# EMPIRICAL LIKELIHOOD ESTIMATION OF AGENT-BASED MODELS

MINH KHOA NGUYEN* STEVE PHELPS† WING LON NG‡

March 2011

**Abstract**

This paper addresses a major methodological issue of Agent-Based (AB) models which has prevented them from being accepted by the mainstream research community. Any scientific model or theory has to be testable for example by comparing predictions which are contingent on the model against empirical observations. If they coincide the model can be considered as provisionally true, else it is falsified and at least one of the model axioms is false. However, in order to derive testable implications a meaningful calibration and estimation procedure is required, and this issue has been given relatively little attention in the AB literature. This paper introduces and examines an estimation procedure for AB models which is based on a statistical methodology called Empirical Likelihood.

*Keywords:* Agent-Based Models, Estimation, Calibration, Empirical Likelihood.

*Corresponding author. mknguy@essex.ac.uk. Centre for Computational Finance and Economic Agents, University of Essex, Colchester, CO4 3SQ, UK.

†sphelps@essex.ac.uk. Essex Business School, University of Essex, Colchester, CO4 3SQ, UK.

‡Centre for Computational Finance and Economic Agents, University of Essex, Colchester, UK.

## 1 Introduction

Although Agent-Based models (AB models) have become increasingly popular in Finance research (see (Chiarella et al., 2009) and (Wagener and Hommes, 2009) for a review) its empirical validation is still an open issue. In general the data generating process of AB model outcomes are not analytically tractable, therefore standard techniques such as empirical likelihood or generalized method of moments cannot be applied for calibration and estimation. In fact, the principal approach to validate a model by assessing its ability to explain empirical phenomena by finding the parameter configuration that best matches empirical facts e.g. by minimising the "distance" of the model outcomes with respect to the empirical counterpart, measured by an objective function, has various methodological issues in the context of AB models. In particular, it is not clear which empirical counterparts and distance measure should be chosen (Fagiolo et al., 2007) and even with a particular choice, there is a fundamental problem comparing AB results with empirical observations. AB models as a Monte Carlo based simulation technique produce multiple realised samples that have to be compared to a single empirical observation and therefore any calibration method of AB models has to deal with Monte Carlo variance introduced by the simulation. As a consequence minimising the distance of an AB model, given a specific distance measure with respect to a specific empirical observation, results in finding an optima of an approximated, stochastic objective function (Winker et al., 2007) which in turn requires robust optimisation techniques. Therefore the following aspects need careful inspection:

1. Choice of the objective function:
   (a) Choice of empirics that the AB model is calibrated against
   (b) Choice of distance measure of the simulation outcome to its empirical counter part
2. Choice of the optimisation technique

For a meaningful calibration or estimation these choices should enable a discrimination of the considered AB model amongst different parameter settings in the presence of Monte

Carlo variance and eventually should help to compare different AB models. Current literature on calibration of financial AB models have proposed two approaches:

1. SMM: Simulated Method of Moments (see (Duffie and Singleton, 1993), (Franke, 2009), (Winker et al., 2007)),
2. IIM: Indirect Inference Method (see (Gourieroux et al., 1993))

All these approaches calibrate AB models against stylized facts of financial time series that are described by some empirical moments or statistics (here we denote $d$ empirical moments shortly with $y_e = (m_1^e, \ldots, m_d^e)$), however using different "distance" measures. Generically a financial AB model produces for each simulation run $i = 1, \ldots, n$ and to a given parameter configuration vector $\gamma$ a sample (price) path

$$x^i_{t\in[0,T]}\left(\gamma\right)$$

or

$$x_{t\in[0,T]}\left(\omega_i\right)\left(\gamma\right),$$

$\omega_i \in \Omega$ stressing that the sample path is also a realization of the random generators embedded in the simulation. Corresponding to the empirical moments or statistics $y_e$ the AB model provides for each simulation run $i = 1, \ldots, n$ and parameter setting $\gamma$ a simulated moment $y_i$:

$$x^i_{t\in[0,T]}\left(\gamma\right) \to y_i\left(\gamma\right) = \left(m_1^i\left(\gamma\right), \ldots, m_d^i\left(\gamma\right)\right).$$

Given this population of simulated moments the SMM minimises

$$\min_{\gamma} \left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i\left(\gamma\right) - y_e\right)\right)' D \left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i\left(\gamma\right) - y_e\right)\right),$$

which is the weighted mean square error of the simulated moments against the empirical moment with respect to $\gamma$ and weight matrix D. In other words the SMM minimises the weighted average distance of the simulated moments to its empirical counterpart. The IIM minimises the distance between simulator outcome and empirical observation using an auxiliary model, that ideally captures all salient features of the data while having a closed form representation, allowing the use of conventional estimation procedures (e.g. maximum likelihood or generalized method of moments). Suppose there is a one to one mapping function $c$ between the auxiliary model parameter vector $\lambda$ and the simulator parameter vector $\gamma$: $\lambda = c(\gamma)$, then the IIM minimises the the distance between the estimate $\lambda_e$, obtained by estimating the auxiliary model with empirical observations, and $\lambda_y$ the estimate of the auxiliary model with simulation outcome with configuration $\gamma$.

