values yield different values of the response, and experiments have to be replicated

[7] In that case, the meta-model refers to *ordinary kriging*, and of simple kriging if the mean is known, contrary to *universal kriging* in the more general case, which is exposed above.

[8] Recall that kriging is an exact interpolator, so that the R^2 coefficient cannot be computed.

several times to obtain a significant evaluation of the response at a given point of the DoE. The kriging model is then applied to the average value of the response over the number of replications (van Beers and Kleijnen 2004). Let $\tilde{y}(\mathbf{x}_i) = y(\mathbf{x}_i) + \epsilon_i$ be the value of the response at point $\mathbf{x}_i$ of the DoE. We assume that $\epsilon_i \hookrightarrow NID(0, \tau_i^2)$, where τ_i^2 is the variance of the observed values of y over replications of the experiment at point $\mathbf{x}_i$ of the DoE. As soon as the process Y and the errors ϵ_i are independent, kriging meta-modeling can very easily be extended to non-deterministic responses. Matrix R in Eqs. (5), (6) and (7), or in (9), is just replaced by $R+\Delta$, where $\Delta \equiv diag(\tau_1^2, \ldots \tau_n^2)$ [see for example (Roustant et al. 2010)]. The only difference is that the kriging-based estimation has now two sources of error: not only the error due to the difference between y and its meta-model Y, but also an experimental error, *i.e.* the noise in the response measurements ϵ. In that case, the variance (6) is higher than in the case of deterministic responses.

Before turning to the application of kriging meta-modeling and DoE to economic models, we conclude this section with a discussion of the main purposes of a meta-model.

2.3.5 Purposes of the Meta-Model

A meta-model is mostly devoted to two purposes. Sensitivity analysis of the computational model is the first one. It aims at identifying how much influence each factor has on the response, and which factors do not significantly affect the response. Rather than interpreting the estimated coefficient θ_g for each factor $g = 1, .., k$, a functional ANOVA of the meta-model Y can be performed, as a proxy of the ANOVA of the true model y [see (Jeong et al. 2005)]. To do so, the total variance of the meta-model Y is decomposed into that of each factor and their interactions. Formally, the decomposition is performed by integrating factors out of the meta-model Y.[9] Let $\mathcal{M}$ be the average and $\mathcal{V}$ be the variance of Y over D:

$$\mathcal{M} \equiv \int Y(\mathbf{x}) \Pi_{g=1}^{k} dx_g \tag{11}$$

$$\mathcal{V} \equiv \int [Y(\mathbf{x}) - \mathcal{M}]^2 \, \Pi_{g=1}^{k} dx_g \tag{12}$$

The main effect of factor x_g on Y (averaged over the other factors) is given by:

$$m(x_g) \equiv \int Y(\mathbf{x}) \Pi_{j \neq g} dx_j - \mathcal{M} \tag{13}$$

and the two-way interaction effect of factors x_g and x_h:

[9] In practice, either the factors have a finite set of values, and the ANOVA is performed using the common formula of the multi-variate analysis of variance with discrete factors, see, for instance, Frey and Patil (2002); or the factors are defined over a continuous domain, and the experimental domain has to be discretized in order to apply these formula. In that case, predictions of the response through the meta-model are evaluated over a k-dimensional grid [see (Welch et al. 1992; Saltelli et al. 1999)].

$$m(x_g, x_h) \equiv \int Y(\mathbf{x}) \Pi_{j \neq g,h} dx_j - m(x_g) - m(x_h) - \mathcal{M} \tag{14}$$

The total sensitivity of the response Y to factor x_g (taking into account its two-way interactions with all other factors) is given by $M(x_g) \equiv m(x_g) + \sum_{g \neq h} m(x_g, x_h)$. The ratio of $\mathcal{M}_g \equiv \frac{M(x_g)}{\mathcal{V}} \in [0, 1]$ indicates the proportion of variation in Y which is due to variation in factor x_g. Such ratio can be plotted as barplot for each factor to sum-up the sensitivity analysis outcomes (see Fig. 6c below for an example). It is usual to plot the sensitivity of the response to each factor in order to highlight the relative magnitude of the effects of the factors, and to indicate interaction effects or non-linearities. For instance, plotting $m(x_g)$ against the values of factor x_g over the experimental domain gives a visual indication of the main effect of factor x_g on Y and, hence, on y (see Fig. 5a–f below). Similarly, plotting $m(x_g, x_h)$ against x_g and x_h yields a three-dimensional surface of the response Y, and highlights the interaction effects of the two factors on the response (see Fig. 5g–i below).

The meta-model can further be optimized, in order to identify the values of the factors which minimize (or maximize) the proxy Y of the true process y.

The next section illustrates the advantages of the kriging-based meta-modeling, and gives practical guidelines to the reader. We show that this approach is more parsimonious that standard Monte-Carlo explorations in both examples, and we use the second example for also introducing the other potentially very interesting use of this approach: finding configurations that are *optimal* in terms of a given criterion (eg. maximizing the social welfare, or minimizing the distance to empirical stylized facts).

3 Applications

3.1 Sensitivity Analysis of Nelson and Winter (1982) Model

3.1.1 A simple model of industry evolution

The industry dynamics model of Nelson and Winter (1982) (Chap. 12–14), has been extensively analyzed in the literature, and its detailed properties have been established using many different types of analysis (the original work of Nelson and Winter uses simple tables and graphics to this end). Many different implementations of this model using different languages and platforms exists on the academic internet (the original code was implemented on a mainframe). We implement this model in NetLogo,[10] which is an open source agent-based modeling platform. We focus on the analysis of the factors that may lead to the emergence of highly concentrated industrial structures.[11]

[10] http://ccl.northwestern.edu/netlogo/

[11] See also Nelson and Winter (1978) for an extensive presentation and discussion of the model. In this paper, we only use this model as a simple example, in order to apply the method previously developed. We adopt values used in the original model for the parameters that we do not include in our experiments.

In this very simple model, each firm uses a single input (physical capital) to produce a unique output. Technical progress results from the R&D activities of each firm and increases the productivity of its capital stock (disembodied technical progress). The physical capital stock of each firm also changes as a consequence of the investment activity of the firm.

The industry is populated by n firms, indexed by $j = 1, \ldots, n$, each producing a quantity $q_{j,t}$ of the good in each period t according to:

$$q_{j,t} = A_{j,t} K_{j,t} \tag{15}$$

where $K_{j,t}$ is firm j's physical capital stock and $A_{j,t}$ is the productivity of its capital. Given their capital stock and their productivity, firms sell the totality of their production to the consumers whose inverse demand function is given by:

$$P_t = \frac{D}{Q_t^{\eta}} \tag{16}$$

where $Q_t \equiv \sum_{j=1}^{n} q_{j,t}$ is the aggregate supply, and the price is consequently adjusted in order to attain the temporary equilibrium of the period on the product market. We use here the values of the demand parameters adopted in the original version of the model ($D = 64$, $\eta = 1$). Given this price, the net profits of firm j equal:

$$\pi_{j,t} = (A_{j,t} P_t - c) K_{j,t} = \left(P_t - \frac{c}{A_{j,t}} \right) q_{j,t} \tag{17}$$

where c is the unit using cost of capital, including R&D costs.

