Blockwise empirical likelihood and efficiency for Markov chains

Ursula U. Müller
Department of Statistics
Texas A&M University
College Station, TX 77843-3143, USA
uschi@stat.tamu.edu

Anton Schick
Department of Mathematics and Statistics
Binghamton University
Binghamton, NY 13902-6000, USA
anton@math.binghamton.edu

Wolfgang Wefelmeyer
Mathematical Institute
University of Cologne
50931 Cologne, Germany

Abstract

Suppose we observe an ergodic Markov chain on an arbitrary state space. The usual nonparametric estimator of a linear functional of the stationary distribution is the empirical estimator. If the stationary distribution obeys finitely many known linear constraints, we can improve the empirical estimator by empirical likelihood weights. Since the observations are dependent, an optimal choice of weights is determined by weighting averages over disjoint blocks of observations with slowly increasing length. We show that the improved empirical estimator is efficient. We also introduce two additively corrected empirical estimators that are asymptotically equivalent to the weighted empirical estimator, hence also efficient.

Keywords: Martingale approximation, perturbation expansion, local asymptotic normality, asymptotically linear estimator, improved empirical estimator, asymptotically efficient estimator.

Mathematics Subject Classification: 60J05, 62M05, 62G20.

1 Introduction

Let $X_0, \dots, X_n$ be observations of an ergodic stationary Markov chain on an arbitrary state space S with σ-algebra $\mathcal{S}$. The distribution of the chain is specified by the transition distribution $Q(x, dx)$ and the stationary distribution $\pi(dx)$. They determine the joint stationary distribution $P(dx, dy) = \pi(dx)Q(x, dy)$ of two successive observations (X_0, X_1).

If the state space S is the real line and if Q has a density $q(x, y)$, then π has a density $p(x)$ and both q and p are determined by the joint density $p(x)q(x, y)$ of two consecutive observations. That density is then determined by expectations

$$Pf = E[f(X_0, X_1)] = \int \pi(dx)Q(x, dy)f(x, y)$$

for a sufficiently large class of (bounded) functions f on S^2.

If the state space is finite, $S = \{1, 2, \dots, m\}$, the transition distribution is an $m \times m$ matrix Q and π is an m-dimensional vector π. In this case the joint distribution is immediately identified by indicators $f(x, y) = \mathbf{1}_{\{(s,t)\}}(x, y)$ with fixed $s, t \in S$. Further probabilities of interest are distribution functions and the distribution of the maximum or minimum, i.e. with f equal to $1(x \le s, y \le t)$, $1\{y \le t\}$, $1\{\min(x, y) \le t\}$ or $1\{\max(x, y) \le t\}$ for given $s, t \in S$. Other functions are nevertheless of interest, e.g. means, moments and covariances.

We are interested in estimating Pf. The simplest estimator is the empirical estimator

$$\mathbb{P}f = \frac{1}{n} \sum_{j=1}^{n} f(X_{j-1}, X_j).$$

It is efficient in the nonparametric model. We do, however, assume more structure, namely that the stationary distribution P fulfills a linear constraint $Ph = 0$ for some vector-valued function h on $S^2 \times [0, \infty)$, e.g. that certain moments or quantiles are *known*. Another example is $h(X_0, X_1) = X_1 - m(X_0)$ with m a known function. Then $E[h(X_0, X_1)] = E[E(X_1|X_0) - m(X_0)] = 0$ is satisfied if $m(X_0) = E(X_1|X_0)$, i.e. the regression of X_1 on X_0 is a known function.

We will show (in Section 3) that we can use this constraint in different ways to improve the empirical estimator $\mathbb{P}f$. Our main focus will be on the blockwise empirical likelihood method, which was introduced by Kitamura (1997) for weakly dependent (discrete-time) processes. It has been applied in particular to various time series models; see e.g. Wu and Cao (2011), Nordman et al. (2013), Kim et al. (2013), Zhang and Shao (2016), Jiang and Wang (2018). We will utilize the version by Schick (2024) for Markov chains.

Here Kitamura's approach leads to the *blockwise weighted* empirical estimator defined in (3.2). A second, simpler, estimator is the *blockwise additively corrected* empirical estimator defined in (3.4). We also look at a version without blocks, the *additively corrected* empirical estimator (3.1), which was first introduced for Markov chains in Müller, Schick and Wefelmeyer (2001). All three estimators are asymptotically equivalent and asymptotically efficient in the sense of a nonparametric version of the convolution theorem of Hájek and Le Cam.

This paper is organized as follows. In Section 2 we characterize efficient estimators. We do so by deriving a version of the convolution theorem for our setting. It leads to a characterization of efficient estimators by means of their influence functions, which must equal the canonical gradient for $Pf = E[(f(X_0, Y_0)]$. The characterization is provided in Theorem 2. In Section 3 we present the results for the three estimators, beginning with the additively corrected estimator in Theorem 3, whose construction is suggested by the efficiency considerations from Section 2. In Theorem 4 we treat the blockwise weighted estimator. The proof of this theorem rests on three lemmas. In one of these lemmas we derive a stochastic expansion, which suggests the third blockwise additively corrected estimator that can be handled using existing results. The efficiency statement for this last estimator is formulated in Theorem 5. A small simulation study for a simple setting with a one-dimensional constraint concludes Section 3. The proofs of the lemmas and of Theorem 4 are in Section 4.

2 A characterization of efficient estimators

We assume that the Markov chain is stationary with stationary distribution π. We want to estimate expectations Pf of unbounded functions f. The constraint $Ph = 0$ also typically involves an unbounded h, for example when we assume that the embedded Markov chain has mean zero. This is why we assume $L_2(\pi)$-ergodicity rather than uniform ergodicity. The results also hold under the more flexible V-ergodicity, for which we refer to Meyn and Tweedie (1993) and Schick and Wefelmeyer (2002). Let $\|g\| = \pi(g^2)^{1/2}$ denote the norm of a function $g \in L_2(\pi)$, and let $\|K\| = \sup\{\|Kg\| : \|g\| = 1\}$ denote the corresponding operator norm of a kernel K on $S \times \mathcal{S}$. With the stationary distribution $\pi(dy)$ we associate the kernel $\Pi(x, dy) = \pi(dy)$ that does not depend on x. Exponential $L_2(\pi)$-ergodicity of the Markov chain is implied by the condition $\|Q - \Pi\| < 1$.

