Journal of Econometrics 185 (2015) 283–304

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

High dimensional generalized empirical likelihood for moment restrictions with dependent data

Jinyuan Chang [a], Song Xi Chen [b,*], Xiaohong Chen [c]

[a] Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, 3010, Australia
[b] Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China
[c] Department of Economics, Yale University, New Haven, CT, 06520, USA

ARTICLE INFO

Article history:
Received 21 May 2013
Received in revised form
23 September 2014
Accepted 10 October 2014
Available online 15 November 2014

JEL classification:
C14
C30
C40

Keywords:
Generalized empirical likelihood
High dimensionality
Penalized likelihood
Variable selection
Over-identification test
Weak dependence

ABSTRACT

This paper considers the maximum generalized empirical likelihood (GEL) estimation and inference on parameters identified by high dimensional moment restrictions with weakly dependent data when the dimensions of the moment restrictions and the parameters diverge along with the sample size. The consistency with rates and the asymptotic normality of the GEL estimator are obtained by properly restricting the growth rates of the dimensions of the parameters and the moment restrictions, as well as the degree of data dependence. It is shown that even in the high dimensional time series setting, the GEL ratio can still behave like a chi-square random variable asymptotically. A consistent test for the over-identification is proposed. A penalized GEL method is also provided for estimation under sparsity setting.

1. Introduction

In economic, financial and statistical applications, econometric models defined with a growing number of parameters and moment restrictions are increasingly employed. Vector autoregressive models, dynamic asset pricing models, dynamic panel data models and high dimensional dynamic factor models are specific examples; see, e.g., Bai and Ng (2002), Stock and Watson (2010) and Fan and Liao (2014). Due to the desire to better capture large scale dynamic fundamental relations, these models with large number of unknown parameters of interest are typically used to for time series data of high dimension due to a large number of variables (relative to the sample size).

The unconditional moment restriction models are the inferential settings of the Generalized Method of Moment (GMM) of Hansen (1982), which is perhaps the most popular econometric method for semiparametric statistical inference. There are two dimensions that play essential roles in this method: the dimension of the moment restrictions and the dimension of the unknown parameters of interest. When both dimensions are fixed and finite, there is a huge established literature on inferential procedures, which include but not restrict to Rothenberg (1973) for the minimum distance, Hansen (1982) and Hansen and Singleton (1982) for the GMM, Owen (1988), Qin and Lawless (1994) and Kitamura (1997) for the empirical likelihood (EL), Smith (1997), Newey and Smith (2004) and Anatolyev (2005) for the generalized empirical likelihood (GEL). Among these methods, some members of the GEL (especially the EL) have the attractive properties of the Wilks theorem (Owen, 1988, 1990; Qin and Lawless, 1994), Bartlett correction (Chen and Cui, 2006, 2007), and a smaller second order bias (Newey and Smith, 2004; Anatolyev, 2005). See Owen (2001), Kitamura (2007) and Chen and Van Keilegom (2009) for reviews.

This paper investigates high dimensional GEL estimation and testing for weakly dependent observations when the dimensions of both the moment restrictions and the unknown parameters of interest may grow with the sample size n. Let p and r denote the dimension of the unknown parameters and the number of moment restrictions, respectively. When $r \geq p$, we investigate the

* Corresponding author. Tel.: +86 10 62760736; fax: +86 10 62760736.
E-mail address: csx@gsm.pku.edu.cn (S.X. Chen).

