Centre for Computational Finance and Economic Agents

WP072-13

Working Paper Series

Minh Khoa Nguyen
Wing Lon Ng
Steve Phelps

Simulation Based Estimation Using Empirical Likelihood

2013

University of Essex

CCFEA
Centre for Computational Finance and Economic Agents

Simulation Based Estimation Using Empirical Likelihood

Minh Khoa Nguyen[a], Wing Lon Ng[a,*], Steve Phelps[a]

[a] *Centre for Computational Finance and Economic Agents (CCFEA), University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK*

Abstract

This paper proposes an empirical likelihood framework for estimating simulation models. Similarly to the simulated moment method, the proposed method matches simulated and empirical time series moments via empirical likelihood. However, in comparison to the simulated moment method, for consistent estimation it does not require some form of continuity of simulation model moments to be continuous with respect to the underlying parameter. These continuity requirements of the simulation model with respect to its parameters is rather a critical assumption in particular for simulation models with non-analytical complex dynamics. For such models it may be hard or even impossible to demonstrate the continuity of their moments. Moreover, the feasibility of the proposed estimation method is demonstrated in a simple simulation exercise with a geometric Brownian motion, where we are able to obtain smaller mean squared errors than the simulated moment method.

Keywords: Empirical Likelihood, Simulation Based Estimation, Unbiased Estimation Equation

1. Introduction

In many areas of science, models which focus on the properties and behavior of individual components and their interactions — so-called individual-based models, or agent-based models — have become increasingly important (Bonabeau, 2002; Farmer and Foley, 2009; Shalizi, 2006). These models are often sufficiently complex that deriving closed-form solutions for quantitative aspects of their macroscopic behavior is often impractical if not impossible. In

*Corresponding author. Tel.:0044-1206-874684

Email addresses: mknguy@essex.ac.uk (Minh Khoa Nguyen), wlng@essex.ac.uk (Wing Lon Ng), sphelps@essex.ac.uk (Steve Phelps)

Preprint submitted to Elsevier *October 17, 2013*

this paper we are particularly interested in simulation models that have non-analytical outputs, and therefore have no information on the moments and the likelihood function or a reduced form. Such models are often analyzed using Monte-Carlo simulation and empirical methods. However, one criticism of these models is the lack of principled methods for estimating their free parameters against empirically-observed data. To address this problem, this paper introduces a new simulation-based estimation approach by matching simulated and empirical time series moments via empirical likelihood (EL).

In many empirical studies, EL is often considered a good choice when the error distribution of the proposed model is asymmetric or censored. More recent work focuses on missing data (e.g. Zhao et al., 2013; Xue, 2013). As a contribution to the existing literature, we propose a simple but effective extension to the EL framework to account for simulation based models.

In fact, former studies in econometrics have proposed several simulation-based estimation techniques, such as the simulated maximum likelihood (Lee, 1992, 1995), the simulated moment method (Lee and Ingram, 1991; Duffie and Singleton, 1993), the indirect inference method (Gourieroux and Monfort, 1991), or the efficient moment method (Gallant and Tauchen, 1996). For a more recent review, see also Yu (2012) and the references therein. However, these estimation procedures face certain difficulties when applied to more complex models. The simulated maximum likelihood method requires that we have some information of the likelihood function, and the efficient moment method as well as the indirect inference method suffer the drawback of using an auxiliary model, the latter inducing a source of arbitrariness of capturing the statistical features of the empirical data.

To avoid these shortcomings, we follow a different approach based on the EL framework introduced by Owen (1990). It employs nonparametric likelihood-based tests that can be applied to various functionals of interest such as the mean or the quantiles of a distribution, or regression parameters in multi-sample problems. It is the non-parametric analogue of the parametric likelihood method and provides efficient estimators and confidence intervals for hypothesis testing. In contrast to the efficient moment method and the indirect inference method, the our proposed EL approach does not need an adequate auxiliary model for approximating the likelihood function. Similarly to the simulated moment method (SMM), the proposed EL method matches simulated and empirical time series moments via empirical likelihood. The idea is to maximize the likelihood of observing the empirical moments as a function of different parameter settings. In fact, the proposed simulation-based estimation procedure maximizes the likelihood ratio of the empirical features across simulation outcomes. Since EL ratios can be used for hypothesis tests and confidence regions, the proposed estimation approach can be interpreted as a series of hypothesis tests with a fixed empirically motivated hypothesis and varying simulated data sets, generated from different configurations, similar to a Monte Carlo setting. For the proposed method we show that: (i) it provides a consistent estimator; (ii) in comparison to the SMM for consistent estimation it does not require that the moments of the simulation model to be continuous with respect to the underlying parameter; (iii) in a simple simulation exercise with a geometric Brownian motion, it is able to obtain smaller mean squared errors than the SMM.

The paper is outlined as follows. In Section 2 we first describe the EL approach. In Section 3 we show that a simple extension to EL provides a consistent estimator for simulation models. In Section 4 we analyze the efficiency of the proposed simulation-based estimation procedure by estimating a geometric Brownian motion process, and show empirically that it converges to the true parameter value. Section 5 concludes.

2. Empirical Likelihood

Owen (1990) defines the empirical likelihood function as follows.

Definition 1. Assume $Y_1, .., Y_n \in \mathbb{R}^d$ are i.i.d. from a common cumulative distribution function F. The non-parametric empirical likelihood of any 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 $P(Y = y) = F(y) - F(y-)$.

$L(F)$ is the probability of obtaining exactly the sample observations of $Y_1, \ldots Y_n$ from F, which resembles exactly the concept of a likelihood function. Let

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

denote the empirical likelihood ratio, where

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i < y\}} \tag{3}$$

represents the empirical cumulative distribution function (ECDF). To avoid $L(F) = 0$ for a continuous F, F must have a positive probability w_i on each observed sample $Y_i, i = 1, \ldots, n$, in order to have a positive non-parametric likelihood. Hence, from Eq. (2), we obtain:

$$R(F) = \prod_{i=1}^{n} \frac{w_i}{\frac{1}{n}}. \tag{4}$$

The enumerator represents the likelihood of a distribution F with weights w_i for the observed data and the denominator is the maximum likelihood estimator given the observations for $Y_1, \ldots, Y_n$.

To obtain the confidence interval for the mean $\mu = E_F[Y]$, Owen (1990) defines the profile empirical likelihood ratio function by

$$R(\mu) = \sup\left\{ \prod_{i=1}^{n} n w_i \middle| \sum_{i=1}^{n} w_i y_i = \mu, \sum_{i=1}^{n} w_i = 1, w_i \geq 0 \right\}. \tag{5}$$

Hence according to Eq. (4), for the given observations $R(\mu)$ is the ratio of (i) the maximum likelihood estimator of all distributions that place non-negative weights on the given observations and (ii) the maximum likelihood estimator of all such distributions that also have the mean μ. The log empirical likelihood function is

$$W(\mu) = \log R(\mu). \tag{6}$$

and the empirical likelihood ratio statistics is

$$-2W(\mu), \tag{7}$$

which can be used to construct asymptotic confidence intervals (CI) for $\mu_0 = E_{F_0}[Y]$, the true parameter with respect to the true distribution F_0 of the observations $Y_1, .., Y_n$.

Owen (1990) has demonstrated that under some regularity conditions

$$-2W(\mu_0) \to \chi^2 \tag{8}$$

as $n \to \infty$, and the $100(1-\alpha)\%$ CI is

$$\{\mu : -2W(\mu) \leq \chi^2(1-\alpha)\}, \tag{9}$$

where $\chi^2(1-\alpha)$ is the $(1-\alpha)^{th}$ quantile of the χ^2- distribution.

From an algorithmic point of view, the profile empirical likelihood ratio function in Eq. (5) involves solving a constrained maximization, which however has only a solution provided that μ is an interior point of the convex hull of $\{y_i, i = 1, ..., n\}$. In order to resolve this convex hull problem, Emerson and Owen (2009) have introduced the balance adjusted empirical likelihood (BAEL). Its profile empirical likelihood ratio function is defined as

$$R(\mu) = \sup\left\{ \prod_{i=1}^{n+2} (n+2)\, w_i \middle| \sum_{i=1}^{n+2} w_i y_i = \mu, \sum_{i=1}^{n} w_i = 1, w_i \geq 0 \right\} \qquad (10)$$

that adds two artificial sample points y_{n+1} and y_{n+2} to the data set and then computes the empirical likelihood ratio statistic on the augmented sample, where y_{n+1} and y_{n+2} are two new sample points around the mean

$$\overline{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$$

in direction u

$$u = \frac{\overline{y}_n - \mathbf{0}}{\|\overline{y}_n - \mathbf{0}\|},$$

where $\|.\|$ is a Euclidean norm. Let $\hat{S}$ denote the sample covariance matrix

$$\hat{S} = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y}_n)(y_i - \overline{y}_n)'$$

then

$$c_u = \left(u' \hat{S}^{-1} u\right)^{-\frac{1}{2}}$$

is the inverse Mahalanobis distance of a unit vector from $\overline{y}_n$ in the direction of u. For a fix $s \in \mathbb{R}$, y_{n+1} and y_{n+2} are defined by

$$y_{n+1} = -s c_u u, \qquad (11)$$

$$y_{n+2} = 2\overline{y}_n + s c_u u. \qquad (12)$$

This results in placing new sample points closer to μ when the covariance in the direction u smaller, and farther when the covariance in that direction is larger.

3. EL and Estimation of Simulation Models

In this section we discuss the use of the EL approach for estimating simulation models by matching simulated and empirical time series moments via empirical likelihood. For this, let us consider a (strictly) stationary and ergodic empirical process denoted by $\{x_t\}$. Moreover, let $\{y_t(\beta)\}$ denote the simulation process with parameter β, which we also assume to be (strictly) stationary in order to ensure that its statistical properties remain stable over the evolution of the output. The moment of interest is given in the form[1]

$$\mu(\beta) = E\left[f\left(y_1(\beta)\right)\right], \tag{13}$$

where f is a measurable real valued function and for estimation our aim is to find the model parameter β_E such that the true simulated and empirical moments coincide, i.e.:

$$\mu(\beta_E) = E\left[f\left(x_1\right)\right], \tag{14}$$

where $E\left[f\left(x_1\right)\right]$ is the true empirical moment. Using the stationarity, Eq. (13) can be written as

$$\mu(\beta) = E\left[\frac{1}{K}\sum_{t=1}^{K} f\left(y_t(\beta)\right)\right], \tag{15}$$

with K is a fix integer. As $\{y_t(\beta)\}_{t=1,\ldots,K}$ can be considered as K-variate random variable $\underline{Y} = \left[y_1(\beta), ..., y_K(\beta)\right]^{'}$, Eq. (15) can be written as

$$\mu(\beta) = E\left[\tilde{f}\left(\underline{Y}(\beta)\right)\right] \tag{16}$$

such that with Eq. (16) and Eq. (14) we get the moment condition

$$E\left[\ddot{g}\left(\underline{Y}(\beta_E), \mu_0\right)\right] = 0, \tag{17}$$

where $\mu_0 = E\left[f\left(x_1\right)\right]$ and

$$\ddot{g}\left(\underline{Y}(\beta), \mu_0\right) = \tilde{f}\left(\underline{Y}(\beta)\right) - \mu_0.$$

However, in the estimation context μ_0 is unknown and only given by an estimate μ_T derived from some sample time series $\{x_t\}_{t=1,\ldots,T}$:

$$\mu_T = \sum_{t=1}^{T} f\left(x_t\right).$$

This motivates the question whether an EL type estimator can be used for estimating β_E. Therefore let us consider

$$\ddot{g}\left(\underline{Y}(\beta), \mu_T\right) = \tilde{f}\left(\underline{Y}(\beta)\right) - \mu_T$$