Because the SMM is merely a curve fitting procedure and the IIM approach suffer a drawback of using an auxiliary model, inducing a source of arbitrariness, we will employ the statistical approach introduced in (Owen, 1990), called empirical likelihood. It is the nonparametric analogue of the parametric likelihood and as its counterpart it provides efficient estimators and confidence intervals for hypothesis testing. Similar to the classic likelihood it allows us to compute the likelihood of a hypothesis $H : \theta = \theta_H$ given a sample $Y_1, .., Y_n$, where the hypothesis can be formulated for any parameter $\theta = T(F)$, where $T$ is a function of the **unknown** distribution $F$. Let $W(\theta_H)$ (as in (7)) denote the empirical likelihood of observing $\theta_H$ given a specific sample $Y_1, .., Y_n$. In the given AB context we get for each $\gamma$ a population of moments: $Y_1(\gamma), .., Y_n(\gamma)$ iid according to a common $F_\gamma$, then $W_\gamma(\theta_H)$ denotes the empirical likelihood of observing $\theta_H$ for a given sample $Y_1(\gamma), .., Y_n(\gamma)$. Naturally, the idea for calibrating or estimating the AB model could be, for example, to maximize the likelihood of the average simulated moment $\mu_\gamma = E_{F_\gamma}[Y(\gamma)]$ to observe the empirical moment $y_e = (m_1^e, \ldots, m_d^e) = \mu_e$; that is we search for a parameter configuration $\hat{\gamma}$ such that:

$$\hat{\gamma} = \underset{\gamma}{argmax}\left[W_\gamma(\mu_e)\right].$$

In other words we look for the configuration $\hat{\gamma}$ of the AB model at which we (on average) most likely observe the empirical moments. In contrast to existing literature, we measure the average distance between simulated and empirical moments in terms of likelihood. The advantage is that the empirical likelihood is general in nature and can be formulated for any $\theta = T(F)$. For example we would not only be interested in the configuration of our AB model that most likely observes the empirical moments or statistics on average but could also require that the simulated moment population has a certain required variance $\sigma_r^2$ . In this case $\theta_H = \left(\mu_e, \sigma_r^2\right)$ and the calibration or estimation problem amounts to

$$\hat{\gamma} = \underset{\gamma}{argmax} \left[W_\gamma \left(\theta_H\right)\right].$$

In fact, as it is shown later (see Theorem 2 and 3), the empirical likelihood and its properties can be extended to any additional information or requirement on $\theta = T(F)$ of the form $g(Y, \theta)$, as long as $E_F [g(Y, \theta)] = 0$ holds, that is we have some unbiased estimation equations $g(Y, \theta)$.

The reminder of this paper is devoted to investigate the capability of the proposed AB estimation, which is from the procedural point of view a sequence of hypothesis tests using a fixed hypothesis $\theta_H$ for different data sets $y_1(\gamma), .., y_n(\gamma)$, generated with an AB model at different configurations $\gamma$. However, before we investigate the potential of this approach in a simple experiment, the next section discusses Empirical Likelihood in detail. Section 3 extends the Balanced Adjusted Empirical Likelihood (BAEL) proposed in (Emerson and Owen, 2009) for the general setting $E_F [g(Y, \theta)] = 0$; BAEL is a modified empirical likelihood that primary allows an algorithmic-stable execution of the proposed estimation procedure in the test case. The results are reported in Section 4 and the last Section concludes..

## 2 Empirical Likelihood

Empirical likelihood was introduced in (Owen, 1990) as a nonparametric analogue to the parametric likelihood estimation, allowing hypothesis testing and confidence interval construction among others for the mean of an unknown distribution $F$. The empirical likelihood function is defined as follows.

**Definition 1.** Given $Y_1, .., Y_n \in \mathbb{R}^d$ iid with common $F$. The nonparametric likelihood of $F$ is

$$L(F) = \prod_{i=1}^{n} \left[F(y_i) - F(y_i-)\right] \tag{1}$$

where $F(y-) = P(Y < y)$ and $F(y) = P(Y \leq y)$ thus $P(Y = y) = F(y) - F(y-)$.

*Remark* 1. $L(F)$ is the probability of getting exactly the sample values $Y_1, \ldots Y_n$ from the cumulative distribution function (CDF) $F$. If $F$ is continuous $L(F) = 0$ and in order to have a positive nonparametric likelihood, a distribution $F$ must place a positive probability on each observed data point $Y_1, \ldots Y_n$.

The nonparametric likelihood is maximized by the empirical cumulative distribution function (ECDF)

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i < y\}}.$$

The likelihood ratio is given by

$$R(F) = \frac{L(F)}{L(F_n)}. \tag{2}$$

For a distribution $F$ that places probability $w_i$ on the value $Y_i$ we get:

$$R(F) = \frac{\prod_{i=1}^{n} w_i}{\prod_{i=1}^{n} \frac{1}{n}} = \prod_{i=1}^{n} n w_i. \tag{3}$$

To obtain the confidence interval for the mean $\mu = E_F[Y]$ we define the profile empirical likelihood ratio function by

$$R(\mu) = \sup \left\{ \prod_{i=1}^{n} n w_i | \sum_{i=1}^{n} w_i y_i = \mu, \sum_{i=1}^{n} w_i = 1, w_i > 0 \right\}. \tag{4}$$

Following the derivation in (3) the denominator profile empirical likelihood ratio function can be interpreted as the likelihood of the observed mean under empirical distribution $\prod_{i=1}^{n} \frac{1}{n}$ and the numerator is the maximized likelihood for a distribution $F$ that is supported on the sample and satisfies $E_F[Y] = \mu$. The log empirical likelihood function is

$$W(\mu) = -2 \log R(\mu). \tag{5}$$

**Theorem 1.** *Let $Y_1, .., Y_n \in \mathbb{R}^d$ iid with some unknown distribution $F_0$. Let $\mu_0$ be the true mean that is $E_{F_0}[Y] = \mu_0$, furthermore $\Sigma = Var_{F_0}[Y] < \infty$ with rank $q > 0$. Then as $n \to \infty$ we have*

$$W(\mu_0) \xrightarrow{d} \chi_q^2$$

*as* $n \to \infty$.