http://dx.doi.org/10.1016/j.jeconom.2014.10.011
0304-4076/

impacts of p and r on the consistency, the rate of convergence and the asymptotic normality of the GEL estimator, the limiting behavior of the GEL ratio statistics as well as the overidentification test. To accommodate the potential serial dependence in the estimating functions induced by the original time series data, the blocking technique is employed. This paper establishes the consistency (with rate) and the asymptotic normality of the GEL estimator under either (fixed) finite or diverging block size M when some suitable restrictions are imposed on r, p, M and n. It is demonstrated that in general the blocking technique with a diverging block size delivers the estimation efficiency. We also discuss the impact of the smallest eigenvalue of the covariance matrix of the averaged estimating function on the consistency and the asymptotic normality of the GEL estimator. We show that, even in high dimensional nonlinear time series setting (with diverging M), the GEL ratio still behaves like a chi-square random variable asymptotically, which echoes a similar result by Fan et al. (2001) for nonparametric regression with iid data. A GEL based over-identification specification test is also presented for high dimensional time series models, which extends that of Donald et al. (2003) for iid data from increasing dimension of moments (r) but fixed finite dimension of parameters (p) to both dimensions are allowed to diverge (as long as $r - p > 0$). Finally, when the parameter space is sparse, a penalized GEL method is proposed to allow for $p > r$, and is shown to attain the oracle property in the selection consistency as well as the asymptotic normality of the estimated non-zero parameters.

There are some studies on the EL and its related methods under high dimensionality of both the moment restrictions and the parameters of interest. Chen et al. (2009) and Hjort et al. (2009) evaluated the EL ratio statistic for the mean under high dimensional setting. Tang and Leng (2010) and Leng and Tang (2012) evaluated a penalized EL when the underlying parameter is sparse in the context of the mean parameters and the general estimating equations, respectively. Fan and Liao (2014) considered penalized GMM estimation under high dimensionality and sparsity assumption. These papers assume independent data. Recently, by allowing for dependent data but losing the self-standardization property of the EL, Lahiri and Mukhopadhyay (2012) proposed a modified EL method by adding a penalty term to the original EL criterion for estimating the high-dimensional mean parameters with $r = p > n$. Lahiri and Mukhopadhyay (2012) did not implement data blocking in their modified (penalized) EL for the means despite the moment equations are serially dependent. The EL ratio statistic based on their modified EL method is no longer asymptotically pivotal. As a result, any inference based on this modified EL has to use data blocking or other HAC long-run variance estimation. The rationale in our paper is to preserve the attractive self-standardization property of the GEL in high dimensional time series setting; doing so makes our allowed dimensionality smaller than that in Lahiri and Mukhopadhyay (2012) but maintains simple GEL inference.

The rest of the paper is organized as follows. Section 2 introduces the high dimensional model framework and the basic regularity conditions. Sections 3 and 4 establish the consistency, the rate of convergence and the asymptotic normality of the GEL estimator. Sections 5 and 6 derive the asymptotic properties of the GEL ratio statistic and the overidentification specification test respectively. Section 7 presents a penalized GEL approach for parameter estimation and variable selection when the unknown parameter is sparse. Section 8 reports some simulation results and Section 9 briefly concludes. Technical lemmas and all the proofs are given in the Appendix.

2. Preliminaries

2.1. Empirical likelihood and its generalization

Let $\{X_t\}_{t=1}^n$ be a sample of size n from an $\mathbb{R}^d$-valued strictly stationary stochastic process, where d denotes the dimension of X_t, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$ be a p-dimensional parameter taking values in a parameter space Θ. Consider a sequence of r-dimensional estimating equation

$$g(X_t, \boldsymbol{\theta}) = (g_1(X_t, \boldsymbol{\theta}), \dots, g_r(X_t, \boldsymbol{\theta}))'$$

for $r \geq p$. The model information regarding the data and the parameter is summarized by moment restrictions

$$E\{g(X_t, \boldsymbol{\theta}_0)\} = \mathbf{0} \tag{1}$$

where $\boldsymbol{\theta}_0 \in \Theta$ is the true parameter. As argued in Hjort et al. (2009), the moment restrictions (1) can be viewed as a triangular array where $r, d, X_t, \boldsymbol{\theta}$ and $g(x, \boldsymbol{\theta})$ may all depend on the sample size n. We will explicitly allow r and/or p grow with n while considering inference for $\boldsymbol{\theta}_0$ identified by (1). Although there is often a connection between d and r which is dictated by the context of an econometrical or statistical analysis, the theoretical results established in this paper are written directly on the growth rates of r and p relative to n. Hence, we will not impose explicit conditions on d which can be either growing or fixed. Certainly, when d diverges, it would indirectly affect the underlying assumptions made in Section 2.3, for instance the moment condition and the rate of the mixing coefficients.

