



*Department of Economics & Finance*

*Chair of Econometric Theory*

# Indirect Inference applied to Financial Econometrics

SUPERVISOR :

Prof. Santucci de Magistris Paolo

CANDIDATE :

Duchesne Jean, 700531

CO-SUPERVISOR :

Prof. Rossetti Claudio

ACADEMIC YEAR 2017-2018

# INDIRECT INFERENCE APPLIED TO FINANCIAL ECONOMETRICS

JEAN DUCHESNE

*Dedicated to Professor Thomas Demuyne, Professor Paolo Santucci de Magistris and my  
parents for their help, guidance and support.*

ABSTRACT. Indirect inference is a simulation-based method for estimating or making inferences on the parameters of economic models. It is very useful for estimating models for which the likelihood function (or any other criterion function that could serve as the basis for the estimate) is analytically insoluble or too difficult to assess. Such models abound in the financial and economic analysis and include non-linear dynamic models, latent (or unobserved) models and models with missing or incomplete data. This thesis exposes the methodology behind indirect inference procedure, gives two examples and finally an empirical application.

## 1. INTRODUCTION

Most of modern researchers are, at one time or another, confronted to the task of building generative models of a process or assemblage on which they work in the hope to grasp a substantial bite of the nature's knowledge. Being careful scientists, they usually do a conscientious job of trying to include their guesses for what are all the most important mechanisms. The result is something that can possibly be set through to produce a simulation of the process of interest. But, (realistic) models often contain some unknown parameters, some hidden truth that we will denominate under the vector  $\theta$ .

It is desirable to tune those unknown features within models to match the data or see if, despite researchers' best efforts, there are aspects of the data which a model just can't match.

Very often, models too loyally built are too complicated for the common practitioner to appeal to any of the usual estimation methods of statistics<sup>1</sup>. Because some models aim for scientific adequacy rather than statistical tractability, it will often happen that there is no way to even calculate the likelihood of a given data set

---

*Date:* June 15, 2018.

*Key words and phrases.* Indirect Inference, Stochastic volatility.

<sup>1</sup>Typically, in finance the underlying mathematical models have become more complex. They take more and more parameters to adapt to the data. We can mention, for example, as phenomena to be taken into account: dependence of volatility on the past, non-linear dynamics, variance, asymmetry to identify the permanent and transient shocks, unobservable factors, leptokurtic effect, etc.

under the unknown parameters in closed form, which would rule out even numerical likelihood maximization.

Indeed, in the over average complicated models, such as stochastic volatility models in continuous time, ARCH factor models, nonlinear random parameter models, the likelihood function is impossible to calculate. Procedures for estimating parameters were then born. We can quote for example, the general method of moments (GMM), the quasi-maximum likelihood (QML) method, Bayesian methods, simulated expectation maximization (SEM) method, variance reduction (Importance Sampling, IS). The results of these different methods are quite diverse. Indeed, the IS method is complicated to implement, GMM and QML methods are not efficient. The SEM and Bayesian methods provide better results.

More generally there has been a division of opinion in the academies between frequentist and Bayesian statisticians regarding the efficiency of the two main inference methods: General Method of Moments and Maximum Likelihood Estimation. Frequentists dismissed the method-of-moments in favor of the MLE, Bayesians never did so. Anyway, both sides have always thoroughly and intentionally focused on the likelihood (frequentists on the location of its maximum and its curvature while Bayesians focus on its entire shape). As sustained by F.Diebold on his blog<sup>2</sup>, this division is what drives the European views regarding estimation methods. Especially American econometricians hold the GMM in high esteem, to which they attribute almost sacred merits, this is not the case of European academicians who found the hype about GMM exaggerated. This is mainly due to how GMM is advertised as potentially useful when there is a likelihood at hand, in other situation the method-of-moments loses its power. Even worse, model moments may also be analytically intractable.

Yet simulation is possible; it seems like there should be some way of saying whether the simulations look like the data. This is where the breaking point comes in. The solution could first come out as an implementation of GMM by simulation: Simulated Method of Moments. By simulating models, in combination with the fine choice of parameters minimizing divergence between simulated and data moments, one can consistently estimate those models' parameters. That is really game-changing: we no longer need to work-out complex likelihoods (even for those that are available in some terms).

But SSM is a peculiar case from of a more general method: Indirect Inference, which itself can be seen as a generalized form of the method of moments of Duffie and Singleton (1993) and has papers of Tony Smith<sup>3</sup> and C.Gourieroux, A.Monfort and E.Renault (1993, J. Applied Econometrics) as founding articles. It introduces a new model, called the "auxiliary model", which is miss-specified and typically not even generative, but is easily fit to the data, and to the data alone. The auxiliary model has its own parameter vector  $\beta$ , with an estimator  $\hat{\beta}$ . These parameters describe aspects of the distribution of observables, and the idea of indirect inference is that we can estimate the generative parameters  $\theta$  by trying to match those aspects of observations. It further bases itself on a metric-set wisely chosen to minimize the

<sup>2</sup>*No Hesitations*, 22 July 2013,

<sup>3</sup>1990, Duke Ph.D. Dissertation, and 1993, J. Applied Econometrics

distance between the auxiliary estimators based on the observed data on one hand and on the other hand the auxiliary estimators of the simulated data.<sup>4</sup> Indirect Inference assumes a relatively easy simulation of the models studied. It allows, from a higher calculation time, a reduction of the bias compared to more traditional methods.

Unlike other methods, the moments that guide the estimation of the parameters of the economic model are themselves the parameters of the auxiliary model. If the auxiliary model comes close to providing a correct statistical description of the structural model, then indirect inference comes close to matching the asymptotic efficiency of maximum likelihood.

Its main advantages lie in its generality. Unlike other bias reduction methods, such as those based on explicit analytical expressions for the bias function or key terms in an asymptotic expansion of the bias, the indirect inference technique calibrates the bias function by simulation and therefore does not require a given explicit form for the bias function or its expansion.

Some sustain that evaluation of the likelihood became as trivial as simulating. As Andrew Harvey and others have emphasized for decades, for any linear model cast in finite-dimensional state-space form one can simply run the Kalman filter and then evaluate the Gaussian likelihood via a prediction-error decomposition. These tools can provide complete likelihood analysis in general non-Gaussian environments. But II remains a more suitable path for cases where model understanding is of interest and is clearer in its implementation. Moreover, II achieves desirable consistency properties under misspecification more easily (as exposed in Rossi and De Magistris 2018). Finally, even-though the more classical approaches like quasi-maximum likelihood method or Bayesian models deliver more satisfying results in statistical terms, there are areas where they simply cannot be used or at least not properly.

For many decades now, interest in analysis and modeling in finance and economics has been growing. Financial data is growing. The emergence of computer tools has made its further development and exploitation possible. The rising of the average computational capacity, following the Moore's law of computing power<sup>5</sup>, has made simulation-based procedures even more competitive than before.

The aim of this thesis is to present the II methodology and implementation (section 2: Indirect Inference, Theoretical aspects) and to discuss it around two examples (that are moving average and Heston-model, in section 3: Examples), followed by an attempt to apply this method on the S&P500 Index (section 4: Application). In this paper we will discover the recipe for an II implementation, the threats to its robustness and solutions to tackle them.

---

<sup>4</sup>The auxiliary estimator may be the Maximum Likelihood Estimator of the auxiliary model (usually simpler), or the estimator corresponding to the approximate likelihood of the initial model.

<sup>5</sup>Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years. The observation is named after Gordon Moore, the co-founder of Fairchild Semiconductor and Intel, whose 1965 paper described a doubling every year in the number of components per integrated circuit ("Moore's Law." Wikipedia, Wikimedia Foundation, 12 June 2018, en.wikipedia.org/wiki/Moore's\_law.)

## 2. INDIRECT INFERENCE, THEORETICAL ASPECTS

## 2.1. Methodology.

The methodology has been extensively developed by Gouriéroux and Montfort, (1992). For the sake of simplicity, only the most important aspects will be exposed, but be sure that you can find proofs and further asymptotic analysis of the method in the founding paper of Gouriéroux and Montfort.

Let (P) be the problem whose parameters we seek to estimate, written in the following form:

$$(2.1) \quad (P) = \begin{cases} y_t = r(y_{t-1}, x_t, u_t, \theta) \\ u_t = \rho(u_{t-1}, \varepsilon_t, \theta), \end{cases}$$

Where  $x_t$  are the observations of exogenous variables, while  $u_t$  and  $\varepsilon_t$  are not observed. Finally  $\theta$  is the parameter we are trying to estimate. His real value will be noted  $\theta_0$ . For example,  $\theta$  could be simple sample means or moments, or regression coefficients, or more generally parameters from some sort of auxiliary model. We will estimate a  $\beta$  by matching  $\hat{\theta}_T$  to  $\hat{\theta}_S^\beta$

Suppose the following hypothesis (H1),

- $\{x_t\}$  is an homogeneous Markov process;
- the process  $\{\varepsilon_t\}$  is a white noise with a known distribution  $G_0$ ;
- the process  $\{y_t, x_t\}$  is stationary.

Let  $f_0(x_t|x^{t-1})$  be the Conditional Density Function. Due to the markovian feature of the process  $\{x_t\}$  it can be written that  $f_0(x_t|x^{t-1}) = f_0(x_t|x_{t-1})$ , with  $x^{t-1} = (x_{t-1}, x_{t-2}, \dots)$ .

By increasing the size of the process, it is possible to take into account higher order processes or processes of reduced form.

$x_t$  being an homogeneous Markov process implies that  $x_t$  is a strongly exogenous process. This means we assume that potential problems due to non-strongly exogenous variables have been solved by considering them as functions of lagged endogenous variables. It is also worth noting that models in which non-strongly exogenous variables appear have the drawback of not being simulable.

## 2.1.1. Problematic.

It is theoretically possible to determine the value of the structural parameters  $\{\theta_1, \theta_2, \dots, \theta_m\}$ , (P), using a Maximum Likelihood Estimator approach. Unfortunately in practice the calculation is very often impossible, so it is necessary to circumvent this impossibility. This can be done numerically using the method studied in this report, namely the Indirect Inference method. In order to calculate the



Finally let's also introduce the mapping/binding function from the  $\theta$ 's to the  $\beta$ 's defined by  $b(F, G, \theta)$ .

Should also be sustained the following additional hypotheses:

- H(2)  $\lim_{T \rightarrow \infty} Q_T(y_t^1, x_t^1, \beta) = Q_\infty(F_0, G_0, \theta_0, \beta)$  the function  $Q_\infty$  being non-stochastic, continuous in  $\beta$  with a unique maximum  $\beta_0$ ;
- H(3) the function  $b(F, G, \theta) = \arg \max_{\beta \in B} Q_\infty(F, G, \theta, \beta)$  is the biding function. Therefore:  $\beta_0 = b(F_0, G_0, \theta_0)$ ;
- H(4)  $b(F_0, G_0, \cdot)$  is calculable term by term and its derivative with respect to the parameter  $\theta$  is of full rank.

Assumptions H(2) to H(4) guarantee that the parameter vector is locally identified when the distance tends to zero. H(2) assumes that the criterion tends asymptotically (and uniformly almost certainly) to a non-stochastic limit. Limit which depends on the unknown auxiliary parameter  $\beta$ , on the characteristics of the true distribution (i.e. the transition distribution  $F_0$  of  $\{x_t\}$ , which is unknown or not totally known at maximum), on the marginal distribution  $G_0$  of  $\{\varepsilon_t\}$  (which is known), and on the true parameter of interest  $\theta_0$  and maybe also on the initial value  $z_0$ . H(3) assumes that the initial conditions have no asymptotic effects. To have global identification  $b(F, G, \theta)$  must be a one-to-one (injective) function, so that it exists only one element of  $\hat{\theta}$  for each respective elements of the true-parameter vector,  $\theta_0$ .

We know from the asymptotic theory exposed by Gallant and White (1998, Chapter 3) that under these assumptions the estimator  $\hat{\beta}_T$  is a consistent estimator of the auxiliary parameter  $\beta_0$ .