*Proof.* (Owen, 1990) □

Therefore a $100(1-\alpha)\,\%$ empirical likelihood confidence interval is formed by taking the values $\mu$ for which $W(\mu) \leq \chi^2_{q,1-\alpha}$. The probability that the true mean $\mu_0$ is in this interval approaches $1-\alpha$ since

$$P\left(W(\mu_0) \leq \chi^2_{q,1-\alpha}\right) \to 1-\alpha$$

as $n \to \infty$, hence the confidence interval is a limiting confidence interval. In fact in (Qin and Lawless, 1994) it is shown that this type of nonparametric hypothesis testing and confidence interval construction can be done for any parameter $\theta = T(F)$ for some function $T$ of $F$, while accounting for additional information on $\theta$ and $F$ given by $l$ unbiased estimation function $g_j(Y,\theta)\,,\ j = 1,\ldots,l$ or in vector form:

$$g(Y,\theta) = (g_1(Y,\theta)\,,\ldots,g_l(Y,\theta))'$$

with

$$E_F\left[g(Y,\theta)\right] = 0.$$

Then the profile empirical likelihood ratio function becomes

$$R(\theta) = \sup\left\{\prod_{i=1}^{n} nw_i | \sum_{i=1}^{n} w_i g(Y_i,\theta) = 0, \sum_{i=1}^{n} w_i = 1, w_i > 0\right\} \tag{6}$$

and

$$W(\theta) = -2\log R(\theta)\,. \tag{7}$$

*Remark* 2. Choosing $Y-\mu$ for $g(Y,\theta)$ we get the original formulation as in (4), (5) . And as mentioned above $\prod_{i=1}^{n} w_i$ is maximized by the ECDF $F_n$ and it follows that $R(\theta)$ is maximized with respect to $\theta$ at $\hat{\theta} = \overline{g}_n = \frac{1}{n}\sum_{i=1}^{n} g_i$ and $W\left(\hat{\theta}\right) = 0$.

The confidence interval for this likelihood ratio function can be found by using

**Theorem 2.** *Let $Y_1, .., Y_n \in \mathbb{R}^d$ iid with some unknown distribution $F_0$. For $\theta \in \Theta \subseteq \mathbb{R}^p$ let $g(Y, \theta) \subseteq \mathbb{R}^l$. Let $\theta_0 \in \Theta$ be the true parameter that satisfies $E_{F_0}[g(Y, \theta_0)] = 0$, furthermore $Var_{F_0}[g(Y, \theta_0)] < \infty$ with rank $q > 0$. Then as $n \to \infty$ we have*

$$W(\theta_0) \xrightarrow{d} \chi_q^2$$

*as $n \to \infty$.*

*Proof.* (Qin and Lawless, 1994) □

In practical terms however the empirical likelihood approach has various issues. As mentioned above since the constructed confidence interval is only a limiting confidence interval, the empirical likelihood needs some improvement for small samples. From the algorithmic point of view difficulties arise due to the fact that computing the empirical likelihood function in equation (4) and (6) involves solving a constrained maximization. In fact the right hand side of (6) only has a solution provided that the zero vector is an interior point of the convex hull of $\{g(Y_i, \theta), i = 1, \ldots, n\}$ and an explicit expression for $R(\theta)$ and $W(\theta)$ can be derived using Lagrange multipliers: the maximum of $\prod_{i=1}^n nw_i$ subject to the constraints $\sum_{i=1}^n w_i g(Y_i, \theta) = 0, \sum_{i=1}^n w_i = 1$ and $w_i > 0, i = 1 : n$ is attained at

$$\hat{w}_i = \left(\frac{1}{n}\right) \frac{1}{(1 + \lambda' g(Y_i, \theta))},$$

where $\lambda \in \mathbb{R}^l$ is the Lagrange multiplier satisfying

$$\sum_{i=1}^n \frac{g(Y_i, \theta)}{(1 + \lambda' g(Y_i, \theta))} = 0, \tag{8}$$

therefore

$$R(\theta) = \prod_{i=1}^n \frac{1}{\{1 + \lambda' g(Y_i, \theta)\}} \tag{9}$$

and

$$W(\theta) = -2 \sum_{i=1}^n \log \left\{1 + \lambda' g(Y_i, \theta)\right\}. \tag{10}$$

In summary, for a given sample $Y_1, .., Y_n$ the empirical likelihood function is only defined over a specific region $\Theta$ when zero is in the convex hull of $\{g(Y_i, \theta), i = 1, \ldots, n\}$, but for any $\theta \notin \Theta$ there is no information available which is problematic when trying to find an initial or the maximum value, which is essential for the estimation purpose. In order to resolve these issues (Chen, 2008), (Emerson and Owen, 2009) have proposed modified empirical likelihood functions. Both rely on the idea adding sample point(s) such that the zero vector becomes an interior point of the convex hull of $\{g(Y_i, \theta), i = 1, \ldots, n\}$, thus the empirical likelihood is defined for every $\theta \in \mathbb{R}^p$, while retaining the distributional convergence of the log empirical likelihood function. Whereas in (Chen, 2008) the approach is developed for the general case (as in (6), (7)),(Emerson and Owen, 2009) focuses on the special case for $g(Y, \theta) = Y - \mu$. Since (Emerson and Owen, 2009) has demonstrated better small sample properties (at least for $g(Y, \theta) = Y - \mu$) the next section is devoted to extend this approach for the general case.