In order to characterize efficient estimators, we show that the distribution of the observations $X_0, X_1, \ldots, X_n$ is *locally asymptotically normal* in the following nonparametric sense. Let

$$U = \{u \in L_2(P) : \ Qu = 0\}.$$

For each $u \in U$ we can construct a perturbation Q_{nu} of Q that is *Hellinger differentiable* with derivative u,

$$P\Big(\Big(\frac{dQ_{nu}}{dQ}\Big)^{1/2} - 1 - \frac{1}{2}n^{-1/2}u\Big)^2 = o(n^{-1}).$$

Write $P^{(n)}$ and $P_{nu}^{(n)}$ for the joint law of the observations under Q and Q_{nu}, respectively. Let N denote a standard normal random variable. The following theorem shows nonparametric local asymptotic normality. Different proofs are in Penev (1991), Bickel (1993), and Greenwood and Wefelmeyer (1995).

Theorem 1. *Assume that* $\|Q - \Pi\| < 1$. *Let* $u \in U$. *Then*

$$\log \frac{dP_{nu}^{(n)}}{dP^{(n)}}(X_0, \dots, X_n) = n^{-1/2}\sum_{j=1}^n u(X_{j-1}, X_j) - \frac{1}{2}Pu^2 + o_p(n^{-1/2})$$

and

$$n^{-1/2}\sum_{j=1}^n u(X_{j-1}, X_j) \Rightarrow (Pu^2)^{1/2}N.$$

In what follows we use the notation

$$Q_x g = \int Q(x, dy)g(x, y), \qquad Q_x^t g = \int Q^t(x, dy)g(x, y),$$

with $Q^{t+1}(x, B) = \int Q(x, dy)Q^t(y, B)$ for $B \in \mathcal{S}$ and $t \geq 1$ a non-negative integer. This gives, in particular,

$$Q^2(x, B) = \int Q(x, dy)Q^1(y, B) = \int Q(x, dy)Q(y, B).$$

From Kartashov (1985a, 1985b, 1996) we obtain the following *perturbation expansion* for $g \in L_2(P)$:

$$n^{1/2}(P_{nu}g - Pg) \to P(uAg) \quad \text{for } u \in U \tag{2.1}$$

with Ag determined by

$$Ag(x, y) = g(x, y) - Q_x g + \sum_{t=1}^\infty (Q_y^t g - Q_x^{t+1} g). \tag{2.2}$$

For $g \in L_2(P)$, the *martingale approximation* of Gordin (1969) and Gordin and Lifšic (1978) is

$$n^{-1/2}\sum_{j=1}^n (g(X_{j-1}, X_j) - Pg) = n^{-1/2}\sum_{j=1}^n Ag(X_{j-1}, X_j) + o_p(1), \tag{2.3}$$

By a central limit theorem for martingales, the right side is asymptotically normal with variance $P[(Ag)^2]$.

Now suppose that the stationary distribution P fulfills the constraint $Ph = 0$ for some d-dimensional vector of functions $h \in L_2^d(P)$. The constraint also holds for the perturbed stationary distribution P_{nu}, and we obtain from (2.1), applied to the components of h, that $P(uAh) = 0$. The perturbations u are therefore constrained to

$$U_h = \{u \in U : \; P(uAh) = 0\}.$$

The stochastic expansion in Theorem 1 involves a norm $(Pu^2)^{1/2}$ on U. By the perturbation expansion (2.1), applied to f, we see that the functional Pf is *differentiable* at P with *gradient* $Af \in U$ in the sense that

$$n^{1/2}(P_{nu}f - Pf) \to P(uAf) \quad \text{for } u \in U.$$

The *canonical gradient* under the constraint $Ph = 0$ is the projection of Af onto U_h. By definition of U_h, the space U has the orthogonal decomposition $U = U_h \oplus [Ah]$, where $[Ah]$ is the linear span of the components of Ah. Hence the canonical gradient of Pf can be written $u_h = Af - u_h^\perp$, where $u_h^\perp$ is the projection of u_h onto $[Ah] = U_h^\perp$. Assuming that $P(AhAh^\top)$ is positive definite, this projection is of the form $u_h^\perp = c_h^\top Ah$ with

$$c_h = P(Ah\,Ah^\top)^{-1} P(Ah\,Af).$$

We obtain the canonical gradient

$$u_h = Af - P(Af\,Ah^\top)P(Ah\,Ah^\top)^{-1}Ah.$$

An estimator $\hat{\vartheta}$ is called *asymptotically linear* for Pf at P with *influence function* v if $v \in U$ and

$$n^{1/2}(\hat{\vartheta} - Pf) = n^{-1/2}\sum_{j=1}^{n} v(X_{j-1}, X_j) + o_p(1).$$

Given the constraint $Ph = 0$, an estimator $\hat{\vartheta}$ is called *regular* for Pf at P with *limit* L if L is a random variable such that

$$n^{1/2}(\hat{\vartheta} - P_{nu}f) \Rightarrow L \quad \text{for } u \in U_h.$$

The convolution theorem of Hájek (1970) then says that $L = P(u_h^2)^{1/2}N + M$ with M independent of N. This justifies calling $\hat{\vartheta}$ *efficient* if $L = P(u_h^2)^{1/2}N$. The convolution theorem also implies the following characterization of efficient estimators. We refer to Bickel, Klaassen, Ritov and Wellner (1998).