Indirect Inference consists of the two following steps:

- (1) The calculation of  $\hat{\beta}_T = \arg \max_{\beta \in B} Q_T(y_T^1, x_T^1, \beta)$  from the observations.  $\hat{\beta}_T$  is an estimator consistent of the auxiliary parameter  $\beta_0$  under the assumptions made previously.
- (2) (a) Simulation of S trajectories  $[\tilde{y}_t^s(\theta, z_0^s), t = 0, \dots, T]; s, \dots, S$  based on independent draws from  $\varepsilon_t$  and on initial values  $z_0^s; s = 1, \dots, S$ .  
 (b) Maximization for each of these trajectories of the criterion function, let us note  $\tilde{\beta}_T^s(\theta, z_0^s) = \arg \max_{\beta \in B} Q_T((\tilde{y}^h)_T^1, x_T^1, \beta)$   
 (c) Resolution of the following minimization problem to obtain an indirect estimator of  $\theta$

$$\min_{\theta \in \Theta} \left[ \hat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \tilde{\beta}_T^s(\theta, z_0^s) \right]' \hat{\Omega}_T \left[ \hat{\beta}_T - \frac{1}{S} \sum_{s=1}^S \tilde{\beta}_T^s(\theta, z_0^s) \right]$$

The minimum, noted  $\tilde{\theta}_T^S(\Omega)$ , is a consistent estimator of  $\theta_0$  under the four previous assumptions.

Note that the second step requires S optimizations for each value of  $\theta$ . It is possible to compile this sequence of optimization to only one.

Indeed, let us consider TS values of  $x$  (exogenous variables) obtained by repeating  $S$  times the values  $x_1, \dots, x_T$ , i.e.  $\tilde{x}_1 = x_1, \dots, \tilde{x}_T = x_T, \tilde{x} = x_1, \dots, \tilde{x}_T = x_T$ . Then it is deduced that

$$\begin{aligned} & \tilde{y}_t(\theta, z_0), t = 0, \dots, TS \\ \text{with } & \begin{cases} \tilde{y}_0(\theta, z_0) = y_0 \\ \tilde{y}_t(\theta, z_0) = r[\tilde{y}_{t-1}(\theta, z_0), \tilde{x}_t, \tilde{u}_t(\theta, z_0), \theta] \\ \tilde{u}_t(\theta, u_0) = \rho(\tilde{u}_{t-1}(\theta, u_0), \tilde{\varepsilon}_t, \theta), \end{cases} \end{aligned}$$

Following this, the second step can be re-written:

- (2) (a) Computation of the  $\tilde{y}_1^s, \dots, \tilde{y}_T^s$  from the initial values

$$z_0, z_T = (y_T, \tilde{u}_T), \dots, z_{T(S-1)} = (y_{T(S-1)}, \tilde{u}_{T(S-1)})$$

- (b) From those simulations, we determine

$$\tilde{\beta}_{TS}(\theta, z_0) = \arg \max_{\beta \in B} Q_T(\tilde{y}_{TS}^1, \tilde{x}_{TS}^1, \beta)$$

- (c) Resolution of the following minimization problem to get the indirect estimator of  $\theta$

$$\min_{\theta \in \Theta} \left[ \hat{\beta}_T - \tilde{\beta}_{TS}(\theta, y_0) \right]' \tilde{\Omega}_T \left[ \hat{\beta}_T - \tilde{\beta}_{TS}(\theta, y_0) \right]$$

This estimator is a consistent estimate of  $\theta_0$  under the very same hypotheses previously stated.

Notice that it is not necessary to calculate the values of  $\tilde{\beta}_T^s(\cdot, z_0^h)$  or  $\tilde{\beta}_{TS}(\cdot, z_0^h)$  for all possible values of  $\theta$  but only for those occurring in the minimization program. Indeed, the minimization is applied to a neighborhood of the starting point of the optimization. This neighborhood will be tightened from one iteration to the next until the minimization constraint is satisfied and that a satisfying optimum is found.

### 2.1.3. Criterion selection.

We can look for a function  $r^*$  approximating  $r$  for which the log-likelihood conditional  $L_T^*(\beta)$  can be easily derivable and we take  $Q_T = \frac{1}{T} L_T^*(\beta)$ . Then approximate  $Q_T$  by  $\frac{1}{T} L_T(\beta)$  with  $L_T(\beta)$  the exact log-likelihood.

The value of the indirect estimator  $\hat{\theta}$  depends on the choice of the matrix  $\tilde{\Omega}_T$ . Now it is a particular frequent case where the value of  $\theta$  is independent of the choice of this matrix : when  $p = \dim(\theta) = \dim(\beta) = q$ .<sup>678</sup> Moreover, under the null hypothesis

<sup>6</sup>Which means exact identification.

<sup>7</sup>With  $p$  representing the size of the vector of true parameters and  $q$  the size of the vector of auxiliary parameters.

<sup>8</sup>When  $p > q$ , the true parameters outnumber the auxiliary parameters. In this case the bridge relation is many-to-one and does not in general permit the construction of adjusted estimates. It is mainly of interest for investigating the effects of misspecification where the auxiliary estimators are constructed under misspecified models, for example. However, in such situations it may be possible to construct consistent estimates for a subset of true parameters, which may be of interest. In other situations, some components of the higher-dimensional true parameter are known or can

that the model is true, the optimal weighting matrix is just the variance-covariance of the moments in the data.

Consider for example the identity matrix as the matrix  $\tilde{\Omega}_T$ . Minimization of  $\left[\hat{\beta}_T - \tilde{\beta}_{TS}(\theta, y_0)\right]' \tilde{\Omega}_T \left[\hat{\beta}_T - \tilde{\beta}_{TS}(\theta, y_0)\right]$  is equivalent to minimize the sum of squares of the elements of the vector  $\hat{\beta}_T - \tilde{\beta}_{TS}(\theta, y_0)$ , and we get in this particular case  $\hat{\beta}_T = \frac{1}{S} \sum_{s=1}^{s=S} \tilde{\beta}_T^s(\theta, z_0^s)$ .<sup>9</sup>

There exists an equivalent method developed by Gallant and Tauchen which bases itself on the partial derivatives of the function  $Q_T$ . This estimator, as opposed to the previous two, only requires one optimization. Because this estimator has been introduced to deal with particular cases<sup>10</sup>, it won't be described in this report any further.

#### 2.1.4. Asymptotic properties.

In this section are exposed the main asymptotic results, whose proofs are given by Gouriéroux and al.<sup>11</sup>. We make the following assumptions:

- H(5)

$$\Xi_T = \sqrt{T} \frac{\partial Q_T}{\partial \beta} [y_T^1 x_T^1, \beta_0] - \sqrt{T} \frac{1}{S} \sum_{s=1}^S \frac{\partial Q_T}{\partial \beta} ([\tilde{y}_T^s(\theta_0, z_0^s)]_T^1, x_T^1, \beta_0)$$

is asymptotically normal with zero mean, with an asymptotic variance-covariance matrix given by

$$W = \lim_{T \rightarrow \infty} V(\Xi_T)$$

and independent of the initial values  $z_0^s, s = 1, \dots, S$ ;

- H(6)  $\lim_{T \rightarrow +\infty} V\left\{\sqrt{T} \frac{\partial Q_T}{\partial \beta} (\tilde{y}_T^s(\theta_0, z_0^s), x_T^1, \beta_0)\right\} = I_0$  is independent of  $z_0^s$ ;
- H(7)  $\lim_{T \rightarrow +\infty} Cov\left\{\sqrt{T} \frac{\partial Q_T}{\partial \beta} (\tilde{y}_T^s(\theta_0, z_0^s), x_T^1, \beta_0), \sqrt{T} \frac{\partial Q_T}{\partial \beta} (\tilde{y}_T^l(\theta_0, z_0^l), x_T^1, \beta_0)\right\} = K_0$  independent of the  $z_0^l$ ;
- H(8)  $\text{plim}_{T \rightarrow +\infty} -\frac{\partial^2 Q_T}{\partial \beta \partial \beta'} [[\tilde{y}_T^s(\theta_0, z_0^s)]_T^1, x_T^1, \beta_0] = -\frac{\partial^2 Q_\infty}{\partial \beta \partial \beta'} (F_0, G_0, \theta_0, \beta_0) = J_0$  is independent of  $z_0^s$ .

---

be estimated from other outside data sources or can be tailored using heuristics. This enables the other components to be consistently estimated by inverting the bridge relation.

<sup>9</sup>Or again  $\hat{\beta}_T = \tilde{\beta}_{TS}(\theta, z_0^h)$

<sup>10</sup>Those cases share the following features: the criterion function is a likelihood function, no exogenous variable, unlimited number of simulations, the model corresponding to the pseudo-likelihood function is asymptotically well specified

<sup>11</sup>See Appendix 1. of C. Gouriéroux, A. Monfort, and E. Renault. Indirect inference. Journal of Applied Econometrics.

Under all the above assumptions and the usual regularity conditions, when  $S$  is fixed and  $T$  goes to infinity, it can be affirmed:

$$\sqrt{T}(\theta_T^S(\Omega) - \theta_0) \rightarrow \mathcal{N}(0, W(S, \Omega))$$

with

$$\begin{aligned} W(S, \Omega) = & \left(1 + \frac{1}{S}\right) \left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0) \Omega \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right)^{-1} \\ & \left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0) \Omega J_0^{-1} (I_0 - K_0)^{-1} J_0^{-1} \Omega \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right) \\ & \left(\frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0) \Omega \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0)\right)^{-1} \end{aligned}$$

Note that due to the conditional independence to exogenous variables of

$$\sqrt{T} \frac{\partial Q_T}{\partial \beta} [[\tilde{y}_T^s(\theta_0, z_0^s)]_T^1, x_T^1, \beta_0]$$

and

$$\sqrt{T} \frac{\partial Q_T}{\partial \beta} [[\tilde{y}_T^l(\theta_0, z_0^l)]_T^1, x_T^1, \beta_0]$$

and because they share the same asymptotic distribution, we can write :

$$\begin{aligned} K_0 &= \lim_{T \rightarrow \infty} V_0 \left( E_0 \left[ \sqrt{T} \frac{\partial Q_T}{\partial \beta'} [y_T^1 x_T^1, \beta_0] / x_T^1 \right] \right) \\ I_0 - K_0 &= \lim_{T \rightarrow \infty} V_0 \left( \sqrt{T} \frac{\partial Q_T}{\partial \beta} [y_T^1 x_T^1, \beta_0] - E_0 \left[ \sqrt{T} \frac{\partial Q_T}{\partial \beta'} [y_T^1 x_T^1, \beta_0] / x_T^1 \right] \right) \end{aligned}$$

The optimal choice for the matrix  $\Omega$  is  $\Omega^* = J_0(I_0 - K_0)^{-1}$ . The asymptotic matrix of variance-covariance is then simplified by

$$W_S^* = \left( \frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0) J_0 (I_0 - K_0)^{-1} J_0 \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0) \right)^{-1}$$

The asymptotic aspects regarding  $K_0$  and  $I_0$  can be interpreted as the need for the  $\beta$ 's to converge towards an optimum. The variance of their drawings should be finite even when  $T \rightarrow \infty$ . If the limit  $K_0$  does not exist, the indirect estimators would not be consistent. More over this should not depend on whatever data the inference is based on.

It is worth noting that the matrix  $W(\infty, \Omega)$  and  $W_S^*$  are not, in general, the asymptotic variance-covariance matrices of the indirect estimators that would be based on the binding function. It is only true when there is no exogenous variable as it is usually the case in time series (see Gourieroux and Montfort, 1992).

We could also write these matrix using the introduced binding function *supra*. We have already mentioned an important special case, when the parameter of interest  $\theta$  and the auxiliary parameter  $\beta$  are of the same size. Indeed we said then that the choice of the matrix  $\Omega$  did not influence the value of the estimator obtained by Indirect Inference, in this case we can choose  $\Omega$  equal to the identity matrix. We obtain a new result concerning this time the variance-covariance asymptotic matrix, namely that  $W(S, \Omega) = W_S^*$ . Let us also note that it is the variance-covariance asymptotic matrix of the estimator solution of  $\hat{\beta} = \tilde{\beta}_{TS}(\theta, z_0)$ .

The calculation of the optimal estimator by Indirect Inference requires a consistent preliminary estimator of  $\Omega^*$  for initializing the calculations. Depending on the case, we may choose such an estimator directly based on the observations or from the simulations by taking the identity matrix in place of the matrix  $\Omega$  in order to determine the resulting estimator, or by taking the quantity

$$\arg \max_{\theta} Q_T \left[ y_T^1, x_T^1, \frac{1}{T} \sum_{s=1}^{s=S} \tilde{y}_T^s(\theta, z_0^s) \right]$$

## 2.2. Monte Carlo Analysis.

To analyse the results of our Indirect Inference estimators, Monte Carlo simulations are conducted. Thanks to that, the distribution of estimators can be drawn and the bias of each estimator from its 'true' counterpart can be calculated.