## 3 Balanced Adjusted Empirical Likelihood for Estimating Equations

Subsequently we define $g_i = g(Y_i, \theta)$ and the following quantities:

$$\overline{g}_n = \frac{1}{n} \sum_{i=1}^{n} g_i,$$

$$S = \frac{1}{n} \sum_{i=1}^{n} g_i g_i^{'},$$

$$v = \overline{g}_n - 0,$$

$$r = \|v\| = \|\overline{g}_n - 0\|,$$

$$u = \frac{v}{r} = \frac{\overline{g}_n - 0}{\|\overline{g}_n - 0\|},$$

where $\|.\|$ is a vector Norm. Following (Emerson and Owen, 2009) we add two sample points

$$g_{n+1} = -sc_u,$$

$$g_{n+2} = 2\overline{g}_n + sc_u$$

with $c_u = \left(u^{'}S^{-1}u\right)^{-\frac{1}{2}}$. Since $c_u$ is the inverse Mahalanobis distance of a unit vector from $\overline{g}_n$ in the direction of $u$, $g_{n+1}$ places a new point near the zero vector when the covariance $S$ in direction $v$ is smaller and further away when the covariance in that direction is larger, thereby insuring that the zero vector is included in the convex hull of $\{g(Y_i, \theta), i = 1, \ldots, n\}$. The second point $g_{n+2}$ is included to maintain the original sample mean since $\frac{1}{n+2}\sum_{i=1}^{n+2} g_i = \overline{g}_n$. The adjusted empirical likelihood function becomes

$$\tilde{R}(\theta) = \sup\left\{\prod_{i=1}^{n+2} nw_i | \sum_{i=1}^{n+2} w_i g_i = 0, \sum_{i=1}^{n} w_i = 1, w_i > 0\right\} \tag{11}$$

and

$$\tilde{W}(\theta) = -2\log\tilde{R}(\theta). \tag{12}$$

Similar as in (9), (10) the explicit expressions for $\tilde{R}(\theta)$ and $\tilde{W}(\theta)$ are derived by using Lagrange multipliers:

$$\tilde{R}(\theta) = \prod_{i=1}^{n+2} \frac{1}{1+\lambda' g_i}, \tag{13}$$

$$\tilde{W}(\theta) = -\sum_{i=1}^{n+2} \log\left\{1 + \lambda^{'} g_i\right\}, \tag{14}$$

while $\lambda \in \mathbb{R}^{l+2}$ must satisfy:

$$\sum_{i=1}^{n+2} \frac{g_i}{1+\lambda' g_i} = 0. \tag{15}$$

The confidence interval for this can be constructed using the following Theorem.

**Theorem 3.** *Let $Y_1, .., Y_n \in \mathbb{R}^d$ with some unknown distribution $F_0$. Let $\theta_0$ be the true parameter that satisfies $E_{F_0}[g(Y, \theta_0) = 0]$, where $g$ is a vector valued function of dimension* *l, furthermore* $\Sigma_g = Var_{F_0}[g(Y,\theta_0)] < \infty$ *with rank* $q > 0$. *Then as* $n \to \infty$ *and for a fixed value* $s$ *we have*

$$\tilde{W}(\theta_0) \xrightarrow{d} \chi_q^2$$

*as* $n \to \infty$.

*Proof.* See Appendix □

## 4 Numerical Experiments

In order to investigate the feasibility of our estimation approach we propose the following experiment. We replace the supposed agent based model by a simulation for which we have the closed form of the data generating process. The advantage is that for any hypothesis $H : \theta = \theta_H$ we pose on the simulated moment samples, we can analytically derive the underlying parameter setting $\hat{\gamma}_a$ that corresponds to it and that can be compared to the parameter setting $\hat{\gamma}$ found by our proposed estimation procedure. In our experiment we will simulate sample price paths following a Geometric Brownian Motion (GBM):

$$P_t = P_0 e^{X_t},$$

where $X_t = (r - \frac{u^2}{2})t + uW_t$ (where $W_t$ is a standard Brownian Motion) and the log returns $r_{\Delta t}$ over an interval $\Delta t$ are normally distributed with $N\left(\left(r - \frac{u^2}{2}\right)\Delta t, u^2\Delta t\right)$. Note, that in this case the outcomes of the simulator are entirely governed by $\gamma = (r, u)$ and the mean and the variance of the simulated returns are functionally related over $r$ and $u$. Following our proposed estimation procedure we seek in this experiment for a configuration $\hat{\gamma}$ at which we on average most likely observe the empirical moments $y_e$. For simplicity we consider here $\theta_H = y_e = (\mu_e, \sigma_e^2)$ the empirical mean and variance of the returns, that is we seek

$$\hat{\gamma} = \underset{\gamma}{argmax}\left[\tilde{W}_\gamma(y_e)\right].$$

In order to calculate $\tilde{W}_\gamma(y_e)$ we first simulate $n$ price paths with the parameter setting $\gamma = (r, u)$ and for each path we compute the sample mean and the variance of the return distribution, resulting in a population $\left\{y_i(\gamma) = \left(\mu_\gamma^i, var_\gamma^i\right), i = 1 : n\right\}$, for which we solve the following the constrained maximization problem

$$\tilde{R}_\gamma\left(y_e\right)=\sup\left\{\prod_{i=1}^{n+2} n w_i \middle| \sum_{i=1}^{n+2} w_i y_i\left(\gamma\right)=y_e, \sum_{i=1}^{n} w_i=1, w_i>0\right\}$$

