ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Empirical likelihood based inference for fixed effects varying coefficient panel data models

Luis A. Arteaga-Molina [a], Juan M. Rodriguez-Poo [a,b,*]

[a] Departamento de Economía, Universidad de Cantabria, Spain
[b] IEE, University of Genève, Switzerland

ARTICLE INFO

Article history:
Received 10 April 2017
Received in revised form 13 November 2017
Accepted 22 November 2017
Available online xxxx

JEL classification:
C14
C12
C23

Keywords:
Nonparametric regression analysis
Varying coefficient panel data model
Fixed effects
Empirical likelihood inference

ABSTRACT

In this paper local empirical likelihood-based inference for non-parametric varying coefficient panel data models with fixed effects is investigated. First, we show that the naive empirical likelihood ratio is asymptotically standard chi-squared when undersmoothing is employed. The ratio is self-scale invariant and the plug-in estimate of the limiting variance is not needed. Second, mean-corrected and residual-adjusted empirical likelihood ratios are proposed. The main interest of these techniques is that without undersmoothing, both also have standard chi-squared limit distributions. As a by product, we propose also two empirical maximum likelihood estimators of the varying coefficient models and their derivatives. We also obtain the asymptotic distribution of these estimators. Furthermore, a non parametric version of the Wilk's theorem is derived. To show the feasibility of the technique and to analyse its small sample properties, using empirical likelihood-based inference we implement a Monte Carlo simulation exercise and we also illustrated the proposed technique in an empirical analysis about the production efficiency of the European Union's companies.

1. Introduction

Recently nonparametric and semiparametric estimation of panel data models has attracted the attention of many researchers in econometrics. The interest to combine panel data techniques, that somehow alleviate the heterogeneity issue, with nonparametric techniques, that weaken considerably the type of assumptions that are necessary to impose in econometric models, has ended up in a vast literature that is surveyed in Su and Ullah (2011). Although the results are rather promising, it is true that the main drawbacks related to nonparametric techniques also appear when we apply them to panel data econometric models. Among others, the curse of dimensionality (e.g. Härdle (1990)) appears as one of the most important problems. In order to overcome this disadvantage varying coefficient models appear as a reasonable specification that encompasses many alternative models. As for the pure nonparametric case, estimation of varying coefficient models with random effects has been already studied in several papers (e.g. Ruckstuhl et al. (2000), Lin and Carroll (2000), Henderson and Ullah (2005), Su and Ullah (2007)). However, under the setting of fixed effects unfortunately much less results are available. In Henderson et al. (2008) direct estimation of the nonparametric components is undertaken through the use of an iterative version of a profile least squares technique. Already in a varying coefficients context a profile least squares approach is proposed in Sun et al. (2009). For differencing estimators in Rodriguez-Poo and Soberón (2014) and Rodriguez-Poo and Soberón (2015) two step backfitting estimators are proposed. Furthermore, a comparison against estimators based in profile

* Corresponding author at: Departamento de Economía, Universidad de Cantabria, Spain.
E-mail address: rodrigjm@unican.es (J.M. Rodriguez-Poo).

https://doi.org/10.1016/j.jspi.2017.11.003

Please cite this article in press as: Arteaga-Molina L.A., Rodriguez-Poo J.M., Empirical likelihood based inference for fixed effects varying coefficient panel data models. J. Statist. Plann. Inference (2017), https://doi.org/10.1016/j.jspi.2017.11.003.

least squares techniques is provided. In Cai and Li (2008) a so called nonparametric generalized method of moments is proposed to estimate the varying coefficients. Finally, in Su and Lu (2013) and Li and Liang (2015) profile least squares results are extended towards dynamic models and smooth backfitting methods are applied to estimate the unknown varying coefficients respectively. Eventually, once we have taken care of the estimation process, the next step would be to concentrate in developing inference tools for this type of models. For statistical inference such as confidence region construction or hypothesis testing the most popular techniques are normal approximations and bootstrap methods. In fact, in all above mentioned papers, asymptotic normal approximations are obtained for the different nonparametric estimators. Unfortunately it is well known that, without undersmoothing, the asymptotic distribution will exhibit a bias and a rather cumbersome expression for the variance term. Hence, if the confidence region that is derived from an asymptotic normal distribution is predetermined to be symmetric a bias correction and a plug-in estimate are needed to make the statistic scale invariant. Furthermore, if one wants to use these confidence bands as a testing device it will be necessary to obtain uniform confidence bands such as in Li et al. (2013).