Monte Carlo simulation uses repeated random sampling to simulate data for our structural models and evaluate their outcomes. In our case, the MC will simulate different time series randomly on base of our modeling. A potential distribution of the parameters of interest, the vector  $\theta$ , can be drawn by simulating the various sources of uncertainty affecting the value of the structural parameters, and then determining the distribution of their value over the range of resultant outcomes.

One rationale for Monte Carlo simulation is its help in our estimations without the need of a tremendous amount of experiments or building thousands of samples.

A Monte Carlo simulation boils down to four simple steps:

- (1) Identify the transfer equation (i.e.: a mathematical model of the activity or process under investigation).
- (2) Define the parameters for each factor in the model.
- (3) Create random data according to those parameters.
- (4) Simulate and analyze the output of the process.

To simplify, if  $V$  represents an observable variable, the estimator  $\hat{V}$  with variance  $E[(\hat{V} - V)^2]$  by the Monte Carlo method for  $M$  paths is given by

$$(2.2) \quad \hat{V} = \frac{1}{M} \sum_{j=1}^M V_j$$

In sum, instead of being satisfied with the conducting of only one estimation for each parameter through indirect inference, we will conduct  $M$  inferences based on  $M$  different simulations. This is the inherent part of randomness in each simulation that will allow us to approach the real value of the structural parameters. The estimators being closer and closer to the value of their underlying variable when  $M \rightarrow \infty$  thanks to the Law of Large Numbers and the Central Limit Theorem. This will allow us to compute the bias and the root mean square errors of our estimates around the structural parameters.

Anyway the Monte Carlo tool will only be useful for simulations-based samples. In different application case, only one sample is known. Especially in the application that will be exposed later, trying to find the parameters of the S&P500 returns, under the assumption that the process follows a specific model, it will not be possible to draw different paths of returns, since only one exists.

## 3. EXAMPLES

## 3.1. Moving Average.

As a first (simple) example, this section will expose an indirect inference estimation of MA(q) parameters.

Let's use a simplistic form of moving average model which will be used to simulate several observations. This will be the structural model. On basis of the literature in Time-Series regarding the relationship between moving-average and auto-regressive models, we will use an AR(q) process as auxiliary model to fit the data and draw the  $\beta$ s.

Here is the equation of a moving average of dimension  $m$ :

$$(3.1) \quad y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_{t-m} \varepsilon_{t-m}$$

Similarly:

$$(3.2) \quad y_t = \varepsilon_t + \sum_{j=2}^m \psi_j \varepsilon_{t-j}$$

Where  $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  and with  $\theta = [\sigma_\varepsilon^2, \psi_1, \dots, \psi_m]$ , the vector of structural parameters (of dimension  $p=m+1$ ).

The auxiliary model will consist of the following Auto-Regressive process of order  $q$ .

$$(3.3) \quad y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_q y_{t-q} + \eta_t = \sum_{j=1}^q \rho_j y_{t-j} + \eta_t$$

With  $\beta = [\rho_1, \dots, \rho_q, \sigma_\eta^2]$ .

As motivated in the previous section, we consider the matrix  $\tilde{\Omega}_T$  as equal to the variance-covariance matrix of the  $\beta$ 's.

## 3.1.1. Invertibility condition.

It should be noted that, just as an infinite-order moving average process can be defined, one could also define an infinite-order auto-regressive process,  $AR(\infty)$ . It turns out that any stationary  $MA(m)$  process can be expressed as an  $AR(\infty)$  process. E.g. suppose we have an  $MA(1)$  process with  $\mu = 0$ .

$$y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1}$$

Implying

$$\varepsilon_i = y_i - \psi_1 \varepsilon_{i-1} = y_i - \psi_1 (y_{i-1} - \psi_1 \varepsilon_{i-2}) = y_i - \psi_1 y_{i-1} + \psi_1^2 \varepsilon_{i-2}$$

And, for  $n$  steps:

$$\varepsilon_i = y_i - \psi_1 y_{i-1} + \psi_1^2 y_{i-2} - \psi_1^3 y_{i-3} + \dots + (-\psi_1)^n y_{i-n} + (-\psi_1)^{n+1} \varepsilon_{i-n-1}$$

Which can be written:

$$\varepsilon_i = \sum_{j=0}^n (-\psi_1)^j y_{i-j} + (-\psi_1)^{n+1} \varepsilon_{i-n-1}$$

As a result we have<sup>12</sup>:

$$\varepsilon_i = y_i - \psi_1 y_{i-1} + \psi_1^2 y_{i-2} - \psi_1^3 y_{i-3} + \dots = \sum_{j=0}^{\infty} (-\psi_1)^j y_{i-j}$$

In other words:

An MA model is invertible if it is algebraically equivalent to a converging infinite order AR model. By converging, it is meant that the AR coefficients decrease to 0 as the process moves back in time.

For a MA(m) model with a specified ACF<sup>13</sup>, there is only one invertible model. The necessary condition for invertibility is that the  $\psi$  coefficients have values such that the equation

$$(3.4) \quad 1 - \psi_1 y - \dots - \psi_1 y^m = 0$$

has solutions for  $y$  that fall outside the unit circle or, equivalently, iff all roots of the characteristic following polynomial constructed following that the most recent error can be written as a linear function of current and past observations

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\psi_1)^j y_{t-j}$$

lie outside unit circle in complex plain.

The essential concept is whether the innovations/noises can be inverted into a representation of past observations. This notion is very much important if one wants to forecast the future values of the dependent variable, a very relevant issue for many financial practitioners and policy makers. Otherwise, the forecasting task will be impossible when the innovations are not invertible (i.e., the innovations in the past cannot be estimated, as it is cannot be observed)<sup>14</sup>. Actually, when the model is not invertible, the innovations can still be represented by observations of the future, this is not helpful at all for forecasting purpose.

---

<sup>12</sup>Equivalently:

$$y_i = \psi_1 y_{i-1} - \psi_1^2 y_{i-2} + \psi_1^3 y_{i-3} - \dots + \varepsilon_i = \sum_{j=0}^{\infty} (-\psi_1)^j y_{i-j} + \varepsilon_i$$

<sup>13</sup>Auto-Correlation Function

<sup>14</sup>For an invertible process,  $|\psi| < 1$  and so the most recent observations have higher weight than observations from the more distant past. But when  $|\psi| > 1$ , the weights increase as lags increase, so the more distant the observations the greater their influence on the current error. When  $|\psi| = 1$ , the weights are constant in size, and the distant observations have the same influence as the recent observations. As neither of these situations make much sense, we prefer the invertible processes.

A relaxation of these constraints can cause the distribution of the parameters' estimates to have unwanted features such as, for example, bi-modality or identification issues.

This imposes us to put on linear nor non-linear constraints in the minimization process. The simulations within the minimization have to take into account that the  $\hat{\theta}$  have to respect the invertibility constraint, otherwise we might end-up with estimates that could lead towards an moving average time series violating the invertibility condition.

### 3.1.2. Simulations' results.

Here are the fitted histograms of the four parameters of a model. It was generated with 500 simulations of MA, AR and looped 500 times for the Monte-Carlo process, with the "true" value of parameters being:  $\theta_1 = \sigma 2_\varepsilon = 0.2, \theta_2 = 0.7, \theta_3 = 0.2$  and  $\theta_4 = -0.3$ .

The bi-modality, which modes appear as distinct peaks (local maxima) of the distributions, is due to the relaxation of the invertibility constraint exposed earlier. The binding function,  $b(F, G, \theta) = \arg \max_{\beta \in B} Q_\infty(F, G, \theta, \beta)$ , is no longer injective and the robustness of the indirect estimators is threatened by the existence of more than one  $\hat{\theta}$  for each  $\theta_0$ . In other words, the histograms should depict a unique modality for each parameter.

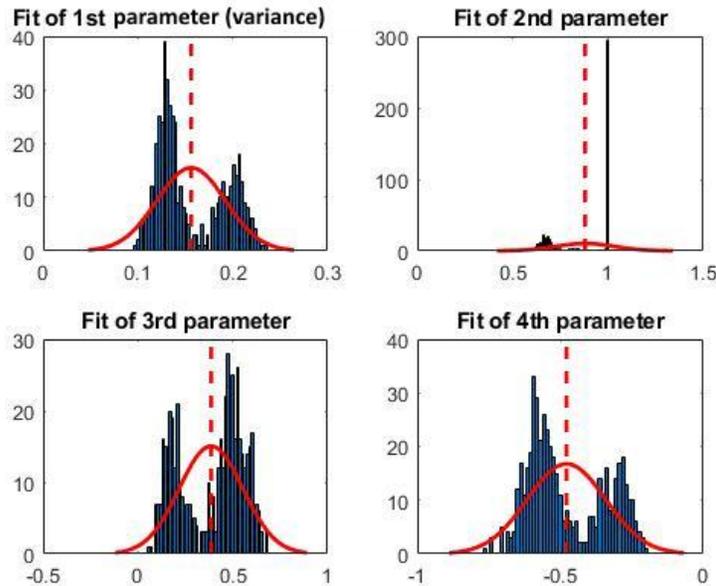


FIGURE 2. Parameters distribution bi-modality

Table 1. brings a synthetic view about the estimates obtained between 2 sets of unconstrained simulations of different length but with the same auxiliary parameter vector and a starting point far from the vector of true parameters ( $S_0^S = [0.07601; 0.16805; 0.03693; 0.01207]$  and  $\theta = [0.3800; 0.84025; 0.018468; 0.06036]$ ). The first column stands the number of over-identifying parameters (with p being the numbers of lags in the structural moving average plus  $\sigma_\varepsilon$  and q the order of the auxiliary auto-regressive model). The results are quiet as expected since we observe that a longer time series allows the indirect estimators to become more efficient. But we could do better since This is mainly due to the fact that these values are derived from unconstrained optimizations.

TABLE 1

$(q - p)$	$\theta$	True value	$\hat{\theta}^{100}$	$\hat{\theta}^{1000}$
5	1	0.3801	0.3650	0.3718
	2	0.8403	0.8918	0.8616
	3	0.01847	0.1881	0.1823
	4	0.0604	0.0518	0.0785

Table 2. shows the invertibility condition actually improves the results. Indeed, the constrained estimates  $\hat{\theta}^c$  outperform the unconstrained estimates  $\hat{\theta}$  at guessing the true values.

TABLE 2

$(q - p)$	$\theta$	True value	$\hat{\theta}^c{}^{100}$	$\hat{\theta}^c{}^{1000}$	$Bias^{100}$	$Bias^{1000}$
5	1	0.3801	0.3799	0.3649	-0.0002 (0.000)	-0.0152 (0.2316)
	2	0.8403	0.8812	0.8784	0.0409 (1.6757)	0.0381 (1.4478)
	3	0.01847	0.0312	0.0065	0.0127 (0.1624)	-0.0120 (0.1430)
	4	0.0604	0.0522	0.0617	-0.0082 (0.0675)	0.0013 (0.0017)

Root mean square errors in parentheses.

Table 3. Depicts the improvements possible by picking an optimization starting point closer ( $S_0^S = [0.3421; 0.7563; 0.01662; 0.05436]$ ) to the true values. It highlights the importance to start with already a good guess of what the parameters should be, and the comparison between table 2. and 3. shows how the estimates are sensible to those starting points.

TABLE 3

$(q-p)$	$\theta$	True value	$\hat{\theta}^c{}^{100}$	$\hat{\theta}^c{}^{1000}$	$Bias^{100}$	$Bias^{1000}$
5	1	0.3801	0.3756	0.3802	-0.0045 (0.0203)	0.0001 (0.0000)
	2	0.8403	0.8665	0.8506	0.0262 (0.6890)	0.0103 (0.2905)
	3	0.01847	0.0214	0.0209	0.0029 (0.0084)	0.0024 (0.0058)
	4	0.0604	0.0466	0.0599	-0.0138 (0.1905)	-0.0005 (0.0002)

Root mean square errors in parentheses, close starting optimization point.

Finally, table 4. shows the positive effect of adding auxiliary parameters on the quality of the estimations. Indeed, estimates drawn on the basis of an AR(4) are less efficient and less consistent than those from an AR(5).

