Efficient Sampling and Metamodeling for Computational Economic Models

Isabelle Salle* & Murat Yıldızoğlu,*

June 20, 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 modelling to approximate the properties of computational models. This paper aims at providing guidelines for the use of a much more effective method, based on an efficient sampling of the parameters space – a *design of experiments (DOE)*, associated with a well-suited interpolation method – a *kriging* metamodel. 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. We analyze two simple economic models using this approach to illustrate the possibilities offered by it. Our appendix gives a sample of the R-project code that can be used to apply this method on other models.

Key-words – Computational Economics; Exploration of Agent-Based Models; Design of Experiments; Metamodeling.

JEL codes – C61; C63; C80; C90.

1 Introduction

Computational models have become a widely used tool in economic research, broadly named as *agent-based computational economics*. 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 works collected in Tesfatsion & Judd (2006). 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 other ones, 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 become prohibitive. For example, with 10 parameters, each of them having 5 potential levels, we need almost 10 million of simulations to cover all configurations, and even more if the model is not deterministic and involves replications. The common solution in economics is to use Monte Carlo simulations, *i.e.* to launch a high number of randomly drawn simulations (typically several thousands) to obtain a representative sample of the mapping between the parameters

*GREThA (UMR CNRS 5113), Université de Bordeaux, Avenue Léon Duguit, F-33608 PESSAC, Corresponding author: Murat Yıldızoğlu yildi@u-bordeaux4.fr

and the model's outcomes. 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 *design of experiments* (hereafter, DoE), coupled with an appropriate interpolation model (namely a *kriging metamodel*) 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.

The intuition behind DoE and kriging metamodeling is perhaps better introduced with an illustrative example. Suppose that a mining company aims at evaluating gold resources on a field to be potentially mined[1]. 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 here arise, 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, and the more sample data, the more precise the resources estimation. As evaluation of resources is costly, it is in the company's interest to minimize the sample size, while maximising 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 metamodels provide an interesting way to handle it, suppose that the rectangle in Figure 1 represents the field, and points 1 to 6 stand for the six sample locations where gold resources have been evaluated.

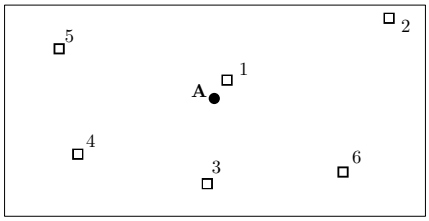


Figure 1: Design of experiments and kriging metamodeling: an illustrative introduction

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. DoE does allow to scatter a minimum number of 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 to 5, 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 metamodels is based on such a principle: values at

[1]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 twenties, for real, non-simulated experiments, see Fisher (1935).

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. 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 metamodel interpolates the model's behaviour 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 & 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 metamodelling to explore the properties of two standard economic models, namely the Nelson & 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 metamodel 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 & Winter (1982) model with only 17 parameter configurations (each being repeated 5 times to take into account the non-deterministic nature of the model), while those effects are only identified using a 10,000 simulation Monte Carlo exercise, and cannot be captured with only 1000. Importantly, the identified effects are in line with the ones discussed in Nelson & Winter (1982). A similar exercise within the Cournot oligopoly model shows that running only 33 parameter values is enough to catch the determinants of the convergence towards the Cournot equilibrium, while those determinants are identified with a 10,000 simulation Monte Carlo sampling. These results clearly highlight the computational gain associated with the use of DoE and kriging metamodeling, and show that no significant informational loss results. Second, we show how a kriging metamodel predicts the behaviour 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 metamodel, and provide the corresponding R Development Core Team (2009) 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 modelling exercise.

The rest of the paper is organized as follows. Section 2 presents the foundations of DoE and kriging metamodeling, Section 3 details the two applications of this method and Section 4 concludes.

2 Method

This section extensively presents the analytical backgrounds of the kriging-based metamodeling 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, while normal characters stand for one dimensional variable.

2.1 Preliminary definitions

Let x_1, ..., 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) $\mathbf{x}_i = (x_{1,i}, ..., x_{k,i})$ 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 $x_1 = x_{1,i}, ..., x_g = x_{g,i}, ..., x_k = x_{k,i}$). 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 $\mathbf{X} \equiv (\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_n)^T \in M_{n,k}(\mathbb{R})$. An experiment (or simulation) is a run of the model, with a particular parameter configuration $\mathbf{x}_i$, $i = 1, ..., n$. Finally, let $y : \mathbf{x} = (x_1, ..., x_k) \in D \in \mathbb{R}^k \rightarrow y(\mathbf{x})$ be the *response* variable.