[1]For a (strictly) stationary process $\{z_t\}$ all moments $E\left[h\left(z_t\right)\right]$ are the same for all t, i.e. $E\left[h\left(z_t\right)\right] = M$ for all t for some constant M. But for covenience in this paper we will refer to $E\left[h\left(z_1\right)\right]$.

and some *iid* simulation outcomes $\underline{Y}_1(\beta), ..., \underline{Y}_T(\beta)$. The corresponding EL profile empirical likelihood ratio of the empirical moment μ_T is given by

$$\ddot{R}(\beta, \mu_T) = \sup\left\{\prod_{i=1}^{T} Tw_i \middle| \sum_{i=1}^{T} w_i \ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right) = 0, \sum_{i=1}^{T} w_i = 1, w_i \geq 0\right\} \tag{18}$$

and

$$-2\ddot{W}(\beta, \mu_T) = -2\log \ddot{R}(\beta, \mu_T). \tag{19}$$

An explicit expression for $-2\ddot{W}(\beta, \mu_T)$ can be derived using Lagrange multipliers (Qin and Lawless, 1994, p. 304) and is given by

$$\ddot{W}(\beta, \mu_T) = -\sum_{i=1}^{T} \log\left(1 + \lambda \ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right)\right), \tag{20}$$

where λ must satisfy

$$\frac{1}{T}\sum_{i=1}^{T} \frac{\ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right)}{1 + \lambda \ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right)} = 0 \tag{21}$$

and the proposed EL type estimator of β_E is given by

$$\hat{\beta}_E = \underset{\beta \in B}{argmin}\left[-2\ddot{W}(\beta, \mu_T)\right]. \tag{22}$$

Hence the estimate $\hat{\beta}_E$ maximizes the (log) empirical likelihood ratio that the simulated moment matches its empirical counterpart μ_T. In principle the estimation approach is a series of hypothesis tests with a fixed hypothesis and varying data sets, generated from different configurations, similar to a Monte Carlo setting. This differs from the standard empirical likelihood estimator for analytical moment conditions, that maximizes the empirical likelihood of the parameter of interest for a given sample (Qin and Lawless, 1994; Newey and Smith, 2004).[2]

Note, here we assume T *iid* simulation outcomes $\underline{Y}_1(\beta), ..., \underline{Y}_T(\beta)$ for $\beta \in B$, (additional to the T empirical observations $\{x_t\}_{t=1,...,T}$). By definition β_E is the parameter for which the true simulation model moment and the true empirical moment coincide. The estimate $\hat{\beta}_E$ in contrast is computed from some simulated and empirical observations. Thus for consistency

[2]For a more detailed discussion of this issue see Appendix Appendix A.1.

$\hat{\beta}_E$ must converge towards β_E while the number of simulated and empirical observations tend to infinity. With the given setting, this is equivalent to

$$\hat{\beta}_E \xrightarrow{p} \beta_E$$

as $T \to \infty$, since by assumption we have for every $\beta \in B$ in total TK simulation observations (i.e. each $\underline{Y}_i(\beta)$ with $i = 1, ..., T$ has K observations). Hence, $TK \to \infty$ as $T \to \infty$.

The consistency of $\hat{\beta}_E$ when estimating simulation models is demonstrated in this section for the simplest case where $\tilde{f}$ is one dimensional and $\beta \in B \subseteq \mathbb{R}$. The proof itself requires us to demonstrate that $-2\ddot{W}(\beta_E, \mu_T)$ is bounded in probability and $-2\ddot{W}(\beta, \mu_T)$ diverges for $\beta \neq \beta_E$ in probability as $T \to \infty$. These two properties are presented in the following subsections.

3.1. Asymptotics of $-2\ddot{W}(\beta, \mu_T)$

In this section we derive for the consistency necessary asymptotic behavior of $-2\ddot{W}(\beta, \mu_T)$. It essentially depends on the properties of the simulated and empirical process and in particular requires that the sample moment μ_T of the empirical process $\{x_t\}$ converges sufficiently quick towards the true process moment. Therefore we consider the class of strongly mixing processes as in Ibragimov and Linnik (1971) that yield a Central Limit Theorem for stationary processes.

Definition 2. Let $\mathcal{F}_k^m = \sigma(x_k, ..., x_m)$. A sequence $\{x_t\}$ is said to be strongly mixing if $\alpha(t) \to 0$ as $t \to \infty$, where

$$\alpha(t) = \sup_{A \in \mathcal{F}_{-\infty}^0 \mathcal{A}, B \in \mathcal{F}_t^\infty} |P(A \cap B) - P(A)P(B)|.$$

The following restates Theorem 18.5.3 in Ibragimov and Linnik (1971).

Theorem 1. *Suppose $\{x_t\}$ is a (strictly) stationary, centered (i.e. $E[x_t] = 0$ for all t) process satisfying the strong mixing condition with mixing coefficient $\alpha(t)$. Moreover, let $E\left[|x_1|^{2+\delta}\right] < \infty$ for some $\delta > 0$. If $\sum_{t=2}^{\infty} \alpha(t)^{\delta/(2+\delta)} < \infty$, then*

$$\sigma^2 = E\left[x_1^2\right] + 2\sum_{k=2}^{\infty} E[x_1 x_k] < \infty.$$

If in addition $\sigma^2 > 0$, then

$$\frac{\sum_{t=1}^{T} x_t}{T^{1/2}\sigma} \xrightarrow{d} N(0, 1)$$

as $T \to \infty$.

Proof. See Ibragimov and Linnik (1971). □

Lemma 1. *Let $\{x_t\}$ be (strictly) stationary and strongly mixing with mixing coefficient $\alpha_x(t)$ and furthermore $\sum_{t=2}^{\infty}\alpha_x(t)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Suppose h is a measurable, real-valued function and define the $z_t = h(x_t)$ with mixing coefficient $\alpha_z(t)$. Then*

$$\alpha_z(t) \to 0$$

as $t \to \infty$ and

$$\sum_{t=2}^{\infty}\alpha_z(t)^{\delta/(2+\delta)} < \infty.$$

Proof. Note, by definition for any mixing coefficient we have $\alpha(t) \geq 0$. Now, as above let $\mathcal{F}_k^m = \sigma(x_k, ..., x_m)$. For a measurable function h and $z_t = h(x_t)$ define $\mathcal{G}_k^m = \sigma(z_k, ..., z_m)$ and it follows $\mathcal{G}_k^m \subseteq \mathcal{F}_k^m$ and thus

$$\alpha_z(t) \leq \alpha_x(t) \tag{23}$$

for all t. [3] As $\alpha_x(t) \to 0$ as $t \to \infty$ we have also $\alpha_z(t) \to 0$ as $t \to \infty$. Consider $h(x) = x^{\delta/(2+\delta)}$ for some $\delta > 0$, then $h'(x) = \frac{\delta}{2+\delta}x^{-2/(2+\delta)} > 0$ for $x \geq 0$. Then with Eq. (23)

$$\sum_{t=2}^{\infty}\alpha_z(t)^{\delta/(2+\delta)} < \sum_{t=2}^{\infty}\alpha_x(t)^{\delta/(2+\delta)} < \infty.$$

□

Lemma 2. *Let $\{x_t\}$ be (strictly) stationary and strongly mixing with mixing coefficient $\alpha_x(t)$ and furthermore $\sum_{t=2}^{\infty}\alpha_x(t)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Let f be a measurable, real-valued function such that $\mu_0 = E[f(x_1)] < \infty$. Moreover, let $z_t = f(x_t) - \mu_0$ with $E\left[|z_1|^{2+\delta}\right] < \infty$ for some $\delta > 0$ and $\sigma_z^2 = E[z_1^2] + 2\sum_{k=2}^{\infty}E[z_1 z_k] > 0$. Then for $\mu_T = \frac{1}{T}\sum_{t=1}^{T}f(x_t)$ we have*

$$\mu_T - \mu_0 = O_p\left(T^{-\frac{1}{2}}\right).$$

Proof. As z_t is a measurable transformation of x_t, it follows that $\{z_t\}$ is stationary with $E[z_t] = 0$ for all t. From Lemma 1 it follows $\{z_t\}$ is strongly

[3] See Jones (2004), p. 305.

mixing with $\sum_{t=2}^{\infty} \alpha_z(t)^{\delta/(2+\delta)} < \infty$ such that all requirements of Theorem 1 are satisfied, thus

$$\frac{\sum_{t=1}^{T} z_t}{T^{1/2}} \xrightarrow{d} N\left(0, \sigma_z^2\right) \tag{24}$$

where $\sigma_z^2 = E\left[z_1^2\right] + 2\sum_{k=2}^{\infty} E\left[z_1 z_k\right] < \infty$. With Eq. (24) we have

$$\begin{aligned} \frac{\sum_{t=1}^{T} \left(f(x_t) - E\left[f(x_1)\right]\right)}{T^{1/2}} &= O_p(1) + o_p(1) \\ \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} f(x_t) - E\left[f(x_1)\right]\right) &= O_p(1) \\ \frac{1}{T}\sum_{t=1}^{T} f(x_t) - E\left[f(x_1)\right] &= O_p\left(T^{-\frac{1}{2}}\right) \\ \mu_T - \mu_0 &= O_p\left(T^{-\frac{1}{2}}\right), \end{aligned} \tag{25}$$

hence $\mu_T = \frac{1}{T}\sum_{t=1}^{T} f(x_t)$ is a consistent estimate of $\mu_0 = E\left[f(x_1)\right]$. □

Lemma 3. *Let $Y_i \geq 0$ iid and suppose that $E\left(Y_1^2\right) < \infty$ then*

$$\max_{i=1,..,n} Y_i = o\left(n^{\frac{1}{2}}\right)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^3 = o\left(n^{\frac{1}{2}}\right)$$

with probability 1 as $n \to \infty$.

Proof. See Owen (1990), p. 98. □

The following provides some quantities and results that are needed to derive the asymptotic behavior of $-2\ddot{W}(\beta, \mu_T)$. We define

$$\begin{aligned} \ddot{g}^*(\beta, \mu_T) &= \max_{i=1,\dots,T} \left|\ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right)\right| \\ &= \max_{i=1,\dots,T} \left|\tilde{f}\left(\underline{Y}_i(\beta)\right) - \mu_T\right|, \end{aligned}$$

$$\begin{aligned} \bar{\ddot{g}}_T(\beta, \mu_T) &= \frac{1}{T}\sum_{i=1}^{T} \ddot{g}\left(\underline{Y}_i(\beta), \mu_T\right) \\ &= \frac{1}{T}\sum_{i=1}^{T} \tilde{f}\left(\underline{Y}_i(\beta)\right) - \mu_T, \end{aligned}$$

and

$$\ddot{\tilde{S}}\left(\beta,\mu_T\right)=\frac{1}{T}\sum_{i=1}^{T}\ddot{g}\left(\underline{Y}_i\left(\beta\right),\mu_T\right)^2.$$

Then the following magnitudes hold.

Lemma 4. *Let z_t be the process defined in Lemma 2 and $y_t\left(\beta\right)$ the stationary simulation process with parameter $\beta\in B$. Let there be a $\beta_E\in B$ such that*

$$E\left[f\left(y_1\left(\beta_E\right)\right)\right]=E\left[f\left(x_1\right)\right]$$

with $\sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}=Var\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)\right]<\infty$. Then

$$\ddot{g}^*\left(\beta_E,\mu_T\right)=o_p\left(T^{\frac{1}{2}}\right)$$

and

$$\ddot{\bar{g}}_T\left(\beta_E,\mu_T\right)=O_p\left(T^{-\frac{1}{2}}\right).$$

Moreover,

$$\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)=O_p\left(1\right)$$

and

$$\frac{1}{T}\sum_{i=1}^{T}\left|\ddot{g}\left(\underline{Y}_i\left(\beta_E\right),\mu_T\right)\right|^3=o_p\left(T^{\frac{1}{2}}\right).$$

Proof. See Appendix Appendix A.2. □

The following theorem demonstrates that $-2\ddot{W}\left(\beta_E,\mu_T\right)$ is bounded in probability as $T\to\infty$.