and finally we get

$$\tilde{W}_\gamma\left(y_e\right)=-2\log\tilde{R}_\gamma\left(y_e\right)$$

and $\hat{\gamma}$ is found by

$$\max_{\gamma=(r,u)}\tilde{W}_\gamma\left(y_e\right).$$

Figure 1-8 depict the surface of $\tilde{W}_\gamma\left(y_e\right)$ over a mesh $(r,u)$, whereas Figure 5-8 plots the surface $\tilde{W}_\gamma\left(y_e\right)$ computed using a Monte Carlo variance reduction technique called Common Random Numbers. In these pictures $\tilde{W}_\gamma\left(y_e\right)$ is calculated for the hypothesis $y_e=(0.015,0.01)$ using $n=300$ price samples over the period $[0,T]=[0,1]$. The moment population is then calculated from the returns $r_{\triangle t}$ with $\triangle t=1/100$. Analytically we know then for the hypothesis $y_e=(0.015,0.01)=\left(\left(r-\frac{u^2}{2}\right)\Delta t, u^2\Delta t\right)$ that the analytical optimal configuration $\hat{\gamma}_a=(2,1)$. As it can be seen in both figures, the optimal region of $\tilde{W}_\gamma\left(y_e\right)$ is around $\hat{\gamma}_a=(2,1)$ and its optimal point is indeed $(2,1)$, which can be explicitly computed using an optimisation algorithm for $\max_{\gamma=(r,u)}\tilde{W}_\gamma\left(y_e\right)$. For Figure 1 we could retrieve successfully the optimal point $\hat{\gamma}_a$ using a Genetic Algorithm (GA), while for Figure 5 a simple gradient decent method was sufficient. Summing up, while maximizing empirical likelihood of the GBM simulator **on average** to reproduce the empirical moments $y_e$ we manage even to retrieve the **exact analytical optimal configuration** $\hat{\gamma}=\hat{\gamma}_a$. The following remark will give an explanation for this.

*Remark* 3. Given $X_{1,\dots,}X_n$ iid with mean $\mu$ and variance $\sigma^2$ then the sample mean defined as $\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i$ is a random variable with expectation

$$E\left[\bar{X}\right]=\frac{1}{n}\sum_{i=1}^{n}E[X_i]=\mu$$

and the sample variance defined as $S^2=\frac{1}{n-1}\sum_{i=1}^n\left(X_i-\bar{X}\right)^2$ is a random variable with expectation

$$E\left[S^2\right]=\sigma^2.$$

**Figure 1.** Empirical Likelihood Surface

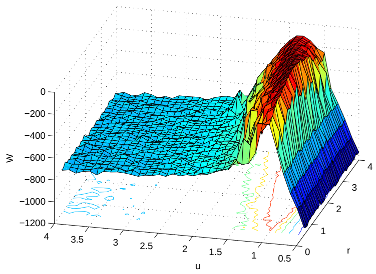



**Figure 2.** Empirical Likelihood Surface r-Axis View

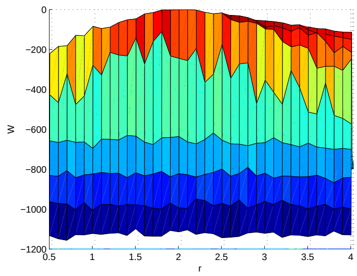



**Figure 3.** Empirical Likelihood Surface u-Axis View

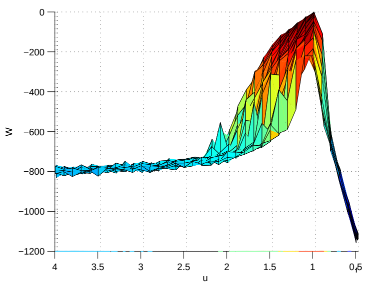



**Figure 4.** Empirical Likelihood Surface Bird Eye's View

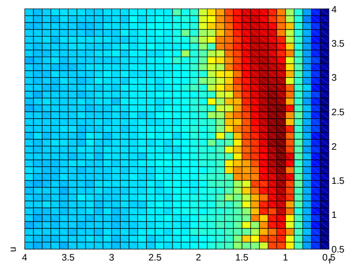


**Figure 5.** Empirical Likelihood Surface

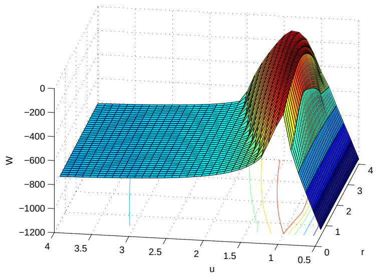



**Figure 6.** Empirical Likelihood Surface r-Axis View

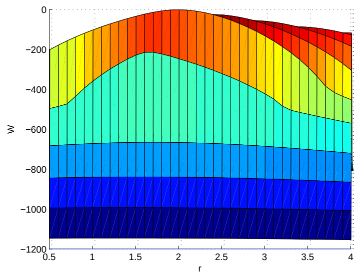



**Figure 7.** Empirical Likelihood Surface u-Axis View

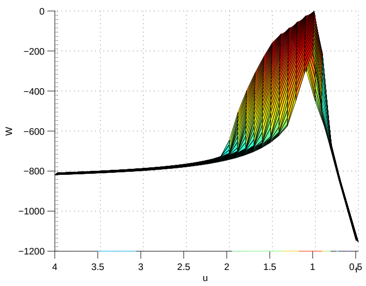
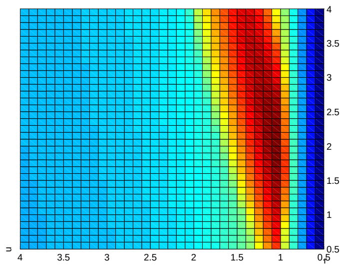



**Figure 8.** Empirical Likelihood Surface Bird Eye's View



Since we maximize the likelihood of the average or expected return sample mean and variance to be $\mu_e$ and $\sigma_e^2$, we have

$$E\left[\bar{X}\right] = \mu = \mu_e$$

and

$$E\left[S^2\right] = \sigma^2 = \sigma_e^2,$$

that is our test estimation actually maximizes the likelihood of the return mean $\mu = \left(r - \frac{u^2}{2}\right)\Delta t$ and variance $\sigma^2 = u^2\Delta t$ to be $\mu_e$ and $\sigma_e^2$.

As a consequence calibrating for $y_e = (0.015, 0.01)$ should give us the analytical solution $\hat{\gamma}_a = (2, 1)$ and that is what we indeed find. Furthermore, we find $\max\limits_{\gamma=(r,u)} \tilde{W}_\gamma\left(y_e\right) = 0$, that means that our simulator observes the moments $y_e = (0.015, 0.01)$ at $\hat{\gamma} = \hat{\gamma}_a = (2, 1)$ with likelihood 1, that is with certainty.