DoE aims at choosing a minimal number n of points $\mathbf{x}_i$, $i = 1, ..., n$ in order to approximate the true model y by a kriging metamodel 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 metamodel, *i.e.* the choice of the form of Y (*see* Wang & Shan (2007)). Two main methods are available (*see* for example Jourdan (2005)): the first one relies on widely used OLS regressions and is outlined in Sub-section **??**; the second one is based on kriging interpolation and requires specific properties of the DoE. This is the subject of Sub-section **??**.

2.2 The classical approach

One can combine a classical DoE with the OLS estimation of a second-order polynomial model (possibly including second-order interactions):

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

where ε is a usual error term. Classical DoE are factorial type DoE (see for example Box & 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 Figure 2a). Consequently, they are not adapted if the response is irregular over the domain. We should use those 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 & 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

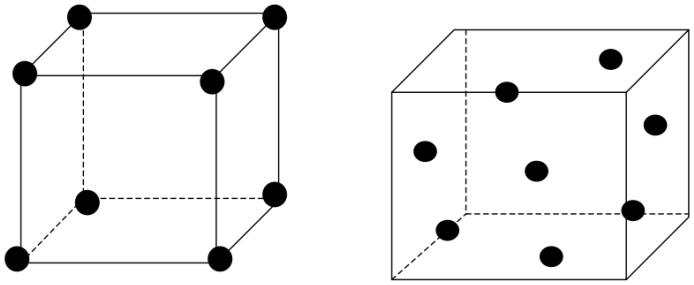
A more accurate metamodel Y over the entire experimental domain can be obtained through the use of a spatial interpolation model, namely a kriging model[2]. It is particularly relevant for the analysis of computer simulations, which produce non-linear dynamics and are highly sensitive to a small change in the parameter values. In that case, the response is not smooth over the experimental domain, and a metamodel accounting for that irregularity should be preferred to the classical approach.

2.3.1 Form of the metamodel

The response y can be predicted through the stochastic process Y (the metamodel):

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

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



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

Figure 2: Examples of DoE (3 factors, 8 points, source: Goupy & Creighton (2007))

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 and a vector $\beta \equiv \{\beta_j\}_{1,...l}$ of coefficients, to be estimated. Z is a stochastic process, representing local deviations of the model from the global trend μ (see Figure 3). The metamodel 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 variance 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 $corr\,(Z(\mathbf{x}_i), Z(\mathbf{x}_j))$. Kriging assumes that the closer 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. That 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:

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

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

Let $\theta \equiv \{\theta_1, ..., \theta_k\}$ be the $1 \times k$ vector of positive values quantifying the relative importance of the k factors on the response y. Element $\theta_g \geq 0$, $g = 1, ..., k$, 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 & 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 metamodel

Parameters to be estimated are the l coefficients β, the vector of k coefficients θ and σ^2. The kriging model estimation involves two steps (Sacks et al. (1989)).

First, the trend coefficients β are estimated using generalized least squares (GLS): $\hat{\beta} = (F' R^{-1} F)^{-1} F' R^{-1} y$. We define $F \equiv (f(\mathbf{x}_1)', ..., f(\mathbf{x}_n)')'$ as the so-called $n \times l$ experimental matrix and $f(x)$ is the vector of the trend values at some point x. Let $\Upsilon(x) \equiv Y(x) - f(x)'\hat{\beta}$ be the detrended process.

Second, residuals are interpolated, making abstraction of the trend. The best linear unbiased predictor (BLUP) of $\Upsilon(x)$ is obtained as a linear combination, with weights $\lambda_{\mathbf{x}} = \{\lambda(\mathbf{x}_1), ..., \lambda(\mathbf{x}_n)\}$, of the n observations of $\Upsilon(\mathbf{x})$ at the n points of the DoE, *i.e.* $\Upsilon(x) = \lambda'_{\mathbf{x}} \Upsilon(\mathbf{x})$. The mean squared error of the estimation is given by $MSE(x) \equiv E[Y(x) - f(x)'\hat{\beta} -$

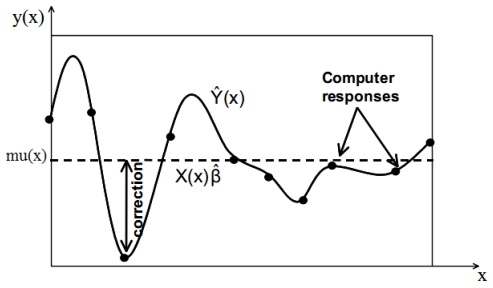


Figure 3: Ordinary kriging metamodel: $\mu(x) = \bar{\mu}$ (source: Jourdan (2005))