Theorem 2. *Let z_t be the process defined in Lemma 2 and $y_t\left(\beta\right)$ a (strictly) stationary simulation process with parameter $\beta\in B$. Let there be a $\beta_E\in B$ such that*

$$E\left[f\left(y_1\left(\beta_E\right)\right)\right]=E\left[f\left(x_1\right)\right]$$

is satisfied with $\sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}=Var\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)\right]<\infty$. Then

$$-2\ddot{W}\left(\beta_E,\mu_T\right)\;=\;O_p\left(1\right).$$

Proof. See Appendix Theorem Appendix A.3. □

The following Theorem proofs that for $\beta\neq\beta_E$ the term $-2\ddot{W}\left(\beta,\mu_T\right)$ diverges as $T\to\infty$.

Theorem 3. *Let z_t be the process defined in Lemma 2 and $y_t(\beta)$ a stationary simulation process with parameter $\beta \in B$. Let there be a $\beta_E \in B$ such that*

$$E\left[f\left(y_1\left(\beta_E\right)\right)\right] = E\left[f\left(x_1\right)\right]$$

is satisfied. Suppose a $\beta \neq \beta_E$ with $E\left[f\left(y_1\left(\beta\right)\right)\right] < \infty$ and $\left|E\left[f\left(y_1\left(\beta\right)\right)\right] - E\left[f\left(x_1\right)\right]\right| > 0$. Moreover, let $\sigma^2_{\tilde{f}(\underline{Y}(\beta))} = Var\left[\tilde{f}\left(\underline{Y}\left(\beta\right)\right)\right] < \infty$. Then

$$-2T^{-\frac{1}{3}}\ddot{W}\left(\beta, \mu_T\right) \xrightarrow{p} \infty$$

as $T \to \infty$.

Proof. See Appendix Appendix A.4. □

3.2. Consistency of $\hat{\beta}_E$

In this section we demonstrate the consistency of $\hat{\beta}_E$ in Eq. (22) for simulation-based estimation problems as in Eq. (14).

Assumption 1.

- Let $\{x_t\}$ be (strictly) stationary and strongly mixing with mixing coefficient $\alpha_x(t)$ and $\sum_{t=2}^{\infty} \alpha_x(t)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Let f be a measurable, real-valued one dimensional function such that $\mu_0 = E\left[f\left(x_1\right)\right] < \infty$. Moreover, let $z_t = f\left(x_t\right) - \mu_0$ with $E\left[\left|z_1\right|^{2+\delta}\right] < \infty$ for some $\delta > 0$ and $\sigma_z^2 = E\left[z_1^2\right] + 2\sum_{k=2}^{\infty} E\left[z_1 z_k\right] > 0$.

- The simulation process $\{y_t(\beta)\}$ is (strictly) stationary and independent from $\{x_t\}$ for $\beta \in B \subseteq \mathbb{R}$. There is a unique parameter $\beta_E \in B$ such that $E\left[f\left(y_1\left(\beta_E\right)\right)\right] = E\left[f\left(x_1\right)\right]$. For all other $\beta \in B$ with $\beta \neq \beta_E$, let $\left|E\left[f\left(y_1\left(\beta\right)\right)\right] - E\left[f\left(x_1\right)\right]\right| > 0$.

- For finite and fix $K \in \mathbb{N}$, $\underline{Y}(\beta) = \left[y_1(\beta), ..., y_K(\beta)\right]' \in \mathbb{R}^K$ and $\tilde{f}$ is defined in Lemma 5 with $l = 1$, let $Var\left[\tilde{f}\left(\underline{Y}\left(\beta\right)\right)\right] < \infty$ for all $\beta \in B$.

Remark 1. Note, a strictly stationary process $\{x_t\}$ that is strong mixing is also ergodic (e.g. see (Lindgren, 2006) on p. 158).

Remark 2. As f is measurable it follows that $\tilde{f}$ is measurable (e.g. see Appendix Appendix A.7).

Theorem 4. *If Assumption 1 is satisfied, then*

$$\hat{\beta}_E \xrightarrow{p} \beta_E$$

as $T \to \infty$.

Proof. With Assumption 1 all requirements of Theorem 2 are satisfied, thus

$$-2\ddot{W}(\beta_E, \mu_T) = O_p(1).$$

By definition for every $\varepsilon > 0$ there exists a constant D such that

$$P\left(\left|-2\ddot{W}(\beta_E, \mu_T)\right| \leq D\right) > 1 - \varepsilon$$

for all $T \in \mathbb{N}$. Note, as[4]

$$-2\ddot{W}(\beta_E, \mu_T) : \Omega \to \mathbb{R}_+, \tag{26}$$

hence for every $\varepsilon > 0$, there exists a constant D such that

$$P\left(-2\ddot{W}(\beta_E, \mu_T) \leq D\right) > 1 - \varepsilon \tag{27}$$

for all $T \in \mathbb{N}$. Now with Assumption 1 satisfying the requirements of Theorem 3, we have for all $\beta \in B$ with $\beta \neq \beta_E$:

$$-2T^{-\frac{1}{3}}\ddot{W}(\beta, \mu_T) \xrightarrow{p} \infty$$

as $T \to \infty$ or

$$-2\ddot{W}(\beta, \mu_T) \xrightarrow{p} \infty \tag{28}$$

as $T \to \infty$, since $T^{-\frac{1}{3}} \leq 1$ for all $T \in \mathbb{N}$. Then for the constant D in Eq. (27), Eq. (28) implies that

$$P\left(-2\ddot{W}(\beta, \mu_T) > D\right) \to 1 \tag{29}$$

as $T \to \infty$ for all $\beta \in B$ with $\beta \neq \beta_E$. Hence, there exists a finite number $T_\beta \in \mathbb{N}$ such that

$$P\left(-2\ddot{W}(\beta, \mu_T) > D\right) \geq 1 - 2\varepsilon \tag{30}$$

for all $T \geq T_\beta$, where $\beta \in B$ and $\beta \neq \beta_E$. Let $B^* := B \backslash \{\beta_E\}$ and $C := \{T_\beta | \beta \in B^*\}$, then from above it follows that all elements of C are finite natural numbers, hence its largest element $T^* = \max_{\beta \in B^*} \{T_\beta\}$ is also a finite natural number such that Eq. (30) holds for all $\beta \in B^*$ if $T \geq T^*$. With Eq. (27) the following inequalities hold with probability at least $1 - 2\varepsilon$:

$$-2\ddot{W}(\beta, \mu_T) > D \geq -2\ddot{W}(\beta_E, \mu_T) \tag{31}$$

[4] See Appendix Appendix A.8.

for all $\beta \in B^*$as $T \geq T^*$. As $\hat{\beta}_E = \underset{\beta \in B}{argmin}\left[-2\ddot{W}\left(\beta, \mu_T\right)\right]$ it follows with Eq. (31)

$$P\left(\hat{\beta}_E = \beta_E\right) \geq 1 - 2\varepsilon$$

and hence

$$P\left(\hat{\beta}_E \neq \beta_E\right) \leq 2\varepsilon \tag{32}$$

or equivalently

$$P\left(\hat{\beta}_E \in B^*\right) \leq 2\varepsilon$$

for all $T \geq T^*$. Now consider the set

$$B_\delta := \left\{\beta \middle| \left|\beta - \beta_E\right| > \delta, \beta \in B\right\}.$$

As by assumption β_E is unique and it follows $B_\delta \subseteq B^*$ for all $\delta > 0$. The latter gives

$$\begin{aligned} P\left(\left|\hat{\beta}_E - \beta_E\right| > \delta\right) &= P\left(\hat{\beta}_E \in B_\delta\right) \\ &\leq P\left(\hat{\beta}_E \in B^*\right) \leq 2\varepsilon \end{aligned}$$

for every $\delta > 0$ and $T \geq T^*$. Finally, as ε and δ can be chosen arbitrarily close to 0, the asserted convergence in probability is established. □

Remark 3. The presented proof of the consistency of $\hat{\beta}_E$ for simulation-based estimation problems as in as in Eq. (14) only addresses the one dimensional case as it employs the univariate CLT for iid variables and strong mixing sequences in Theorem 2. These CLTs guarantee that the simulated and empirical sample moment, i.e. $\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta\right)\right)$ and $\frac{1}{T}\sum_{t}^{T}f\left(x_t\right)$, converge with a rate of $T^{-\frac{1}{2}}$ to their true moment. Therefore using the multivariate CLT for *iid* variables and strong mixing sequences should demonstrate consistency of $\hat{\beta}_E$ when estimating simulation models with multivariate simulated and empirical sample moment. Note, multivariate CLT can be derived from the univariate CLT using the Cramer-Wold Lemma (see Van der Vaart (1998), p. 16).

4. Simulation and Estimation Experiment

In this section we illustrate the application of our simulation-based estimation approach in Section 3 for general problems of the form

$$\mu\left(\beta_E\right) = E\left[f\left(x_1\right)\right],$$

where $\mu(\beta_E)$ is the true moment of the simulation model and $E[f(x_1)]$ is the corresponding true moment of the empirical observation that is only given by an estimate μ_T derived from some sample time series $\{x_t\}_{t=1,\ldots,T}$:

$$\mu_T = \frac{1}{T}\sum_{t=1}^{T} f(x_t).$$

Thus we compute parameter estimates derived from simulated (log GBM increment) sample paths and pseudo empirical moments, which are generated from some pre-specified GBM setting and analyze its performance in comparison to the AEL (Chen et al., 2008) and the SMM approach (Lee and Ingram, 1991; Duffie and Singleton, 1993).

Consider the stochastic differential equation

$$dS_t = \alpha S_t dt + \delta S_t dW_t$$

and its Itô solution

$$S_t = S_0 \exp\left(\left(\alpha - \frac{\delta^2}{2}\right)t + \delta W_t\right),$$

yielding

$$\log(S_t) = \log(S_0) + \left(\alpha - \frac{\delta^2}{2}\right)t + \delta W_t.$$

For a fix Δt and $j \in \mathbb{N}$ define the log return process $\{r_j\}_{j\in\mathbb{N}}$ by

$$\begin{aligned} r_j &= \log(S_{j\Delta t}) - \log\left(S_{(j-1)\Delta t}\right) \\ &= \left(\alpha - \frac{\delta^2}{2}\right)\Delta t + \delta\left(W_{j\Delta t} - W_{(j-1)\Delta t}\right). \end{aligned} \tag{33}$$

Since W_t has i.i.d. normal innovations, we have $W_{j\Delta t} - W_{(j-1)\Delta t} \sim N(0, \Delta t)$ and the log return process $\{r_j\}_{j\in\mathbb{N}}$ in Eq. (33) follows

$$r_j \sim N\left(\left(\alpha - \frac{\delta^2}{2}\right)\Delta t, \delta^2\Delta t\right). \tag{34}$$