TABLE 4

$(q-p)$	$\theta$	$\hat{\theta}^c{}^{100}$	$\hat{\theta}^c{}^{1000}$	$Bias^{100}$	$Bias^{1000}$
4	1	0.3721	3784	-0.0080 (0.0646)	-0.0017 (0.0028)
	2	0.8669	8575	0.0266 (0.7084)	0.0172 (0.2955)
	3	0.0389	0.0209	0.0204 (0.0084)	0.0024 (0.0058)
	4	0.0489	0.0605	-0.0115 (0.1327)	0.0001 (0.0000)
5	1	0.3756	0.3802	-0.0045 (0.0203)	0.0001 (0.0000)
	2	0.8665	0.8506	0.0262 (0.6890)	0.0103 (0.2905)
	3	0.0214	0.0209	0.0029 (0.0084)	0.0024 (0.0058)
	4	0.0466	0.0599	-0.0138 (0.1905)	-0.0005 (0.0002)

Root mean square errors in parentheses, close starting optimization point.

The complete tables from these simulations can be found in Appendix A.. They compare values for more (q-m) and more times series lengths.

### 3.2. Stochastic Volatility.

#### 3.2.1. The structural model.

As second example, we will aim to estimate the volatility dynamic of simulated time series of stock prices.

The Heston model we use is similar to the Geometric Brownian Motion (GBM) process<sup>15</sup> used in the classic Black-Scholes model, such a model assumes that the volatility of the asset is not constant, nor even deterministic, but follows a random process.

Stochastic Differential Equation (SDE):

$$(3.5) \quad P_t = p + \int_0^t \mu(P_t)dt + \int_0^t \sigma(P_t)dW_t,$$

with  $t \in [0, T], p \in \mathbb{R}^n$ .

or equivalently:

$$(3.6) \quad SDE = \begin{cases} dP_t = \mu_{P_t}dt + \sigma_{P_t}dW_t \\ d\sigma_t^2 = \kappa(\omega - \sigma_t^2)dt + \eta\sigma_t dW_t, \end{cases}$$

The first equation represents of (3.6) the stock value where  $P_t$  is observed (i.e.:  $\mu$ =rate of returns,  $\sigma$ =volatility,  $W_t$ = standard Brownian Motion<sup>16</sup>). The second equation is not observed and represents the volatility features of the stock prices.

The variables of interest are therefore  $\kappa, \omega$  and  $\eta$ .

With this formulation, the variance dynamics follow a square-root diffusion, also referred to as a Cox-Ingersoll-Ross (CIR) process, or is a  $CIR(\kappa, \omega, \eta)$ . That square-root diffusion process was first developed by Cox, Ingersoll, and Ross (1985) and was used for interest rate modeling. In the Heston stochastic volatility setting, the CIR process defines the variance dynamics. Modeled in this way, the variance has a mean-reverting structure, with  $\kappa$  being the speed of mean-reversion and  $\omega$  being the long term level of variance. Intuitively, having  $\kappa > 0$  and  $\omega > 0$ , makes sure that if the variance is higher than its usual values, then it will be pulled down towards them and vice versa, if the variance is too low then the quantity  $\omega - \sigma^2$  will be positive and the process will be dragged up towards its long term level.

The speed of mean-reversion, or  $\kappa$ , is a parameter that takes account of the empirically observed clustered volatility. Also log returns are usually leptokurtic<sup>17</sup> and controlled by  $\eta$ , being the variance volatility. High  $\eta$  values provide higher peaks,

<sup>15</sup>Or also named Wiener process

<sup>16</sup>Which is used to model noisy fluctuations of stocks.

<sup>17</sup>As opposed to mesokurtic distributions, a distribution is leptokurtic when its kurtosis value is a large positive number.

and in particular, when  $\eta = 0$  the variance is deterministic<sup>18</sup>. For the variance process to be strictly positive the following condition is needed:  $2\kappa\omega > \eta^2$ , with  $\kappa > 0$ ,  $\omega > 0$  and  $\eta > 0$ .

What should be considered in the first place is a discretization of the SDE. This method is most useful when you want to compute the path between  $P_0$  and  $P_t$ , i.e. we want to know all the intermediary points  $P_i$  for  $0 \leq i \leq t$ . More over it is a condition to make simulations and apply Monte Carlo analysis.

Space discretization: The interval  $[0, T]$  is divided into  $N$  equally sized subintervals of length  $\Delta$ . The price of the underlying asset will take values in the unbounded interval  $[0, \infty)$ .

Fix  $p \in \mathbb{R}$  and let  $\Delta^{(N)}$  be the equidistant mesh of size  $N$ , i.e.  $t_i = \frac{i}{N}T, i = 0, \dots, N$ . Set  $\Delta t_i$  and  $\Delta W_t = W_{t_{i+1}} - W_{t_i}, i = 0, \dots, N - 1$ . Define the Euler-Maruyama<sup>19</sup> approximation to  $P_T$  by  $\bar{P}_0^N = p$  and

$$(3.7) \quad \bar{P}_{i+1}^{(N)} = \bar{P}_i^{(N)} + \mu(\bar{P}_i^{(N)})\Delta t_i + \sigma(\bar{P}_i^{(N)})\Delta W_i,$$

with  $i = 0, \dots, N - 1$ . Or, more explicitly:

$$(3.8) \quad \begin{cases} P_t = P_{t-1} + \mu P_{t-1} dt + P_{t-1} \sqrt{\sigma_{t-1}} W_t^1 \\ \sigma_t = \sigma_{t-1} + \kappa(\omega - \sigma_{t-1}) dt + \eta \sqrt{\sigma_{t-1}} \sqrt{dt} W_t^2 \end{cases}$$

In these simulations, we will assume  $\mu = 0$  and that we have daily sampled prices (which means  $dt = 1$ , and  $N = T$ ).

### 3.2.2. The auxiliary model.

As suggested by C.Monfardini (1997), we will use a Gaussian auto-regressive representation of a given order (i.e. an AR(m)) for  $\ln y_t^2$ , in order to approximate the ARMA nature of the Heston model.

$$(3.9) \quad \ln y_t^2 = \beta_0 + \beta_1 \ln y_{t-1}^2 + \beta_2 \ln y_{t-2}^2 + \dots + \beta_m \ln y_{t-m}^2 + \varepsilon_t$$

With  $\varepsilon_t \sim I.I.N(0, \tau^2)$  and  $y$  being the stock returns. By posing:  $x_t = \ln y_t^2$ ;

$$\begin{aligned} \underline{x} &= (x_{m+1}, x_{m+2}, \dots, x_T)' \\ \underline{x}_{-l} &= (x_{m+1-l}, x_{m+2-l}, \dots, x_{T-l}), \forall l \in l, \dots, m \\ \underline{X}_{-m} &= (1, \underline{x}_{-1}, \underline{x}_{-2}, \dots, \underline{x}_{-m}) \\ \beta &= (\beta^*, \tau^2)'. \end{aligned}$$

<sup>18</sup>In other words, the log-returns will be normally distributed.

<sup>19</sup>The Euler-Maruyama scheme for discretization of SDEs is simple to understand and implement, but suffers from a low order of convergence, especially in the strong sense.

The criterion function can be written as following:

$$\varphi_T(\underline{x}, \beta) = -\frac{1}{2} \ln(2\pi\tau^2) - \frac{1}{2\tau^2(T-m)} (\underline{x} - X_{-m}\beta^*)' (\underline{x} - X_{-m}\beta^*)$$

Leading to the estimators:

$$\begin{aligned} \beta^* &= (X'_{-m} X_{-m})^{-1} X'_{-m} \underline{x} \\ \hat{\tau}^2 &= \frac{\hat{\epsilon}' \hat{\epsilon}}{T-m} \text{ with } \hat{\epsilon} = \underline{x} - X_{-m} \hat{\beta}^* \end{aligned}$$

The matrix  $\Omega^*$  used is approached by a converging estimator  $\hat{\Omega}_T^* = \hat{J}_T \hat{I}_T^{-1} \hat{J}_T$  with

$$\begin{aligned} \hat{I}_T^{-1} &= \hat{V}_T^0 + \sum_{k=1}^K (\hat{V}_T^k + \hat{V}_T^{-k}) \left(1 - \frac{k}{K+1}\right) \\ \hat{V}_T^k &= \frac{1}{T-m} \sum_{t=k+1}^{T-m} \frac{\partial q_t}{\partial \beta} \frac{\partial q_{t+k}}{\partial \beta'} \\ q_t &= -\frac{1}{2} (\ln(2\pi\tau^2) + \frac{1}{\tau^2} (x_t - \beta_0 - \beta_1 x_{t-1} - \dots - \beta_m x_{t-m})^2) \end{aligned}$$

Parameters  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_m)$  and  $\tau^2$  can be easily estimated through the Maximum Likelihood method, based on the sequential factorization of the density of  $\ln y_t^2$  given its past and conditioning on the first  $m$  observations. The model considers the stock returns rather than directly the stock prices because the price time-series under stochastic volatility are not stable but their returns are.

### 3.2.3. Simulations' results.

As far as the sample sizes  $T$  considered are concerned, it should be emphasized that stochastic volatility inference is quite demanding in terms of sample information required, due to the presence of a latent structure governing the volatility of the modality. That is why we will use longer time series than in the previous example.

We took the assumption that  $\mu = 0$ , which simplifies the simulations. It can be noticed that this assumption is quite often relied on, notably for European option pricing since this assumption of  $\mu = 0$  has no impact on the results.

Of course, we should simulate the true model with parameters respecting  $2\kappa\omega > \eta^2$ . This is will be done by imposing a non-linear constraint in the minimization process, as we did for the invertibility condition for the moving average example.

In table 2 can be find the results from a simulated Heston process with  $T = 2000$  and  $3000$ ,  $dt = 1$  (daily sampling) implying  $N = 2000$  and with  $[\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3] = [\hat{\kappa}, \hat{\omega}, \hat{\eta}]$ . The true values were  $\kappa = 0.01$ ,  $\omega = 0.0004$  and  $\eta = 0.0015$ . For the starting point of the optimization,  $z_0^{TS}$ , we will consider  $z_2^0 = \text{var}(y_t)$  as the starting value for  $\omega$  (it would seem that the standard measure of variance could approximate fairly the long-term volatility). We could complete our first guess by trial-simulations. Finally we took  $z_1^0 = 0.005$ ,  $z_2^0 = 0.00000486$  and  $\theta_3^0 = 0.00075$  as starting points of the optimization.

The results are as expected and in a similar fashion than in the previous example. Bias And the Rmse (Root mean square of errors) are reduced with the number of auxiliary parameters. In sum, over-identified auxiliary performs better in terms of Bias and Rmse and estimates  $\hat{\theta}$ 's get closer to their true values with higher q's, at least for  $\hat{\theta}_1$  and  $\hat{\theta}_3$ . But those improvements are limited, adding an infinite amount of over-identifying parameters will not get us infinitely closer. Finally, as previously, longer time series allow us to efficiently reduce the root mean square error, but does not impact Bias as much.

TABLE 5

$(q - p)$	$\theta$	$\hat{\theta}^{1000}$	$\hat{\theta}^{2000}$	$Bias^{1000}$	$Bias^{2000}$
10	1	0.0558	0.0284	0.0458 (2.0938)	0.0184 (0.3370)
	2	0.0000	0.0001	-0.0004 (0.0002)	-0.0003 (0.0001)
	3	0.0002	0.0002	-0.0013 (0.0017)	-0.0013 (0.0017)
20	1	0.0367	0.0262	0.0267 (0.7135)	0.0162 (0.2620)
	2	0.0013	0.0011	0.0009 (0.0008)	0.0007 (0.0006)
	3	0.0002	0.0014	-0.0013 (0.0016)	-0.0001 (0.0000)

Root mean square errors in parentheses, close starting optimization point.

## 4. EMPIRICAL APPLICATION

In this section, we will try to estimate the underlying parameters of the S&P500 Stock Market Index returns, under the assumption it follows a Heston process.

The S&P500 observations will stand in place of the simulations developed in previous sections. We therefore do not "use" a structural model *per se*, the original structural model is the one that generated those observations and thus is unknown.

## 4.1. Tests of hypotheses on the parameters of interest.

In this application case, no Monte Carlo analysis can be conducted since we are not able to draw different paths of data. Monte Carlo uses randomized samples in order to approximate a deterministic process. In this application, the observations are already "determined".

We cannot longer rely on the benefits of Monte Carlo simulations to ensure the indirect estimates robustness. An *ex-post* test is needed to verify that the parameters are different from zero. Since we would like to reject the hypothesis that  $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$  are  $\neq 0$ , We therefore need a Wald test.