Firms invest in each period a fixed proportion of their capital stock on imitative and innovative R&D. These R&D investments respectively determine the probabilities for the firm of being successful in imitation (and imitate the technique with the highest current productivity in the industry), and in innovation. When successful in innovation, the exact new technique that can be discovered by the firm depends on the properties of the technological regime characterizing this industry. Nelson and Winter consider different technological regimes in their original work. We consider here only the simplest technological regime with science–based innovations. In that case, the new technique discovered after a successful innovation depends on the latent productivity that increases at a constant exponential rate g (the advance of scientific knowledge). In each period, with a probability $P(innov = 1) = 0.0025 K_{j,t}$, a firm can be successful in innovation, and discover a new technology $\tilde{A}_{j,t}$, assuming $log(\tilde{A}_{j,t}) \hookrightarrow \mathcal{N}(A_0 + g.t, \sigma^2)$. With a probability $P(imit = 1) = \tau_{im} K_{j,t}$, a firm can also be successful to imitate the best technology in the industry: $A_t^* = \max_{j=1,\ldots n} A_{j,t}$. In $t + 1$, firm j's technology is therefore given by the best of these technologies:

$$A_{j,t+1} = \max \left(A_{i,t}, innov \cdot \tilde{A}_{i,t}, imit \cdot A_t^* \right) \tag{18}$$

Gross investment I of a firm is constrained depending on whether the firm is making economic profits or not. Formally,

$$\bar{I}_{j,t} = \begin{cases} \pi_{j,t} & \text{if } \pi_{j,t} < 0 \\ (1+B)\cdot\pi_{j,t} & \text{if } \pi_{j,t} \geq 0 \end{cases} \tag{19}$$

where $B > 0$ denotes the external financing of firms. Each firm j has a target mark-up, defined as $\mu_{j,t} = \frac{\epsilon}{\epsilon - s_{j,t}}$, where ϵ is the perceived demand elasticity [η, equals to $1/n$ in the original model, see Eq. (16)], and $s_{j,t} \equiv \frac{q_{j,t}}{Q_t}$ is the market share of firm j. Firms desire positive net investment $I^T_{j,t}$ if the ratio of price to unit cost exceeds a target markup factor. Formally:

$$I^T_{j,t} = \left(1 - \mu_{j,t}\frac{c}{A_{j,t+1}P_t}\right) \tag{20}$$

Finally, actual gross investment $I_{j,t}$ is given by $\max\left[0, \min(\bar{I}_{j,t}, I^T_{j,t})\right]$.

Technical progress drives the performance of the firms and dynamics of their capital stock. We focus here on the dynamics of resulting market concentration, measured as the concentration of the capital stock among firms. Since we do not have exit in this model, the evolutionary dynamics will correspond to higher growth rate for the successful firms in the industry, and as a consequence, the concentration of capital will change. Depending on the technological regime, technical progress can be unevenly distributed between firms, and yield an increasing market concentration. We measure the final state of this concentration through the value in the last period of a normalized Herfindahl index, denoted by $h_t \in [0, 1]$:

$$h_t = \frac{\sum_{j=1}^{n}\left(\frac{K_{j,t}}{\sum_{j=1}^{n}K_{j,t}}\right)^2 - \frac{1}{n}}{1 - \frac{1}{n}} \tag{21}$$

When the market shares tend to be evenly distributed across the n firms $\left(K_j = K, \forall j\right)$, the standard Herfindhal index $\sum_{j=1}^{n}\left(\frac{K_{j,t}}{\sum_{j=1}^{n}K_{j,t}}\right)^2$ tends to $1/n$, and the normalized index h tends to zero. When one firm tends to corner the whole market, the index, and its normalized value, tend to unity. Values close to zero therefore indicate a competitive market and values close to one stand for a monopolistic industry. The normalized value is used in order to compare simulations with different numbers of firms.

3.1.2 Comparing Monte Carlo Sampling Versus the NOLH and Kriging Approach

Two Simulation Protocols We follow Nelson and Winter (1978) by defining a period t a quarter, setting $c = 0.16$, $A_{j,0} = A_0 = 0.16$, $\forall j$, and defining $k = 6$ factors: $\tau_{im} \in [0.000625, 0.005]$, reflecting different levels of the difficulty of imitation; $n \in [\![2, 32]\!]$; $\epsilon \in [0.8, 1000]$; $g \in [0.25, 1.5]\%$; $B \in [1, 3.5]$; $\sigma^2/g \in [\![4, 12]\!]$.

We compare two alternative methods to explore this experimental domain and to determine the effect of these 6 factors on the response h. The first one involves 1, 000

Monte Carlo simulations,[12] over which we adjust a polynomial regression model of the form (1) with two-way interaction effects.

The second one implements the NOLH DoE that we have introduced in the previous section, over which we estimate a kriging meta-model of h, denoted by H. We then perform an ANOVA of the kriging meta-model. As we need to discretize the experimental domain to generate the NOLH DoE, we follow the values investigated by Nelson and Winter (1978), and consider $\tau_{im} = \{0.000625; 0.00125; 0.0025; 0.005\}$, $n = \{2, 4, 8, 16, 32\}$, $\epsilon = \{0.8; 1; 1,000\}$, $g \in [0.25, 1.5]$ % by 0.25 step, $B = \{1; 1.5; 2; 2.5; 3; 3.5\}$ and $\frac{\sigma^2}{g} = \{4, 8, 12\}$. The corresponding DoE from Sanchez (2005) involves only 17 points and is given in Table 3. Following Nelson and Winter (1982), we repeat each non-deterministic run 5 times, i.e. we launch 85 simulations. The analysis is performed using JMP (Cary 2010, Chap. 14), *see* also Oeffner (2008) for an application to a macroeconomic agent-based model.[13] We use ordinary kriging (i.e the trend μ is assumed to be a constant), and the correlation function is Gaussian.

Results Figure 5 reports the ANOVA table of the kriging meta-model H as well as the plots of marginal and interaction effects,[14] and Tables 1 and 2 depict the results of the polynomial regression models (for the 1, 000 and the 10, 000 simulation samples), in which cross-products have been introduced (*see* second column) in order to allow for comparisons with the ANOVA table of the kriging meta-model.

The overall picture is fairly the same: the number of firms n and the rate of imitation τ_{im} are the main determinants of market concentration, while parameters g, ϵ and B do not significantly influence the structure of the industry. This can be seen by looking at the plots of the marginal effects of the response to values of g, ϵ and B, which clearly show that the response is invariant to these parameters' values (cf. Fig. 5c, e, f). Accordingly, the ANOVA table reports negligible or even zero marginal effects for these parameters (see the third column $M(.)$). By contrast, Fig. 5a, b show that the estimated Herfindhal index is highly sensitive to the values of n and τ_{im}, and the ANOVA table reports non-zero marginal effects (respectively 0.6402 and 0.566 for n and τ_{im}). The more firms, or the less frequent imitation, the more concentrated the industry (cf. Fig. 5a, b). As the size of the market is fixed [see Eq. (16)], the selective pressure is strengthened and the decrease in price is faster, *ceteris paribus*, as the number of firms n increases. This mechanism intuitively explains the predominant role of n. Moreover, the interaction term between n and τ_{im} is significant (the ANOVA table reports a non-zero interaction effect, precisely 0.2246, and the response surface in Fig. 5g is sensitive to values of τ_{im} and n). Imitation affects the industry all the more that n is large. When the number of firms increases, the selective pressure on lagging firms increases, and imitation becomes the major tool for decreasing this pressure, by catching up with the technological leader. The highest degree of concentration (corresponding to the highest value of H) is obtained with scarce imitation and a high

[12] We also consider a 10, 000 simulations Monte Carlo sample for robustness checks.