Using this property of the GBM, the log GBM return time series samples can be simulated by

$$\begin{aligned} \begin{pmatrix} \underline{r}_1(\beta) \\ \vdots \\ \underline{r}_n(\beta) \end{pmatrix} &= \begin{pmatrix} r_{1,1}(\beta) & \ldots & r_{1,K}(\beta) \\ \vdots & & \vdots \\ r_{n,1}(\beta) & \ldots & r_{n,1}(\beta) \end{pmatrix} \\ &= \left(\alpha - \frac{\delta^2}{2}\right)\Delta t + \sqrt{\delta^2\Delta t}\begin{pmatrix} r^*_{1,1} & \ldots & r^*_{1,K} \\ \vdots & & \vdots \\ r^*_{n,1} & \ldots & r^*_{n,1} \end{pmatrix}, \end{aligned} \tag{35}$$

where $r^*_{i,j}$ are independent $N(0,1)$ for all $i = 1, ..., n$ and $j = 1, ..., K$. As pseudo empirical moments we use

$$\mu_T = \frac{1}{T}\sum_{t=1}^{T}\left(\begin{array}{c} x_t \\ x_t^2 \end{array}\right),$$

that is the first two moments generated from some sample GBM log return time series $\{x_t\}_{t=1,...,T}$ with setting

$$\beta_E = (\alpha_E, \delta_E) = (2, 1)$$

and $\Delta t = 1/10$, where the true moments are

$$E\left[\left(\begin{array}{c} x_1 \\ x_1^2 \end{array}\right)\right] = \left(\begin{array}{c} 0.15 \\ 0.10225 \end{array}\right).$$

Consider

$$g_i = \left(\begin{array}{c} \frac{1}{K}\sum_{j=1}^{K} r_{i,j}(\beta) \\ \frac{1}{K}\sum_{j=1}^{K} (r_{i,j}(\beta))^2 \end{array}\right).$$

Given some T simulated log GBM return time series samples of length K and motivated by Section 3 a BAEL estimate of β_E is given with a fixed $s \in \mathbb{R}$ by

$$\hat{\beta}_E^{BAEL} = \underset{\beta \in B}{argmin}\left[-2\tilde{W}_\beta(\mu_T)\right] \tag{36}$$

$$= \underset{\beta \in B}{argmin}\left[-2\log \tilde{R}_\beta(\mu_T)\right], \tag{37}$$

where

$$\tilde{R}_\beta(\mu_T) = \sup\left\{\prod_{i=1}^{T+2} nw_i \middle| \sum_{i=1}^{T+2} w_i g_i = \mu_T, \sum_{i=1}^{T+2} w_i = 1, w_i \geq 0\right\}. \tag{38}$$

Note, in Section 3 we have only demonstrated the consistency of a simulation estimator based on standard EL. Here however, in order to avoid the convex hull problem of the empirical likelihood ratio function, we experimentally consider the BAEL version (see Section 2). The AEL type equivalent of $\hat{\beta}_E^{BAEL}$ is given by

$$\hat{\beta}_E^{AEL} = \underset{\beta \in B}{argmin}\left[-2W^*_\beta(\mu_T)\right],$$

where $W^*_\beta(\mu_T) = \log\left(R^*_\beta(\mu_T)\right)$ with

$$R^*_\beta(\theta_e) = \sup\left\{\prod_{i=1}^{T+1} nw_i \middle| \sum_{i=1}^{T+1} w_i g_i = \mu_T, \sum_{i=1}^{T+1} w_i = 1, w_i \geq 0\right\} \tag{39}$$

including one artificial sample point

$$g_{T+1}(\beta) = -\frac{\log(n)}{2}\frac{1}{T}\sum_{i=1}^{T} g_i.$$

Finally, the SMM type estimator is given by

$$\hat{\beta}_E^{SMM} = \underset{\beta \in B}{argmin}\, \hat{Q}_T(\beta), \tag{40}$$

where

$$\hat{Q}_T(\beta) = G_T^*(\beta)' \hat{W} G_T^*(\beta),$$

and

$$G_T^*(\beta) = \left(\left(\begin{array}{c} \frac{1}{TK}\sum_{t=1}^{TK} r_t(\beta) \\ \frac{1}{TK}\sum_{t=1}^{TK} r_t(\beta)^2 \end{array} \right) - \left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T} x_t \\ \frac{1}{T}\sum_{t=1}^{T} x_t^2 \end{array} \right) \right)$$

with

$$\frac{1}{TK}\sum_{t=1}^{TK} r_t(\beta) = \frac{1}{TK}\sum_{i=1}^{T}\sum_{j=1}^{K} r_{i,j}(\beta). \tag{41}$$

As in Franke (2009), the we set $\hat{W} = \hat{\Omega}^{-1}$ for the weight matrix, where

$$\hat{\Omega} = \Gamma_0 + \sum_{j=1}^{P}\left(1 - \frac{j}{p+1}\right)\left(\Gamma_j + \Gamma_j'\right) \tag{42}$$

and[5]

$$\Gamma_j = \frac{1}{T}\sum_{t=j+1}^{T}\left(\left(\begin{array}{c} x_t \\ x_t^2 \end{array}\right) - \mu_T\right)\left(\left(\begin{array}{c} x_{t-j} \\ x_{t-j}^2 \end{array}\right) - \mu_T\right)'.$$

With this choice of the weight matrix, the SMM estimate has the smallest asymptotic covariance (e.g. see Lee and Ingram (1991) or Duffie and Singleton (1993)).

For the experiment we consider $T = [10, 25, 50, 100, 250]$ and compute 10^3 estimates of $\hat{\beta}_E^{BAEL}$, $\hat{\beta}_E^{AEL}$ and $\hat{\beta}_E^{SMM}$. Each estimator is derived from a pseudo empirical moment μ_T (derived from a GBM time series with $\beta_E = (2, 1)$) and some T simulated log return GBM series samples of length $K = 5$ and $\Delta t = 1/10$.

Table 1 presents some statistics of the pseudo empirical moment samples, that are used for estimation. Note, within each estimation we fix the seed

[5]Similarly to Franke (2009), the bandwidth p we set as the smallest integer greater than or equal to $T^{1/4}$.

T	First moment samples		Second moment samples	
	Mean	Variance	Mean	Variance
10	1.5222E-1	9.5472E-3	1.2447E-1	3.0204E-3
25	1.5096E-1	3.7129E-3	1.2397E-1	1.1295E-3
50	1.5066E-1	1.8461E-3	1.2289E-1	5.4831E-4
100	1.4991E-1	9.7118E-4	1.2225E-1	2.8179E-4
250	1.4976E-1	3.8969E-4	1.2242E-1	1.0915E-4

Table 1: This table displays the mean and the variance of the moment samples a 1000 different of GBM log return time series $\{x_t\}_{t=1,...,T}$ with length T, generated with $\beta_E = (\alpha_E, \delta_E) = (2, 1)$ and $\Delta t = 1/10$, where the true moments are given by $E[x_1] = 0.15$ and $E[x_1^2] = 0.1225$.

of the sequence of random numbers such that the difference between two time series $\{r_j(\beta_1)\}$ and $\{r_j(\beta_2)\}$ can be attributed to the differences in the parameters β_1 and β_2. For each estimation methodology (BAEL, AEL and SMM) we use the same set of seeds for the sequences of random numbers, such that for a given seed all methodologies would operate on the same samples $r^*_{i,j}$ for all $i = 1, ..., T$ and $j = 1, ..., K$ (see Eq. (35)). Therefore, the difference of the resulting $\hat{\beta}_E^{BAEL}, \hat{\beta}_E^{AEL}$ and $\hat{\beta}_E^{SMM}$ are a product of the method (SMM, BAEL and AEL) and the optimization algorithm in use to find the minimum. In order to ease replication of our method we employ for this experiment some common ready-implemented optimization routines in Matlab[6]: Nelder-Mead, interior-point, sequential-quadratic-programming and active-set algorithm.

Figure 1 displays the MSE of α and δ of 10^3 BAEL, AEL and SMM estimation results, where the true configuration is given with $\alpha_E = 2$ and $\delta_E = 1$. For the BAEL estimation approach a range of s values are considered with $s = [1, 50, 100]$. Each estimator is computed from a pseudo empirical moment μ_T (derived from a GBM time series with $\beta_E = (2, 1)$) and some T simulated log return GBM series samples of length $K = 5$ and $\Delta t = 1/10$. Across estimation methodologies (SMM, BAEL and AEL) the experiment uses the same set of fixed seeds such that for a given seed all methodologies would operate on the same samples $r^*_{i,j}$ for all $i = 1, ..., T$ and $j = 1, ..., K$. Therefore the difference of the resulting $\hat{\beta}_E^{BAEL}, \hat{\beta}_E^{AEL}$ and $\hat{\beta}_E^{SMM}$ is due to the difference of the surface generated by the considered methodologies. In particular the optimization algorithm is run on the quadrant $[0.01, 3] \times [0.01, 4]$

[6]http://www.mathworks.nl/help/optim/ug/fmincon.html#brh041i

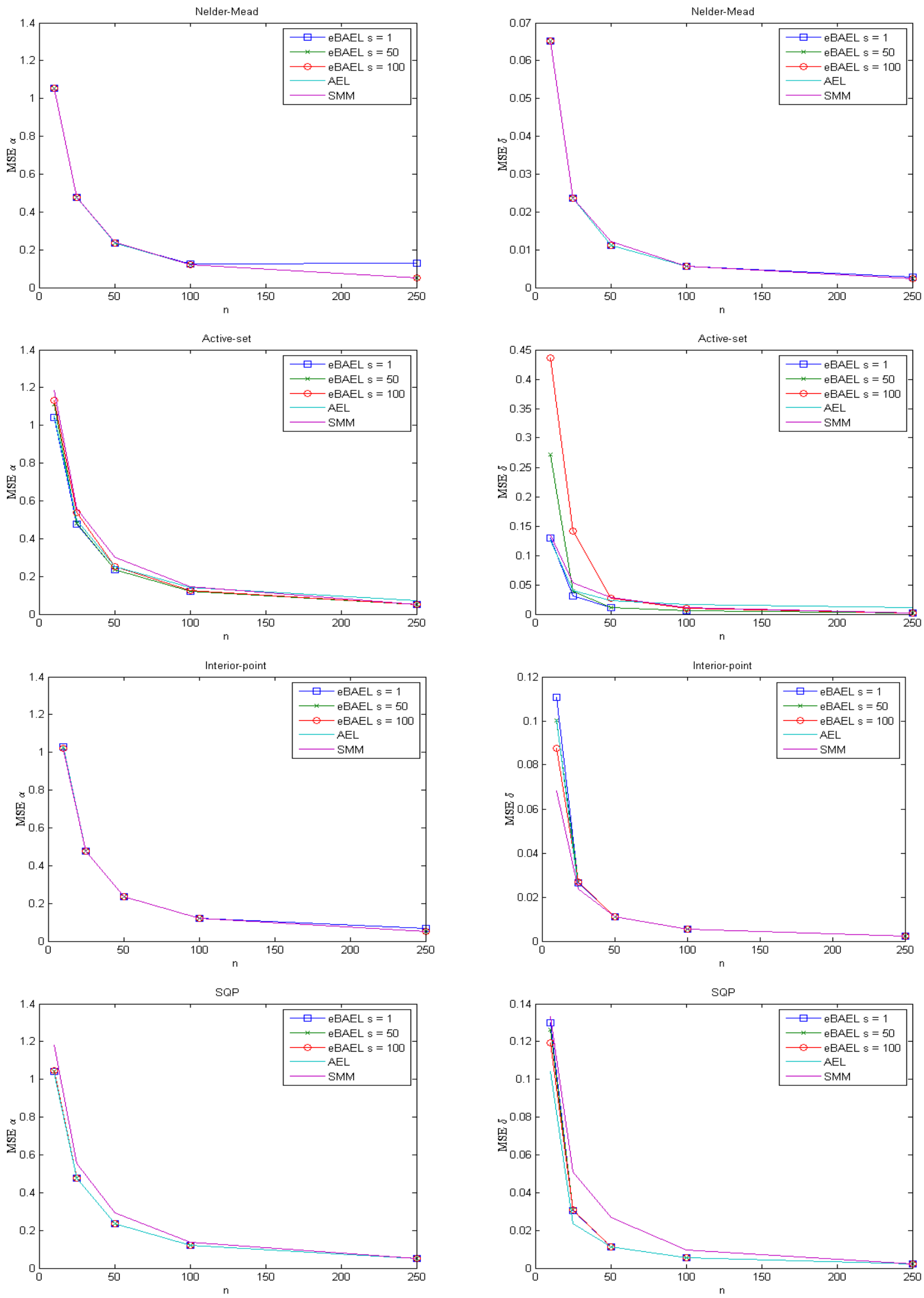

Figure 1: MSE of α and δ of 10^3 BAEL, AEL and SMM estimation results using different algorithms.