## 5 Conclusions

Our proposed estimation procedure minimises the distance between the average simulated moments and the empirical moments using likelihood as a *distance* measure and the underlying idea is to maximize the likelihood to observe the empirical moments with respect to the configuration of our AB model. In principal we execute a series of hypothesis tests with a fixed hypothesis and varying data sets, generated at different configurations. We have tested the proposed approach in a simple setting using a GBM simulator, where we aimed at maximizing the empirical likelihood of the GBM simulator on average to observe a given empirical moment $y_e = \left(\mu_e, \sigma_e^2\right)$, which in this case is equivalent to directly estimating the mean and variance of the GBM return distribution. Correspondingly the results of our experiment show that the proposed procedure was capable to retrieve the exact analytical optimal configuration $\hat{\gamma} = \hat{\gamma}_a$ as desired. Furthermore the displayed surface of the objective function$\tilde{W}_\gamma\left(y_e\right)$ was able to capture the underlying functional relationship of the parameters $r$ and $u$ in the GBM even in the presence of Monte Carlo variance and therefore is potentially useful in discriminating arbitrary AB models among different parameter settings. The latter and the extend to which $\tilde{W}_{\gamma}(y_e)$ is capable to compare different AB models is subject to further research. Note, the limitations of $\tilde{W}_{\gamma}(y_e)$ is closely related to limitations of the optimisation algorithm used to find its optimal value. In general we cannot expect the surface of $\tilde{W}_{\gamma}(y_e)$ to be well behaved. It might have multiple optima, indifference regions or discontinuities. The first two points imply that the considered AB model has parameter regions at which its results are qualitatively indistinguishable and the last point might correspond to algorithmic unstable regions. Therefore the quality of the results obtained by the proposed AB estimation depends on the robustness of the optimisation algorithm in use to find the best configuration.

## References


Chen, J.: 2008, 'Adjusted Empirical Likelihood and its Properties'. *Journal of Computational and Graphical Statistics* **17**(2), 426–443.

Chiarella, C., R. Dieci, and X.-Z. He: 2009, *Heterogeneity,Market Mechanisms and Asset Price Dynamics*, Chapt. 5. Elsevier.

Duffie, D. and K. Singleton: 1993, 'Simulated moments estimation of Markov models of asset prices'. *Econometrica* **61**(4), 929–952.

Emerson, S. C. and A. B. Owen: 2009, 'Calibration of the empirical likelihood method for a vector mean'. *Electronic Journal of Statistics* **3**, 1161–1192.

Fagiolo, G., A. Moneta, and P. Windrum: 2007, 'A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems'. *Computational Economics* **30**(3), 195–226.

Franke, R.: 2009, 'Applying the method of simulated moments to estimate a small agent-based asset pricing model'. *Journal of Empirical Finance* **16**(5), 804–815.

Gourieroux, C., A. Monfort, and E. Renault: 1993, 'Indirect inference'. *Journal of Applied Econometrics* **8**, 85–118.

Owen, A. B.: 1990, 'Empirical Likelihood Ratio Confidence Region'. *The Annals of StatisticsStatistics* **18**(1), 90–120.

Qin, J. and J. Lawless: 1994, 'Empirical Likelihood and General Estimating Equations'. *The Annals of Statistics* **22**(1), 300–325.

Wagener, F. and C. Hommes: 2009, *Complex Evolutionary Systems in Behavioral Finance Cars Hommes*, pp. 217–276. Elsevier.

Winker, P., M. Gilli, and V. Jeleskovic: 2007, 'An objective function for simulation based inference on exchange rate data'. *Journal of Economic Interaction and Coordination* **2**(2), 125–145.

## Appendix

Before we come to the main proof of Theorem 3 we need to introduce some quantities and a Lemma that will be needed later on. For the following let $g_{n+2} = b_n\overline{g}_n + sc_u$ ,$b_n$ a constant and:

$$\tilde{S} = \frac{1}{n+2}\sum_{i=1}^{n+2} g_i g_i^{'},$$

$$\tilde{\overline{g}}_n = \frac{1}{n+2}\sum_{i=1}^{n+2} g_i,$$

and $g^* = \max_{i=1:n}\|g_i\|$, $\tilde{g}^* = \max_{i=1:n+2}\|g_i\|$. Note for $b_n = 2$ it is

$$\tilde{\overline{g}}_n = \frac{1}{n+2}\left[n\overline{g}_n + 2\overline{g}_n\right] = \overline{g}_n.$$

Furthermore the following magnitudes hold: i)[1] $g^* = o_p\left(n^{\frac{1}{2}}\right)$ , ii)[2] $g_{n+1} = O_p(1)$ and iii)[3] $\overline{g}_n = O_p\left(n^{-\frac{1}{2}}\right)$. It follows

$$g_{n+2} = O_p\left(b_n n^{-\frac{1}{2}}\right)\begin{cases} \le o_p\left(n^{\frac{1}{2}}\right) & ,b_n \le o(n) \\ > o_p\left(n^{\frac{1}{2}}\right) & ,b_n > o(n) \end{cases}$$

that is $g_{n+2}$ and $\tilde{g}^*$ are of order $o_p\left(n^{\frac{1}{2}}\right)$ as long as $b_n$ is of order $o(n)$.