[13] R Development Core Team (2013) software can also be used but the package `effects`, which computes ANOVA marginal effects, is not directly connected to the `DiceKriging` package, which performs kriging estimation and the modeler has to use the package `sensitivity`, which delivers less detailed results [*see* (Roustant et al. 2010)].

[14] These figures are built using the principles given in Sub-subsection 2.3.5.

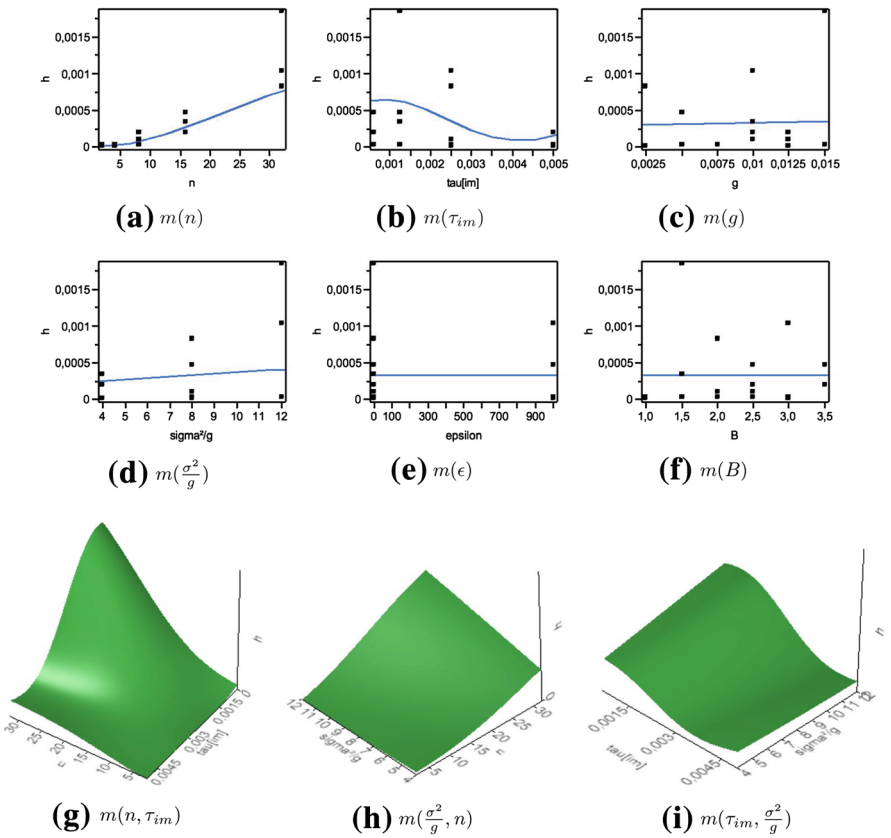


Fig. 5 ANOVA table of the kriging model H

number of firms (see Fig. 5g). Intuitively, among a lot of firms, a firm is more likely to gain a striking competitive edge if it cannot be imitated easily.

When we compare these results with the Monte-Carlo simulations, we observe that not all of them can be captured by the 1, 000 simulation experiment. In fact, in more extensive tests, we observed that we need at least 2, 000 Monte-Carlo simulations to be able to systematically catch[15] the direct effects of n and τ_{im}. Otherwise, one of these effects can remain hidden in some samples. Concerning the two-way interactions, only the one between these two main variables can be observed in the 1, 000 simulations experiment. More runs are necessary to observe the role of the easiness to discover new technologies (determined by g and σ^2/g). The role of this technological dimension only appears in experiments with a considerably higher number of simulations (10, 000 simulations are necessary in this case), and the linear regressions catch only the multiplicative components of this dimension.

[15] We consider that we catch these effects in a robust way if they appear as significant in each of the 100 regressions obtained from 100 random sets of 2, 000 simulations.

Table 1 OLS regression of h over the 1, 000 points Monte Carlo sample

	Estimate	Std. error	Pr(> |t|)		Estimate	Std. error	Pr(> |t|)
cst.	0.312	0.116	0.0082**	$n.\frac{\sigma^2}{g}$	0.001	0.000	0.035*
n	0.965	0.003	0.000***	$n.\tau_{im}$	6.752	0.4284	0.000***
$\frac{\sigma^2}{g}$	0.001	0.012	0.90	$\frac{\sigma^2}{g}.\tau_{im}$	1.926	2.0052	0.338
τ_{im}	−84.13	24.262	0.001***	$n.B$	−0.001	0.001	0.144
B	−0.0214	0.034	0.528	$\frac{\sigma^2}{g}.B$	0.001	0.003	0.804
g	5.54	8.186	0.499	$\tau_{im}.B$	9.1775	5.4153	0.1
ϵ	0.000	0.000	0.347	$n.g$	−0.1469	0.1479	0.321
				$\frac{\sigma^2}{g}.g$	−0.5928	0.6673	0.375
				$\tau_{im}.g$	1056.295	1297.656	0.416
				$B.g$	−0.66	1.836	0.719
				$n.\epsilon$	−0.000	0.000	0.179
				$\frac{\sigma^2}{g}.\epsilon$	−0.000	0.000	0.517
				$\tau_{im}.\epsilon$	−0.001	0.01	0.364
				$B.\epsilon$	0.000	0.000	0.922
				$g.\epsilon$	−0.000	0.003	0.917

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1
Adjusted $R^2 = 0.9997$

Table 2 OLS regression of h over the 10, 000 points Monte Carlo sample

	Estimate	Std. error	Pr(> |t|)		Estimate	Std. eror	Pr(> |t|)
cst.	0.3580	0.0408	0.0000***	$n.\frac{\sigma^2}{g}$	−0.0001	0.0001	0.2408
n	0.9551	0.0011	0.0000****	$n.\tau_{im}$	7.1698	0.1494	0.0000***
$\frac{\sigma^2}{g}$	0.0047	0.0037	0.2062	$\frac{\sigma^2}{g}.\tau_{im}$	−0.1254	0.5763	0.8278
τ_{im}	−66.4845	7.6013	0.0000***	$n.B$	−0.0001	0.0002	0.6389
B	0.0002	0.0107	0.9883	$\frac{\sigma^2}{g}.B$	0.0001	0.0008	0.9190
g	3.4004	2.6649	0.2020	$\tau_{im}.B$	0.5617	1.5445	0.7161
ϵ	0.0001	0.0000	0.0565	$n.g$	−0.2076	0.0525	0.0001***
				$\frac{\sigma^2}{g}.g$	−0.3193	0.2025	0.1150
				$\tau_{im}.g$	836.0536	368.2086	0.0232*
				$B.g$	0.0559	0.5426	0.9180
				$n.\epsilon$	0.0000	0.0000	0.6882
				$\frac{\sigma^2}{g}.\epsilon$	−0.0000	0.0000	0.0607
				$\tau_{im}.\epsilon$	−0.0060	0.0047	0.1944
				$B.\epsilon$	−0.0000	0.0000	0.9122
				$g.\epsilon$	−0.0031	0.0016	0.0565