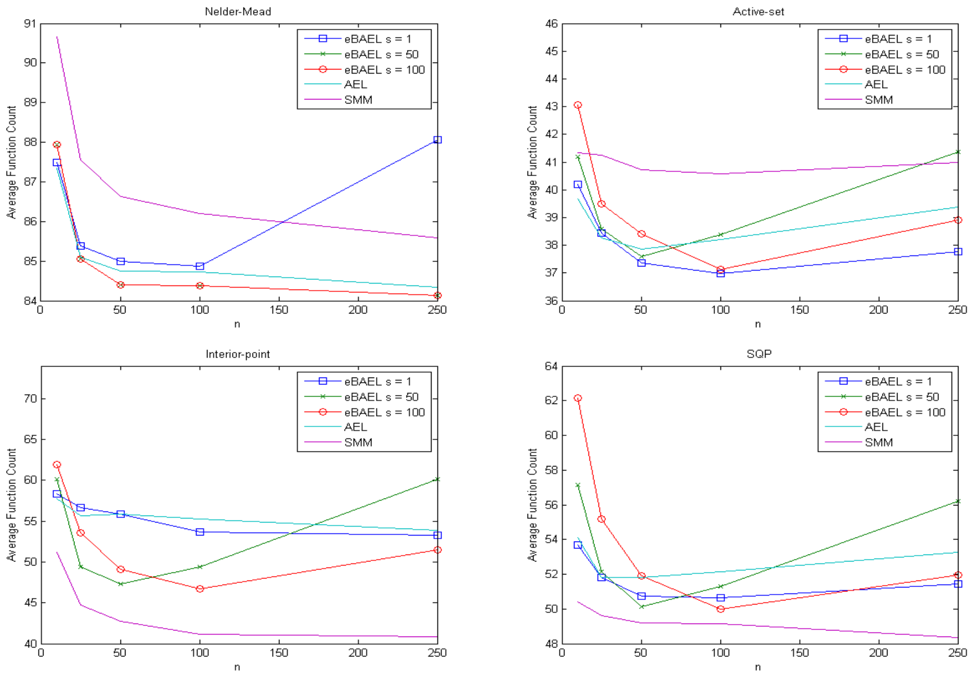

Figure 2: Estimation, average function count.

This figure presents the average function number of evaluation of $\tilde{W}_{\beta}(\theta_e)$, $W^*_{\beta}(\theta_e)$ and $\hat{Q}_T(\beta)$ with respect to β during the 1000 estimation experiments of Figure 1.

with random initial starting point.

Figure 1 displays the resulting MSE (mean square error) of 1000 repetitions of the estimation experiment across the considered optimization routines and Figure 2 displays the corresponding average function evaluation of the optimization routines. These figure demonstrate that for the active-set algorithm, the BAEL method with $s = 1$ performs best in terms of MSE and efficiency. However, different values of s seem to have an effect on the MSE of δ as well as the efficiency (low average function count). While the largest effect of s on the MSE of δ holds for $T \leq 50$, the effect of s on the average function count is convex in T. The AEL and SMM methodologies have slightly worse MSEs than the BAEL method with $s = 1$. In terms of efficiency, the SMM method is less efficient than EL approaches. In the case of the interior-point algorithm, the SMM methodology works best in terms of accuracy and efficiency. While for α the MSE of all methods are almost the same, for δ there is some difference in the MSE for $T \leq 50$ between the SMM and the EL approaches, the latter being slightly higher. In terms of efficiency, the EL approaches are worse than the SMM, while BAEL is mostly better than the AEL method. For the Nelder-Mead algorithm all MSE are roughly the same except for BAEL with $s = 1$ at $T = 250$, where Figure 2 suggest that in some occasion the algorithm has been trapped in a local minimum. In general, the EL approaches have been more efficient than the SMM method. In the sequential-quadratic-programming case, the EL approaches are performing better in terms of MSE, however they are less efficient than the SMM.

Overall, for the majority of cases there exists an BAEL version that has a better MSE than SMM.[7]

5. Conclusion

In this paper we introduced empirical likelihood for estimating simulation models with complex dynamics that have non-analytical outputs, and therefore have no information on the moments and the likelihood function or a reduced form. Similarly to the SMM (Lee and Ingram, 1991; Duffie and Singleton, 1993), the proposed method matches simulated and empirical time series moments via empirical likelihood. For one dimensional estimation equations of the mean, we showed that the proposed method converges to the true parameter value. Moreover, as the consistency of this EL simulation-based estimator should be straightfowardly extendable to the mulitvariate case (see Remark 3), the main differences of our estimation approach to the SMM are: (i) it does not allow for overidentifying restrictions, such that the number of moment conditions can be greater than the dimension of the parameter; (ii) it has a stronger requirement on the empirical process $\{x_t\}$ being stationary and strongly mixing; (iii) in terms of the simulation process however, the EL methodology does not require some form of continuity[8] of the model moments with respect to its parameters or the ergodicity of the simulation process. Note, the ergodicity assumption of the simulation process in the SMM approach is not very critical. Ergodicity allows to estimate the true simulation moments with both sample and time averages. Therefore, even the if the simulation process is not ergodic, one should be able to replace the time average estimates in the SMM with sample average estimates, while retaining at least consistency of the methodology. However, the continuity requirements of the simulation model with respect to its parameters is rather a critical assumption for simulation models with non-analytical complex dy-

[7]Note, however there is no clear indication on the best choice of s. and there is no clear pattern of the efficiency effect of s.

[8]For the consistency requirements as the SMM estimator and in particular the continuity requirement see Appendix Appendix A.9

namics. For such models it maybe hard or even impossible to demonstrate the continuity of their moments. Empirically, the feasibility of the proposed estimation method was demonstrated in a simple simulation exercise with a geometric Brownian motion, where we were able to obtain smaller mean squared errors than the SMM.

Overall our work shows the promise of empirical likelihood as a general-purpose tool for estimating complex simulation models, such as agent-based models, which are increasingly being used in economics and finance, as well as other social sciences and the physical sciences. Towards this goal we have shown EL has several necessary properties which make it suitable as the basis for such an estimation method. In future work we will extend these results by analyzing the asymptotic distribution of the estimation error (e.g. $\hat{\beta}_E - \beta_E$), and provide an extension for over-identifying restrictions including its test.

References

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States*, 99(Suppl. 3):7280–7287.

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

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

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

Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.

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.

Gallant, A. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12(04):657–681.

Gourieroux, C. and Monfort, A. (1991). Simulation Based Inference in Models with Heterogeneity. *Annals of Economics and Statistics*, (20-21):69–107.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff.

Jones, G. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320.

Lee, B.-S. and Ingram, B. F. (1991). Simulation estimation of time-series models. *Journal of Econometrics*, 47(2-3):197–205.

Lee, L. (1992). On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. *Econometric Theory*, 8(4):518–552.

Lee, L. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 11(3):437–483.

Lindgren, G. (2006). Lectures on Stationary Stochastic Processes. Technical report, Lund University.

Meintrup, D. and Schäffler, S. (2005). *Stochastik Theorie und Anwendung.* Springer, Berlin, Heidelberg.

Newey, W. and Smith, R. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

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

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

Shalizi, C. R. (2006). Methods and Techniques of Complex Systems Science: An Overview. In Deisboeck, T. S. and Kresh, J. Y., editors, *Compex Systems Science in Biomedicine*, Topics in Biomedical Engineering International Book Series, chapter 2, pages 33–114. Springer US.

Van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Xue, L. (2013). Estimation and empirical likelihood for single-index models with missing data in the covariates. *Computational Statistics & Data Analysis*, 68(0):82 – 97.

Yu, J. (2012). Simulation Based Estimation Methods for Financial Time Series Models. In Duan, J.-C., Härdle, W. K., and Gentle, J. E., editors, *Handbook of Computational Finance*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Zhao, H., Zhao, P.-Y., and Tang, N.-S. (2013). Empirical likelihood inference for mean functionals with nonignorably missing response data. *Computational Statistics & Data Analysis*, 66(0):101 – 116.

Appendix A. Appendix

Appendix A.1. Discrimination of the Estimation Problem in Eq. (17) in Contrast to Standard EL Moment Condition Problem

For moment conditions of the form

$$\begin{aligned} 0 &= E\left[g\left(y, \beta_0\right)\right] \\ &= \int g\left(y, \beta_0\right) dP_y, \end{aligned} \tag{A.1}$$

Newey and Smith (2004) and Newey and Smith (2004) have demonstrated that under some regularity conditions, a consistent empirical likelihood estimator is given by

$$\begin{aligned} \hat{\beta} &= \underset{\beta \in B}{argmin} \left[W\left(\beta\right)\right] \\ &= \underset{\beta \in B}{argmin} \left[\sup_{\lambda} \sum_{i=1}^{n} \log\left(1 + \lambda' g\left(y_i, \beta\right)\right)\right], \end{aligned}$$

where $y_1, ..., y_n$ are some *iid* samples and β is some q-dimensional parameter $\beta = C\left(dP_y\right)$, that is expressed as a functional C of the unknown distribution dP_y of y. The estimation problem in this paper however is of the form

$$\begin{aligned} E\left[\ddot{g}\left(\underline{Y}\left(\beta, \mu_0\right)\right)\right] &= 0 \\ E\left[\tilde{f}\left(\underline{Y}\left(\beta\right)\right)\right] &= \mu_0 \\ \int \tilde{f}\left(\underline{Y}\right) dP_{\underline{Y}(\beta)} &= \mu_0, \end{aligned}$$

where β is not just some q-dimensional parameter that is a functional of the unknown distribution $dP_{\underline{Y}}$ of $\underline{Y}$ but in fact the parameter β that we consider in this paper does also effect the distribution, which is denoted by $dP_{\underline{Y}(\beta)}$.

Appendix A.2. Proof of Lemma 4

Proof. First note that from Lemma 2, it follows that $|\mu_T| = O_p(1)$. [9] From $\sigma^2_{\tilde{f}(\underline{Y}(\beta_E))} = E\left[\tilde{f}(\underline{Y}(\beta_E))^2\right] - E\left[\tilde{f}(\underline{Y}(\beta_E))\right]^2 < \infty$, it follows $E\left[\left|\tilde{f}(\underline{Y}(\beta_E))\right|^2\right] < \infty$. Now, since f is measurable with Lemma 5 we have that $\tilde{f}$ is measurable. Hence for *iid* $\underline{Y}_i(\beta_E)$ with $i = 1, ..., T$ it follows $\left|\tilde{f}(\underline{Y}_i(\beta_E))\right|$ are iid. Thus all conditions of Lemma 3 are satisfied such that

$$\max_{i=1,...,T} \left|\tilde{f}(\underline{Y}_i(\beta_E))\right| = o_p\left(T^{\frac{1}{2}}\right),$$

and we have

$$\begin{aligned} \ddot{g}^*(\beta_E, \mu_T) &= \max_{i=1,...,T} \left|\tilde{f}(\underline{Y}_i(\beta_E)) - \mu_T\right| \\ &\leq \max_{i=1,...,T} \left|\tilde{f}(\underline{Y}_i(\beta_E))\right| + |\mu_T| \\ &= o_p\left(T^{\frac{1}{2}}\right) + O_p(1) \\ &= o_p\left(T^{\frac{1}{2}}\right). \end{aligned} \tag{A.2}$$

For $\bar{\ddot{g}}_T(\beta_E, \mu_T)$ we can write

$$\begin{aligned} \bar{\ddot{g}}_T(\beta_E, \mu_T) &= \frac{1}{T}\sum_{i=1}^{T} \ddot{g}(\underline{Y}_i(\beta_E), \mu_T) \\ &= \frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_T \\ &= \frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0 + \mu_0 - \mu_T. \end{aligned}$$

As the assumptions of the CLT are satisfied[10], we get[11] $\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0 = O_p\left(T^{-\frac{1}{2}}\right)$. Lemma 2 gives $\mu_0 - \mu_T = O_p\left(T^{-\frac{1}{2}}\right)$ (see Eq. (25)) and

[9] As $\mu_T \xrightarrow{p} \mu_0$ we have $|\mu_T - \mu_0| = o_p(1)$. Using the inverse triangle inequality, it follows $|\mu_T| - |\mu_0| \leq |\mu_T - \mu_0| = o_p(1)$, hence $|\mu_T| \leq O(1) + o_p(1) = O_p(1)$.