Theorem 2. *Assume that* $\|Q - \Pi\| < 1$. *Let* $f \in L_2(P)$ *and* $h \in L_2(P)$ *with* $Ph = 0$ *and* $P(Ah\,Ah^\top)$ *positive definite. Under the constraint* $Ph = 0$, *an estimator* $\hat\vartheta$ *for* Pf *is efficient at* P *if and only if* $\hat\vartheta$ *is asymptotically linear for* Pf *at* P *with influence function equal to the canonical gradient* u_h,

$$n^{1/2}(\hat\vartheta - Pf) = n^{-1/2}\sum_{j=1}^n u_h(X_{j-1}, X_j) + o_p(1). \tag{2.4}$$

The asymptotic variance of such an estimator $\hat\vartheta$ is

$$P(u_h^2) = P((Af)^2) - P(Af\,Ah^\top)P(Ah\,Ah^\top)^{-1}P(Ah\,Af).$$

This should be compared with the asymptotic variance of the empirical estimator $\mathbb{P}f = (1/n)\sum_{j=1}^n f(X_{j-1}, X_j)$, which is $P((Af)^2)$. In the next section we construct three efficient estimators for Pf under the constraint $Ph = 0$.

3 Three efficient estimators

By the characterization of efficient estimators for Pf under the constraint $Ph = 0$, given in Theorem 2, equation (2.4), an efficient estimator must be asymptotically equivalent to

$$\begin{aligned}
&Pf + \frac{1}{n}\sum_{j=1}^n u_h(X_{j-1}, X_j)\\
&\quad = Pf + \frac{1}{n}\sum_{j=1}^n \Big(Af(X_{j-1}, X_j) - P(Af\,Ah^\top)P(Ah\,Ah^\top)^{-1}Ah(X_{j-1}, X_j)\Big)\\
&\quad = \frac{1}{n}\sum_{j=1}^n \Big(f(X_{j-1}, X_j) - P(Af\,Ah^\top)P(Ah\,Ah^\top)^{-1}h(X_{j-1}, X_j)\Big) + o_p(n^{-1/2})\\
&\quad = \mathbb{P}f - \gamma^\top\Sigma^{-1}\frac{1}{n}\sum_{j=1}^n h(X_{j-1}, X_j) + o_p(n^{-1/2}).
\end{aligned}$$

The unknown expectations $\gamma = P(Af\,Ah$ and $\Sigma = P(Ah\,Ah^\top)$ can be estimated with empirical estimators. This is the approach of Müller, Schick and Wefelmeyer (2001), to which we refer for the proof. They introduce the *additively corrected empirical estimator* of Pf

$$\mathbb{P}_a f = \mathbb{P}f - \hat\gamma_a^\top\hat\Sigma_a^{-1}\frac{1}{n}\sum_{j=1}^n h(X_{j-1}, X_j). \tag{3.1}$$

Here $\hat{\gamma}_a$ is an empirical estimator for $P(Ah\,Af)$ and $\hat{\Sigma}_a$ is an empirical estimator for $P(Ah\,Ah^\top)$,

$$\begin{aligned}\hat{\gamma}_a &= \frac{1}{n}\sum_{j=1}^{n} h(X_{j-1},X_j)f(X_{j-1},X_j)\\ &\quad+\sum_{k=1}^{m}\frac{1}{n-k}\sum_{j=1}^{n-k}\big(h(X_{j-1},X_j)f(X_{j+k-1},X_{j+k})+h(X_{j+k-1},X_{j+k})f(X_{j-1},X_j)\big),\\ \hat{\Sigma}_a &= \frac{1}{n}\sum_{j=1}^{n} h(X_{j-1},X_j)h^\top(X_{j-1},X_j)+2\sum_{k=1}^{m}\frac{1}{n-k}\sum_{j=1}^{n-k}h(X_{j-1},X_j)h(X_{j+k-1},X_{j+k}),\end{aligned}$$

with m tending to infinity more slowly than the sample size n.

Theorem 3. (**First estimator**) *Assume that $\|Q-\Pi\|<1$. Let $f\in L_2(P)$ and $h\in L_2^d(P)$ with $Ph=0$ and $P(Ah\,Ah^\top)$ positive definite. Under the constraint $Ph=0$, the additively corrected estimator $\mathbb{P}_a f$ from (3.1) is asymptotically linear in the sense of (2.4) for Pf at P with influence function u_h, and therefore regular and efficient for Pf.*

A different improvement of the empirical estimator $\mathbb{P}f$ consists in weighting it appropriately. The empirical likelihood of Owen (1988, 2001) was developed for independent observations. For (weakly) dependent observations we can use the blockwise empirical likelihood introduced by Kitamura (1997). Decompose (the initial section of) the time points $1,\dots,n$ into $\nu=\lfloor n/m\rfloor$ disjoint blocks of length m, where m tends slowly to infinity with the sample size n. For the sake of simplicity we assume in the following that ν is exactly n/m, and therefore $n=\nu m$. For $i=1,\dots,m$ and $k\in L_2(P)$ we introduce the i-th block average

$$B_i k=\frac{1}{m}\sum_{l=1}^{m}k(X_{(i-1)m+l-1},X_{(i-1)m+l})$$

and set

$$F_i=B_i f\quad\text{and}\quad H_i=\begin{pmatrix}B_i h_1\\ \vdots\\ B_i h_d\end{pmatrix}.$$

Then the empirical estimator for Pf based on $n=\nu m$ observations can be written $\mathbb{P}f=\nu^{-1}\sum_{i=1}^{\nu}F_i$. The *blockwise weighted empirical estimator* for Pf is

$$\mathbb{P}_w f=\frac{1}{\nu}\sum_{i=1}^{\nu}\frac{F_i}{1+\zeta_\nu^\top H_i}\tag{3.2}$$

with random vector ζ_ν chosen such that $1+\zeta_\nu^\top H_i > 0$, $i = 1, \dots, \nu$, and

$$\sum_{i=1}^{\nu} \frac{H_i}{1+\zeta_\nu^\top H_i} = 0.$$

Such a random vector exists on an event with probability tending to one. On its complement we set ζ_ν equal to zero. On that event $\mathbb{P}_w f$ coincides with the empirical estimator $\mathbb{P}f$.