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1
Adjusted $R^2 = 0.996$

These results also show that we may over-estimate direct effects if we do not use a sufficient number of simulations. We also observe that a higher concentration tends to emerge when research outcomes are strongly dispersed, all the more that there are many firms. It should be noted that the kriging analysis underlines the individual effects of g and σ^2/g (see Fig. 5c, d), as well as the interactions of σ^2/g with n and τ_{imit} (see Fig. 5h, i), while the least squares regression model based on 10, 000 data significantly reports the interactions of g with n and τ_{imit}. Despite this minor discrepancy, the two models deliver the same message, and highlight the role of innovation draws: concentration is higher if innovations are diverse, especially if imitation is rare, and firms are numerous. In such a context, a firm is more likely to gain a competitive edge because it is more likely to draw a leading innovation, without being imitated in subsequent periods. While this result is intuitively appealing, it should be noted that the effects are quite small, as the ANOVA table reports non-zero but much smaller marginal effects (0.0341 for $\frac{\sigma^2}{g}$, 0.0102 for the interaction between n and $\frac{\sigma^2}{g}$ and 0.0103 for the interaction between τ_{im} and $\frac{\sigma^2}{g}$).

The possibility of catching the main effects of the parameters, and their interactions, with only 85 simulations, instead of 2, 000 or 10, 000 clearly show the frugality of the approach proposed in this article. Importantly, note that these results are completely consistent with those of Nelson and Winter (1978, 1982).

3.2 Kriging-Based Optimization Within a Cournot Oligopoly Model

We now propose a two-stage analysis of a Cournot oligopoly model with adaptively learning firms. First, as in the preceding section, we develop a sensitivity analysis using a kriging meta-model, in order to demonstrate the possibility of such an analysis exclusively using R-Project (R Development Core Team 2013) software.[16] The exact form of the kriging meta-model (the correlation function and the trend) is chosen according to the Q^2 predicticity coefficient and the external validation procedure discussed in Sect. 2.3.4. We then compare these sensitivity results with those obtained using a simple econometric analysis based on Monte Carlo simulations. At a second stage, we use the meta-model to identify the configuration of factors that yields a minimal distance of firms to the Cournot equilibrium. This configuration is hence the one that favors the convergence of learning firms to this equilibrium.

3.2.1 A Baseline Oligopoly Model

We define a simple oligopoly game with $n > 1$ firms, calibrated as in Vallée and Yıldızoğlu (2009):

$$P(q_j, Q_{-j}) = 256 - 2(q_j + Q_{-j}) \tag{22}$$

$$C(q_j) = 56q_j + q_j^2 \tag{23}$$

[16] The complete code used in this section is provided in Appendix 1.

where P denotes the aggregate price, q_j firm j's supply, $Q_{-j} = \sum_{i \neq j} q_i$ and $C(.)$ is the cost function. We assume $n = 30$, and the game has two symmetric equilibria, $\forall j$:

- *Cournot-Nash equilibrium* (CE): $q_j \simeq 3.125$ and the profit equals $\pi_j^c = 68.5$
- *Walrasian equilibrium* (WE): $q_j = q^w \simeq 3.2258$ and the profit equals $\pi_j^w = 62.45 < \pi_j^c$

Firms update their supply q_j according to a learning mechanism. With a probability P_{im}, for each period, firms can imitate the strategy of the firm which is making the highest profit in the industry. Otherwise, they use their individual mental model, represented by an artificial neural network (ANN).[17] Each firm is endowed with a one-hidden layer ANN with $hid > 1$ hidden nodes. This ANN is fed in each period with 4 inputs (price evolution, evolution of individual sales, variation of individual costs and evolution of individual profit), and each firm has a population of 40 quantity strategies. In each period, each firm selects the strategy among this population which maximizes the expected discounted profit flow predicted by the ANN over the $fL+1$ future periods (with a discount factor set to 0.99). As new observations become available, firms' ANNs are trained by back propagating the errors on the ANN coefficients: $epoch$ iterations are performed to reduce each time a proportion δ of the error between the predicted and the actual profit flow. Finally, every γ_{GA} periods, the population of strategies is modified by operators selection, mutation (with a probability P_{mut}) and average crossover (with a probability P_{co}).

In this model, we investigate the design of the learning algorithm which allows the industry to converge towards the CE. We have $k = 8$ factors depicting firms' learning, with the following associated variation domains: imitation probability $P_{im} \in [0, 0.25]$; mutation probability, $P_{mut} \in [0.01, 0.2]$; crossover probability, and $P_{co} \in [0.05, 0.4]$; number of hidden nodes, $hid \in [\![2, 4]\!]$; forward looking horizon of the firms, $fL \in [\![0, 12]\!]$; length of the training period, $epoch \in [\![20, 50]\!]$; error correction rate, $\delta \in [0.05, 1]$; GA frequency, $\gamma_{GA} \in [\![1, 30]\!]$. The response variable is the absolute distance of aggregate supply to its CE value, $d \equiv \mid \sum_{j=1}^{n} q_j - nq^c \mid$, evaluated at the end of a 1, 000 period run, and its kriging-based approximation is denoted by D. We aim at determining the factor configuration for which the distance, d, is minimized. We sample the 8-dimensional parameter space with Sanchez (2005) NOLH DoE given in Table 4, which defines $n = 33$ deterministic experiments. Each experiment is repeated 20 times in order to account for the non-deterministic nature of the learning model, and we apply kriging over the average response in each experiment. The combination of a reduced number of simulations with a kriging meta-model turns out to be very useful indeed for models involving algorithms such as ANNs, which are quite time consuming to run.

3.2.2 Sensitivity Analysis and Optimization with a Kriging Meta-Model

Model Selection The first stage of our analysis concerns the choice of the form of the kriging meta-model. The R-project package `DiceKriging` [see

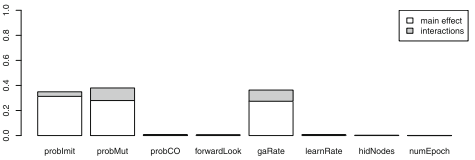
[17] See Masters (1993) for a general statement, see Yıldızoğlu (2001) and Yıldızoğlu et al. (2012) for the precise description of the learning algorithm.

	cov. matèrn		Gaussian cov.		Exp. cov.	
	$\bar{\mu}$	1st order	$\bar{\mu}$	1st order	$\bar{\mu}$	1st order
RMSE	1.5787	1.6335	**1.3208**	2.4772	2.1608	2.1458
Q^2	0.6865	0.7543	0.7144	0.6703	0.3913	0.6457

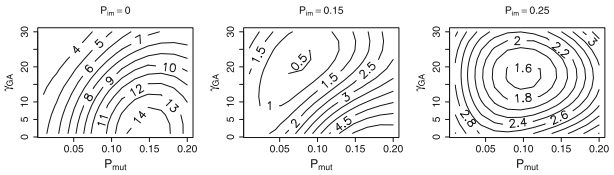
(a) Comparison and selection of kriging meta-models based on the NOLH DoE.