[10] The assumptions of the CLT are satisfied by $\tilde{f}(\underline{Y}_i(\beta_E))$ iid, $E\left[\tilde{f}(\underline{Y}(\beta_E))\right] = \mu_0$ and $\sigma^2_{\tilde{f}(\underline{Y}(\beta_E))} < \infty$.

[11] From the CLT it follows $\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0\right) \xrightarrow{d} N\left(0, \sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}\right)$, that is $\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0\right) = N\left(0, \sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}\right) + o_p(1) = O_p(1) + o_p(1) = O_p(1)$. Hence, $\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0 = O_p\left(T^{-\frac{1}{2}}\right)$.

therefore

$$\ddot{\bar{g}}_T(\beta_E,\mu_T) = O_p\left(T^{-\frac{1}{2}}\right) + O_p\left(T^{-\frac{1}{2}}\right) = O_p\left(T^{-\frac{1}{2}}\right).$$

For $\ddot{\tilde{S}}(\beta_E,\mu_T)$ we have

$$\begin{aligned}\ddot{\tilde{S}}(\beta_E,\mu_T) &= \frac{1}{T}\sum_{i=1}^{T}\ddot{g}\left(\underline{Y}_i(\beta_E),\mu_T\right)^2 \\ &= \frac{1}{T}\sum_{i=1}^{T}\left(\tilde{f}\left(\underline{Y}_i(\beta_E)\right)-\mu_T\right)^2 \\ &= \frac{1}{T}\sum_{i=1}^{T}\left(\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\right)^2 - 2\mu_T\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\right)+\mu_T^2.\end{aligned}$$

Using the Continuous Mapping Theorem (CMT) and the Slutsky theorem repeatedly, it follows[12]

$$\ddot{\tilde{S}}(\beta_E,\mu_T) \xrightarrow{p} E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)^2\right] - 2E\left[f(x_1)\right]E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right] + \left(E\left[f(x_1)\right]\right)^2 \tag{A.3}$$

$$= E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)^2\right] - E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right]^2$$

$$= Var\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right], \tag{A.4}$$

where the second last line follows from the definition of β_E. By assumption $Var\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right] < \infty$, thus $\ddot{\tilde{S}}(\beta_E,\mu_T) = O_p(1)$. For $\frac{1}{T}\sum_{i=1}^{T}\left|\ddot{g}\left(\underline{Y}_i(\beta_E),\mu_T\right)\right|^3$ we can write

$$\begin{aligned}\frac{1}{T}\sum_{i=1}^{T}\left|\ddot{g}\left(\underline{Y}_i(\beta_E),\mu_T\right)\right|^3 &\leq \max_{i=1,\dots,T}\left|\ddot{g}\left(\underline{Y}_i(\beta_E),\mu_T\right)\right|\left(\frac{1}{T}\sum_{i=1}^{T}\ddot{g}\left(\underline{Y}_i(\beta_E),\mu_T\right)^2\right) \\ &= \ddot{g}^*(\beta_E,\mu_T)\,\ddot{\tilde{S}}(\beta_E,\mu_T) \\ &= o_p\left(T^{\frac{1}{2}}\right)O_p(1) \\ &= o_p\left(T^{\frac{1}{2}}\right),\end{aligned}$$

where we have used Eq. (A.2) and $\ddot{\tilde{S}}(\beta_E,\mu_T) = O_p(1)$. □

[12] For details see Section Appendix A.5 in Appendix Appendix A.

Appendix A.3. Proof of Theorem 2

Proof. For the proof, first we derive that $|\lambda| = O_p\left(T^{-\frac{1}{2}}\right)$. Knowing that, we show $\lambda = \ddot{\bar{S}}\left(\beta_E, \mu_T\right)^{-1} \bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) + o_p\left(T^{-\frac{1}{2}}\right)$. Plugging this expression for λ into the profile empirical log likelihood ratio statistic $-2\ddot{W}\left(\beta_E, \mu_T\right)$ and verifying that its elements are bounded in probability completes the proof. For the proof let us write shortly $\ddot{g}_i\left(\beta_E, \mu_T\right) = \ddot{g}\left(\underline{Y}_i\left(\beta_E\right), \mu_T\right)$.

The following shows that $|\lambda| = O_p\left(T^{-\frac{1}{2}}\right)$. By using $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ and $\hat{\lambda} = \lambda/\rho$, $\rho = |\lambda|$ in Eq. (21), it follows

$$\begin{aligned}
0 &= \frac{\hat{\lambda}}{T}\sum_{i=1}^{T}\frac{\ddot{g}_i\left(\beta_E, \mu_T\right)}{1+\lambda\ddot{g}_i\left(\beta_E, \mu_T\right)} \\
&= \frac{\hat{\lambda}}{T}\sum_{i=1}^{T}\ddot{g}_i\left(\beta_E, \mu_T\right) - \frac{\hat{\lambda}}{T}\sum_{i=1}^{T}\frac{\lambda\ddot{g}_i\left(\beta_E, \mu_T\right)^2}{1+\lambda\ddot{g}_i\left(\beta_E, \mu_T\right)} \\
&= \hat{\lambda}\bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) - \frac{\rho\hat{\lambda}^2}{T}\sum_{i=1}^{T}\frac{\ddot{g}_i\left(\beta_E, \mu_T\right)^2}{1+\rho\hat{\lambda}\ddot{g}_i\left(\beta_E, \mu_T\right)} \\
&\leq \hat{\lambda}\bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) - \frac{\rho}{1+\rho\ddot{g}^*\left(\beta_E, \mu_T\right)}\hat{\lambda}^2\ddot{\bar{S}}\left(\beta_E, \mu_T\right).
\end{aligned} \tag{A.5}$$

As $\ddot{\bar{S}}\left(\beta_E, \mu_T\right) \xrightarrow{p} \sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}$ (see Eq. (A.4)),

$$\hat{\lambda}^2\ddot{\bar{S}}\left(\beta_E, \mu_T\right) \geq (1-\varepsilon)\,\sigma^2_{\underline{Y}(\beta_E)}$$

in probability for some $1 > \varepsilon > 0$. Using Eq. (A.5) gives

$$\frac{\rho}{1+\rho\ddot{g}^*\left(\beta_E, \mu_T\right)} \leq \frac{\hat{\lambda}\bar{\ddot{g}}_T\left(\beta_E, \mu_T\right)}{(1-\varepsilon)\,\sigma^2_{\underline{Y}(\beta_E)}}. \tag{A.6}$$

With the order of $\bar{\ddot{g}}_T\left(\beta_E, \mu_T\right)$ and $\ddot{g}^*\left(\beta_E, \mu_T\right)$ of Lemma 4, it follows that $\frac{\hat{\lambda}\bar{\ddot{g}}_T(\beta_E,\mu_T)}{(1-\varepsilon)\sigma^2_{\underline{Y}(\beta_E)}} = O_p\left(T^{-\frac{1}{2}}\right)$ and with Eq. (A.6) we get

$$\rho = |\lambda| = O_p\left(T^{-\frac{1}{2}}\right). \tag{A.7}$$

Next we show $\lambda = \ddot{\bar{S}}\left(\beta_E, \mu_T\right)^{-1} \bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) + o_p\left(T^{-\frac{1}{2}}\right)$. Let $\kappa_i = \lambda\ddot{g}_i\left(\beta_E, \mu_T\right)$. Having established an order bound for $|\lambda|$ and with $\ddot{g}^*\left(\beta_E, \mu_T\right) = o_p\left(T^{\frac{1}{2}}\right)$, it is

$$\max_{i=1,\ldots,T}|\kappa_i| = O_p\left(T^{-\frac{1}{2}}\right) o_p\left(T^{\frac{1}{2}}\right) = o_p(1). \tag{A.8}$$

Using $\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}$ in Eq. (21) we get

$$
\begin{aligned}
0 &= \frac{1}{T}\sum_{i=1}^{T} \frac{\ddot{g}_i\left(\beta_E, \mu_T\right)}{1+\lambda \ddot{g}_i\left(\beta_E, \mu_T\right)} \\
&= \bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) - \ddot{S}\left(\beta_E, \mu_T\right)\lambda + \frac{1}{T}\sum_{i=1}^{T} \frac{\lambda^2 \ddot{g}_i\left(\beta_E, \mu_T\right)^3}{1+\lambda \ddot{g}_i\left(\beta_E, \mu_T\right)}. \qquad \text{(A.9)}
\end{aligned}
$$

The last term is bounded above by norm

$$
\frac{1}{T}\sum_{i=1}^{T} \frac{\lambda^2 \ddot{g}_i\left(\beta_E, \mu_T\right)^3}{1+\lambda \ddot{g}_i\left(\beta_E, \mu_T\right)} \leq |\lambda|^2 \frac{1}{T}\sum_{i=1}^{T} \left|\ddot{g}_i\left(\beta_E, \mu_T\right)\right|^3 \left|1+\lambda \ddot{g}_i\left(\beta_E, \mu_T\right)\right|^{-1} \qquad \text{(A.10)}
$$

With the given order of $|\lambda|$, $\frac{1}{T}\sum_{i=1}^{T} \left|\ddot{g}\left(\underline{Y}_i\left(\beta_E\right), \mu_T\right)\right|^3$ in Lemma 4 and Eq. (A.8), the order of Eq. (A.10) becomes

$$
\left(O_p\left(T^{-\frac{1}{2}}\right)\right)^2 o_p\left(T^{\frac{1}{2}}\right) O_p(1) = o_p\left(T^{-\frac{1}{2}}\right). \qquad \text{(A.11)}
$$

Using the latter in Eq. (A.9) gives

$$
\lambda = \ddot{S}\left(\beta_E, \mu_T\right)^{-1} \bar{\ddot{g}}_T\left(\beta_E, \mu_T\right) + o_p\left(T^{-\frac{1}{2}}\right). \qquad \text{(A.12)}
$$

Now we show that the empirical log likelihood ratio statistic $-2\ddot{W}\left(\beta_E, \mu_T\right)$ is bounded in probability. By Eq. (A.8) we use the expansion

$$
\log\left(1+\kappa_i\right) = \kappa_i - \frac{1}{2}\kappa_i^2 + \eta_i, \qquad \text{(A.13)}
$$

where for some finite $B > 0$,

$$
P\left(\left|\eta_i\right| \leq B\left|\kappa_i\right|^3, 1 \leq i \leq T\right) \to 1 \qquad \text{(A.14)}
$$

as $T \to \infty$. Substituting Eq. (A.13) in Eq. (20) we get

$$
\begin{aligned}
-2\ddot{W}\left(\beta_E, \mu_T\right) &= 2\sum_{i=1}^{T} \log\left(1+\kappa_i\right) \qquad \text{(A.15)} \\
&= 2\sum_{i=1}^{T} \kappa_i - \sum_{i=1}^{T} \kappa_i^2 + 2\sum_{i=1}^{T} \eta_i. \qquad \text{(A.16)}
\end{aligned}
$$