We will need the following assumption throughout.

Assumption 1. Assume that $\|Q-\Pi\| < 1$. Let the length of the blocks m and the number of blocks $\nu = n/m$ tend to infinity such that $m^2 = O(\nu)$ as $n \to \infty$.

Theorem 4. (**Second estimator**) *Assume that Assumption 1 is satisfied. Let $f \in L_2(P)$ and $h \in L_2^d(P)$ with $Ph = 0$ and $P(Ah\,Ah^\top)$ positive definite. Then the blockwise weighted empirical estimator $\mathbb{P}_w f$ from (3.2) is asymptotically linear in the sense of (2.4) for Pf at P with influence function u_h. In particular, $\mathbb{P}_w f$ is regular and efficient.*

The proof of Theorem 4 is provided in Section 4. It is based on the following three lemmas, which are also proved in Section 4.

Lemma 1. *Assume that Assumption 1 is satisfied and that $h \in L_2^d(P)$ with $Ph = 0$. Then*

$$m^{1/2} \max_{1\le i\le \nu} \|H_i\| = o_p(\nu^{1/2}).$$

Lemma 2. *Assume that Assumption 1 is satisfied and that $k_1, k_2 \in L_2(P)$ with $Pk_1 = Pk_2 = 0$. Then*

$$\frac{m}{\nu}\sum_{i=1}^{\nu} B_i k_1 B_i k_2 \xrightarrow{P} P(Ak_1\,Ak_2).$$

Lemma 3. *Assume that Assumption 1 is satisfied. Let $f \in L_2(P)$ and $h \in L_2^d(P)$ with $Ph = 0$. Under the constraint $Ph = 0$, the estimator $\mathbb{P}_w f$ admits the stochastic expansion*

$$\mathbb{P}_w f = \mathbb{P}f - \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i + o_p(n^{-1/2}). \tag{3.3}$$

We finally consider our third estimator, the *blockwise additively corrected empirical estimator*, which is suggested by the above expansion (3.3),

$$\mathbb{P}_b f = \mathbb{P}f - \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i. \tag{3.4}$$

That the estimator $\mathbb{P}_b f$ is also regular and efficient, is an immediate consequence of (3.3) and Theorem 4. We provide this result in the following theorem.

Theorem 5. (**Third estimator**) *Suppose the assumptions of Theorem 4 are satisfied. Then the blockwise additively corrected empirical estimator $\mathbb{P}_b f$ of Pf given in (3.4) is regular and efficient under the constraint $Ph = 0$.*

The two additively corrected estimators $\mathbb{P}_a$ and $\mathbb{P}_b$ from Theorem 3 and Theorem 5 have a similar form and are essentially the same. To see this consider the first estimator

$$\mathbb{P}_a f = \mathbb{P}f - \hat{\gamma}_a^\top \hat{\Sigma}_a^{-1} \frac{1}{\nu} \sum_{i=1}^{\nu} h(X_{i-1}, X_i),$$

which involves two (consistent) empirical estimators $\hat{\gamma}_a$ and $\hat{\Sigma}_a$ to estimate $\gamma = P(AfAh)$ and $\Sigma = P(AhAh^\top)$, respectively. The third estimator has the same structure, but uses instead $\hat{\gamma}_b = m/\nu \sum_{i=1}^{\nu} F_i H_i$ and $\hat{\Sigma}_b = m/\nu \sum_{i=1}^{\nu} H_i H_i^\top$ in place of $\hat{\gamma}_a$ and $\hat{\Sigma}_a$. In the proof of Theorem 4 we show with the help of Lemma 2 that these estimators are also consistent for γ and Σ, so the two approaches are indeed asymptotically equivalent. The estimators $m/\nu \sum_{i=1}^{\nu} F_i H_i$ and $m/\nu \sum_{i=1}^{\nu} H_i h_i^\top$ remain consistent if we replace H_i by $H_i - \bar{H}$ with $\bar{H} = 1/\nu \sum_{i=1}^{\nu} H_i = \frac{1}{n} \sum_{j=1}^{n} h(X_{j-1}, X_j)$. We denote this version of $\mathbb{P}_b f$ by $P_c f$, i.e.,

$$\mathbb{P}_c f = \mathbb{P}f - \frac{1}{\nu} \sum_{i=1}^{\nu} F_i (H_i - \bar{H})^\top \Big(\frac{1}{\nu} \sum_{i=1}^{\nu} (H_i - \bar{H})(H_i - \bar{H})^\top \Big)^{-1} \frac{1}{\nu} \sum_{i=1}^{\nu} H_i. \tag{3.5}$$

In our limited simulations (see Table 1 below) $\mathbb{P}_c f$ performed better than $P_b f$.

In order to examine the finite sample behavior of our proposed approach we conducted a small simulation study in a simple autoregressive model of order 1, $X_i = \rho X_{i-1} + \varepsilon_i$, with $\rho = 0.5$, $\varepsilon_i \sim N(0,1)$, $i = 1, \dots, n = 200$, initial value $X_0 \sim N(0, 1/(1-\rho^2))$ and one-dimensional constraint

$$E[h(X_{i-1}, X_i)] = E[X_i - \rho X_{i-1}] = E[\varepsilon_i] = 0.$$

This constraint is automatically incorporated since the innovations are generated from a standard normal distribution.

We compared the empirical estimator $\mathbb{P}f$ with the blockwise weighted empirical estimator $\mathbb{P}_w f$ that uses empirical likelihood weights and with the blockwise additively corrected estimators $\mathbb{P}_b f$ and $\mathbb{P}_c f$. We did not include estimator $\mathbb{P}_a f$ in the study since $\mathbb{P}_b f$ and $\mathbb{P}_c f$ are of the same type and are easier to compute.