$E(D) = \bar{\mu}$	5.0279	$\hat{\sigma}^2$	9.9231
$\theta_{P_{im}}$	0.0998	$\theta_{\gamma_{GA}}$	19.4
$\theta_{P_{mu}}$	0.0791	θ_{δ}	1.9
$\theta_{P_{co}}$	0.7	θ_{hid}	4
θ_{fL}	24	θ_{epoch}	60

(b) Ordinary kriging estimation of d (Gaussian correlation function)



(c) Sensitivity analysis of the selected meta-model



(d) Estimated response surfaces of the kriging meta-model, the other 5 parameters are set at the middle of their variation domains. P_{mut} on the x-axis, γ_{GA} on the y-axis, for three values of $P_{im} = \{0, 0.15, 0.25\}$.

Fig. 6 Kriging model D

(Roustant et al. 2010)] allows this flexibility. We compare three different correlation functions (Gaussian, exponential and Matèrn $\nu = 5/2$, which is the default function in this package), and two specifications of the trend μ (a constant, and a first-order polynomial).[18] In order to discriminate between these potentially valid specifications, we use both external validation, and cross-validation (see paragraph 2.3.4). For external validation, we evaluate the kriging meta-model's predictions at 7 additional experimental points, that we randomly choose over the whole experimental domain (see Table 4). Figure 6a reports the results of this comparison. The two criteria broadly give the same insights into the meta-model's predictive power. Yet, as we only have a small set of points ($n = 33$), we rely on external validation, and choose the form of the meta-model which minimizes the root-mean-square error (RMSE) between the predicted and the observed response values at the 7 additional points. Accordingly, we select the ordinary kriging meta-model (i.e. in which the trend μ is only a constant term) with the Gaussian correlation function. Table 6b gives the estimated values of the coefficients of the selected kriging meta-model.

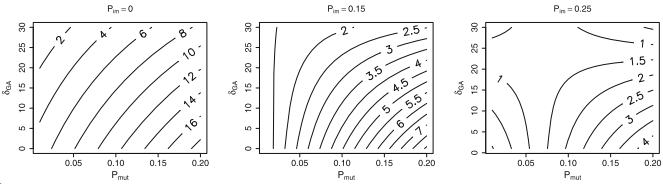
Sensitivity Analysis In a second stage, the sensitivity analysis of the kriging meta-model is performed. Figure 6c summarizes the effects of each factor on the response, and identifies three factors which drive the model's dynamics, with the same order of magnitude (about one third of the response variability is due to each of these three

[18] Higher order polynomials would involve too many parameters to be estimated, considering only 33 observations.

	Estimate	Std. Error	Pr(> |t|)		Estimate	Std. Error	Pr(> |t|)
(Intercept)	5.0125	0.8573	0.0000***	P_{mut}:γ_{GA}	-1.0281	0.0500	0.0000***
P_{im}	-23.5296	2.5362	0.0000***	P_{mut}:δ	-0.9211	1.6241	0.5706
P_{mut}	72.0525	3.1213	0.0000***	P_{mut}:*hid*	1.0042	0.5483	0.0671
P_{co}	-1.1500	1.7586	0.5131	P_{mut}:*epoch*	-0.1028	0.0500	0.0400 .
fL	-0.0053	0.0482	0.9131	P_{co}:fL	0.0352	0.0694	0.6117
γ_{GA}	-0.1331	0.0201	0.0000***	P_{co}:γ_{GA}	-0.0200	0.0290	0.4910
δ	0.6246	0.6470	0.3344	P_{co}:δ	-1.3020	0.9434	0.1676
hid	-0.3029	0.1942	0.1187	P_{co}:*hid*	0.3694	0.3140	0.2395
epoch	-0.0169	0.0176	0.3365	P_{co}:*epoch*	0.0338	0.0289	0.2423
P_{im}:P_{mut}	-241.3283	6.1985	0.0000***	fL:γ_{GA}	-0.0015	0.0008	0.0578
P_{im}:P_{co}	-2.9659	3.5752	0.4068	fL:δ	0.0248	0.0253	0.3264
P_{im}:fL	-0.2741	0.0964	0.0045**	fL:*hid*	0.0021	0.0085	0.8062
P_{im}:γ_{GA}	1.0552	0.0405 .	0.0000	fL:*epoch*	0.0007	0.0008	0.3640
P_{im}:δ	2.0076	1.3233	0.1293	γ_{GA}:δ	-0.0139	0.0106	0.1910
P_{im}:*hid*	-0.1325	0.4436	0.7652	γ_{GA}:*hid*	0.0046	0.0035	0.1922
P_{im}:*epoch*	0.0796	0.0407	0.0506	γ_{GA}:*epoch*	0.0000	0.0003	0.9822
P_{mut}:P_{co}	1.0917	4.4441	0.8060	δ:*hid*	-0.0796	0.1158	0.4919
P_{mut}:fL	0.1270	0.1197	0.2890	δ:*epoch*	-0.0069	0.0106	0.5184
				hid:*epoch*	0.0026	0.0036	0.4718

Adjusted $R^2 = 0.61$ Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1

(a) OLS regression of d over the 10,000 point Monte Carlo sample.



(b) Estimated response surfaces of the OLS model (estimated on the 10,000 Monte Carlo sample), the other 5 parameters are set at the middle of their variation domains. P_{mut} on the x-axis, γ_{GA} on the y-axis, for three values of $P_{im} = \{0, 0.15, 0.25\}$.

Fig. 7 OLS regression model over the 10,000 Monte Carlo sample

factors):[19] the probability of imitation P_{im} (*probImit*), the probability of mutation P_{mut} (*probMut*) and the rate of application of the genetic algorithm γ_{GA} (*gaRate*), which modifies the population of firms' strategies.

We can again compare the results of the NOLH and kriging–based DoE to the ones obtained using a standard Monte-Carlo analysis using an OLS estimation of the determinants of the distance d. Figure 7a reports the results obtained with a sample of 10, 000 Monte Carlo simulations. The results show that the ANOVA-based sensitivity analysis of the meta-model is again able to catch the role of the main three factors: P_{im}, P_{mut} and γ_{GA}. We should note that the Monte-Carlo analysis becomes able to systematically catch the direct effects only after 3800 runs, while at least 4600 runs are necessary to catch also all two-ways interactions. Compared with the 660 simulations we run when using NOLH and kriging, Monte-Carlo sampling clearly appears more costly.