$\lambda_{\mathbf{x}}' \Upsilon(\mathbf{x}))^2]$. As the MSE is convex, the solution exists, is unique and is given by $\lambda_{\mathbf{x}}^* \equiv R^{-1} r(\mathbf{x})$, where $R = R(\mathbf{x}_i, \mathbf{x}_j)_{1 \leqslant i,j \leqslant n}$ is the correlation matrix of $Y(\mathbf{x})$ and $r(\mathbf{x}) = R(x, \mathbf{x}_i)_{1 \leqslant i \leqslant n}$ is the vector of the correlations between $Y(x)$ and $Y(\mathbf{x})$. By replacing the stochastic vector $\Upsilon(\mathbf{x})$ by its observed value $y(\mathbf{x}) - F\hat{\beta}$ and $\lambda_{\mathbf{x}}$ by the optimal weights $\lambda_{\mathbf{x}}^*$ in the expression $\lambda_{\mathbf{x}}' \Upsilon(\mathbf{x})$, one obtains the so-called mean prediction of y at any point $x \in D$:

$$E(Y(x)) = f'(x)\hat{\beta} + r(\mathbf{x})' R^{-1} [y(\mathbf{x}) - F\hat{\beta}] \tag{4}$$

Similarly, by plugging $\lambda_{\mathbf{x}}^*$ in the expression of the MSE, one obtains the mean squared error of the predictor (the so-called kriging variance):

$$s_Y(x) = \sigma^2 \left(1 - r(\mathbf{x})' R^{-1} r(\mathbf{x})\right) \tag{5}$$

$$+ \left(f(\mathbf{x})' - r(\mathbf{x})' R^{-1} F\right)' \times \left(F' R^{-1} F\right)^{-1} \left(f(x)' - r(\mathbf{x})' R^{-1} F\right) \tag{6}$$

σ^2 is estimated as follows:

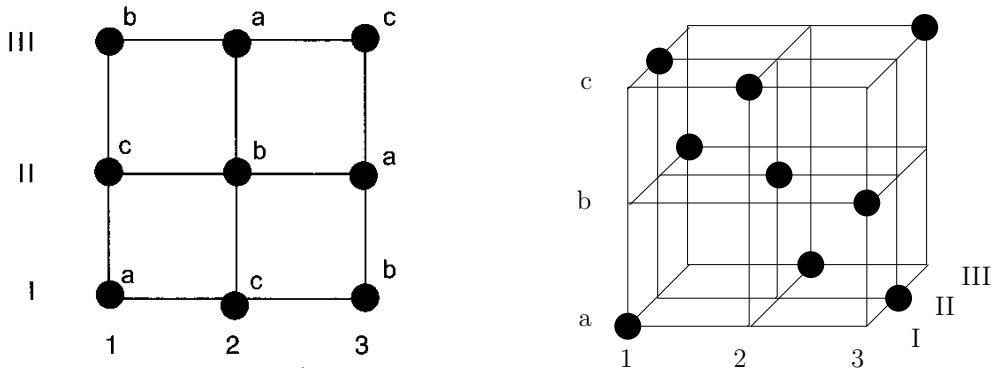
$$\hat{\sigma^2} = \frac{1}{n} (Y - F\hat{\beta})' R^{-1} (Y - F\hat{\beta}) \tag{7}$$

and parameters θ are mostly estimated using the maximum of likelihood, under Gaussian assumptions. Consequently, the following expression measures the quality of the model, where smaller values represent a better fit ($\mathbf{1}$ is a k-dimensional unit vector):

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

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

Properties of the DoE for the kriging approach In classical DoE, points are set on the extremities of the domain. That characteristic prevents the analysis from accurately estimating parameters θ 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, that is why we speak about *uniform designs* (see Figure 2b). That 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



(a) 2D-view (b) 3D-view

Figure 4: A latin cube, with 3 factors and 3 levels (source: Goupy & Creighton (2007))

factors and the response (Fang et al. (2000)). *Latin hypercubes* are often used to this end (see Goupy & Creighton (2007)). Figure 4 illustrates the way such a DoE is constructed in a case with 3 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 hence possible[3], and we randomly take one. The DoE involves 9 points, over the 27 initially to be tested. The DoE can be represented in 3 dimensions, we speak about a cube. When more than 3 dimensions are involved, we speak about an hypercube. Hypercubes ensure that the non-collapsing criteria is fulfilled: each combination of the values 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 & 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. That feature can create multicolinearity issues in the estimation of the metamodel. 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) develops an efficient trade-off between both properties, while limiting the size of the DoE. The orthogonality criteria 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 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.

Additional issues in the metamodel choice One can choose more complex forms of the correlation function R but a more sophisticated correlation function requires more observations to accurately estimate its parameters.

[3]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.

In practice, the trend μ is often reduced to a constant $\bar{\mu}$, which is thus interpreted as the mean[4] of the process Y. In that case, equations (4) and (5) are reduced to:

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

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 θ (Jourdan (2005)).

One may choose the form of the trend, or the correlation function, according to an optimality criterion (see Salle et al. (2012) for an application to an economic model). Either one can use cross-validation, or external validation.