Recall from the asymptotic properties that when S is fixed (and equal to 1 in this case) and T goes to infinity, it can be affirmed:

$$\sqrt{T}(\theta_T^S(\Omega) - \theta_0) \rightarrow \mathcal{N}(0, W(S, \Omega))$$

with

$$W_S^* = \left( \frac{\partial b'}{\partial \theta}(F_0, G_0, \theta_0) J_0 (I_0 - K_0)^{-1} J_0 \frac{\partial b}{\partial \theta'}(F_0, G_0, \theta_0) \right)^{-1}$$

The test therefore assess whether the binding function  $b(\theta)$  does not  $\theta \rightarrow \frac{1}{S} \sum_{i=1}^3 \beta_i$  for values of  $\frac{1}{S} \sum_{i=1}^3 \beta_i = 0$ . It further analyzes the  $W(S, \Omega)$  which should be equal to  $J(\hat{\theta}) \cdot \Omega \cdot J(\hat{\theta})'$  which is a  $p \times p$  matrix ( $p$  is the dimension of the vector of structural parameters), with  $J(\hat{\theta}) = \frac{\partial b(\theta)}{\partial \theta} |_{\theta=\hat{\theta}}$ . As exposed in the theoretical background, this matrix, under some general conditions, is the variance-covariance matrix of the

elements of  $\theta$  :

$$\begin{bmatrix} \sigma_1^2 & \dots & cov(\hat{\theta}_1, \hat{\theta}_p) \\ \vdots & \ddots & \vdots \\ cov(\hat{\theta}_p, \hat{\theta}_1) & \dots & \sigma_p^2 \end{bmatrix}.$$

The Wald test can be used to test a single hypothesis on multiple parameters, as well as to test jointly multiple hypotheses on single/multiple parameters. We will use the indirect optimal estimator obtained under constraint with the condition

$$\theta_1=0, \text{ that will be noted } (\hat{\theta}^{0S}) = \begin{bmatrix} 0 \\ \theta_{2T}^{\hat{S}} \end{bmatrix}.$$

We can introduce the score statistic built as a Wald statistic  $\xi_T^W$ :

$$\xi_T^W = T(\theta_{1T}^{\hat{S}})' \hat{W}_1^{*-1} (\theta_{1T}^{\hat{S}})$$

Where  $\hat{W}_1^*$  is a consistent estimator of the asymptotic covariance-variance matrix. This test statistic has the distribution  $\chi^2(p)$ , with  $p=\dim\theta$ .

Hopefully we can approximate the same result by using a z-test. the z-test should take the following form:

$$(4.1) \quad z_t = \frac{\hat{\theta}}{\text{standard-error}}$$

with

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

Especially, we desire to reject the hypothesis that  $\theta = 0 \Rightarrow H_0 : \theta = 0$ . The statistic of test under the null-hypothesis is the following:

$$Z = \sqrt{T} \frac{\hat{\theta}}{\sigma}$$

If  $|z_t|$ , the realization of the test statistic, is higher than the quantile of order  $1 - \frac{\alpha}{2}$  of  $N(0,1)$ , with  $\alpha$  being the confidence order,  $H_0$  can be rejected. *In Extensia*,  $H_0$  is rejected if  $z_i > 1.96$ .

This test could be directional. However, commonly the Wald test is generalized to test multiple parameters by squaring the z statistics and this should have a chi-squared distribution, asymptotically. So p parameters can be tested by a  $\chi^2$  statistic equal to  $z_1^2 + z_2^2 + \dots + z_p^2$  comparing the result to a chi-squared distribution with p degree of freedom.<sup>20</sup>

#### 4.2. Block-Bootstrapping.

Despite our deception due to the fact that we cannot rely on Monte Carlo analysis anymore, we can try to circumvent this situation by artificially summon different samples on which we can infer.

We can inspire ourselves by what has been done in the literature on time-series.

It comes out that, if the data is i.i.d, one could treat the sample data as the population, and do sampling with replacement of observations and this would allow to get multiple simulations of some statistic. This is called boot-strapping. In this case of time series, we are clearly not allowed to do so due to the likely existence of auto-correlations between observations. More over, a time series is essentially a sample of size 1 issued from a stochastic process, re-sampling the same original sample will bring no new information. Therefore, re-sampling of a time series requires new ideas.

As developments around this purpose we can consider the re-sampling method introduced by Efron (1979), which was designed for i.i.d. univariate data but are easily extended to multivariate data. As discussed by the author, in the case where  $x_1, x_2, \dots, x_n$  is a sample of vectors, in the aim to guarantee to keep the covariance

---

<sup>20</sup>With only one parameter, a test with 1 degree of freedom is equivalent to the z test because the critical value  $+/- 1.96$  for z with  $\alpha = .05$  is the square root of the critical value for  $\chi^2$  with 1 degree of freedom (3.84). So you can get a directional Wald test by taking the square root of a  $\chi^2$  test statistic iff it has 1 degree of freedom.

structure of the data untouched. It is not immediately obvious whether one can re-sample a time series, but there are two ways:

- Model-based re-sampling, which is easily adopted to time series. The new samples are obtained by simulating the time series model.
- Model-free re-sampling of time series, which is accomplished by block re-sampling, also called block bootstrap.

Variations on the 'block bootstrap' are intuitive. Here, depending on the method, we select stretches of the time series, either overlapping or not and of fixed length or random, which can guarantee stationarity in the samples (Politis and Romano, 1991) then stitch them back together to create re-sampled times series on which we can compute complementary statistics.

In itself, the simulations we conducted previously for our examples were a kind of Model-based re-sampling procedures. For this section, even-though we started by saying that we relied on the assumption that the observed data follow a Heston process, we should highlight the fact that Heston models are simplistic forms of stochastic volatility. For the boot-strapping we will thus not rely on the fact that the time series have a specific structural form, since we want to estimate its structural parameters.

For this application, we will start from an original sample of the S&P500 Stock Market Index value on a time frame starting the 09/06/2008 and finishing the 09/06/2018 (with daily sampling) we thus end up with 2518 observations. Relying on the results obtained from Heston-based simulations in previous sections, we will take a window of 2000 observations as the fixed length for our bootstrapping. These stretches of time will be overlapping, letting the window to "roll" over the original sample. This mean we will construct  $2518-2000=518$  new blocks of length 2000. This method is known as the *Moving Block Boot-strap* or MBB in short.

Let be  $\{SP_t\}_{t \in \mathbb{N}}$  the observed Index time series, what are observed are the prices  $\{P_1, \dots, P_n\} \equiv SP_n$ , and consider the log-returns by applying the following transformation:  $r_t = \log(P_t) - (P_{t-1})$  and  $y_t = \log(r_t^2)$ . This transformation is needed to have stationarity.

Let  $l$  be an integer satisfying  $1 \leq l < n$ , here  $n = 2518$ . We define the overlapping blocks  $B_1, \dots, B_N$  ( $N = 518$ ) of length  $l = 2000$  contained in  $SP_{2518}$  as

$$\begin{aligned} B_1 &= (y_1, y_2, \dots, y_l), \\ B_2 &= (y_2, y_3, \dots, y_{l-1}), \\ &\vdots \\ B_N &= (y_{n-l+1}, \dots, y_n) \end{aligned}$$

where  $N = n - l + 1$ .

After that, we will be able to draw the fitted histograms of the realizations of  $\hat{\theta}$  on the 518 new samples and improve our estimations. Obviously this method does not unveil totally new information because the window considered is quite big regarding the original sample. The MBB's improvements increase when the blocks' stretches of time become shorter (relatively to the original sample) and when the original sample becomes larger.

### 4.3. Results.

We use as original sample, the S&P500 Stock Market Index value on a time frame starting the 09/06/2008 and finishing the 09/06/2018 (with daily sampling), with 2518 observations.

We conducted a first indirect estimation of the S&P500 parameters with a starting point arbitrarily chosen by looking at the literature on Heston-S&P500 calibration. This is a tricky point in the process. Indeed the results are heavily influenced by the starting optimization point, which is one of the drawbacks of the technique. Anyway, we can extrapolate the value of parameters with simple statistics, at least for  $\omega$ . We will consider  $\frac{0}{2} = var(r_t)$ . More over, as previously, we should take as starting points parameters respecting  $2\omega > \eta^2$ . With all these considerations, we considered to use  $[0.05; 1.451 \cdot 10^{-4}; 0.002]$  as starting point of the optimization:  $\theta_0$ .

TABLE 6. Hypotheses testing

	$\hat{\theta}$	z-stat	$H_0$ rejected
$\kappa$	0.0080	3.8593	Yes
$\omega$	0.00062792	0.3031	No
$\eta$	0.0034	1.6282	No

With std-error = 0.1039.

The hypothesis was rejected only for  $\kappa$ , we cannot confirm that  $\omega$  and  $\eta$  are different from zero, even-though we were close to reject  $H_0$  for  $\eta$ .

Unfortunately the MBB did not allow us to have strictly different guesses regarding the parameters' value. We found  $\hat{\theta} = [0.01126, 0.00012333, 0.00363]$ , but the fitted histogram in figure 3. clearly show a problem of consistency.

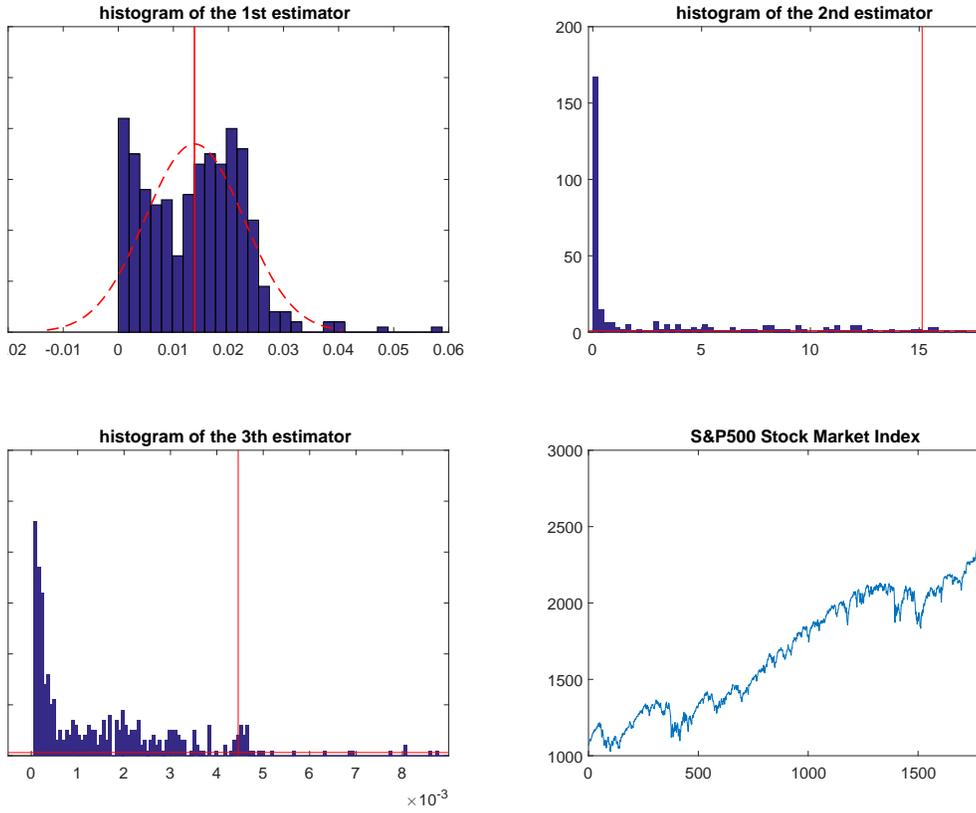


FIGURE 3. Histograms of bootstrapped estimators

## 5. CONCLUSION

Through this work, we exposed the main features of the Indirect Inference methodology and its underlying assumptions and hypotheses. In a nutshell, we need to use a link-function between the auxiliary parameters and the structural parameters that guarantees the uniqueness of the distance-minimization's solutions and a certain level of consistency. Thanks to the two examples and the application, we demonstrated that a simulation's process quality is required to implement II properly. That is interesting, because simulation ability is a fine litmus test of model understanding. Therefore, to attain quality estimates, one has to consider every possible restrictions applying on the parameters values.