Convergence to the Cournot Equilibrium Graphics in Figure 6d give more details on the structure of the estimated response surface (D), based on the kriging meta-

[19] It should be noted that applying sensitivity analysis to the other forms of kriging identifies the same determinants, which indicates that the overall picture of the meta-model is not sensitive to the specification.

model according to the values of P_{im}, P_{mut} and γ_{GA}. The main insight is the primary role of the social dimension of learning (approximated by the rate of imitation) in the convergence towards the CE: in the absence of imitation, the model remains far from the CE (D equals almost 4 units in the best case, see left panel of Figure 5d), and mutation has to be very scarce for this distance to be minimized (recall that mutation of strategies arises every γ_{GA} periods, with a probability P_{mut} for each strategy). Convergence is much stronger ($D < 0.5$) when moderate imitation ($P_{im} = 0.15$) combines with moderate mutation ($P_{mut} \simeq 0.07$, and rather infrequent modification of the strategy population $\gamma_{GA} \simeq 20$, *see* the middle panel). Moderate rates of imitation and mutation balance the risk of premature convergence to a sub-optimal situation in terms of profits when imitation is frequent, and mutation is too rare to introduce enough innovative strategies among the population of firms. Too much imitation prevents firms from sufficiently using individual learning through their ANN, exposing to the risk of premature disappearance of the diversity of strategies, and hinders convergence towards the CE. It should be noted that the negative role of social learning in the convergence toward the CE within a Cournot oligopoly has been extensively discussed in the related literature (see, notably, Vallée and Yıldızoğlu (2009)).

In comparison, plotting the estimated value of the distance, D, based on the OLS regression estimated on the Monte Carlo sample fails to deliver these insights about the role of imitation (Figure 7b).

Optimal Configuration In a third stage, we determine the factor values that minimize the estimated value of the distance (D). Any optimization algorithm can be used to perform such task, but we draw the attention on the `rgenoud` R-project package (*R-GENetic Optimization Using Derivatives*, see Mebane and Sekhon (2011)), connected with the `DiceOptim` package [see (Roustant et al. 2010)]. It carries out a powerful optimization function that combines evolutionary algorithm methods for global purposes, with a gradient-based method for a local search of optima.[20] By using this algorithm, we are able to determine the optimal design of the learning algorithm that favors the convergence towards the CE:

$$P_{im}^* = 0.1454, hid^* = 3, fL^* = 9, epoch^* = 37, \delta^* = 0.8745,$$
$$\gamma_{GA}^* = 23, P_{mut}^* = 8\ \%, P_{co}^* = 5\ \%$$

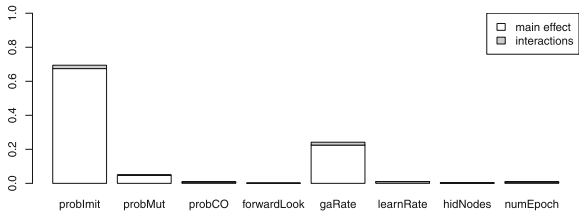
for a minimum value $D^* = 0.1165$, which is very small given that D measures the total distance over 30 firms. Test simulations performed with this optimal configuration actually report very small values of the distance, which proves that the kriging estimation is accurate and reliable.

Importance of Being Consistent In a last stage, we run a comparison exercise, in order to highlight the importance of selecting the sample points with the NOLH DoE in obtaining an accurate prediction of the response surface. Instead of using this DoE, we completely randomly draw 33 points over the experimental domain,

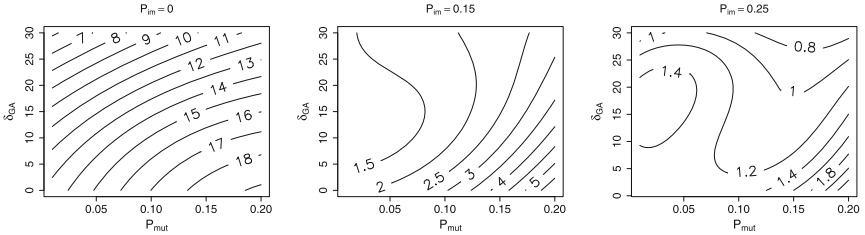
[20] See also Salle et al. (2012) for an application of this function to the minimization of a Central Bank's loss function in a macroeconomic agent-based model.

	cov. matèrn		Gaussian cov.		Exp. cov.	
	$\bar{\mu}$	1st order	$\bar{\mu}$	1st order	$\bar{\mu}$	1st order
RMSE	2.757	2.411	2.545	**2.134**	5.534	4.595
Q^2	0.852	0.9	0.901	0.921	0.547	0.772

(a) Comparison and selection of kriging models



(b) Sensitivity analysis of the selected meta-model



(c) Estimated response surfaces of the kriging meta-model, the other 5 parameters are set at the middle of their variation domains. P_{mut} on the x-axis, γ_{GA} on the y-axis, for three values of $P_{im} = \{0, 0.15, 0.25\}$.

Fig. 8 Robustness check: kriging meta-model based on a random 33 points DoE

and repeat the estimation procedure of the kriging meta-model based on this random sample. Figure 8a reports the predictivity coefficient values and the RMSE of external validation over a 7 points sample. The kriging meta-model with a Gaussian correlation function and a first order trend minimizes the RMSE, but the corresponding values are much higher than those previously obtained through the use of the NOLH DoE data, and they clearly indicate that the meta-model is less accurate. Figure 8a shows that, when based on a random DoE, the kriging meta-model fails to identity all significant effects on the response. In particular, the probability of mutation, P_{mut}, does not stand out, while two-third of the total variability of the distance d is attributed to the sole probability of imitation. Figure 8b shows the same response surfaces as in Figure 8c for the random DoE-based estimation. We observe a strong discrepancy between these two sets of graphics, especially for high levels of imitation. By optimizing the kriging meta-model using a random DoE, we obtain a negative minimal distance (-4.805), which is clearly not relevant (as the distance is computed as a sum of absolute terms). This exercise highlights the primary role of the choice of the correct DoE to obtain convincing results.

4 Concluding Remarks

This paper presents and illustrates a protocol proposed as an alternative to Monte Carlo exploration of computer simulations that involve many parameters, and a high computational cost. We give guidelines for the implementation of a parsimonious method for sampling the parameter space, and reliably predicting the response surface over the whole experimental domain. We show, using two example models, that such a parsimonious protocol can give very interesting results. In our first application, for example, we show that 10, 000 Monte-Carlo simulations are necessary to capture all the effects that we can capture, with the proposed method, using only 85 simulations. We also indicate how it can be combined for searching a specific configuration of the model that would comply with a given criteria (stylized facts, or optimal economic performance). Many agent based models in economics, management, and in other social sciences could hence benefit from such an approach.

Acknowledgments We are grateful to two anonymous referees for their comments and suggestions, and to the participants of the Lipari summer school on "Data mining and modelling of complex techno-socio-economic systems" (July 2012, Italy) for useful comments and discussions. We are responsible for all remaining errors.

Appendix

A DoE

See appendix Tables 3 and 4.