Lemma 4 and Eq. (A.14) give an order bound for the last term

$$
\begin{aligned}
2\left|\sum_{i=1}^{T} \eta_i\right| &\leq 2B|\lambda|^3 \sum_{i=1}^{T} \left|\ddot{g}_i\left(\beta_E, \mu_T\right)\right|^3 \\
&= 2BO_p\left(T^{-\frac{1}{2}}\right)^3 o_p\left(T^{\frac{3}{2}}\right) = o_p(1). \qquad \text{(A.17)}
\end{aligned}
$$

Let us rewrite Eq. (A.12) by

$$\lambda = \ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)+\beta \tag{A.18}$$

with $|\beta| = o_p\left(T^{-\frac{1}{2}}\right)$. Using Eq. (A.18), Eq. (A.17) and re-substituting $\kappa_i = \lambda\ddot{g}_i\left(\beta_E,\mu_T\right)$ in Eq. (A.16) gives

$$\begin{aligned}
-2\ddot{W}\left(\beta_E,\mu_T\right) &= 2\sum_{i=1}^{T}\lambda\ddot{g}_i\left(\beta_E,\mu_T\right)-\sum_{i=1}^{T}\left(\lambda\ddot{g}_i\left(\beta_E,\mu_T\right)\right)^2+o_p\left(1\right)\\
&= 2T\left[\left(\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)^2\right)+\beta\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)\right]\\
&\quad -T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)^2-2T\beta\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)\\
&\quad -T\beta^2\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)+o_p\left(1\right)\\
&= T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)^2-T\beta^2\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)+o_p\left(1\right)\\
&= T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)^2+o_p\left(1\right). \qquad (A.19)
\end{aligned}$$

The last equality holds because

$$T\beta^2\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)=O\left(T\right)\left(o_p\left(T^{-\frac{1}{2}}\right)\right)^2O_p\left(1\right)=o_p\left(1\right).$$

The expression in Eq. (A.19) can be written as

$$\begin{aligned}
&T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\bar{\ddot{g}}_T\left(\beta_E,\mu_T\right)^2\\
=\;&T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)-\mu_0\right)^2 \qquad (A.20)\\
&+2T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\left(\mu_0-\mu_T\right)\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)-\mu_0\right) \qquad (A.21)\\
&+T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\left(\mu_0-\mu_T\right)^2. \qquad (A.22)
\end{aligned}$$

With Eq. (25) the order of Eq. (A.22) is

$$\begin{aligned}
T\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)^{-1}\left(\mu_0-\mu_T\right)^2 &= O\left(T\right)O_p\left(1\right)O_p\left(T^{-\frac{1}{2}}\right)^2\\
&= O_p\left(1\right).
\end{aligned}$$

Note, that with the Slutsky theorem and the CLT we get

$$\frac{\sigma_{\underline{Y}(\beta_E)}}{\sqrt{\ddot{\tilde{S}}\left(\beta_E,\mu_T\right)}}\frac{\sqrt{T}\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)-\mu_0\right)}{\sigma_{\underline{Y}(\beta_E)}}\xrightarrow{d}1N\left(0,1\right) \tag{A.23}$$

since $\ddot{\tilde{S}}(\beta_E, \mu_T) \xrightarrow{p} \sigma^2_{\tilde{f}(\underline{Y}(\beta_E))}$ as $T \to \infty$ and it follows for Eq. (A.20)

$$\frac{\sigma^2_{\underline{Y}(\beta_E)}}{\ddot{\tilde{S}}(\beta_E, \mu_T)} \frac{T\left(\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0\right)^2}{\sigma^2_{\underline{Y}(\beta_E)}} \xrightarrow{d} 1\ \chi^2_1,$$

as $T \to \infty$, hence its order is

$$\begin{aligned} \frac{T\left(\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0\right)^2}{\ddot{\tilde{S}}(\beta_E, \mu_T)} &= O_p(1) + o_p(1) \\ &= O_p(1). \end{aligned}$$

The order of Eq. (A.21) is with Eq. (25) and the CLT

$$\begin{aligned} T2\ddot{\tilde{S}}(\beta_E, \mu_T)^{-1}(\mu_0 - \mu_T)\left(\frac{1}{T}\sum_{i=1}^{T} \tilde{f}(\underline{Y}_i(\beta_E)) - \mu_0\right) &= O(T)\, O_p(1)\, O_p\left(T^{-\frac{1}{2}}\right) O_p\left(T^{-\frac{1}{2}}\right) \\ &= O_p(1). \end{aligned}$$

Thus overall we get

$$\begin{aligned} -2\ddot{W}(\beta_E, \mu_T) &= O_p(1) + O_p(1) + O_p(1) + o_p(1) \\ &= O_p(1) \end{aligned}$$

□

Appendix A.4. Proof of Theorem 3

Proof. As f is measurable with Lemma 5 we have $\tilde{f}$ is measurable, hence for iid $\underline{Y}_i(\beta)$ with $i = 1, ..., T$ it follows $\left|\tilde{f}(\underline{Y}_i(\beta))\right|$ are iid. By assumption it follows $E\left[\tilde{f}(\underline{Y}(\beta))^2\right] < \infty$, thus all conditions of Lemma 3 in Appendix Appendix A are satisfied such that

$$\max_{i=1,\ldots,T}\left|\tilde{f}(\underline{Y}_i(\beta))\right| = o_p\left(T^{\frac{1}{2}}\right).$$

With Lemma 2 we have $|\mu_T| = O_p(1)$. Therefore

$$\begin{aligned} \ddot{g}^*(\beta, \mu_T) &= \max_{i=1,\ldots,T} |g_i(\beta, \mu_T)| \\ &= \max_{i=1,\ldots,T}\left|\tilde{f}(\underline{Y}_i(\beta)) - \mu_T\right| \\ &\leq \max_{i=1,\ldots,T}\left|\tilde{f}(\underline{Y}_i(\beta))\right| + |\mu_T| \\ &= o_p\left(T^{\frac{1}{2}}\right) + O_p(1) \\ &= o_p\left(T^{\frac{1}{2}}\right). \end{aligned}$$

For $\bar{\ddot{g}}_T(\beta, \mu_T)$, it follows with the Slutsky Theorem

$$\bar{\ddot{g}}_T(\beta, \mu_T) \xrightarrow{p} E\left[\tilde{f}\left(\underline{Y}(\beta)\right)\right] - \mu_0 \quad \text{(A.24)}$$

as $T \to \infty$, hence $\bar{\ddot{g}}_T(\beta, \mu_T) = O_p(1)$. Note, by assumption we have $\left|E\left[\tilde{f}\left(\underline{Y}(\beta)\right)\right] - \mu_0\right| > 0$. Now, using the CMT for Eq. (A.24) with $f(x) = x^2$, it follows

$$\bar{\ddot{g}}_T(\beta, \mu_T)^2 \xrightarrow{p} A \quad \text{(A.25)}$$

as $T \to \infty$ with $A > 0$, since $A = \left(E\left[\tilde{f}\left(\underline{Y}(\beta)\right)\right] - \mu_0\right)^2$. Moreover, using the Slutsky Theorem repeatedly[13] we get

$$\frac{1}{T}\sum_{i=1}^{T} g_i(\beta, \mu_T)^2 \xrightarrow{p} E\left[\tilde{f}\left(\underline{Y}(\beta)\right)^2\right] - \mu_0 E\left[\tilde{f}\left(\underline{Y}(\beta)\right)\right] + \mu_0^2 \quad \text{(A.26)}$$

as $T \to \infty$, hence $\frac{1}{T}\sum_{i=1}^{T} g_i(\beta, \mu_T)^2 = O_p(1)$. Now let $\dot{\lambda} = T^{-\frac{2}{3}}\bar{\ddot{g}}_T(\beta, \mu_T) M$, where $M > 0$ is a constant. The order of $\left|\dot{\lambda}\right|$ is[14] $T^{-\frac{2}{3}}O_p(1)$ and therefore

$$\begin{aligned}
\max_{i=1,\ldots,T}\left|\dot{\lambda}\ddot{g}_i(\beta, \mu_T)\right| &= \left|\dot{\lambda}\right| \ddot{g}^*(\beta, \mu_T) && \text{(A.27)}\\
&= T^{-\frac{2}{3}}O_p(1)\, o_p\left(T^{\frac{1}{2}}\right) && \text{(A.28)}\\
&= o_p(1). && \text{(A.29)}
\end{aligned}$$

With Eq. (A.29) it is $1 + \dot{\lambda} g_i(\beta, \mu_T) > 0$ with probability going to 1. Hence using the Taylor expansion

$$\log(1+x) = x - \frac{x^2}{2(1+\xi)} \quad \text{(A.30)}$$

for some ξ between 0 and x and the duality of the maximization problem, it is

$$\begin{aligned}
\ddot{W}(\beta, \mu_T) &= -\sup_{\lambda}\left\{\sum_{i=1}^{T}\log\left(1 + \lambda g_i(\beta, \mu_T)\right)\right\}\\
&\leq -\sum_{i=1}^{T}\log\left(1 + \dot{\lambda} g_i(\beta, \mu_T)\right)\\
&= -\left[\sum_{i=1}^{T}\dot{\lambda} g_i(\beta, \mu_T) - \frac{1}{2}\sum_{i=1}^{T}\frac{\left(\dot{\lambda} g_i(\beta, \mu_T)\right)^2}{(1+\xi_i)}\right]. && \text{(A.31)}
\end{aligned}$$

[13] Compare to Section Appendix A.5 in Appendix Appendix A.
[14] For details see Section Appendix A.6 in Appendix Appendix A.

Note, from Eq. (A.29) all ξ_i are within $o_p(1)$ neighborhood of 0 uniformly. Therefore the second term Eq. (A.31) is no larger than

$$\begin{aligned}\sum_{i=1}^{T}\left(\dot{\lambda}g_i(\beta,\mu_T)\right)^2 &= T\dot{\lambda}^2\frac{1}{T}\sum_{i=1}^{T}g_i(\beta,\mu_T)^2\\ &= O(T)\,O_p\left(T^{-\frac{4}{3}}\right)O_p(1)=o_p(1),\end{aligned}$$

where we have used Eq. (A.26). The first term in Eq. (A.31) is with Eq. (A.25)

$$\begin{aligned}\sum_{i=1}^{T}\dot{\lambda}g_i(\beta,\mu_T) &= \dot{\lambda}T\bar{\bar{g}}_T(\beta,\mu_T)\\ &= T^{-\frac{2}{3}}MT\bar{\bar{g}}_T(\beta,\mu_T)^2\\ &= T^{\frac{1}{3}}M(A+o_p(1)).\end{aligned}$$

Therefore Eq. (A.31) gives

$$\ddot{W}(\beta,\mu_T)\leq -T^{\frac{1}{3}}MA+o_p(1). \tag{A.32}$$

Since M can be arbitrarily large, we have for $\beta\neq\beta_E$ that $-2T^{-1/3}\ddot{W}(\beta,\mu_T)\to\infty$ in probability. □

Appendix A.5. Annotation of Eq. (A.4)

$$\ddot{\tilde{S}}(\beta_E,\mu_T) = \frac{1}{T}\sum_{i=1}^{T}\left(\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\right)^2 - 2\mu_T\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\right)+\mu_T^2. \tag{A.33}$$

From Lemma 2 we know $\mu_T\xrightarrow{p}\mu_0$ as $T\to\infty$. As $f(x)=x^2$ is a continuous function with the CMT it follows $\mu_T^2\xrightarrow{p}\mu_0^2$ as $T\to\infty$. Note, μ_0 and $E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right]$ are constants. With the LLN we have $\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\xrightarrow{p}E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right]$ as $T\to\infty$. Therefore with the Slutsky Theorem[15] we get

$$\mu_T\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i(\beta_E)\right)\right)\xrightarrow{p}\mu_0E\left[\tilde{f}\left(\underline{Y}(\beta_E)\right)\right]$$

[15] See for example Van der Vaart (1998) on p. 11.

as $T \to \infty$. Using the Slutsky Theorem again

$$2\mu_T\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)\right)+\mu_T^2 \xrightarrow{p} \mu_0 E\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)\right]+\mu_0^2$$

as $T \to \infty$, where $\mu_0 E\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)\right]+\mu_0^2$ is a constant again. Applying the Slutsky Theorem a last time results in

$$\begin{gathered}\frac{1}{T}\sum_{i=1}^{T}\left(\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)\right)^2-2\mu_T\left(\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)\right)+\mu_T^2\\ \xrightarrow{p}\\ E\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)^2\right]+\mu_0 E\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)\right]+\mu_0^2\end{gathered}$$

as $T \to \infty$, where the first term on the right side is due to the LLN, as $\frac{1}{T}\sum_{i=1}^{T}\tilde{f}\left(\underline{Y}_i\left(\beta_E\right)\right)^2 \xrightarrow{p} E\left[\tilde{f}\left(\underline{Y}\left(\beta_E\right)\right)^2\right]$ for $T \to \infty$.