We observed that longer time series can improve the consistency of indirect estimators but do not reduce their bias as much. Reduction of bias can be attained through the addition of auxiliary parameters. Even though the marginal impact of an auxiliary is decreasing and become negative as some point (likely for very long auxiliary vectors of parameters).

In sum, we highlighted the main advantage of Indirect Inference: its generality and the ease to adapt it to specific and complex models like stochastic volatility. Nevertheless, its drawbacks and flaws should not be left aside. Firstly, the indirect estimates demonstrated a high sensitivity to initial optimization point. This is a necessity, before the implementation of II, to have fair and plausible guesses about what the  $\{\theta\}$  should look like. Also, the auxiliary model only fits the data, the estimator obtain are also heavily influenced by the sample. Globally, this warns us to take good care in the elaboration of the process and to remain vigilant about the results.

More over, there has been a new trend these last years concerning the requirements researchers have to face to get published. Quantity of Journals now impose the necessity to have replicable results, for the sake of science. Researches involving indirect estimation cannot totally satisfy this point since simulations within the minimization are base on a part of randomness. For example, in the stochastic volatility model exposed in the second example, each simulation uses a Brownian motion that is different from the precedent. Still this does not call into question the results candidly drawn with indirect inference, but one should be careful to data-mining practices that might deliver conclusions based on providential results more than scientific ones.

What should keep the interest of academicians and researchers in the future is the fact that II is an easy-access door to an area of experimentation and empirical opportunities that was previously in-explorable for more-classical methods. II is a flexible skeleton on which can attached other statistical tools in the aim to construct efficient and consistent estimators. If correctly applied, there is no doubt that indirect inference can bring satisfying results for whatever problem where methods relying on likelihood functions cannot find solutions.

## APPENDIX A. COMPLETE TABLES

TABLE 7. Results from the simulations of an MA(3) matched with AR(q), unconstrained with remote starting point

(q-p)	$\theta$	$\hat{\theta}^{1000}$	T=100	T=200	T=500	T=1000
1	1	0.3776	-0.0149 (0.2206)	-0.0137 (0.1884)	-0.0020 (0.0042)	-0.0025 (0.0064)
	2	0.8487	0.0468 (2.1919)	0.0260 (0.6760)	0.0136 (0.1839)	0.0085 (0.0717)
	3	0.1807	0.0286 (0.8176)	0.0010 (0.0010)	0.0113 (0.1281)	-0.0040 (0.0161)
	4	0.0627	0.0058 (0.0336)	-0.0031 (0.0099)	0.0030 (0.0093)	0.0023 (0.0052)
2	1	0.3682	-0.0092 (0.0855)	-0.0185 (0.3427)	-0.0024 (0.0058)	-0.0119 (0.1406)
	2	0.8605	0.0300 (0.9001)	0.0495 (2.4503)	0.0127 (0.1604)	0.0203 (0.4111)
	3	0.1959	0.0123 (0.1515)	0.0111 (0.1232)	0.0004 (0.0002)	0.0112 (0.1265)
	4	0.0454	0.0039 (0.0152)	0.0128 (0.1651)	-0.0014 (0.0019)	-0.0150 (0.2248)
3	1	0.3718	0.0322 (1.0349)	-0.0124 (0.1529)	-0.0109 (0.1181)	-0.0083 (0.0689)
	2	0.8726	0.0536 (2.8692)	0.0528 (2.7911)	0.0232 (0.5384)	0.0324 (1.0472)
	3	0.1898	0.0089 (0.0791)	0.0023 (0.0052)	0.0011 (0.0012)	0.0051 (0.0262)
	4	0.0748	0.0213 (0.4537)	0.0018 (0.0033)	-0.0008 (0.0007)	0.0144 (0.2076)
4	1	0.3678	-0.0126 (0.1590)	-0.0146 (0.2133)	-0.0073 (0.0531)	-0.0123 (0.1501)
	2	0.8710	0.0508 (2.5837)	0.0287 (0.8254)	0.0285 (0.8149)	0.0307 (0.9444)
	3	0.1768	0.0286 (0.8192)	0.0006 (0.0004)	0.0018 (0.0032)	-0.0079 (0.0618)
	4	0.0723	0.0147 (0.2171)	0.0050 (0.0251)	0.0240 (0.5777)	0.0119 (0.1426)
5	1	0.3718	-0.0151 (0.2273)	-0.0121 (0.1466)	-0.0095 (0.0910)	-0.0083 (0.0686)
	2	0.8616	0.0516 (2.6582)	0.0307 (0.9442)	0.0183 (0.3365)	0.0214 (0.4574)
	3	0.1823	0.0034 (0.0117)	-0.0136 (0.1850)	-0.0048 (0.0234)	-0.0023 (0.0055)
	4	0.0785	-0.0086 (0.0737)	0.0238 (0.5687)	0.0199 (0.3977)	0.0182 (0.3297)

With  $\theta = [0.3800; 0.84025; 0.018468; 0.06036]$ , Root mean square errors in parentheses.

TABLE 8. Results from the simulations of an MA(3) matched with AR(q), constrained with remote starting point

(q-p)	$\theta$	$\hat{\theta}^{100}$	T=100	T=200	T=500	T=1000
1	1	0.3789	-0.0012 (0.0014)	-0.0029 (0.0083)	-0.0095 (0.0895)	-0.0045 (0.0200)
	2	0.8663	0.0260 (0.6784)	0.0081 (0.0648)	0.0157 (0.2463)	0.0209 (0.4255)
	3	0.0585	0.0401 (1.6050)	0.0030 (0.0090)	0.0021 (0.0044)	0.0002 (0.000)
	4	0.0641	0.0037 (0.0137)	-0.0191 (0.3663)	-0.0099 (0.0980)	-0.0057 (0.0327)
2	1	0.3843	0.0042 (0.0177)	-0.0070 (0.0486)	-0.0112 (0.1253)	-0.0130 (0.1700)
	2	0.8593	0.0190 (0.3605)	0.0300 (0.8995)	0.0258 (0.6645)	0.0355 (1.2601)
	3	0.0462	0.0277 (0.7673)	0.0264 (0.6987)	0.0017 (0.0029)	0.0037 (0.0138)
	4	0.0533	-0.0071 (0.0501)	-0.0005 (0.0002)	-0.0170 (0.2876)	0.0010 (0.0009)
3	1	0.3676	-0.0125 (0.1555)	-0.0148 (0.2205)	-0.0109 (0.1194)	-0.0097 (0.0932)
	2	0.8705	0.0302 (0.9123)	0.0350 (1.2241)	0.0297 (0.8816)	0.0230 (0.5287)
	3	0.0635	0.0450 (2.0232)	0.0335 (1.1251)	-0.0120 (0.1431)	-0.0051 (0.0258)
	4	0.0793	0.0189 (0.3580)	0.0008 (0.0007)	-0.0073 (0.0539)	-0.0016 (0.0025)
4	1	0.3739	-0.0062 (0.0387)	-0.0092 (0.0843)	-0.0113 (0.1282)	-0.0099 (0.0978)
	2	0.8717	0.0314 (0.9853)	0.0375 (1.4055)	0.0350 (1.2264)	0.0325 (1.0594)
	3	0.0242	0.0057 (0.0329)	-0.0023 (0.0052)	0.0130 (0.1701)	-0.0066 (0.0435)
	4	0.0538	-0.0066 (0.0434)	-0.0009 (0.0007)	0.0062 (0.0388)	0.0023 (0.0052)
5	1	0.3799	-0.0002 (0.000)	-0.0168 (0.2823)	-0.0091 (0.0828)	-0.0152 (0.2316)
	2	0.8818	0.0409 (1.6757)	0.0395 (1.5628)	0.0295 (0.8719)	0.0381 (1.4478)
	3	0.0312	0.0127 (0.1624)	0.0093 (0.0868)	-0.0027 (0.0074)	-0.0120 (0.1430)
	4	0.0522	-0.0082 (0.0675)	0.0020 (0.0042)	-0.0007 (0.0005)	0.0013 (0.0017)

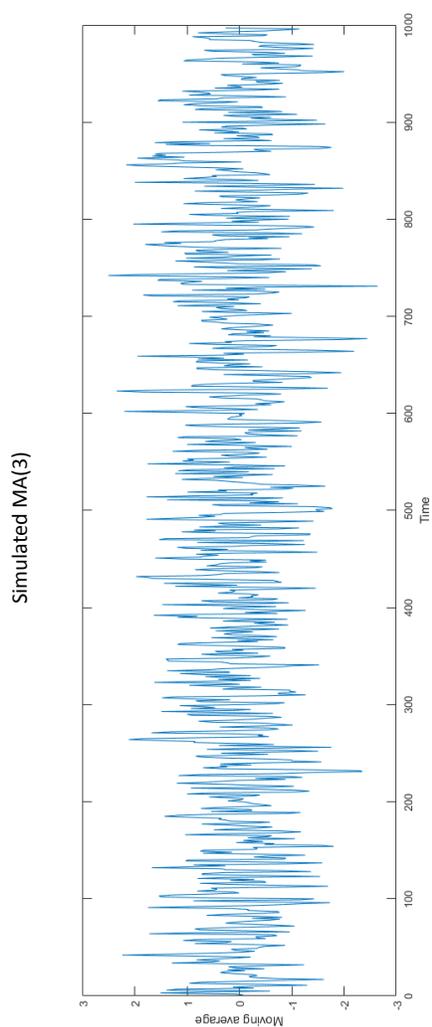
With  $\theta = [0.3800; 0.84025; 0.018468; 0.06036]$ , Root mean square errors in parentheses.

TABLE 9. Results from the simulations of an MA(3) matched with AR(q), constrained with close starting point

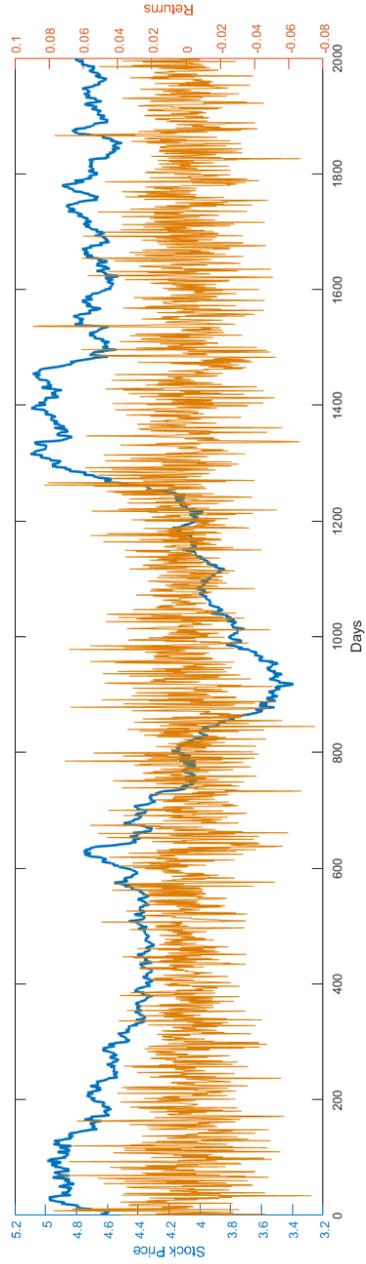
(q-p)	$\theta$	$\hat{\theta}^{100}$	T=100	T=200	T=500	T=1000
1	1	0.3769	-0.0032 (0.0103)	0.0057 (0.0324)	-0.0003 (0.0001)	-0.0049 (0.0237)
	2	0.8440	0.0037 (0.0136)	-0.0077 (0.0600)	0.0045 (0.0207)	0.0193 (0.3714)
	3	0.0561	0.0376 (1.4130)	0.0120 (0.1436)	0.0029 (0.0084)	0.0074 (0.0544)
	4	0.0399	-0.0205 (0.4217)	-0.0256 (0.6549)	-0.0134 (0.1807)	-0.0050 (0.0248)
2	1	0.3786	-0.0015 (0.0022)	-0.0077 (0.0585)	-0.0010 (0.0009)	-0.0076 (0.0578)
	2	0.8559	0.0146 (0.2117)	0.0237 (0.5635)	0.0100 (0.0999)	0.0210 (0.403)
	3	0.0512	0.0327 (1.0685)	0.0145 (0.2112)	0.0082 (0.0671)	0.0020 (0.0039)
	4	0.0484	-0.0120 (0.1440)	-0.0052 (0.0273)	-0.0106 (0.1118)	0.0013 (0.0018)
3	1	0.3704	-0.0097 (0.0949)	-0.0057 (0.0330)	-0.0042 (0.0175)	-0.0011 (0.0013)
	2	0.8411	0.0008 (0.0006)	0.0126 (0.1598)	0.0117 (0.1363)	0.0144 (0.202)
	3	0.0351	0.0166 (0.2756)	-0.0045 (0.0200)	0.0079 (0.0619)	0.0055 (0.0305)
	4	0.0533	-0.0071 (0.0508)	-0.0139 (0.1926)	-0.0076 (0.0574)	0.0004 (0.0002)
4	1	0.3721	-0.0080 (0.0646)	-0.0050 (0.0250)	-0.0074 (0.0541)	-0.0017 (0.0028)
	2	0.8669	0.0266 (0.7084)	0.0247 (0.6081)	0.0228 (0.5197)	0.0172 (0.2955)
	3	0.0389	0.0204 (0.4161)	0.0115 (0.1314)	0.0001 (0.0000)	0.0024 (0.0058)
	4	0.0489	-0.0115 (0.1327)	-0.0068 (0.0466)	0.0013 (0.0017)	0.0001 (0.0000)
5	1	0.3756	-0.0045 (0.0203)	-0.0020 (0.0041)	-0.0046 (0.0215)	0.0001 (0.0000)
	2	0.8665	0.0262 (0.6890)	0.0183 (0.3354)	0.0170 (0.2905)	0.0103 (0.1067)
	3	0.0214	0.0029 (0.0084)	-0.0002 (0.0001)	0.0004 (0.0001)	0.0024 (0.0058)
	4	0.0466	-0.0138 (0.1905)	-0.0018 (0.0033)	0.0023 (0.0051)	-0.0005 (0.0002)

With  $\theta = [0.3800; 0.84025; 0.018468; 0.06036]$ , Root mean square errors in parentheses.