Table 3 DoE, Nelson and Winter (1982), $k = 6$ factors, $n = 17$ experiments

exp.	τ_{im} %	n	$\frac{\sigma^2}{g}$	ϵ	B	g %
1	0.125	32	12	1	1.5	1.5
2	0.0625	4	12	1	1	0.75
3	0.0625	8	4	1	2.5	1.25
4	0.125	16	8	1,00	2.5	0.5
5	0.25	32	8	0.8	2	0.25
6	0.05	4	8	1,00	1	1.25
7	0.25	4	12	1	3	0.75
8	0.25	32	12	1,00	3	1
9	0.25	8	8	1	2.5	1
10	0.25	2	4	1	3	0.25
11	0.5	16	4	1	3.5	1
12	0.5	8	12	1,00	2	0.5
13	0.25	8	8	0.8	2	1.25
14	0.125	2	8	1,00	2.5	1.5
15	0.0625	16	8	0.8	3.5	0.5
16	0.125	16	4	1	1.5	1
17	0.125	4	8	0.8	1.5	0.75

Table 4 DoE, oligopoly model with learning, $k = 8$ factors, $n = 33$ experiments

exp. DoE	δ	P_{im}	P_{co}	P_{mut}	fL	hid	γ_{GA}	$epoch$
1	1.00	0.02	0.20	0.05	11	3	21	34
2	0.91	0.25	0.09	0.08	6	2	23	29
3	0.88	0.11	0.37	0.04	0	3	22	21
4	0.58	0.22	0.40	0.09	11	2	25	22
5	0.94	0.01	0.21	0.05	8	3	13	37
6	0.97	0.23	0.16	0.06	5	2	6	46
7	0.70	0.12	0.39	0.06	0	3	12	47
8	0.55	0.17	0.38	0.08	11	3	7	50
9	0.67	0.06	0.13	0.11	9	3	1	26
10	0.76	0.16	0.15	0.14	3	3	4	31
11	0.73	0.05	0.31	0.19	4	2	5	25
12	0.79	0.18	0.28	0.19	9	4	15	32
13	0.61	0.04	0.12	0.12	7	2	29	43
14	0.85	0.15	0.18	0.18	2	3	28	42
15	0.64	0.05	0.35	0.18	5	2	20	43
16	0.82	0.16	0.26	0.20	10	4	17	40
17	0.53	0.13	0.23	0.11	6	3	16	35
18	0.05	0.23	0.25	0.16	2	3	10	36
19	0.14	0.00	0.36	0.13	6	4	8	41
20	0.17	0.14	0.08	0.17	12	3	9	49
21	0.47	0.03	0.05	0.12	1	4	6	48
22	0.11	0.24	0.24	0.16	4	3	18	33
23	0.08	0.02	0.29	0.15	7	4	25	24
24	0.35	0.13	0.06	0.15	12	3	19	23
25	0.50	0.08	0.07	0.13	1	4	24	20
26	0.38	0.19	0.32	0.10	3	3	30	44
27	0.29	0.09	0.30	0.07	9	3	27	39
28	0.32	0.20	0.14	0.02	8	4	26	45
29	0.26	0.07	0.17	0.02	3	2	16	38
30	0.44	0.21	0.33	0.09	5	4	2	27
31	0.20	0.10	0.27	0.03	10	3	3	28
32	0.41	0.20	0.10	0.03	8	4	11	28
33	0.23	0.09	0.19	0.01	2	2	14	30
Additional points for (external) validation								
1	0.75	0.21	0.10	0.06	7	3	5	39
2	0.86	0.15	0.35	0.03	6	2	19	42
3	0.21	0.12	0.15	0.07	10	2	13	25
4	0.52	0.06	0.18	0.09	3	3	16	46
5	0.41	0.09	0.27	0.04	9	4	23	36
6	0.30	0.19	0.31	0.06	2	3	26	28
7	0.63	0.04	0.23	0.02	4	4	9	31

R Codes for Kriging

B R codes for kriging

```
# X is in data.frame format, and contains the DoE (n=33 rows, k=8 columns), column names are probImit,
#probMut, probCO, forwardLook, gaRate, learnRate, hidNodes, numEpoch.
# XValid is in data.frame format, and contains the values of the factors at the additional points (n=7 rows,
#k=8 columns)
# y is in data.frame format, and contains the values of the response d at the 33 points of the DoE, averaged
 #over the 20 replications (column is named totDist, 33 rows)
# yValid is in data.frame format, and contains the values of the response at the 7 additional points (averaged
 #over the 20 replications ).
# DataVar is a column of n=33 rows, with contains the variance of the response d over the 20 replications of
 #each 33 experiments.
# Downloading kriging packages (see Roustant et al. (2010))
library(DiceKriging)
library(DiceEval)
library(DiceOptim)
library(rgenoud)

# Creating a function calculQ to compute the Q2 coefficient for any kriging model m:
calculQ <- function (m) { error <- (leaveOneOut.km(m, type='UK')$mean - y)^2
x <- 1
cumul <- 0
while (x < 34) {
cumul <- cumul + error[x,]
x <- x +1
}
cumul
devi <- (y - mean(y))^2
denom <- 0
i <- 1
while (i < 34) {
denom <- denom + devi[i,]
i <- i +1
}
Q2 <- 1 - (cumul / denom)
Q2
 }

# Estimating the 6 kriging models and corresponding Q2 with:
#mean (ordinary kriging) and matern 5/2 covariance:
m1 <- km(~ 1, design=X, response=y, covtype="matern5_2", noise.var=DataVar$totDist)
m1
calculQ(m1)
#a first-order polynomial trend and matern 5/2 covariance:
m2 <- km(~ ., design=X, response=y, noise.var=DataVar$totDist, covtype="matern5_2")
m2
calculQ(m2)
#mean (ordinary kriging) and gaussian covariance:
m3 <- km(~ 1, design=X, response=y, covtype="gauss", noise.var=DataVar$totDist)
m3
calculQ(m3)
#a first-order polynomial trend and gaussian covariance:
m4 <- km(~ ., design=X, response=y, noise.var=DataVar$totDist, covtype="gauss")
m4
calculQ(m4)
#mean (ordinary kriging) and exponential covariance:
m5 <- km(~ 1, design=X, response=y, covtype="exp", noise.var=DataVar$totDist)
m5
calculQ(m5)
#a first-order polynomial trend and exponential covariance:
m6 <- km(~ ., design=X, response=y, noise.var=DataVar$totDist, covtype="exp")
```

```
m6
calculQ(m6)

# Computing the associated RMSE of the 6 kriging models
test1 <- predict(m1, newdata=XValid, type="UK")
test2 <- predict(m2, newdata=XValid, type="UK")
test3 <- predict(m3, newdata=XValid, type="UK")
test4 <- predict(m4, newdata=XValid, type="UK")
test5 <- predict(m5, newdata=XValid, type="UK")
test6 <- predict(m6, newdata=XValid, type="UK")
RMSE1 <- RMSE(Valid$totDist, test1$mean)
RMSE2 <- RMSE(Valid$totDist, test2$mean)
RMSE3 <- RMSE(Valid$totDist, test3$mean)
RMSE4 <- RMSE(Valid$totDist, test4$mean)
RMSE5 <- RMSE(Valid$totDist, test5$mean)
RMSE6 <- RMSE(Valid$totDist, test6$mean)

# The selected model is m3 (ordinary kriging with Gaussian correlation), computing m3 sensitivity analysis:
library( sensitivity )
kriging.mean3 <-function(X,m3) predict.km(m3,X,"UK",se.compute=FALSE)$mean
SA.metamodel3 <-fast99(model=kriging.mean3,factors=c("probImit", "probMut", "probCO",
 "forwardLook", "gaRate", "learnRate", "hidNodes", "numEpoch"), q.arg=list(list(min=0,max=0.1),
 list(min=0.01, max=0.1), list(min=0.05, max=0.4), list(min=0, max=12), list(min=1, max=30),
 list(min=0.05, max=1), list(min=2, max=4), list(min=20, max=50)), m=m3)
plot(SA.metamodel3)

# Drawing the response surface of the kriging model m3, as a function of probMut and learnRate values)
n.grid <- 12
x.grid <- seq(0.01,0.1,length=n.grid)
y.grid <- seq(0.01,1,length=n.grid)
X.grid <- expand.grid(probImit=0.05,probMut=x.grid,probCo=0.4, fL=6, gaRate=15, learnRate=y.grid,
hidNodes=3, numEpoch=30)
pred.m3 <- predict(m3, X.grid, "UK")
contour(x.grid, y.grid, matrix(pred.m3$mean, n.grid, n.grid), 12, xlab=expression(prob[im]), ylab=
expression(delta), main="Kriging mean (OK)")

#optimizing the kriging model m3:
x_star <- max_EI(m3, lower=c(0,0.01,0.05, 0, 1, 0.01, 2, 20), upper=c(0.25,0.2, 0.4, 12, 30, 1, 4, 50),
control =list(pop.size=100, max.generations=50, wait.generations=50))
opt1 <- data.frame(x_star$par[1], x_star$par[2], x_star$par[3], x_star$par[4], x_star$par[5],
x_star$par[6], x_star$par[7], x_star$par[8])
opt1
pred.m3 <- predict(m3, opt1, "SK")
pred.m3$mean
```