Cross validation consists in removing one or several points of the DoE and reestimating the model, and then comparing the error between the estimations and the observed values at the removed points, based on the Q^2 *predictivity coefficient* (see Durrande et al. (2012)). That criterion is a proxy of the R^2 of standard linear regressions and is computed as[5]:

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

where $\bar{y}$ is the mean of y over the n observations, $\tilde{Y}(\mathbf{x}_i)$ is the predicted value of y at the sampled point $\mathbf{x}_i$, when the metamodel Y is estimated with the $n-1$ points $\mathbf{x}_{-i}$. Values close to zero denote a weak explanatory power, whereas values close to one indicate better fit.

External validation requires the estimation of the response at additional points, outside the DoE, and the comparison of the estimation error. The 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, all the more that the DoE involves a small number of points.

In the case of non-deterministic responses, the same experiment can yield different values of the response and experiments have to be replicated several times to get a significant evaluation of the response. The kriging model is then applied to the average value of the response over the number of replications (van Beers & Kleijnen (2004)). Let $\tilde{y}(x_i) = y(x_i) + \epsilon_i$ be the value of the response at point $x_i \in D$. We assume $\epsilon_i \hookrightarrow NID(0, \tau_i^2)$. As soon as the process Y and the errors ϵ_i are independent, the model can very easily be extended to non-deterministic responses. Matrix R in equations (4), (5) and (7), or in (9), is just replaced by $R + \Delta$, where $\Delta \equiv diag(\tau_1^2, ...\tau_n^2)$ (see for example Roustant et al. (2010)). The only difference is that the model has now two sources of errors, not only the error due to the difference between y and its metamodel Y, but also an experimental error, *i.e.* the noise in the response measures ϵ. As a consequence, the variance (5) will be higher than in the case of deterministic responses.

Purposes of the metamodel The kriging metamodel is mostly devoted to two purposes. Sensitivity analysis of the 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. A variance analysis (ANOVA) of the metamodel Y can be performed, as a proxy of the ANOVA of the true model y (see Jeong et al. (2005)). Formally, let $\mathcal{M}$ be the average of y over D and $\mathcal{V}$ be the variance of the model y as follows:

$$\mathcal{M} \equiv \int ... \int y(x) dx_1 ... dx_k \tag{11}$$

$$\mathcal{V} \equiv \int ... \int \left[y(x) - \mathcal{M} \right]^2 dx_1 ... dx_k \tag{12}$$

[4] We speak in that case of *ordinary kriging*, and of simple kriging if the mean is known, contrary to *universal kriging*, in the more general case, which is exposed above.

[5] Recall that kriging is an exact interpolator, that is why we cannot compute the R^2.

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

$$m(x_g) \equiv \int ... \int y(x) dx_1 ... dx_{g-1} dx_{g+1} ... dx_k - \mathcal{M} \tag{13}$$

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

$$m(x_g, x_h) \equiv \int ... \int y(x) dx_1 ... dx_{g-1} dx_{g+1} ... dx_{h-1} dx_{h+1} ... dx_k - m(x_g) - m(x_h) \tag{14}$$

and the total sensitivity of the response to factor x_g (and their interactions with the other factors) is given by $M(x_g) \equiv m(x_g) + \sum_{g \neq h} m(x_g, x_h)$.

The metamodel can also 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 provides two simple applications of the kriging-based metamodeling and hence, gives practical guidelines to the reader.

3 Applications

First, we perform an ANOVA of a very simple economic model to illustrate how kriging-based metamodeling can be used as a sensitivity analysis tool. To that purpose, we compare the results obtained with the design generated with the NOLH, with those of a Monte Carlo design. Second, we apply an optimization procedure to a kriging metamodel, to give an overview of the new results which can be obtained with that method.

3.1 Sensitivity analysis of Nelson & Winter (1982) model

3.1.1 A simple model of industry evolution

We consider the model of Nelson & Winter (1982, Chap. 12 to 14) and investigate, in that model, the factors that could lead to the emergence of concentration[6]. The industry is populated by n firms, indexed by $j = 1, ..., n$, 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. The inverse demand function is given by:

$$P_t = \frac{64}{Q_t} \tag{16}$$

where $Q_t \equiv \sum_{j=1}^{n} q_{j,t}$ is the aggregate supply. 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 variable cost factor per unit of capital, including R & D costs.