In our scenario with one-dimensional constraint $E[h(X_{i-1}, X_i)] = E[\varepsilon_i] = 0$, the block average H_i in block i becomes the block average of the innovations in block i, $H_i = \bar{\varepsilon}_i$, and the estimator $\mathbb{P}_b f$ and $P_c f$ simplify to

$$\mathbb{P}_b f = \mathbb{P}f - \frac{\nu^{-1} \sum_{i=1}^{\nu} F_i \bar{\varepsilon}_i}{\nu^{-1} \sum_{i=1}^{\nu} \bar{\varepsilon}_i^2} \frac{1}{\nu} \sum_{i=1}^{\nu} \bar{\varepsilon}_i, \qquad \mathbb{P}_c f = \mathbb{P}f - \frac{\nu^{-1} \sum_{i=1}^{\nu} F_i (\bar{\varepsilon}_i - \bar{\varepsilon})}{\nu^{-1} \sum_{i=1}^{\nu} (\bar{\varepsilon}_i - \bar{\varepsilon})^2} \frac{1}{\nu} \sum_{i=1}^{\nu} \bar{\varepsilon}_i,$$

with $\bar{\varepsilon} = n^{-1} \sum_{j=1}^{n} \varepsilon_j$ the grand average of the innovations. In order to compare the above estimators of expectations Pf we consider five functions f. The simulated means and the simulated mean squared errors of the estimators are given in Table 1.

$f(x,y)$		$\mathbb{P}f$	$\mathbb{P}_w f$	$\mathbb{P}_b f$	$\mathbb{P}_c f$
$1(y<0)$	mean	0.500	0.500	0.475	0.500
	n*MSE	0.572	0.113	0.315	0.112
$1(\min(x,y)>0)$	mean	0.334	0.333	0.315	0.331
	n*MSE	0.637	0.191	0.285	0.191
$y/\sqrt{1+x^2}$	mean	0.000	0.000	0.000	0.000
	n*MSE	1.681	1.681	0.069	0.080
$\max(x,y)$	mean	0.462	0.460	0.438	0.461
	n*MSE	4.123	4.011	0.438	0.299
xy	mean	0.667	0.665	0.604	0.633
	n*MSE	4.580	4.569	5.163	4.815

Table 1: The table entries are the simulated means and the simulated mean squared errors (multiplied with the sample size $n = 200$) for the empirical estimator $\mathbb{P}f$, the blockwise weighted estimator using empirical likelihood weights $\mathbb{P}_w f$ and the blockwise additively corrected estimators $\mathbb{P}_b f$ and $\mathbb{P}_c f$. The number of blocks is $\nu = 20$ and the block size is $m = 10$. The simulation results are based on $100,000$ repetitions.

The simulation results in Table 1 show that the proposed approaches perform about as well or are better than the empirical estimator $\mathbb{P}f$. The figures for the first two functions are especially good: the blockwise weighted estimator and the additively corrected estimator all have clearly smaller MSE's than the empirical estimator. The results for the third and fourth function are clearly in favor of the additively corrected estimators, whereas the blockwise weighted and the empircial estimator perform similarly. The fifth function is an example where all estimators perform in a similar way, i.e. weighting or adding a correction term to $\mathbb{P}f$ does not lead to an improvement. Overall, estimator $\mathbb{P}_c f$ seems to work best in the considered scenario.

4 Proofs.

Proof of Theorem 4. By the martingale approximation (2.3) we have

$$\mathbb{P}f = \frac{1}{n} \sum_{j=1}^{n} f(X_{j-1}, X_j) = Pf + \frac{1}{n} \sum_{j=1}^{n} Af(X_{j-1}, X_j) + o_p(n^{-1/2})$$

and

$$(4.1)\qquad \frac{1}{\nu}\sum_{i=1}^{\nu} H_i = \frac{1}{n}\sum_{j=1}^{n} h(X_{j-1},X_j) = \frac{1}{n}\sum_{j=1}^{n} Ah(X_{j-1},X_j) + o_p(n^{-1/2}) = O_p(n^{-1/2}).$$

In view of these two expansions and Lemma 3, the desired result follows if we verify the following,

$$(4.2)\qquad \frac{m}{\nu}\sum_{i=1}^{\nu} F_i H_i = P(AfAh) + o_p(1),$$

$$(4.3)\qquad \frac{m}{\nu}\sum_{i=1}^{\nu} H_i H_i^{\top} = P(AhAh^{\top}) + o_p(1).$$

It follows from Lemma 2, applied with $k_1 = f - Pf$ and $k_2 = h_a$, $a = 1,\dots,d$, that $m/\nu\sum_{i=1}^{\nu} B_i(f-Pf)B_i h_a$, which is the a-th coordinate of $m/\nu\sum_{i=1}^{\nu}(F_i-Pf)H_i$, converges in probability to $P(A(f-Pf)Ah_a)$, which equals the a-th coordinate of $P(AfAh)$ because $APf = 0$. This and (4.1) show

$$\frac{m}{\nu}\sum_{i=1}^{\nu} F_i H_i = \frac{m}{\nu}\sum_{i=1}^{\nu}(F_i - Pf)H_i + Pf\Big(\frac{m}{\nu}\sum_{i=1}^{\nu} H_i\Big) = P(AfAh) + o_p(1).$$

Applying Lemma 2 with $k_1 = h_a$ and $k_2 = h_b$ for $a, b = 1,\dots,d$ establishes (4.3).