**Lemma 1.**

$$\tilde{S} - S \to 0$$

*in probability as $n \to \infty$.*

*Proof.*

$$\begin{aligned}\tilde{S} &= \frac{1}{n+2}\left[\sum_{i=1}^{n} g_i g_i^{'} + g_{n+1}g_{n+1}^{'} + g_{n+2}g_{n+2}^{'}\right] = \frac{1}{n+2}\left[nS + s^2c_u^2uu^{'} + \left(b_n\overline{g_n} + sc_u\right)^2 uu^{'}\right] \\ &= \frac{n}{n+2}S + \frac{s^2c_u^2 + \left(b_n\overline{g_n} + sc_u\right)^2}{n+2}uu^{'}\end{aligned}$$

[1](Chen, 2008) p.22

[2]$S \to \Sigma$ in probability as $n \to \infty$, therefore is bounded above $O_p(1)$

[3](Qin and Lawless, 1994) p.19

Since S converges in probability to $\Sigma_g$ for $n \to \infty$, $c_u$ is of order $O_p(1)$ , $s$ as a constant is of order $O(1)$ and $\overline{g}_n$ is order of $O_p\left(n^{-\frac{1}{2}}\right)$, the order of the last term is

$$\begin{aligned}
&\left[O(1)\,O_p(1) + \left(O_p\left(b_n n^{-\frac{1}{2}}\right) + O(1)\,O_p(1)\right)^2\right] O\left(n^{-1}\right) \\
&\quad = \left[O_p(1) + \left(O_p\left(b_n n^{-\frac{1}{2}}\right)\right)^2\right] O\left(n^{-1}\right) \\
&\quad = \left[O_p(1) + O_p\left(b_n^2 n^{-1}\right)\right] O\left(n^{-1}\right) \\
&\quad = O_p\left(n^{-1}\right) + O_p\left(\left(\frac{b_n}{n}\right)^2\right) \\
&\quad = O_p\left(n^{-1}\right) \quad + \begin{cases} o_p\left(n^{-1}\right) & , b_n \le o(\sqrt{n}) \\ o_p\left(\left(\frac{b_n}{n}\right)^2\right) & , b_n > o(\sqrt{n}). \end{cases}
\end{aligned}$$

Therefore as long as $b_n$ is of order $o(\sqrt{n})$ the error term is of order[4] $O_p\left(n^{-1}\right)$ and as $n \to \infty$, $\widetilde{S} - S \to 0$ in probability. □

The following proofs Theorem 3.

*Proof.* Without loss of generality let $\sigma_1^2 \le \ldots \le \sigma_m^2$ be the EV (Eigenvalues) of $\Sigma_g = Var_{F_0}\left[g(Y,\theta_0)\right]$ with $\sigma_1^2 = 1$ . For $\theta = \theta_0$ using $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ and $\hat{\lambda} = \lambda/\rho$, $\rho = \|\lambda\|$ in (8) we get:

$$\begin{aligned}
0 = &\quad \frac{\hat{\lambda}'}{n}\sum_{i=1}^{n+2}\frac{g_i}{1+\lambda' g_i} = \frac{\hat{\lambda}'}{n}\sum_{i=1}^{n+2} g_i - \frac{\rho}{n}\sum_{i=1}^{n+2}\frac{\left(\hat{\lambda}' g_i\right)^2}{1+\hat{\lambda}' g_i} \\
\le &\quad \hat{\lambda}'\overline{g}_n\left(1 + b_n/n\right) - \frac{\rho}{n}\sum_{i=1}^{n}\frac{\left(\hat{\lambda}' g_i\right)^2}{1+\hat{\lambda}' g_i} \\
= &\quad \hat{\lambda}'\overline{g}_n + n^{-1} b_n \hat{\lambda}'\overline{g}_n - \frac{\rho}{n(1+\rho\tilde{g}^*)}\sum_{i=1}^{n}\left(\hat{\lambda}' g_i\right)^2 \\
= &\quad \hat{\lambda}'\overline{g}_n + O_p\left(b_n n^{-\frac{3}{2}}\right) - \frac{\rho}{n(1+\rho\tilde{g}^*)}\sum_{i=1}^{n}\left(\hat{\lambda}' g_i\right)^2 \\
\le &\quad \hat{\lambda}'\overline{g}_n - \frac{\rho(1-\varepsilon)}{(1+\rho\tilde{g}^*)} + O_p\left(b_n n^{-\frac{3}{2}}\right).
\end{aligned}$$

The last inequality holds because $\sum_{i=1}^{n}\left(\hat{\lambda}' g_i\right)^2 \ge (1-\varepsilon)\,\sigma_1^2 = (1-\varepsilon)$ for some $\varepsilon > 0$. Therefore we get

$$\frac{\rho}{(1+\rho\tilde{g}^*)} \le \frac{\hat{\lambda}'\overline{g}_n}{1-\varepsilon} + O_p\left(b_n n^{-\frac{3}{2}}\right) \tag{16}$$

with

$$O_p\left(b_n n^{-\frac{3}{2}}\right)\begin{cases} \le o_p\left(n^{-\frac{1}{2}}\right) & , b_n \le o(n) \\ > o_p\left(n^{-\frac{1}{2}}\right) & , b_n > o(n) \end{cases}$$

since $\frac{\hat{\lambda}'\overline{g}_n}{1-\varepsilon}$ is of order $o_p\left(n^{-\frac{1}{2}}\right)$ it follows from equation (16)

$$\rho = \lambda = o_p\left(n^{-\frac{1}{2}}\right) \tag{17}$$

[4]From the definition of $o$ and $O$ it follows: $f = o(n) \Rightarrow f = O(n)$ and therefore $o(n) + O(n) = O(n)$.

as long as $b_n = o(n)$. Using $\frac{1}{1+x} = 1 - x - \frac{x^2}{1+x}$ in (8) we have:

$$
\begin{aligned}
0 = \frac{1}{n+2} \quad \textstyle\sum_{i=1}^{n+2} \frac{g_i}{1+\lambda' g_i} &= \textstyle\frac{1}{n+2} \sum_{i=1}^{n+2} g_i \left(1 - \lambda' g_i + \frac{\left(\lambda' g_i\right)^2}{1+\lambda' g_i}\right) \\
&= \textstyle\frac{1}{n+2} \left(\sum_{i=1}^{n+2} g_i - \lambda \sum_{i=1}^{n+2} g_i g_i^{'} + \sum_{i=1}^{n+2} \frac{g_i \left(\lambda' g_i\right)^2}{1+\lambda' g_i}\right). \qquad (18)
\end{aligned}
$$