We consider here only the simple technological regime with science–based innovations. In that case, latent productivity increases at a constant exponential rate g. Firms can discover a more productive technique A_j either by innovation or by imitation. In each period, with a probability $P(innov = 1) = 0.0025 K_{j,t}$, a firm can innovate and draws a new technology $log(\tilde{A}_{j,t}) \hookrightarrow \mathcal{N}(A_0 + g.t, \sigma^2)$. With a probability $P(imit = 1) = \tau_{im} K_{j,t}$, a firm can also imitate the best technology in the industry $A_t^* = \max_{j=1,...n} A_{j,t}$. In $t+1$, firm j's technology is therefore given by the best technology over:

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

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

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

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

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 (its value in the original model is 1, see Equation (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 $\max\left[0, \min(\bar{I}_{j,t}, I^T_{j,t})\right]$.

We study the determinants of the response variable $h_t \in [0,1]$, the normalized Herfindhal index of capital concentration in the last period of the simulation, values close to zero denote a competitive framework and values close to one stand for a monopolistic industry.

3.1.2 Comparing Monte Carlo sampling *versus* the NOLH and kriging approach

Simulation protocols We follow Nelson & 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; $n \in [\![2, 32]\!]$; $\epsilon \in [0.8, 1000]$; $g \in [0.25, 1.5]\%$; $B \in [1, 3.5]$ and σ^2 is related to g with a factor from 4 to 12.

We consider two alternative methods to explore that experimental domain, and to determine the effect of those 6 factors on the response h. The first one involves 1000 Monte Carlo simulations[7], over which we adjust a polynomial regression model of the form (1) with two-interactions effects.

The second one implements the NOLH DoE that we have introduced in the previous section, over which we estimate a kriging model of h, denoted by H. We then perform an ANOVA of that kriging metamodel. As we need to discretize the experimental domain to generate the NOLH DoE, we follow the values investigated by Nelson & Winter (1978), and consider $\tau_{im} = \{0.000625; 0.00125; 0.0025; 0.005\}$, $n = \{2, 4, 8, 16, 32\}$, $\epsilon = \{0.8; 1; 1000\}$, $g \in [0.25, 1.5]\%$ 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 4. Following Nelson & 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)[8]. 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 model H as well as the plots of marginal and interaction effects, and Tables 1 and 2 depict the results of the polynomial regression models (for 1000 and 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 model.

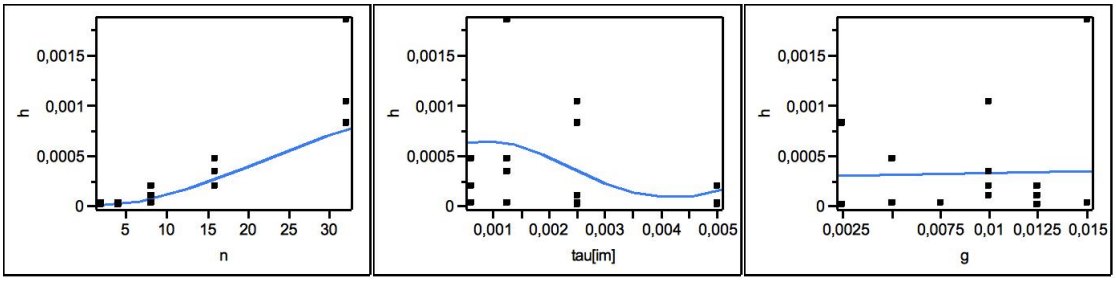
The overall picture is fairly the same: the number of firms n and the rate of imitation τ_{im} are the main determinants of concentration, while parameters B, ϵ and g do not significantly influence the structure of the industry (cf. Figures 5f, 5e, 5c). The more firms, or the less

[7] We also consider a 10 000 simulation Monte Carlo sample for robustness checks.

[8] R Development Core Team (2009) 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)).

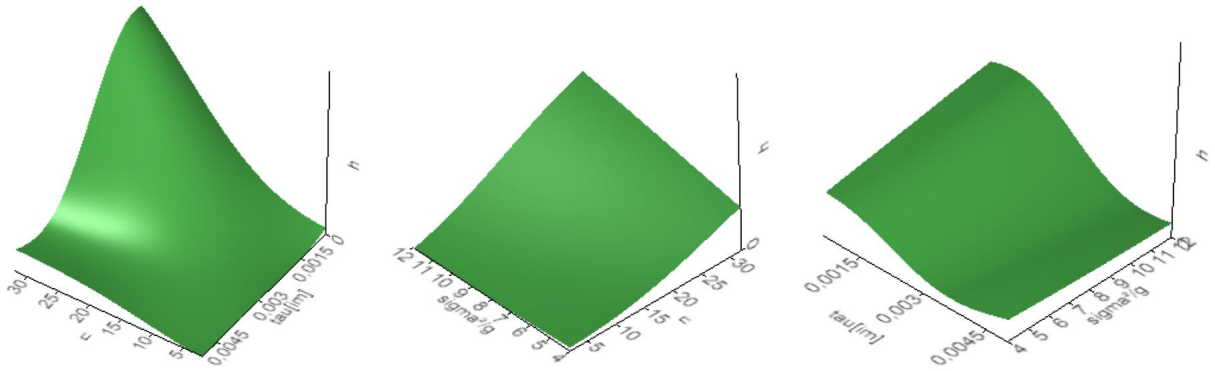
factors	θ	$M(.)$	$m(.)$	$m(., n)$	$m(., g)$	$m(., \frac{\sigma^2}{g})$	$m(., \tau_{im})$	$m(., B)$	$m(., \epsilon)$
n	0.0015	0.6402	0.4054	.	0.0000	0.0102	0.2246	0	0
g	14.3371	0.0006	0.0006	0.0000	.	0.0000	0.0000	0	0
$\frac{\sigma^2}{g}$	0.002	0.0341	0.0136	0.0102	0.0000	.	0.0103	0	0
τ_{im}	176794.16	0.566	0.331	0.2246	0.0000	0.0103	.	0	0
B	0	0	0	0	0	0	0	.	0
ϵ	0	0	0	0	0	0	0	0	.
$\bar{\mu} = E(H) = 0.005$			$\hat{\sigma^2} = 0.0000$			$-2 \ln Lik. = -246,6999$			