Proof of Lemmas 1 and 2. Let $k \in L_2(P)$ with $Pk = 0$ and set $l = Ak$ with Ak defined in (2.2). For $i = 1,\dots,n$, set

$$K_i = m^{-1/2}\sum_{j=1}^{m} k(X_{(i-1)m+j-1}, X_{(i-1)m+j}),$$
$$L_i = m^{-1/2}\sum_{j=1}^{m} l(X_{(i-1)m+j-1}, X_{(i-1)m+j}).$$

In Schick (2024, Theorem 5) a Wilks' type theorem for a blockwise empirical likelihood for Markov chains under a conditional Lindeberg condition is presented. In the proof the following results are derived under Assumption 1.

$$E(L_i|X_0,\dots,X_{(i-1)m}) = 0, \quad j = 1,\dots,\nu,$$

$$(4.4)\qquad \frac{1}{\nu}\sum_{i=1}^{\nu} E(L_i^2|X_0,\dots,X_{(i-1)m}) = P(l^2) + o_p(1),$$

(4.5) $$\frac{1}{\nu}\sum_{i=1}^{\nu} E(L_i^2\mathbf{1}[|L_i| > \eta\nu^{1/2}]|X_0,\dots,X_{(i-1)m}) = o_p(1), \quad \eta > 0,$$

(4.6) $$\frac{1}{\nu}\sum_{i=1}^{\nu}(K_i - L_i)^2 = o_p(1).$$

It follows from (4.4) and (4.5) that $\nu^{-1}\sum_{i=1}^{\nu} L_i^2 = P(l^2) + o_p(1)$, and this and (4.6) imply

(4.7) $$\frac{1}{\nu}\sum_{i=1}^{\nu} K_i^2 = P(l^2) + o_p(1).$$

It follows from (4.5) that $\max_{1\le i\le\nu}|L_i| = o_p(\nu^{1/2})$ and from (4.6) that $\max_{1\le i\le\nu}|K_i - L_i|^2 \le \sum_{i=1}^{\nu}(K_i - L_i)^2 = o_p(\nu)$. Thus we have

(4.8) $$\max_{1\le i\le\nu}|K_i| = o_p(\nu^{1/2}).$$

The conclusion of Lemma 1 follows from the inequality

$$m^{1/2}\max_{1\le i\le\nu}\|H_i\| \le \sum_{a=1}^{d}\max_{1\le i\le\nu}|m^{-1/2}\sum_{j=1}^{n} h_a(X_{(i-1)m+j-1}, X_{(i-1)m+j})|$$

and (4.8) applied with $k = h_1,\dots,k = h_d$. To obtain the conclusion of Lemma 2, write $K_{i,a}$ for K_i if $k = k_a$, $a = 1,2$, set $l_a = Ak_a$, and form the matrices

$$M_n = \frac{1}{\nu}\sum_{i=1}^{\nu}\begin{bmatrix} K_{i,1}^2 & K_{i,1}K_{i,2} \\ K_{i,2}K_{i,1} & K_{i,2}\end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} P(l_1^2) & P(l_1l_2) \\ P(l_2l_1) & P(l_2^2)\end{bmatrix}.$$

Fix a unit vector $u = (u_1,u_2)^\top$ in $\mathbb{R}^2$. Applying (4.7) with $k = u_1k_1 + u_2k_2$, we derive

$$u^\top M_n u = \frac{1}{\nu}\sum_{i=1}^{\nu} K_i^2 = P(Ak)^2) = u^\top M u.$$

Since this holds for all unit vectors, we conclude that $M_n = M + o_p(1)$ and this gives

$$\frac{1}{\nu}\sum_{i=1}^{\nu} K_{i,1}K_{i,2} = \frac{m}{\nu}\sum_{i=1}^{\nu} B_ik_1B_ik_2 = P(l_1l_2) + o_p(1)$$

which is the conclusion of Lemma 2. □

Proof of Lemma 3. For the proof of the expansion provided in this lemma we will use several technical arguments provided in Peng and Schick (2013) Chapters 5 and 6 and adapt them to our scenario. Recall that $H_i = m^{-1}\sum_{k=1}^m h(X_{(i-1)m+k-1}, X_{(i-1)m+k})$ with $h = (h_1, \dots, h_d)^\top$. Write

$$H_\nu^* = \max_{1\le i\le \nu} \|H_i\|, \qquad \bar H_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} H_i, \qquad H_\nu^{(k)} = \sup_{|u|=1}\Big(\frac{1}{\nu}\sum_{i=1}^{\nu}(u^\top H_i)^k\Big), \quad k = 1, 2, \dots$$

$$S_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top \quad \text{and} \quad T_\nu = \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i.$$

Further let λ_ν and Λ_ν denote the smallest and largest eigenvalue of S_ν,

$$\lambda_\nu = \inf_{|u|=1} u^\top S_\nu u, \quad \Lambda_\nu = \sup_{|u|=1} u^\top S_\nu u.$$

Peng and Schick (2013) consider random vectors with fixed and increasing dimension as the sample size n tends to infinity. In contrast to that paper, we consider d-dimensional vectors of block averages, with d fixed, where the number of blocks ν and also the length of the blocks m increases as $n \to \infty$. Instead of assumptions (A1)–(A3) in Peng and Schick (2013) Section 6 we will therefore need the following three conditions that are adapted to the different asymptotics considered here.

(C1) $m^{1/2} H_\nu^* = o_p(\nu^{1/2})$,

(C2) $\bar H_\nu = O_p(n^{-1/2})$,

(C3) $m S_\nu = P(AhAh^\top) + o_p(1)$.

The first condition is Lemma 1. Conditions (C2) and (C3) are derived in the proof of Theorem 4 and follow from Lemma 2.

Condition (C3) guarantees that there are numbers a and b, $0 < a < b$ with $P(a \le m\lambda_\nu \le m\Lambda_\nu \le b) \to 1$. Therefore the probability of the event $\{\lambda_\nu > 5H_\nu^* \|\bar H_\nu\|\}$ tends to one. On this event the assumptions of Lemma 5.2 in Peng and Schick (2013) are satisfied, which guarantees that a unique d-dimensional random vector ζ_ν exists with $1 + \zeta_\nu^\top H_i > 0$, $i = 1, \dots, \nu$, and

$$\frac{1}{\nu}\sum_{i=1}^{\nu} \frac{H_i}{1 + \zeta_\nu^\top H_i} = 0. \tag{4.9}$$

We refer to Peng and Schick (2013) for details.

