BAYESIAN ESTIMATION OF TIME SERIES WITH EMPIRICAL LIKELIHOOD

Juraj Panek

February 12, 2017

Czech Technical University in Prague
Faculty of Nuclear Sciences and Physical Engineering
Department of Mathematics

Our goal is to study the use of empirical likelihood for time series in Bayesian setting. We will cover:

1. Introduction to the concept of empirical likelihood
2. Possible modifications of empirical likelihood to treat the dependence in data (time series)
3. Proposition of Bayesian samplers from the posterior distribution with the empirical likelihood
    - 3.1 Direct weighting of the prior via empirical likelihood.
    - 3.2 Importance resampling with empirical likelihood weights.
    - 3.3 Markov chain Monte Carlo via empirical likelihood.
4. Examples on selected models of time series

EMPIRICAL LIKELIHOOD

Empirical likelihood is a nonparametric method of statistical inference without the assumptions about the distribution family of observed data.
Suppose $X_1, \ldots, X_n$ is a iid sample from a distribution F, given the data a likelihood function would be

$$L(F) = \prod_{i=1}^{n} [F(X_i) - F(X_i-)] = \prod_{i=1}^{n} p_i. \tag{1}$$

The concept of EL is based on the ratio of the nonparametric likelihood L

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

where F is the CDF and F_n is the empirical CDF.

The empirical likelihood is defined via the **profile empirical likelihood ratio** defined as

$$\mathcal{R}(\theta) = \sup\{R(F) | T(F) = \theta, F \in \mathcal{F}\}, \tag{3}$$

where T is a function of distribution F.
Equation 3 can be rewritten as

$$\mathcal{R}(\theta) = \max\left\{\prod_{i=1}^{n} nw_i \Big| \sum_{i=1}^{n} w_i m(X_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\right\}, \tag{4}$$

where $m(X_i, \theta)$ is an **estimating equation**.
Under **mild** and **iid** conditions $-2\log\mathcal{R}(\theta)$ **converges in distribution** to $\chi^2_{(p)}$ as $n \to \infty$.

The framework of empirical likelihood was designed for iid data. In **time series** context the **limit distribution theorem fails**.

EL modifications are required to accommodate the dependence in the data. Possible options are:

- **Based on time series model.** Assumes a structural model for time series (e.g. AR, ARCH, GARCH)
- **Block-wise empirical likelihood.** No underlying model assumptions are required, uses blocks of data as a remedy to dependence.

Method known from parametric statistical inference. To tackle dependence a parametric model for time series is considered. Method expresses the observations in terms of variables assumed to be independent.
For auto-regressive model AR(p)

$$X_t = \psi_1 X_{t-1} + \cdots + \psi_p X_{t-p} + \epsilon_t, \tag{5}$$

for iid ϵ_t with $\mathrm{E}[\epsilon_t] < \infty$ the theoretical distribution $T(F) = \theta$ can be constrained as

$$g(X_t, \ldots,_{t-p}; \psi) = \left\{X_t - \sum_{i=1}^{p} \psi_i X_{t-i}\right\} \cdot (X_t, \ldots, X_{t-p})' \in \mathbb{R}^p, \quad t = p, \ldots, n. \tag{6}$$

It can be proved that under certain conditions the χ^2 limit distribution property holds.

BLOCK-WISE EMPIRICAL LIKELIHOOD

For weakly dependent time series the individual data are replaced by blocks of consecutive observations.

For given block of data of length l the distribution constraint is expressed via block average $M_i(\theta) = l^{-1} \sum_{j=i}^{i+l-1} m(X_j, \theta)$.

Profiling the empirical likelihood ratio yields

$$\mathcal{R}_b(\theta) = \max \left\{ \prod_{i=1}^{n} n w_i \Big| \sum_{i=1}^{n} w_i M(X_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}. \quad (7)$$

Analogously as with the iid case, it can be shown that

$$-2 l^{-1} \log \mathcal{R}(\theta) \to \chi^2_{(p)}. \quad (8)$$

Addition of the term l^{-1} is required to account for the overlap in the data after blocking.

BAYESIAN COMPUTATION VIA EL

EMPIRICAL LIKELIHOOD WEIGHTED SAMPLER

Original proposition was employing the empirical likelihood as importance sampling weights, referred to as the $\mathrm{BC}_{\mathrm{el}}$.

1. Sample $\theta' \sim \pi(\theta)$.
2. Calculate the empirical likelihood ratio $\mathcal{R}(\theta)$.
3. Save θ' with weight $\xi = \mathcal{R}(\theta)$.