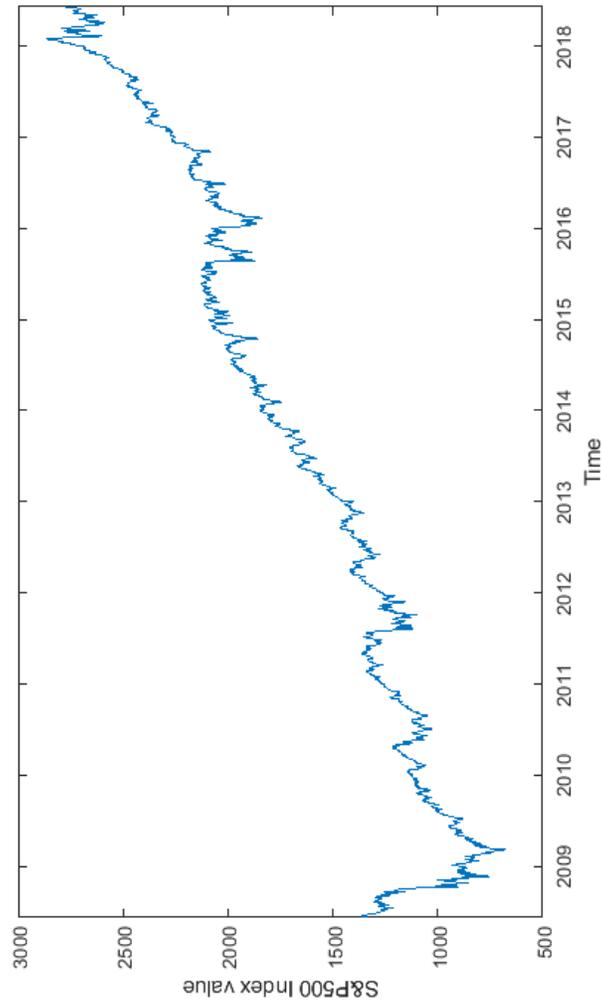
APPENDIX B. FIGURES



Simulated Stock Price and Returns with Heston model



S&P500 Stock Market Index, from 09/06/2008 to 09/06/2018



## APPENDIX C. MATLAB CODES

**MA artificial generation**

```
function y=MASIM_GEN(theta,z)

m=length(theta)-1;
sigma2_epsilon=theta(1);
epsilon=z*sqrt(sigma2_epsilon);
theta_MA=theta(2:end);
[-,eps_lags]=newlagmatrix(epsilon,m);
y=eps_lags*theta_MA+epsilon(m+1:end);
```

*Published with MATLAB® R2017b*

## MA Main file

## Setting of simulation parameters

```

global v
v=[100,200,500,1000]; % time-length of each sample
global m;
m=3; % number of lags
global Q
q=m+1; % number of auxiliary parameters
Q=[q,q+1,q+2,q+3,q+4];
BVMA=zeros(m+1,length(V),length(Q));
global M;
M=100; % number of Monte-Carlo iterations

theta_true=[0.3801,0.8403,0.01847,0.0604]';
theta=0.2*theta_true; % Starting point of each optimization
theta_hat=zeros(rows(theta_true),M);

```

## Generation of the simulations and estimations

```

for v=1:length(V)
for q=1:length(Q)
global T
T=V(v);
hbar = parfor_progressbar(M,sprintf('Progress for sample size = %d and q = %d',v(v),q(q)));

parfor jj=1:M
z=randn(T,1);
y=MASTM_GEN(theta_true,z);

% Fitting with AR(q) process:

[y_new,y_lags]=newlagmatrix(y,Q(q));
results=ols(y_new,y_lags);
beta_hat=[results.beta;results.sige]; % Estimation of beta by the auxiliary model

% Simulations:

S=200; % Number of optimization
E=randn(T,S);
options=optimset('Display','iter');

weight=w_Matrix(beta_hat,y);

% Boundaries are needed for the constraint optimisation:
lb=[0;-1;-1;-1]; % Sets the lower bounds of the structural parameters
ub=[1;1;1;1]; % Sets the upper bounds of the structural parameters

theta_hat(:,jj)=fmincon('II_GENMA_AR',theta,[],[],[],[],lb,ub,NONLCON(theta,z),options,beta_hat,E,weight);

hbar.iterate(1);
end
close(hbar);
toc
[Bias,Rmse]=MCA(theta_hat,theta_true);
BVMA(:,v,q)=Bias;

```

**II's simulations**

```
function Q=II_GENMA_AR(theta,beta_hat,E,w)
beta_S=zeros(cols(E),rows(beta_hat));
q=rows(beta_hat)-1;
for i=1:cols(E)
    y_i=MASIM_GEN(theta,E(:,i));
    [y_new,y_lags]=newlagmatrix(y_i,q);
    results=ols(y_new,y_lags);
    beta_S(i,:)=[results.beta;results.sige]';
end
```

## Generation of the distance matrix

```
beta_bar=mean(beta_S)';
differ=beta_bar-beta_hat;
Q=differ'*w*differ; % Criterion
```

*Published with MATLAB® R2017b*

## Simulation of discretized trajectories from the Heston model

```

function [p,sigma2]=SV(theta,S0,T,N,w);

% S0=initial condition;
% theta= parameters;
% eta= vol-of-vol parameter
% N=number of steps;
% M=number of paths;
% output:
% S: N+1xM vector of simulated trajectories of prices
% sigma2: N+1xM vector of simulated trajectories of variances

kappa=theta(1);
omega=theta(2);
eta=theta(3);
dt=T/N; % size of small interval
p=zeros(N+1,1); % initialize matrix of S
sigma2=zeros(N+1,1); % initialize matrix of variances
p(1)=log(S0); % S(0)

sigma2(1)=omega; % sigma(0)
u1=w(:,1); % innovations to prices
u2=w(:,2); % innovations to variances
minvalue=0.0000001;

i = 1;
while i <= N; % loop over N
pi = p(i);
sigma2i = sigma2(i);
sigma2(i+1) = sigma2i + kappa*(omega-sigma2i)*dt + sqrt(dt)*eta*sqrt(sigma2i).*u2(i+1);
sigma2(i+1)=max(sigma2(i+1),minvalue);
p(i+1) = pi + sqrt(dt)*sqrt(sigma2(i+1))*u1(i+1);
i = i + 1;
end;

```

*Published with MATLAB® R2017b*

The codes for the rest of the second example is similar to its equivalent from the first example on MA.

## Application, without Block-Bootstrapping

```
load('SP'); % original sample
dt=1; % daily sampling
```

### Setting & Fitting of the auxiliary model

```
logP=log(SP);
x=diff(logP);
x(x==0)=[];
x_mean=mean(x);
x_mean=x_mean*ones(2516,1);
omega_zero=(1/(length(SP))*sum((x-x_mean).^2));
theta=[0.005, omega_zero, 0.005]'; % starting point of the optimization
y=log(x.^2);
T=length(y);
n_lags=10;
[y_new,y_lags]=newlagmatrix(y,n_lags,1);
results=ols(y_new,y_lags);

beta_hat=[results.beta;results.sige]; % Beta by the auxiliary model
```

### Simulation & Minimization

```
S=200; % Number of optimizations
weight=w_Matrix(beta_hat,y);%w_Matrix(beta_hat,y);
E=randn(T+1,2,S);
% Boundaries are needed for the constraint optimisation:
lb=[eps;eps;eps]; % Sets the lower bounds of the structural parameters
ub=[1;1;1]; % Sets the upper bounds of the structural parameters
options=optimset('Display','iter');
global s;
s=ones(length(theta),n_lags);
theta_hat=fmincon('II_SP',theta,[],[],[],[],lb,ub,[],options,beta_hat,E,weight,T,SP);
```

### z-test

```
z_t=(sqrt(T))*(theta_hat/mean(s(:,1)))
```

*Published with MATLAB® R2017b*

## Application, with Block-Bootstrapping

### Setting the Simulation and Bootstrapping's parameters

```

Load('SP'); % original sample
dt=1; % daily sampling
logP=log(SP);
x=diff(logP);
x(x==0)=[];
x_mean=mean(x);
x_mean=x_mean*ones(length(x),1);
omega_zero=(1/(length(SP))*sum((x-x_mean).^2));
theta=[0.005, omega_zero, 0.005]'; % starting point of the optimization

window=2000;
P=ones(window,1);
steps=length(SP)-window; % number of subsamples
theta_hat=zeros(3,steps);

```

### Simulations & Bootstrapping

```

for ii=1:steps
    jj=1;
    while jj<=window
        P(jj,1)=SP(ii+jj,1);
        jj=jj+1;
    end
    logP=log(P);
    x=diff(P);
    x(x==0)=[];
    y=log(x.^2);
    T=length(y);
    n_lags=20;
    [y_new,y_lags]=newlagmatrix(y,n_lags,1);
    results=ols(y_new,y_lags);

    beta_hat=[results.beta;results.sige];

    weight=w_Matrix(beta_hat,y);
    E=randn(T+1,2,200);

    lb=[eps;eps;eps]; % Sets the lower bounds of the structural parameters
    ub=[1;1;1]; % Sets the upper bounds of the structural parameters
    options=optimset('Display','iter');
    global s;
    s=ones(length(theta),n_lags,ii);
    theta_hat(:,ii)=fmincon('II_SP',theta,[],[],[],[],lb,ub,[],options,beta_hat,E,weight,T,P);
end
mean=mean(theta_hat');

```

Published with MATLAB® R2017b

## APPENDIX D. SUMMARY

Most of modern researchers are, at one time or another, confronted to the task of building generative models of a process or assemblage on which they work in the hope to grasp a substantial bite of the nature's knowledge. Being careful scientists, they usually do a conscientious job of trying to include their guesses for what are all the most important mechanisms. The result is something that can possibly be set through to produce a simulation of the process of interest. But, (realistic) models often contain some unknown parameters, some hidden truth that we will denominate under the vector  $\theta$ .

It is desirable to tune those unknown features within models to match the data or see if, despite researchers' best efforts, there are aspects of the data which a model just can't match.

Very often, models too loyally built are too complicated for the common practitioner to appeal to any of the usual estimation methods of statistics<sup>21</sup>. Because some models aim for scientific adequacy rather than statistical tractability, it will often happen that there is no way to even calculate the likelihood of a given data set under the unknown parameters in closed form, which would rule out even numerical likelihood maximization.

Indeed, in the over average complicated models, such as stochastic volatility models in continuous time, ARCH factor models, nonlinear random parameter models, the likelihood function is impossible to calculate. Procedures for estimating parameters were then born. We can quote for example, the general method of moments (GMM), the quasi-maximum likelihood (QML) method, Bayesian methods, simulated expectation maximization (SEM) method, variance reduction (Importance Sampling, IS). The results of these different methods are quite diverse. Indeed, the IS method is complicated to implement, GMM and QML methods are not efficient. The SEM and Bayesian methods provide better results.