Appendix A.6. Annotation of Eq. (A.27)

From Eq. (A.24) we have

$$\bar{\ddot{g}}_T\left(\beta,\mu_T\right) \xrightarrow{p} B$$

as $T \to \infty$, where $B=\left|E\left[\tilde{f}\left(\underline{Y}\left(\beta\right)\right)\right]-\mu_0\right|>0$, a constant, and therefore

$$\left|\bar{\ddot{g}}_T\left(\beta,\mu_T\right)-B\right|=o_p\left(1\right).$$

With the latter we get

$$\left|\bar{\ddot{g}}_T\left(\beta,\mu_T\right)\right| \leq \left|B\right|+o_p\left(1\right)=O_p\left(1\right).$$

For $\dot{\lambda}=T^{-\frac{2}{3}}\bar{\ddot{g}}_T\left(\beta,\mu_T\right)M$ with $M>0$ a constant, it follows then

$$\begin{aligned}\left|\dot{\lambda}\right| &= T^{-\frac{2}{3}}M\left|\bar{\ddot{g}}_T\left(\beta,\mu_T\right)\right|\\ &= T^{-\frac{2}{3}}O\left(1\right)O_p\left(1\right)\\ &= T^{-\frac{2}{3}}O_p\left(1\right).\end{aligned}$$

Appendix A.7. Measurability of $\tilde{f}$

Theorem 5. *Let $(\Omega,\mathcal{F})$ be a measure space and $k=(k_1,...,k_l):\Omega\to\mathbb{R}^l$. Then k is $\mathcal{F}-\mathcal{B}^l$ measurable if and only if $k_s:\Omega\to\mathbb{R}$ for $s=1,...,l$ are $\mathcal{F}-\mathcal{B}$ measurable.*

Proof. See Meintrup and Schäffler (2005) on p. 18. □

Lemma 5. *Let $\underline{Y}$ be a real-valued K-variate random vector and suppose $f_1, ..., f_l$ are measurable functions mapping from $\mathbb{R}$ into $\mathbb{R}$. Then $\tilde{f}$ is a $\mathcal{B}^K - \mathcal{B}^l$ measurable function where*

$$\tilde{f} = \begin{pmatrix} \tilde{f}_1 \\ \vdots \\ \tilde{f}_l \end{pmatrix} = \begin{pmatrix} \frac{1}{K}\sum_{t=1}^{K} f_1(y_t) \\ \vdots \\ \frac{1}{K}\sum_{t=1}^{K} f_l(y_t) \end{pmatrix}.$$

Proof. First note that for $s = 1, ..., l$ we can write

$$\begin{aligned} \tilde{f}_s &= \frac{1}{K}\sum_{t=1}^{K} f_s(y_t) \\ &= \frac{1}{K}\sum_{t=1}^{K} f_s \circ p_t(\underline{Y}) \\ &= \frac{1}{K}\sum_{t=1}^{K} f_{s,t}(\underline{Y}), \end{aligned}$$

where $p_t : \mathbb{R}^K \to \mathbb{R}$ denotes the projection on the t-th component. Knowing[16] that p_t is a $\mathcal{B}^K - \mathcal{B}$ measurable function and f_s is a $\mathcal{B} - \mathcal{B}$ measurable, it follows that $f_{s,t}$ is also $\mathcal{B}^K - \mathcal{B}$ measurable function since $f_{s,t} = f_s \circ p_t$. By definition $\tilde{f}_s : \mathbb{R}^K \to \mathbb{R}$ is a sum of $\mathcal{B}^K - \mathcal{B}$ measurable functions, hence $\tilde{f}_s$ is itself a $\mathcal{B}^K - \mathcal{B}$ measurable function for all $s = 1, ..., l$. Finally, with Theorem 5 it follows that $\tilde{f} = \left(\tilde{f}_1, ..., \tilde{f}_l\right)' : \mathbb{R}^K \to \mathbb{R}^l$ is a $\mathcal{B}^K - \mathcal{B}^l$ measurable function. □

Appendix A.8. Annotation of Eq. (26)

Let us consider

$$V = \sup\left\{\sum_{i=1}^{T} \log(Tw_i) \,|\, \sum_{i=1}^{T} w_i = 1, w_i \geq 0\right\}. \quad \text{(A.34)}$$

The Lagrange of V is

$$\mathcal{L} = \sum_{i=1}^{T} \log(w_i) + \tau\left(1 - \sum_{i=1}^{T} w_i\right) + T\log(T).$$

[16] The fact that p_t is a $\mathcal{B}^K - \mathcal{B}$ measurable function is demonstrated in Meintrup and Schäffler (2005) on p. 18. On the same page it is also proven the composition and the sum of measurable functions is measurable again.

With the first order condition for $\mathcal{L}$

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{1}{w_i} - \tau = 0 \tag{A.35}$$

we get

$$\sum_{i=1}^{T} w_i \frac{\partial \mathcal{L}}{\partial w_i} = T - \tau = 0$$

that is $\tau = T$. Using the latter in Eq. (A.35) the optimal weights w_i^*, $i = 1, ..., T$ are given by

$$w_i^* = \frac{1}{T}. \tag{A.36}$$

Hence $V = \sum_{i=1}^{T} \log (T w_i^*) = 0$. Now by definition

$$\ddot{W}(\beta, \mu_T) = \sup \left\{ \sum_{i=1}^{T} \log (T w_i) \,|\, \sum_{i=1}^{T} w_i \tilde{f}\left(\underline{Y}_i(\beta)\right) = \mu_T, \sum_{i=1}^{T} w_i = 1, w_i \geq 0 \right\},$$

that is $\ddot{W}(\beta, \mu_T)$ is a constrained version of V (see Eq. (A.34)) and therefore

$$\ddot{W}(\beta, \mu_T) \leq V = 0$$

or

$$-2\ddot{W}(\beta, \mu_T) \geq 0.$$

Appendix A.9. SMM Consistency Requirements

Lee and Ingram (1991) introduces the simulated moment method for stationary and ergodic time series and Duffie and Singleton (1993) deals with geometrically ergodic processes, that are not initially drawn from their stationary and ergodic distribution but will converge to it. For stationary and ergodic time series Lee and Ingram (1991) present the following requirements for the consistency of the SMM estimator.

Assumption 2.

- $\{x_t\} \in \mathbb{R}^m$ and $\{y_t(\beta)\} \in \mathbb{R}^m$ are (strictly) stationary and ergodic but independent of each other for all $\beta \in B \subseteq \mathbb{R}^p$.
- B is compact and $h : \mathbb{R}^m \times B \to \mathbb{R}^l$, for every $\beta \in B$ the simulated moment $E\left[|h(y_1(\beta))|\right] < \infty$.
- Let $l \geq p$ and there exists a unique β_0 such that $\{y_t(\beta_0)\}$ is drawn from the distribution as $\{x_t\}$.

- The function $h_1(\beta) = h(y_1(\beta))$ is continuous in the mean for all $\beta \in B$ i.e.

$$\lim_{\delta \to 0} E\left[\sup_{\grave{\beta} \in K(\beta,\delta)} \left\| h_1\left(\grave{\beta}\right) - h_1(\beta) \right\| \right] = 0.$$

- The weight matrix $\hat{W}$ satisfies $\hat{W} \xrightarrow{p} W$, where W is a positive semi-definite matrix.

For geometrically ergodic processes Duffie and Singleton (1993) give the following requirements for the consistency of the SMM estimator.

Assumption 3.

- For each $\beta \in B \subseteq \mathbb{R}^p$, $\left\{ \left\| f_t^{\beta} \right\|_{2+\delta} : t = 1, 2, ... \right\}$ is bounded for some $\delta > 0$. The family $\left\{ f_t^{\beta} \right\}$ is Lipschitz, uniformly in probability and $\beta \to E\left[f_{\infty}^{\beta}\right]$ is continuous, where $f_t^{\beta} = f\left(y_t^{\beta}\right)$ is a measurable function mapping into $\mathbb{R}^l$ with $l \geq p$.
- For all $\beta \in B$, the process $\left\{ y_t^{\beta} \right\}$ is geometrically ergodic.
- Σ_0 is nonsingular and $\hat{W} \to W_0 = \Sigma_0^{-1}$ almost surely, where (for any t)

$$\Sigma_0 = \sum_{j=-\infty}^{\infty} E\left[\left(f_t^* - E\left[(f_t^*)\right]\right) \left(f_{t-j}^* - E\left[\left(f_{t-j}^*\right)\right]\right)' \right]$$

- (Uniqueness of minimizer) $C(\beta_0) < C(\beta)$ with $\beta \in B$, $\beta \neq \beta_0$ where

$$C(\beta) = G_{\infty}(\beta)' W_0 G_{\infty}(\beta)$$

  with $G_{\infty}(\beta) = \left(E\left[f_{\infty}^*\right] - E\left[f_{\infty}^{\beta}\right)\right]$

The following lemma demonstrates that both Lee and Ingram (1991) and Duffie and Singleton (1993) require the moments of the simulation process to be continuous with respect to the underlying parameter β.

Lemma 6. *Let $h_1(\beta)$ be first moment continuous at β, then $E[h_1(\beta)]$ is continuous in β.*

Proof. A function k is continuous in at some point β if

$$\lim_{\delta\to 0,\grave{\beta}\in K(\beta,\delta)} \left\| k\left(\grave{\beta}\right) - k\left(\beta\right) \right\| = 0,$$

where $K(\beta,\delta)$ an the open ball around β with radius δ. Now consider

$$\begin{aligned}
\lim_{\delta\to 0,\grave{\beta}\in K(\beta,\delta)} \left\| E\left[h_1\left(\grave{\beta}\right)\right] - E\left[h_1\left(\beta\right)\right] \right\| &= \lim_{\delta\to 0,\grave{\beta}\in K(\beta,\delta)} \left\| E\left[h_1\left(\grave{\beta}\right) - h_1\left(\beta\right)\right] \right\| \\
&\leq \lim_{\delta\to 0,\grave{\beta}\in K(\beta,\delta)} E\left[\left\| h_1\left(\grave{\beta}\right) - h_1\left(\beta\right) \right\|\right] \\
&\leq \lim_{\delta\to 0} E\left[\sup_{\grave{\beta}\in K(\beta,\delta)} \left\| h_1\left(\grave{\beta}\right) - h_1\left(\beta\right) \right\|\right] \\
&= 0.
\end{aligned}$$

The last line holds due to first moment continuity. Setting $k(\beta) = E[h_1(\beta)]$ it follows with the definition that $E[h_1(\beta)]$ is continuous in β. □