[1] For each of the 6 factors (by row), the second column reports the associated value of θ (see Equation (3)), the third one gives the total sensitivity, which fall into the main effect (fourth column, see Equation (13)) and the two-interaction effects with the others factors (all remaining columns, see Equation (14)).



(a) $m(n)$ (b) $m(\tau_{im})$ (c) $m(g)$

(d) $m(\frac{\sigma^2}{g})$ (e) $m(\epsilon)$ (f) $m(B)$



(g) $m(n, \tau_{im})$ (h) $m(\frac{\sigma^2}{g}, n)$ (i) $m(\tau_{im}, \frac{\sigma^2}{g})$

Figure 5: ANOVA table of the kriging model H.

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

Table 1: OLS regression of h over the 1000 points Monte Carlo sample.

	Estimate	Std. Error	Pr(> |t|)		Estimate	Std. Error	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
Adjusted $R^2 = 0.996$				Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1			

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

frequent imitation, the more concentrated the industry (cf. Figures 5a and 5b). As the size of the market is fixed (see Equation (16)), the selective pressure is strengthened and the decrease in price is faster, *ceteris paribus*, as the number of firms n increases. That mechanism intuitively explains the salient role of n. Moreover, the interaction term between n and τ_{im} is highly significant (see Figure 5g): imitation affects the industry all the more that n is large, and in that case, scarce imitation leads to a high degree of concentration. Intuitively, it means that among a lot of firms, a firm is more likely to gain a striking competitive edge, all the more that it cannot be imitated easily. When the number of firms increase, the selective pressure on lagging firms increases, and the imitation becomes the major tool for decreasing this pressure, by catching up with the technological leader. Using the $10,000$ simulation sample allows to identify further effects, which are not identified with a 1000 simulation sample but are highlighted by the ANOVA of the kriging model. Nevertheless, the degree of concentration is affected in a less sizeable way, compared to the effect of parameters n and τ_{imit}. The variability of research outcomes (σ^2/g) and the growth rate of latent productivity g have a weak positive effect on concentration (see Figures 5d and 5h): concentration tends

	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

Table 3: Comparison and selection of kriging models.

to emerge when research outcomes are strongly dispersed, all the more that there is a lot of firms. It should be noted that the kriging analysis underlines the individual effects of g and σ^2/g (see Figures 5c and 5d), as well as the interactions of σ^2/g with n and τ_{imit} (see Figures 5h and 5i), while the LS model based on $10,000$ data only significantly reports the interactions of g with n and τ_{imit}. Despite that minor discrepancy, the two models deliver the same message, and highlight the role of innovation draws : concentration is higher if innovations are drawn in a wide range, all the more that imitation is rare and firms are numerous. Here again, in such a context, a firm is more likely to gain a competitive edge. While that result is intuitively appealing, it should be noted that the effects are quite small. The possibility of catching the main effects of the parameters, and their interactions, with only 85 simulations, instead of 1000 or $10,000$ clearly show the frugality of the approach proposed in this article.

Note finally that those results are completely consistent with those of Nelson & Winter (1978, 1982).

3.2 Kriging-based optimization

3.2.1 A baseline oligopoly model

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

$$P(q_j, Q_{-j}) = 256 - 2(q_j + Q_{-j}) \tag{21}$$

$$C(q_j) = 56q_j + q_j^2 \tag{22}$$

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)[9]. Each firm is endowed with a one-hidden layer ANN with $hid > 1$ hidden nodes, which is fed for each period with 4 inputs (price evolution, individual sales evolution, variation of individual costs and evolution of individual profit), and each firm has a population of 40 quantity strategies. For each period, each firm selects the strategy among that 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' ANN 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 *gaRate* periods, the population of strategies is modified by operators selection, mutation (with a probability P_{mut} every γ_{GA} periods) and average crossover (with a probability P_{co}).