The parameter estimates can be obtained directly by

$$\mathrm{E}_{\hat{\pi}}[\theta] = \sum_{i=1}^{N} \xi_i^* \theta_i, \tag{9}$$

where ξ_i^* are self normalized weights.
$\mathrm{BC}_{\mathrm{el}}$ suffers from the increasing variance of weights as any importance sampling algorithm.

EMPIRICAL LIKELIHOOD IMPORTANCE SAMPLING RESAMPLING

We proposed to use the importance resampling technique combined with $\mathrm{BC_{el}}$.

1. Obtain samples $(\theta_1, \dots, \theta_N)$ with corresponding EL weights $(\xi_1, \dots \xi_N)$ by $\mathrm{BC_{el}}$.
2. Sample N values from $(\theta_1, \dots, \theta_N)$ according to their normalised weights ξ_i^*.
3. Add a random walk with small variance to newly populated samples.
4. Recalculate the EL weights.

The resampling from Step 2 was implemented according to multinomial distribution with parameters N and $(\xi_1^*, \dots, \xi_N^*)$. Importance resampling significantly improved the $\mathrm{BC_{el}}$ algorithm in the variance of weights and parameter estimates.

We proposed to sample direclty from the empirical likelihood approximation of posterior distribution

$$\pi_{\mathrm{el}}(\theta|\mathrm{x}) = \frac{\mathcal{R}(\theta)\pi(\theta)}{\int_{\Theta} \mathcal{R}(\theta)\pi(\theta)\mathrm{d}\theta}. \tag{10}$$

Sampling directly from the $\pi_{\mathrm{el}}(\theta|\mathrm{x})$ can be still classified as likelihood-free.
The normalizing constant $\int_{\Theta} \mathcal{R}(\theta)\pi(\theta)\mathrm{d}\theta$ is computationally difficult to obtain.
Therefore we propose to combine the Markov chain Monte Carlo methods with empirical likelihood.

The proposed MCMC via EL algorithm is of the following form:

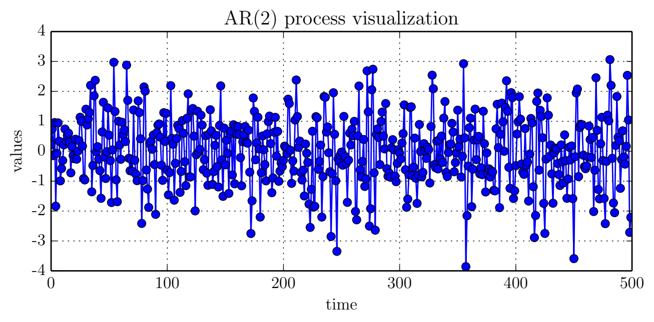
1. If currently at θ propose a move to θ' according to the transition kernel $q(\theta \to \theta')$.
2. Calculate
$$h = \min\left(1, \frac{\mathcal{R}(\theta')\pi(\theta')q(\theta' \to \theta)}{\mathcal{R}(\theta)\pi(\theta)q(\theta \to \theta')}\right)$$
3. Move to θ' with probability h, otherwise remain at θ.

It can be shown that the stationary distribution of the resulting Markov chain $(\theta_1, \theta_2, \dots)$ is indeed from the posterior $\pi_{el}(\theta|x)$.

EXAMPLES

To illustrate the proposed methods of inference, consider a autoregressive model of the second order.

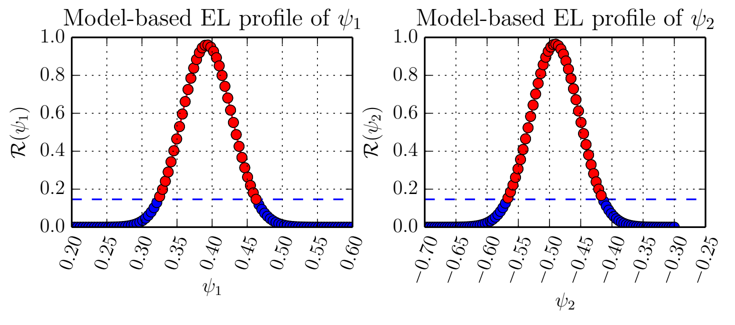
$$X_t = \psi_1 X_{t-1} + \psi_2 Xt - 2 + \epsilon_t. \tag{11}$$



parameter	estimate	standard error	95% Confidence Interval
ψ_1	0.390	0.039	$[0.313, 0.467]$
ψ_2	-0.489	0.039	$[-0.566, -0.412]$

Table: Parametric MLE estimates.