The last term is bounded above by norm:

$$
\begin{aligned}
\frac{1}{n+2} \sum_{i=1}^{n+2} \frac{g_i \left(\lambda' g_i\right)^2}{1+\lambda' g_i} &\leq \max_{i=1:n+2} \|g_i\| \frac{1}{n+2} \sum_{i=1}^{n+2} \|\lambda\|^2 \|g_i\|^2 \left|1+\lambda' g_i\right|^{-1} \\
&= \frac{(\tilde{g}^*)}{n+2} \sum_{i=1}^{n+2} \|\lambda\|^2 \|g_i\|^2 \left|1+\lambda' g_i\right|^{-1}. \qquad (19)
\end{aligned}
$$

From now on we use $b_n = 2$ as in the definition of $\tilde{W}(\theta_0)$ and the order of equation (19) becomes[5]

$$
o_p\left(n^{\frac{1}{2}}\right)\left(o_p\left(n^{-\frac{1}{2}}\right)\right)^2 O_p(1) = o_p\left(n^{-\frac{1}{2}}\right)
$$

with equation (18) we get:

$$
0 = \frac{1}{n+2} \sum_{i=1}^{n+2} g_i - \frac{\lambda}{n+2} \sum_{i=1}^{n+2} g_i g_i^{'} + o_p\left(n^{-\frac{1}{2}}\right)
$$

therefore,

$$
\begin{aligned}
\lambda &= \left[\frac{1}{n+2} \sum_{i=1}^{n+2} g_i\right] / \left[\frac{1}{n+2} \sum_{i=1}^{n+2} g_i g_i^{'} + o_p\left(n^{-\frac{1}{2}}\right)\right] \\
&= \tilde{\bar{g}}_n \tilde{S}^{-1} + o_p\left(n^{-\frac{1}{2}}\right). \qquad (20)
\end{aligned}
$$

Now using $\log(1+Y_i) = Y_i - \frac{1}{2} Y_i^2 + \eta_i$, $b_n = 2$ and (20) in $\tilde{W}(\theta_0)$

---

[5] As long as $b_n$ is of order $o(n)$ we know from above $\hat{g}^*$ is of order $o_p\left(n^{\frac{1}{2}}\right)$, the order of $\lambda$ is given in equation (17). For $\theta = \theta_0$ it is $\frac{1}{n+2} \sum_{i=1}^{n+2} \|g_i\|^2 = \frac{1}{n} \sum_{i=1}^{n} \|g_i\|^2$ therefore $\frac{1}{n} \sum_{i=1}^{n} \|g_i\|^2 dF_0(y_i) \to \int_{\mathbb{R}^d} \|g(y)\|^2 dF_0(y) = Var[g(y, \theta_0)] = \Sigma_g < \infty$ for $n \to \infty$ and with equation (15) the last two terms are of order $O_p(1)$.

$$
\begin{aligned}
\tilde{W}(\theta_0) &= 2\sum_{i=1}^{n+2}\log\left(1+\lambda^{'}g_i\right) = 2\sum_{i=1}^{n+2}\lambda^{'}g_i - \sum_{i=1}^{n+2}\left(\lambda^{'}g_i\right)^2 + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\lambda^{'}\tilde{\overline{g}}_n - (n+2)\lambda^{'}\tilde{S}\lambda + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\left(\tilde{\overline{g}}_n\tilde{S}^{-1} + o_p\left(n^{-\frac{1}{2}}\right)\right)^{'}\tilde{\overline{g}}_n - (n+2)\lambda^{'}\tilde{S}\lambda + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\left[\tilde{\overline{g}}_n^{'}\tilde{S}^{-1}\tilde{\overline{g}}_n\right] + 2(n+2)\,o_p\left(n^{-\frac{1}{2}}\right)\tilde{\overline{g}}_n + (n+2)\lambda^{'}\tilde{S}\lambda + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + 2(n+2)\,o_p\left(n^{-\frac{1}{2}}\right)\overline{g}_n + (n+2)\lambda^{'}\tilde{S}\lambda + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + O(n)\,o_p\left(n^{-\frac{1}{2}}\right)o_p\left(\overline{g}_n\right) + (n+2)\lambda^{'}\tilde{S}\lambda + 2\sum_{i=1}^{n+2}\eta_i \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + O(n)\,o_p\left(n^{-\frac{1}{2}}\right)o_p\left(n^{-\frac{1}{2}}\right) + (n+2)\lambda^{'}\tilde{S}\lambda + o_p(1) \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + o_p(1) + O(n)\,o_p\left(n^{-\frac{1}{2}}\right)O_p(1)\,o_p\left(n^{-\frac{1}{2}}\right) + o_p(1) \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + o_p(1) + o_p(1) + o_p(1) \\
&= 2(n+2)\left[\overline{g}_n^{'}\tilde{S}^{-1}\overline{g}_n\right] + o_p(1).
\end{aligned}
$$

Because $W(\theta_0) = n\left(\overline{g}_n^{'}S^{-1}\overline{g}_n\right) + o_p(1)$ , $\tilde{W}(\theta_0) - W(\theta_0) \to 0$ in probability, since $\tilde{S} \to S$ in Lemma 1 and $\frac{n}{n+2} \to 1$ as $n \to \infty$. Furthermore it is shown in (Qin and Lawless, 1994) that $W(\theta_0) \to \chi_q^2$ in probability as $n \to \infty$, therefore we have $\tilde{W}(\theta_0) \to \chi_q^2$. $\square$