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

In that model, we investigate the design of the learning algorithm which allows the industry to converge towards CE. We have $k = 8$ factors depicting firms' learning, with the associated variation domains: $P_{im} \in [0, 0.25]$, $hid \in [\![2, 4]\!]$, $fL \in [\![0, 12]\!]$, $epoch \in [\![20, 50]\!]$, $\delta \in [0.05, 1]$, $\gamma_{GA} \in [\![1, 30]\!]$, $P_{mut} \in [0.01, 0.2]$ and $P_{co} \in [0.05, 0.4]$. The response variable is the absolute distance of aggregate supply to CE, $d \equiv \mid \sum_{j=1}^{n} q_j - nq^c \mid$, and the kriging-based approximation is denoted by D. We aim at determining the configuration(s) of the factors, for which that distance is minimized. We sample the 8-dimensional parameters space with Sanchez (2005) DoE given in Table 5, which defines $n = 33$ non-deterministic experiments, each is repeated 20 times and we apply kriging over the average response in each experiment. Efficient sampling is especially useful for models involving algorithms such as ANN, which are very time consuming to run.

3.2.2 Optimization of the kriging metamodel

We perform all the analysis with R Development Core Team (2009) software. The corresponding code is provided in Appendix B.

The first stage is to choose the form of the kriging model. We compare three forms of correlation functions (Gaussian, exponential and Matèrn $\nu = 5/2$, which is the default function in the package `DiceKriging`, see Roustant et al. (2010)), and two specifications of the trend μ, a constant and a first-order polynomial [10]. To that purpose, we use both external validation and cross-validation. For external validation, we evaluate the model at 7 additional experimental points, which we randomly choose over the whole experimental domain (see Table 5). Table 3 reports the result of the comparison. Both criteria broadly set the models in the same order, but we have only few points ($n = 33$), that is why we rather rely on external validation. Accordingly, we retain the form of the kriging metamodel which minimizes the RMSE between the predicted response and the effectively measured one at the 7 additional points. We therefore select the ordinary kriging model (*i.e.* in which the trend μ is only a constant term) with the Gaussian correlation function.

Table 6a then gives the estimations of the coefficients of the selected model. Figure 6b 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[11]: 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. Figures 6c displays the estimated response D according to the values of those factors. The main insight is the primary role of social learning in the convergence towards CE: in the absence of imitation, the model remains far from CE (D equals almost 4 units) and mutation has to be very scarce (recall that mutation of strategies arises every γ_{GA} period, with a probability P_{mut} for each strategy). Convergence is much better ($D < 0.5$) with a combination of moderate imitation ($P_{im} = 0.15$) and moderate mutation ($P_{mut} \simeq 0.07$, $\gamma_{GA} \simeq 20$). That configuration of learning allows efficient exploration and diffusion of interesting strategies in terms of profit, and favors the convergence towards CE, which delivers a higher collective profit. Nevertheless, too much imitation prevents firms from sufficiently using individual learning through their ANN, and hinders convergence towards CE. Note that the negative role of social learning in the convergence of the Cournot oligopoly has been extensively discussed in the literature (see notably Vallée & Yıldızoğlu (2009)).

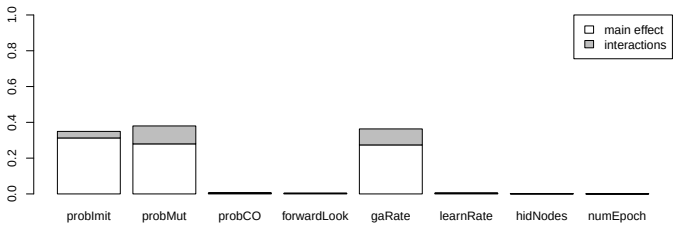
In a last stage, we determine the factors configuration $(P_{im}^*, hid^*, fL^*, epoch^*, \delta^*, \gamma_{GA}^*, P_{mut}^*, P_{co}^*)$ which minimizes the estimated value of the distance D, denoted by D^*. Any optimization algorithms can be used, but we draw the attention on the package `rgenoud` (*R-GENetic Optimization Using Derivatives*, see Mebane & Sekhon (2011)), connected with `DiceOptim` (see Roustant et al. (2010)) provided by R Development Core Team (2009). That carries out a

[10] Higher order polynomials would involve too many parameters to estimate, considering only 33 observations.