We assume the dependence in the time series $\{X_t\}$ satisfies the α-mixing condition (Doukhan, 1994). Specifically, let $\mathscr{F}_u^v = \sigma(X_t : u \leq t \leq v)$ be the σ-field generated by the data from a time u to a time v for $v \geq u$. Then, the α-mixing coefficients are defined as

$$\alpha_X(k) = \sup_d \sup_{A \in \mathscr{F}_{-\infty}^0, B \in \mathscr{F}_k^\infty} |P(A \cap B) - P(A)P(B)| \quad \text{for each } k \geq 1.$$

The α-mixing condition means that $\alpha_X(k) \to 0$ as $k \to \infty$. When $\{X_t\}$ are independent, $\alpha_X(k) = 0$ for all $k \geq 1$.

We employ the blocking technique (Hall, 1985; Carlstein, 1986; Künsch, 1989) to preserve the dependence among the underlying data. Let M and L be two integers denoting the block length and separation between adjacent blocks, respectively. Then, the total number of blocks is $Q = \lfloor (n-M)/L \rfloor + 1$, where $\lfloor \cdot \rfloor$ is the integer truncation operator. For each $q = 1, \dots, Q$, the qth data block $B_q = (X_{(q-1)L+1}, \dots, X_{(q-1)L+M})$. The average of the estimating equation over the qth block is

$$\phi_M\left(B_q, \boldsymbol{\theta}\right) = \frac{1}{M} \sum_{m=1}^{M} g(X_{(q-1)L+m}, \boldsymbol{\theta}). \tag{2}$$

Clearly, $E\{\phi_M(B_q, \boldsymbol{\theta}_0)\} = \mathbf{0}$. For any n and $\boldsymbol{\theta} \in \Theta$, $\{\phi_M(B_q, \boldsymbol{\theta})\}_{q=1}^Q$ is a new stationary sequence. The blockwise EL (Kitamura, 1997) is defined as

$$\mathscr{L}(\boldsymbol{\theta}) = \sup\left\{\prod_{q=1}^{Q} \pi_q \middle| \pi_q > 0, \sum_{q=1}^{Q} \pi_q = 1, \sum_{q=1}^{Q} \pi_q \phi_M(B_q, \boldsymbol{\theta}) = \mathbf{0}\right\}. \tag{3}$$

Employing the routine optimization procedure for the blockwise EL leads to

$$\mathscr{L}(\boldsymbol{\theta}) = \prod_{q=1}^{Q}\left\{\frac{1}{Q} \frac{1}{1 + \widehat{\lambda}(\boldsymbol{\theta})'\phi_M(B_q, \boldsymbol{\theta})}\right\}, \tag{4}$$

where $\widehat{\lambda}(\boldsymbol{\theta})$ is a stationary point of the function $q(\lambda) = -\sum_{q=1}^Q \log\{1 + \lambda'\phi_M(B_q, \boldsymbol{\theta})\}$.

The EL estimator for $\boldsymbol{\theta}_0$ is $\widehat{\boldsymbol{\theta}}_{EL} = \arg\max_{\boldsymbol{\theta} \in \Theta} \log \mathscr{L}(\boldsymbol{\theta})$. The maximization in (3) can be carried out more efficiently by solving the corresponding dual problem, which implies that $\widehat{\boldsymbol{\theta}}_{EL}$ can be obtained as

$$\widehat{\boldsymbol{\theta}}_{EL} = \arg\min_{\boldsymbol{\theta} \in \Theta} \max_{\lambda \in \widehat{\Lambda}_n(\boldsymbol{\theta})} \sum_{q=1}^{Q} \log\{1 + \lambda'\phi_M(B_q, \boldsymbol{\theta})\}, \tag{5}$$