References

Besag, J., & Clifford, P. (1991). Sequential monte carlo p-values. *Biometrika*, *78*(2), 301–304.

Booth, J., & Butler, R. (1999). An importance sampling algorithm for exact conditional tests in log-linear models. *Biometrika*, *86*(2), 321–332.

Box, G., & Draper, N. (1987). *Empirical model building and responses surfaces*. New York: Wiley.

Cary, N. (2010). *JMP®9 modeling and multivariate methods*. Cary: SAS Institute Inc.

Cioppa, T., (2002). *Efficient nearly orthogonal and space-filling experimental designs for high-dimensional complex models*. Naval postgraduate school: Doctoral Dissertation in philosophy in operations research.

Durrande, N., Ginsbourger, O., & Roustant, O. (2012). Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté des Sciences de Toulouse Tome*, *21*(3), 481–499.

Fang, K., Lin, D., Winker, P., & Zhang, Y. (2000). Uniform design: theory and application. *Technometrics*, *42*(3), 237–248.

Fisher, R. A. (1935). *The design of experiments* (9th ed.). New York: Macmillan.

Frey, C., & Patil, S. (2002). Identification and review of sensitivity analysis methods. *Risk Analysis*, *22*(3), 553.

Goupy, J., & Creighton, L. (2007). *Introduction to design of experiments with JMP examples* (3rd ed.). Cary: SAS Institute Inc.
Herbst, E., & Schorfheide, F. (2013). *Sequential monte carlo sampling for DSGE models, working paper 19152*. National Bureau of Economic Research.
Iman, R., & Helton, J. (1988). An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Analysis*, *8*, 71–90.
Jeong, S., Murayama, M., & Yamamoto, K. (2005). Efficient optimization design method using kriging model. *Journal of Aircraft*, *42*, 413–420.
Jourdan, A. (2005). Planification d'experiences numeriques. *Revue MODULAD*, *33*, 63–73.
Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, *52*(6), 119–139.
Masters, T. (1993). *Practical neural network recipes in C++*. New York: Academic Press.
Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*, 1246.
Mebane, W. J., & Sekhon, J. (2011). Genetic optimization using derivatives: the rgenoud package for R. *Journal of Statistical Software*, *42*(11), 1–26.
Miller, J., & Page, S. (2007). *Complex adaptive systems*. Princeton: Princeton University Press.
Nelson, R. R., & Winter, S. G. (1978). Forces generating and limiting concentration under schumpeterian competition. *Bell Journal of Economics*, *9*(2), 524–548.
Nelson, R. R., & Winter, S. G. (1982). The schumpeterian tradeoff revisited. *American Economic Review*, *72*(1), 114–132.
Oeffner, M. (2008). *Agent-based Keynesian macroeconomics—an evolutionary model embedded in an agent-based computer simulation*. Doctoral dissertation: Bayerische Julius - Maximilians Universitat, Wurzburg.
R Development Core Team. (2013). R: A language and environment for statistical computing, R Foundation for statistical computing, Vienna. ISBN 3-900051-07-0. http://www.R-project.org
Roustant, O., Ginsbourger, D., & Deville, Y. (2010). DiceKriging, diceOptim: two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, *55*(2), 100.
Sacks, J., Welch, W., Mitchell, T., & Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, *4*(4), 409.
Salle, I., Sénégas, M., & Yıldızoğlu, M. (2012). *How transparent should a Central Bank be? an ABM assessment*. avril: mimeo, Bordeaux University.
Saltelli, A., Tarantola, S., & Chan, K. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, *41*(1), 39–56.
Sanchez, S. M., (2005). Work smarter, not harder: Guidelines for designing simulation experiments, In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, & J. A. Joines (eds.) *Proceedings of the 2005 Winter Simulation Conference*. Software available at http://harvest.nps.edu/linkedfiles/nolhdesigns_v4.xls
Silva, I., Assuncao, R., & Costa, M. (2009). Power of the sequential monte carlo test. *Sequential Analysis*, *28*(2), 163–174.
Tesfatsion, L., & Judd, K. L. (Eds.). (2006). *Handbook of computational economics. Agent-based computational economics* (Vol. 2). Amsterdam: North-Holland.
Vallée, T., & Yıldızoğlu, M. (2009). Convergence in the finite Cournot oligopoly with social and individual learning. *Journal of Economic Behavior & Organization*, *72*(2), 670–690.
van Beers, W., & Kleijnen, J. (2004). Kriging interpolation in simulation: a survey. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), *Handbook of statistics*. New York: Elsevier.
Wang, G., & Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design*, *129*, 370.
Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., & Morris, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics*, *34*(1), 15–25.
Ye, K. (1998). Orthogonal column latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association*, *93*(444), 1430–1439.
Yıldızoğlu, M. (2001). Connecting adaptive behaviour and expectations in models of innovation: The potential role of artificial neural networks. *European Journal of Economic and Social Systems*, *15*(3), 51–65.
Yıldızoğlu, M., Sénégas, M.-A., & Zumpe, M. (2012). Learning the optimal buffer-stock consumption rule of Carroll. *Macroeconomic Dynamics*, *5*, 255.