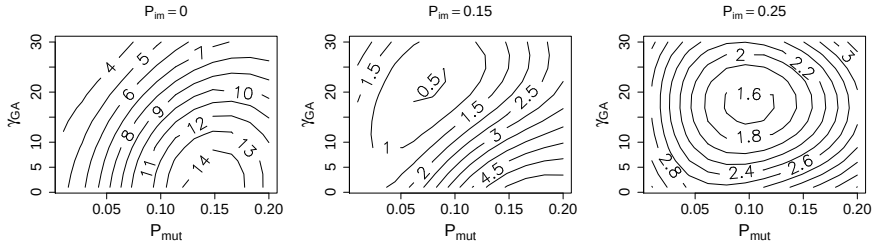
[11] Note that applying sensitivity analysis to the other forms of kriging identifies the same determinants, which indicates that the overall picture of the metamodel is not sensitive to the specification.

$E(D) = \bar{\mu}$	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

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



(b) Sensitivity analysis of the selected metamodel



(c) Estimated response surfaces of the kriging 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 $P_{im} = \{0, 0.15, 0.25\}$.

Figure 6: Kriging model D.

quite powerful optimization function that efficiently combines evolutionary algorithm methods for global purpose with a derivative-based method for local search of optima (see also Salle et al. (2012) for an application of that function to the minimization of a Central Bank's loss function in a macroeconomic agent-based model). With that algorithm, we obtain the optimal design of the learning algorithm to converge towards CE:

$$(P^*_{im}, hid^*, fL^*, epoch^*, \delta^*, \gamma^*_{GA}, P^*_{mut}, P^*_{co}) = (0.1454, 3, 9, 37, 0.8745, 23, 0.08, 0.05)$$

for a minimum value $D^* = 0.1165$, which is very small given that D measures the total distance for 30 firms. Simulations performed with that optimal configuration effectively report very small values of the distance, which proves that the kriging estimation is accurate.

4 Concluding remarks

This paper presents and illustrates an alternative to Monte Carlo exploration of computer simulation models involving many parameters and a high computational cost. We give guidelines for the implementation of an efficient and time-saving method for sampling the parameters space and optimally predicting the response over the whole experimental domain. We show, using two example frameworks, that such a parsimonious model can give very interesting results. Many AMBs in economics and management could hence benefit from such an approach.

References

Box, G. & Draper, N. (1987), *Empirical Model Building and Responses Surfaces*, Wiley. New York.

Cary, N. (2010), *JMP®9 Modeling and Multivariate Methods*, SAS Institute Inc.

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

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

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

Fisher, R. A. (1935), *The Design of Experiments*, 9 edn, Macmillan.

Goupy, J. & Creighton, L. (2007), *Introduction to Design of Experiments with JMP Examples (Third Edition)*, third edn, SAS Institute Inc., Cary, NC, USA.

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**, pp. 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++*, Academic Press, New York.

Matheron, G. (1963), 'Principles of geostatistics', *Economic Geology* **58**.

Mebane, W. J. & Sekhon, J. (2011), 'Genetic Optimization Using Derivatives: The rgenoud Package for R', *Journal of Statistical Software* **42**(11), pp. 1–26.

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), 114–32.

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 (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *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* **VV**(II).

Sacks, J., Welch, W., Mitchell, T. & Wynn, H. (1989), 'Design and analysis of computer experiments', *Statistical Science* **4**(4).

Salle, I., Sénégas, M. & Yıldızoğlu, M. (2012), How Transparent Should a Central Bank be? an ABM assessment. mimeo, Bordeaux University, avril.

Sanchez, S. (2005), Nolh designs spreadsheet . Software available online via http://diana.cs.nps.navy.mil/SeedLab/.

Tesfatsion, L. & Judd, K. L., eds (2006), *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*, North-Holland, Amsterdam.

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*, R .G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds, chapter Computer experiments. In Handbook of Statistics.

Wang, G. & Shan, S. (2007), 'Review of Metamodeling Techniques in Support of Engineering Design Optimization', *Journal of Mechanical Design* **129**.

Ye, K. (1998), 'Orthogonal column latin hypercubes and their application in computer experiments', *Journal of the American Statistical Association* **93**(444), pp. 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., Salle, I. & Zumpe, M. (2012), 'Learning the optimal buffer-stock consumption rule of Carroll', *Macroeconomic Dynamics* p. forthcoming.

A DoE

exp.	$\tau_{im}\%$	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	1000	2.5	0.5
5	0.25	32	8	0.8	2	0.25
6	0.05	4	8	1000	1	1.25
7	0.25	4	12	1	3	0.75
8	0.25	32	12	1000	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	1000	2	0.5
13	0.25	8	8	0.8	2	1.25
14	0.125	2	8	1000	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, Nelson & Winter (1982), $k = 6$ factors, $n = 17$ experiments

exp.	δ	P_{im}	P_{co}	P_{mut}	fL	hid	γ_{GA}	$epoch$
				DoE				
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

Table 5: DoE, oligopoly model with learning, $k = 8$ factors, $n = 33$ experiments

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
```