In this paper, we propose to use empirical likelihood techniques to construct confidence intervals/regions. These techniques have acquired importance since they were introduced in Owen (1988, 1990) because of the advantages of this method over other methods such as normal approximation and bootstrap; for instance, empirical likelihood methods adjust to the true shape of the underlying distribution and do not require the estimation of scale, skewness (Hall and La Scala, 1990) or limiting variance as the studentization is carried out internally via optimization. Therefore, the confidence regions are reliable, range preserving and transformation respecting (Hall and La Scala, 1990). Another advantage is the method's flexibility, as it can be used when the data is incomplete, distorted or tied. Also, DiCiccio et al. (1991) have proved that empirical likelihood regions are Bartlett correctable; thus, it has advantages over the bootstrap and the jackknife methods. Finally, it combines the reliability of non-parametric methods with the effectiveness of the likelihood approach and it has good asymptotic properties and power (Owen, 1990). In fact, empirical likelihood techniques have been already applied to obtain confidence bands for longitudinal data varying coefficient models with random effects (e.g. Xue and Zhu (2007)) but unfortunately these type of results are not available for the fixed effects case. For the fixed effect case, in Zhang et al. (2011) confidence bands based in empirical likelihood techniques are derived under a partially linear model specification. They obtain, under rather restrictive assumptions, maximum empirical likelihood estimators of both parametric and nonparametric components. Furthermore, they obtain an empirical likelihood ratio that is biased if the optimal bandwidth is used.

In this paper, and starting from a fixed effects varying coefficient model, we obtain maximum empirical likelihood estimators of both the varying parameters and their derivatives. This last result is very interesting for testing constancy of parameter variation. Furthermore, we develop empirical likelihood ratios and we derive a non-parametric version of the Wilks' theorem. In order to obtain an unbiased ratio, we propose two modifications of the empirical likelihood ratio: the mean corrected and the residual adjusted empirical likelihood ratios. Based on these results, we can build up confidence regions for the parameter of interest through a standard chi squared approximation. The rest of this paper is organized as follows. In Section 2 we propose to construct the confidence bands for the unknown functions and their derivatives by using what we call a naive empirical likelihood technique. This technique shows as main drawback sub-optimal rates of convergence. In Section 3, as a byproduct, we provide two alternative maximum empirical likelihood estimators of the fixed effect nonparametric varying parameters model and their derivatives. In Section 4, and using the estimators that were previously derived, we propose two alternative techniques that enables us to obtain optimal nonparametric rates: Mean corrected and residual-adjusted empirical likelihood ratios. In Section 5 we provide a Monte Carlo experiment and in Section 6 we undertake an empirical study about the production efficiency of the European Union's companies. Finally Section 7 concludes. The proofs of the main results are collected in Appendix A.

2. Naive empirical likelihood

Consider the following varying coefficient panel data regression model

$$Y_{it} = X_{it}^{\top} m(Z_{it}) + \mu_i + v_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{1}$$

where Y_{it} is the response, Z_{it} and X_{it} are vectors of covariates of dimension q and d respectively, and $m(z) = (m_1(z), \ldots, m_d(z))^{\top}$ is a $d \times 1$ vector of unknown functions; here μ_i stands for heterogeneity of unknown form, that is, individual characteristic that are not observed, and v_{it} are random errors that do variate along time and across individuals. On this econometric model we impose the following standard assumptions,

Assumption 2.1. Let $(Y_{it}, X_{it}, Z_{it})_{i=1,\ldots,N;\ t=1,\ldots,T}$ be a set of independent and identically distributed (i.i.d.) $\mathbb{R}^{d+q+1}$ random variables in the subscript i for each fixed t and strictly stationary over t for a fixed i.

Assumption 2.2. The random errors v_{it} are independent and identically distributed, with 0 mean and homoscedastic variance $\sigma_v^2 < \infty$. They are also independent of X_{it} and Z_{it} for all i and t. Furthermore, $E|v_{it}|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption 2.3. Let μ_i can be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.