The error term in (3.3) is

$$\begin{aligned} R_n &= \mathbb{P}_w f - \mathbb{P} f + \frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top \Big(\frac{1}{\nu}\sum_{i=1}^{\nu} H_i H_i^\top\Big)^{-1} \frac{1}{\nu}\sum_{i=1}^{\nu} H_i \\ &= \frac{1}{\nu}\sum_{i=1}^{\nu}\Big(\frac{F_i}{1 + \zeta_\nu^\top H_i} - F_i + F_i H_i^\top S_\nu^{-1} \bar H_\nu\Big). \end{aligned}$$

We need to show that it has order $o_p(n^{-1/2})$. As a first step, we bound its absolute value by a sum of two terms,

$$|R_n| \le \Big|\frac{1}{\nu}\sum_{i=1}^{\nu}\Big(\frac{F_i}{1+\zeta_\nu^\top H_i} - F_i + F_i H_i^\top \zeta_\nu\Big)\Big| + \Big|\frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top (S_\nu^{-1}\bar{H}_\nu - \zeta_\nu)\Big|. \tag{4.10}$$

In order to obtain the desired rate $o_p(n^{-1/2})$ for both terms, we will use some auxiliary results from Peng and Schick (2013) Chapter 5. We begin with a brief review of these results, adapted to our setting.

Let u be a unit vector such that $\zeta_\nu = \|\zeta_\nu\| u$. From (4.9) we obtain

$$0 = \frac{1}{\nu}\sum_{i=1}^{\nu}\frac{u^\top H_i(1+\zeta_\nu^\top H_i - \zeta_\nu^\top H_i)}{1+\zeta_\nu^\top H_i} = u^\top \bar{H}_\nu - \|\zeta_\nu\|\frac{1}{\nu}\sum_{i=1}^{\nu}\frac{(u^\top H_i)^2}{1+\zeta_\nu^\top H_i}.$$

This and the inequality

$$\lambda_\nu \le u^\top S_\nu u = \frac{1}{\nu}\sum_{i=1}^{\nu}(u^\top H_i)^2 \le \frac{1}{\nu}\sum_{i=1}^{\nu}\frac{(u^\top H_i)^2(1+\|\zeta_\nu\|H_\nu^*)}{1+\zeta_\nu^\top H_i}$$

yield $\lambda_\nu\|\zeta_\nu\| \le (1+\|\zeta_\nu\|H_\nu^*)u^\top \bar{H}_\nu \le (1+\|\zeta_\nu\|H_\nu^*)\|\bar{H}_\nu\|$ and thus

$$\|\zeta_\nu\| \le \frac{\|\bar{H}_\nu\|}{\lambda_\nu - \|\bar{H}_\nu\|H_\nu^*}. \tag{4.11}$$

From this we derive

$$\|\zeta_\nu\|H_\nu^* \le \frac{\|\bar{H}_\nu\|H_\nu^*}{\lambda_\nu - \|\bar{H}_\nu\|H_\nu^*} < \frac{1}{4}, \tag{4.12}$$

$$\max_{1\le i\le\nu}\frac{1}{1+\zeta_\nu^\top H_i} \le \frac{1}{1+\|\zeta_\nu\|H_\nu^*} < \frac{4}{3}, \tag{4.13}$$

$$\frac{1}{\nu}\sum_{i=1}^{\nu}(\zeta_\nu^\top H_i)^2 = \zeta_\nu^\top S_\nu \zeta_\nu \le \Lambda_\nu\|\zeta_\nu\|^2 \le \frac{\Lambda_\nu\|\bar{H}_\nu\|^2}{(\lambda_\nu - \|\bar{H}_\nu\|H_\nu^*)^2}.$$

We can write $1/(1+d) - 1 + d = d^2/(1+d)$. This identity with $d = \zeta_\nu^\top H_i$ and (4.13) yield for vectors $r_1, \dots, r_\nu$ of the same dimension

$$\Big\|\frac{1}{\nu}\sum_{i=1}^{\nu}\Big(\frac{r_i}{1+\zeta_\nu^\top H_i} - r_i + r_i H_i^\top \zeta_\nu\Big)\Big\| \le \Big\|\frac{1}{\nu}\sum_{i=1}^{\nu} r_i \frac{(\zeta_\nu^\top H_i)^2}{1+\zeta_\nu^\top H_i}\Big\| \tag{4.14}$$

Taking $r_i = S_\nu^{-1} H_i$ we obtain in view of (4.9)

$$\|\zeta_\nu - S_\nu^{-1}\bar{H}_\nu\| \le \Big\|\frac{1}{\nu}\sum_{i=1}^{\nu} S_\nu^{-1} H_i \frac{(\zeta_\nu^\top H_i)^2}{1+\zeta_\nu^\top H_i}\Big\| = \sup_{\|u\|=1}\frac{1}{\nu}\sum_{i=1}^{\nu} u^\top S_\nu^{-1} H_i \frac{(\zeta_\nu^\top H_i)^2}{1+\zeta_\nu^\top H_i}.$$

where we used the formula $\|x\| = \sup_{\|u\|=1} u^\top x$ in the last step. We now use the Cauchy-Schwarz inequality, (4.12) and the identity $(u^\top S_\nu^{-1} H_i)^2 = u^\top S_\nu^{-1} H_i H_i^\top S_\nu^{-1} u$ to obtain the bound

$$\|\zeta_\nu - S_\nu^{-1}\bar{H}_\nu\|^2 \le \sup_{\|u\|=1} u^\top S_\nu^{-1} u \, \frac{1}{\nu}\sum_{i=1}^{\nu} \frac{(\zeta_\nu^\top H_i)^4}{(1+\zeta_\nu^\top H_i)^2} \le \frac{16}{9\lambda_\nu}\|\zeta_\nu\|^4 H_\nu^{(4)}.$$

The eigenvalues $m\lambda_\nu$ and $m\Lambda_\nu$ of mS_ν are finite and bounded away from zero on an event tending to 1 in probability. Inequality (4.11) combined with rates $m^{1/2}H_\nu^* = o_p(\nu^{1/2})$ and $\bar{H}_\nu = O_p(n^{-1/2})$ specified in (C1) and (C2) yield

$$\|\zeta_\nu\| \le \frac{m\|\bar{H}_\nu\|}{m\lambda_\nu - m\|\bar{H}_\nu\| H_\nu^*} = O_p(mn^{-1/2}). \tag{4.15}$$