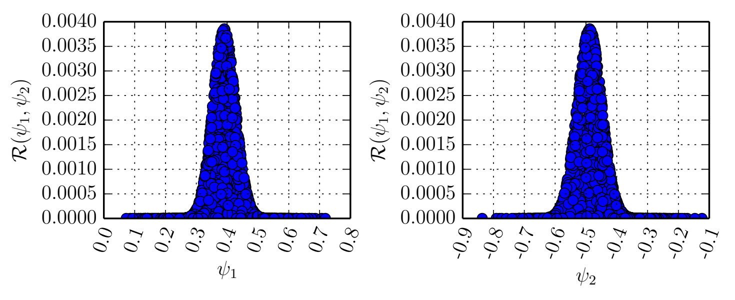
Non Bayesian model-based EL inference.



parameter	mean	standard error	95% Confidence Interval
ψ_1	0.393	0.036	$[0.325, 0.463]$
ψ_2	-0.491	0.040	$[-0.567, -0.417]$

Table: Estimates via the block-based empirical likelihood.

Consider a prior $\pi(\psi_1) = \mathcal{N}(0.394, 0.5^2), \pi(\psi_2) = \mathcal{N}(-0.491, 0.5^2)$.

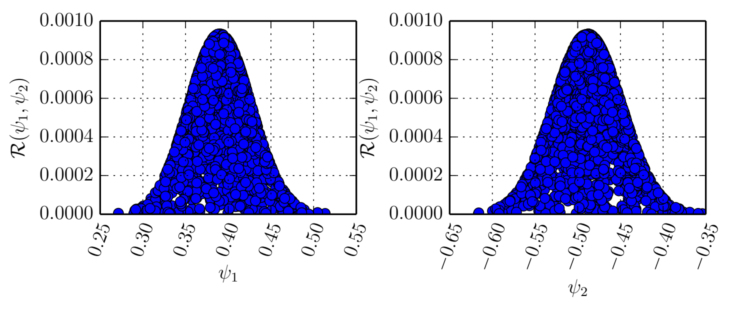


parameter	mean	standard error	ESS
ψ_1	0.389	0.002	479
ψ_2	-0.485	0.002	479

Table: Estimates of the Bayesian weighted sampler via model-based EL.

WEIGHTED EMPIRICAL LIKELIHOOD IMPORTANCE RESAMPLING

Employment of a single run of the multinomial resampling.

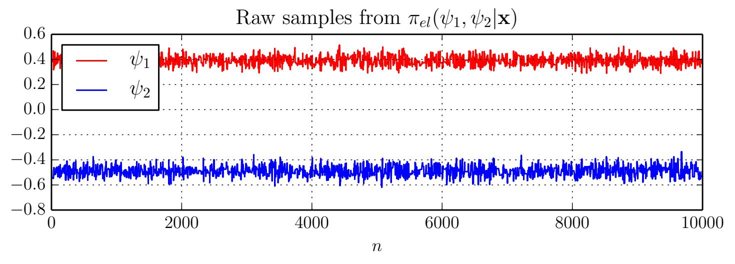


parameter	mean	standard error	ESS
ψ_1	0.391	0.001	1552
ψ_2	-0.487	0.001	1552

Table: Parameter estimates after a single run of multinomial resampling.

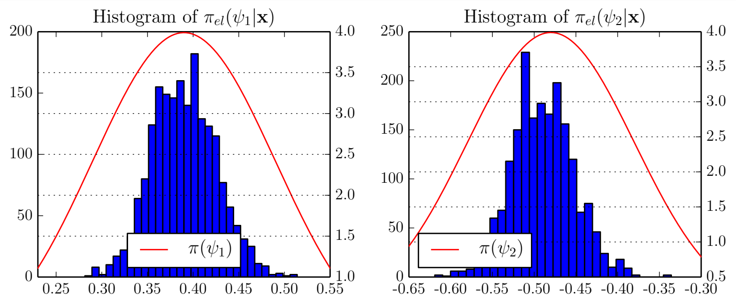
MCMC WITH MODEL-BASED EL

- Consider sampling directly from the posterior $\pi_{\text{el}}(\psi_1, \psi_2 | \mathbf{x})$ using the modified MCMC algorithm.
- The Markov chain is required to be well mixing, i.e. does not get stuck rejecting successive samples.
-



Resulting MC is usually correlated and does not reach its stationarity right away. Use PACF for detection. We apply burn-in period and thinning.

- Taking every 5th sample suppresses autocorrelation.
- Burn-in period set to 400 samples.