More generally there has been a division of opinion in the academies between frequentist and Bayesian statisticians regarding the efficiency of the two main inference methods: General Method of Moments and Maximum Likelihood Estimation. Frequentists dismissed the method-of-moments in favor of the MLE, Bayesians never did so. Anyway, both sides have always thoroughly and intentionally focused on the likelihood (frequentists on the location of its maximum and its curvature while Bayesians focus on its entire shape). As sustained by F.Diebold on his blog<sup>22</sup>, this division is what drives the European views regarding estimation methods. Especially American econometricians hold the GMM in high esteem, to which they attribute almost sacred merits, this is not the case of European academicians who

---

<sup>21</sup>Typically, in finance the underlying mathematical models have become more complex. They take more and more parameters to adapt to the data. We can mention, for example, as phenomena to be taken into account: dependence of volatility on the past, non-linear dynamics, variance, asymmetry to identify the permanent and transient shocks, unobservable factors, leptokurtic effect, etc.

<sup>22</sup>*No Hesitations*, 22 July 2013,

found the hype about GMM exaggerated. This is mainly due to how GMM is advertised as potentially useful when there is a likelihood at hand, in other situation the method-of-moments loses its power. Even worse, model moments may also be analytically intractable.

Yet simulation is possible; it seems like there should be some way of saying whether the simulations look like the data. This is where the breaking point comes in. The solution could first come out as an implementation of GMM by simulation: Simulated Method of Moments. By simulating models, in combination with the fine choice of parameters minimizing divergence between simulated and data moments, one can consistently estimate those models' parameters. That is really game-changing: we no longer need to work-out complex likelihoods (even for those that are available in some terms).

But SSM is a peculiar case from of a more general method: Indirect Inference, which itself can be seen as a generalized form of the method of moments of Duffie and Singleton (1993) and has papers of Tony Smith<sup>23</sup> and C.Gourieroux, A.Monfort and E.Renault (1993, J. Applied Econometrics) as founding articles. It introduces a new model, called the "auxiliary model", which is miss-specified and typically not even generative, but is easily fit to the data, and to the data alone. The auxiliary model has its own parameter vector  $\beta$ , with an estimator  $\hat{\beta}$ . These parameters describe aspects of the distribution of observables, and the idea of indirect inference is that we can estimate the generative parameters  $\theta$  by trying to match those aspects of observations. It further bases itself on a metric-set wisely chosen to minimize the distance between the auxiliary estimators based on the observed data on one hand and on the other hand the auxiliary estimators of the simulated data.<sup>24</sup> Indirect Inference assumes a relatively easy simulation of the models studied. It allows, from a higher calculation time, a reduction of the bias compared to more traditional methods.

Unlike other methods, the moments that guide the estimation of the parameters of the economic model are themselves the parameters of the auxiliary model. If the auxiliary model comes close to providing a correct statistical description of the structural model, then indirect inference comes close to matching the asymptotic efficiency of maximum likelihood.

Its main advantages lie in its generality. Unlike other bias reduction methods, such as those based on explicit analytical expressions for the bias function or key terms in an asymptotic expansion of the bias, the indirect inference technique calibrates the bias function by simulation and therefore does not require a given explicit form for the bias function or its expansion.

Some sustain that evaluation of the likelihood became as trivial as simulating. As Andrew Harvey and others have emphasized for decades, for any linear model cast in finite-dimensional state-space form one can simply run the Kalman filter and then evaluate the Gaussian likelihood via a prediction-error decomposition. These

<sup>23</sup>1990, Duke Ph.D. Dissertation, and 1993, J. Applied Econometrics

<sup>24</sup>The auxiliary estimator may be the Maximum Likelihood Estimator of the auxiliary model (usually simpler), or the estimator corresponding to the approximate likelihood of the initial model.

tools can provide complete likelihood analysis in general non-Gaussian environments. But II remains a more suitable path for cases where model understanding is of interest and is clearer in its implementation. Moreover, II achieves desirable consistency properties under misspecification more easily (as exposed in Rossi and De Magistris 2018). Finally, even-though the more classical approaches like quasi-maximum likelihood method or Bayesian models deliver more satisfying results in statistical terms, there are areas where they simply cannot be used or at least not properly.

For many decades now, interest in analysis and modeling in finance and economics has been growing. Financial data is growing. The emergence of computer tools has made its further development and exploitation possible. The rising of the average computational capacity, following the Moore's law of computing power<sup>25</sup>, has made simulation-based procedures even more competitive than before.

The aim of this thesis is to present the II methodology and implementation (section 2: Indirect Inference, Theoretical aspects) and to discuss it around two examples (that are moving average and Heston-model, in section 3: Examples), followed by an attempt to apply this method on the S&P500 Index (section 4: Application). In this paper we will discover the recipe for an II implementation, the threats to its robustness and solutions to tackle them.

Through this work, we exposed the main features of the Indirect Inference methodology and its underlying assumptions and hypotheses. In a nutshell, we need to use a link-function between the auxiliary parameters and the structural parameters that guarantees the uniqueness of the distance-minimization's solutions and a certain level of consistency. Thanks to the two examples and the application, we demonstrated that a simulation's process quality is required to implement II properly. That is interesting, because simulation ability is a fine litmus test of model understanding. Therefore, to attain quality estimates, one has to consider every possible restrictions applying on the parameters values.

We observed that longer time series can improve the consistency of indirect estimators but do not reduce their bias as much. Reduction of bias can be attained through the addition of auxiliary parameters. Even though the marginal impact of an auxiliary is decreasing and become negative as some point (likely for very long auxiliary vectors of parameters).

In sum, we highlighted the main advantage of Indirect Inference: its generality and the ease to adapt it to specific and complex models like stochastic volatility. Nevertheless, its drawbacks and flaws should not be left aside. Firstly, the indirect estimates demonstrated a high sensitivity to initial optimization point. This is a necessity, before the implementation of II, to have fair and plausible guesses about what the  $\{\theta\}$  should look like. Also, the auxiliary model only fits the data, the

---

<sup>25</sup>Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years. The observation is named after Gordon Moore, the co-founder of Fairchild Semiconductor and Intel, whose 1965 paper described a doubling every year in the number of components per integrated circuit ("Moore's Law." Wikipedia, Wikimedia Foundation, 12 June 2018, en.wikipedia.org/wiki/Moore's\_law.)

estimator obtain are also heavily influenced by the sample. Globally, this warns us to take good care in the elaboration of the process and to remain vigilant about the results.

More over, there has been a new trend these last years concerning the requirements researchers have to face to get published. Quantity of Journals now impose the necessity to have replicable results, for the sake of science. Researches involving indirect estimation cannot totally satisfy this point since simulations within the minimization are base on a part of randomness. For example, in the stochastic volatility model exposed in the second example, each simulation uses a Brownian motion that is different from the precedent. Still this does not call into question the results candidly drawn with indirect inference, but one should be careful to data-mining practices that might deliver conclusions based on providential results more than scientific ones.

What should keep the interest of academicians and researchers in the future is the fact that II is an easy-access door to an area of experimentation and empirical opportunities that was previously in-explorable for more-classical methods. II is a flexible skeleton on which can attached other statistical tools in the aim to construct efficient and consistent estimators. If correctly applied, there is no doubt that indirect inference can bring satisfying results for whatever problem where methods relying on likelihood functions cannot find solutions.

## REFERENCES

- Aiř-Sahalia, Yacine. Handbook of Financial Econometrics. Amsterdam: Elsevier NH, 2011.
- Azencott, Robert, and Yutheeka Gadhyan. "Accuracy of Maximum Likelihood Parameter Estimators for Heston Stochastic Volatility SDE." *Journal of Statistical Physics*, vol. 159, no. 2, 2015, pp. 393âĀĶ420., doi:10.1007/s10955-014-1120-x.
- Broto, Carmen, and Esther Ruiz. "Estimation Methods for Stochastic Volatility Models: a Survey." *Journal of Economic Surveys*, vol. 18, no. 5, 2004, pp. 613âĀĶ649., doi:10.1111/j.1467-6419.2004.00232.x.
- Corsi, Fulvio, and Roberto Renò. "Discrete-Time Volatility Forecasting with Persistent Leverage Effect and the Link with Continuous-Time Volatility Modeling." *SSRN Electronic Journal*, 2010, doi:10.2139/ssrn.1316953.
- Davies, Robert, et al. "The Application of Time Series Modelling and Monte Carlo Simulation: Forecasting Volatile Inventory Requirements." *Applied Mathematics*, vol. 05, no. 08, 2014, pp. 1152âĀĶ1168., doi:10.4236/am.2014.58108.
- Diebold, Francis. "GMM, the Strange American Estimator." *No Hesitations*, 22 July 2013, fxdiebold.blogspot.com/2013/07/gmm-strange-american-estimator.html.
- Drăgulescu, Adrian A, and Victor M Yakovenko. "Probability Distribution of Returns in the Heston Model with Stochastic Volatility\*." *Quantitative Finance*, vol. 2, no. 6, 2002, pp. 443âĀĶ453., doi:10.1080/14697688.2002.0000011.
- Efron, Bradley, and Robert J. Tibshirani. "Efficient Bootstrap Computations." *An Introduction to the Bootstrap*, 1993, pp. 338âĀĶ357., doi:10.1007/978-1-4899-4541-9\_23.
- Gallant, A. Ronald, and George Tauchen. "Which Moments to Match?" *Econometric Theory*, vol. 12, no. 04, 1996, p. 657., doi:10.1017/s0266466600006976.
- Gouriéroux, Christian, and Alain Monfort. "Indirect Inference." *Simulation-Based Econometric Methods*, 1997, pp. 61âĀĶ92., doi:10.1093/0198774753.003.0004.
- Ghysels, et al. "Simulation Based Inference in Moving Average Models." *Annales D'économie Et De Statistique*, no. 69, 2003, p. 85., doi:10.2307/20076364.
- Hamilton, James D. *Time Series Analysis*. Levant Books, 2012.
- Harvey, A., et al. "Multivariate Stochastic Variance Models." *The Review of Economic Studies*, vol. 61, no. 2, Jan. 1994, pp. 247âĀĶ264., doi:10.2307/2297980.
- Kloeden, Peter E., and Eckhard Platen. "Numerical Solution of Stochastic Differential Equations." 1992, doi:10.1007/978-3-662-12616-5.
- Kreiss, Jens-Peter, and Soumendra Nath Lahiri. "Bootstrap Methods for Time Series." *Time Series Analysis: Methods and Applications Handbook of Statistics*, 2012, pp. 3âĀĶ26., doi:10.1016/b978-0-444-53858-1.00001-6.
- Monfardini, Chiara. "Estimating Stochastic Volatility Models Through Indirect Inference." *The Econometrics Journal*, vol. 1, no. 1, 1998, pp. 113128., doi:10.1111/1368-423x.11007.
- Pastorello, Sergio, and al. "Statistical Inference for Random-Variance Option Pricing." *Journal of Business amp; Economic Statistics*, vol. 18, no. 3, 2000, pp. 358âĀĶ367., doi:10.1080/07350015.2000.10524876.

Rossi, Eduardo, and Paolo Santucci De Magistris. "Indirect Inference with Time Series Observed with Error." SSRN Electronic Journal, 2018, doi:10.2139/ssrn.3106274.

Rutledge, Ann, and Sylvain Raynes. Elements of Structured Finance. Oxford University Press, USA, 2010.

Shephard, Neil. "Statistical Aspects of ARCH and Stochastic Volatility." Time Series Models, 1996, pp. 1-67., doi:10.1007/978-1-4899-2879-5\_1.

Shu, Jinghong, and Jin E. Zhang. "Pricing S&p 500 Index Options Under Stochastic Volatility With The Indirect Inference Method." Journal of Derivatives Accounting, vol. 01, no. 02, 2004, pp. 171-186., doi:10.1142/s021986810400021x.

The New Palgrave Dictionary of Economics. Palgrave Macmillan, 2018.

*-Bibliography built on the basis of the Chicago Manual of Style 16th edition (full note), formatted by BibMe.org.-*

*Email address*, LUISS Guido Carli, Rome: `jean.duchesne@studenti.luiss.it`

*Email address*, Solvay Brussels School, Brussels: `jean.duchesne@ulb.ac.be`