Since the term $H_\nu^{(4)}$ is bounded by $\Lambda_\nu (H_\nu^*)^2 = o_p(\nu/m^2) = o_p(n/m^3)$, we have

$$\|\zeta_\nu - S_\nu^{-1}\bar{H}_\nu\|^2 = O_p(m) O_p\Big(\frac{m^4}{n^2}\Big) o_p\Big(\frac{n}{m^3}\Big) = o_p\Big(\frac{m^2}{n}\Big). \tag{4.16}$$

We are now in a position to derive the desired order $o_p(n^{-1/2})$ for the error term in (3.3) and return to the two terms on the right-hand side of (4.10). Let us consider the second term first. It follows from (4.2) that $\|T_\nu\| = O_p(1/m)$. This and (4.16) give

$$\Big|\frac{1}{\nu}\sum_{i=1}^{\nu} F_i H_i^\top (S_\nu^{-1}\bar{H}_\nu - \zeta_\nu)\Big| \le \|T_\nu\| \|S_\nu^{-1}\bar{H}_\nu - \zeta_\nu\| = o_p(n^{-1/2}).$$

The first term on the right-hand side of (4.10) is bounded by the sum

$$\Big|\frac{1}{\nu}\sum_{i=1}^{\nu}\Big(\frac{F_i - Pf}{1+\zeta_\nu^\top H_i} - (F_i - Pf) + (F_i - Pf) H_i^\top \zeta_\nu\Big)\Big| + |Pf| \|\bar{H}_\nu\| \|\zeta_\nu\|$$

The second summand is of order $O_p(m/n) = o_p(n^{-1/2})$ in view of (C2) and (4.15). We use inequality (4.14), this time with $\widetilde{F}_i = F_i - Pf$ in place of r_i, and the Cauchy-Schwarz inequality to bound the square of the first summand by

$$\frac{1}{\nu}\sum_{i=1}^{\nu}\widetilde{F}_i^2 \frac{1}{\nu}\sum_{i=1}^{\nu}\frac{(\zeta_\nu^\top H_i)^4}{(1+\zeta_\nu^\top H_i)^2} \le \frac{1}{\nu}\sum_{i=1}^{\nu}\widetilde{F}_i^2 \frac{16}{9}\|\zeta_\nu\|^4 H_\nu^{(4)} = O_p\Big(\frac{1}{m}\Big) O_p\Big(\frac{m^4}{n^2}\Big) o_p\Big(\frac{n}{m^3}\Big) = o_p\Big(\frac{1}{n}\Big).$$

The order follows from (4.15), $H_\nu^{(4)} = o_p(n/m^3)$ and

$$E\Big(\frac{1}{\nu}\sum_{i=1}^{\nu}\widetilde{F}_i^2\Big) = E(\widetilde{F}_1^2) = O\Big(\frac{1}{m}\Big).$$

□

References

[1] Bickel, P.J. (1993). Estimation in semiparametric models. In: *Multivariate Analysis: Future Directions* (C.R. Rao, ed.), 55–73, North-Holland, Amsterdam.

[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York.

[3] Gordin, M.I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10**, 1174–1176.

[4] Gordin, M.I. and Lifšic, B.A. (1978). The central limit theorem for stationary Markov processes. *Soviet Math. Dokl.* **19**, 392–394.

[5] Greenwood, P.E. and Wefelmeyer, W. (1995). Efficiency of empirical estimators for Markov chains, *Ann. Statist.*, **23**, 132–143.

[6] Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14**, 323–330.

[7] Jiang, F. and Wang, L. (2018) Adjusted blockwise empirical likelihood for long memory time series models. *Stat. Methods Appl.* **27**, 319–332.

[8] Kartashov, N.V. (1985a). Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theory Probab. Math. Statist.* **30**, 71–89.

[9] Kartashov, N.V. (1985b). Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Theory Probab. Appl.* **30**, 247–259.

[10] Kartashov, N.V. (1996). *Strong Stable Markov Chains.* VSP, Utrecht.

[11] Kim, Y.M., Lahiri, S.N. and Nordman, D.J. (2013). A progressive block empirical likelihood method for time series. *J. Amer. Statist. Assoc.* **108**, 1506–1516.

[12] Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084–2102.

[13] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

[14] Müller, U.U., Schick, A. and Wefelmeyer, W. (2001). Improved estimators for constrained Markov chain models. *Statist. Probab. Lett.* **54**, 427–435.

[15] Nordman, D.J., Bunzel, H. and Lahiri, S.N. (2013). A nonstandard empirical likelihood for time series. *Ann. Statist.* **41**, 3050–3073.

[16] Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

[17] Owen, A.B. (2001). *Empirical Likelihood.* Monographs on Statistics and Applied Probability 92, Chapman & Hall/CRC, London.

[18] Penev, S. (1991). Efficient estimation of the stationary distribution for exponentially ergodic Markov chains. *J. Statist. Plann. Inference* **27**, 105–123.

[19] Peng, H. and Schick, A. (2013). Empirical likelihood approach to goodness of fit testing. *Bernoulli* **19**, 954–981.

[20] Schick, A. (2024). Empirical likelihood with martingale differences. To appear in *J. Time Ser. Anal.*, https://doi.org/10.1111/jtsa.12800.

[21] Schick, A. and Wefelmeyer, W. (2002). Estimating joint distributions of Markov chains. *Stat. Inference Stoch. Process.* **5**, 1–22.

[22] Wu, R. and Cao, J. (2011). Blockwise empirical likelihood for time series of counts. *J. Multivariate Anal.* **102**, 661–673.

[23] Zhang, X. and Shao, X. (2016). On the coverage bound problem of empirical likelihood methods for time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78**, 395–421.