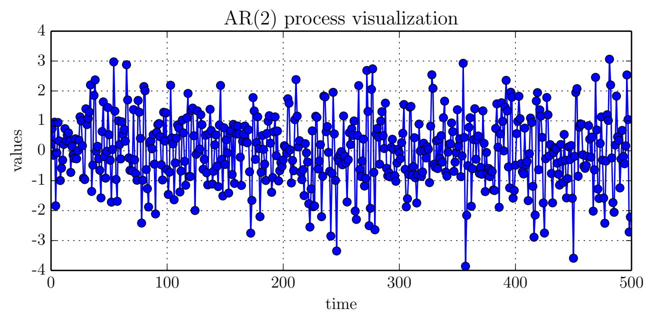


$E[\psi_1]$	$Var[\psi_1]$	$E[\psi_2]$	$Var[\psi_2]$
0.391	0.035	-0.490	0.036

Table: Estimates obtained by the modified MCMC algorithm.

Consider an inference on the mean of the stationary ARMA(2,2) process

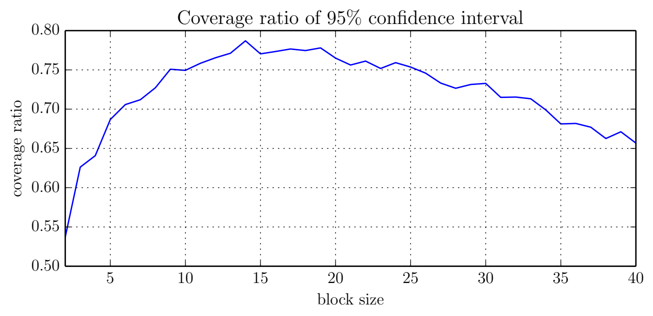
$$X_t = \psi_1 X_{t-1} + \psi_2 X_{t-2} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}, \qquad (12)$$



Parameter	ψ_1	ψ_2	θ_1	θ_2	T	σ^2	$\bar{X}_n$	s^2
Value	0.3	0.5	-0.5	0.1	100	1.0	0.632	1.272

BLOCK-WISE EMPIRICAL LIKELIHOOD

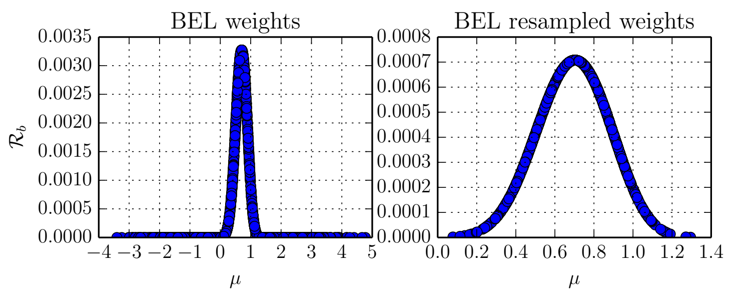
- The block-wise EL requires an additional parameter l defining the size of a block.
- Theoretically little is known about the optimal choice.
- The choice of l is usually explored by the coverage ratio for a given problem. Block length 15 was chosen.

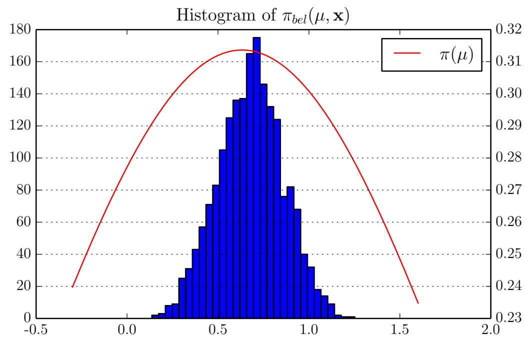


Once the block size is set, the inference follows a similar pattern to the previous example.

For Bayesian inference we place a prior on the process mean

$$\pi(\mu) = \mathcal{N}(\bar{X}_n, s^2) = \mathcal{N}(0.632, 1.272). \tag{13}$$





	method	mean	std. error	ESS
μ	WBEL	0.696	0.034	426
μ	ReWBEL	0.704	0.019	1743
μ	MCMC BEL	0.680	0.179	-

Table: Bayesian BEL estimates of μ.

CONCLUSIONS

- Model based treatment of time series via empirical likelihood is limited to the construction of valid estimating equations. Applications are limited to AR models and its modifications.
- Block-wise empirical likelihood does not require model specification however little is still known about the optimal block length selection.
- Block structure can be only applied to limited class of models that satisfy the requirement of weak dependence.

Thank you for your attention