Economics Letters 120 (2013) 437–441

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

On a general class of long run variance estimators[☆,☆☆]

Xianyang Zhang [*], Xiaofeng Shao

University of Missouri-Columbia, United States
University of Illinois at Urbana-Champaign, United States

HIGHLIGHTS

- We propose a general class of LRV estimators in the GMM framework.
- The LRV estimator includes some recently developed estimators as special cases.
- First order asymptotics of the Wald statistics based on general LRV estimators.

ARTICLE INFO

Article history:
Received 12 April 2013
Accepted 30 May 2013
Available online 10 June 2013

JEL classification:
C12
C13

Keywords:
Fixed-smoothing asymptotics
Generalized method of moments
Long run variance matrix

ABSTRACT

This note proposes a class of estimators for estimating the asymptotic covariance matrix of the generalized method of moments (GMM) estimator in the stationary time series models. The proposed estimator is general enough to include the traditional heteroskedasticity and autocorrelation consistent (HAC) covariance estimator and some recently developed estimators, such as the cluster covariance estimator and projection-based covariance estimator, as special cases. We also study the first order asymptotics of the Wald statistics based on the general covariance estimators when the underlying smoothing parameter is held fixed.

1. Introduction

In stationary time series models, the asymptotic covariance matrix of the generalized method of moments (GMM) estimator is usually estimated nonparametrically by the kernel-based methods, where the bandwidth parameter is assumed to grow slowly with the sample size in the asymptotics (see Newey and West, 1987; Andrews, 1991). Recent studies on heteroskedasticity and autocorrelation consistent (HAC) based robust inference have developed alternative first order asymptotic theory (as compared to the traditional χ^2-based approximation), which was shown to provide more accurate approximation to the sampling distributions of the associated test statistics. For example, Kiefer and Vogelsang (2005, KV, hereafter) developed a first order asymptotic theory where the proportion of the bandwidth involved in the HAC estimator to the sample size T, denoted as b, is held fixed in the asymptotics. Using the higher-order Edgeworth expansions, Jansson (2004), Sun et al. (2008), Sun (2010) and Zhang and Shao (forthcoming) rigorously proved that the fixed-b asymptotics provides a high order refinement over the traditional small-b asymptotics in the Gaussian location model. Sun (2013) developed a procedure for hypothesis testing in time series models by using the nonparametric series method. The basic idea is to project the time series onto a space spanned by a set of Fourier basis functions (see Phillips, 2005, for an early development) and construct the covariance matrix estimator based on the projection vectors with the number of basis functions held fixed. Also see Sun (2011) for the use of a similar idea in the inference of the trend regression models. Ibragimov and Müller (2010) proposed a subsampling based t-statistic for robust inference where the unknown dependence structure can be in the temporal, spatial or other forms. In their paper, the number of non-overlapping blocks is held fixed. The t-statistic based approach was extended by Bester et al. (2011) to the inference of spatial and panel data with group structure. In the context of misspecification testing, Chen and Qu (forthcoming) proposed a modified M test of Kuan and Lee (2006) which

☆☆ This note is drawn from an early working paper entitled "Fixed-smoothing asymptotics for time series" (arXiv:1204.4228). A substantial part of the working paper appeared in Zhang and Shao (forthcoming), which has little overlap with this note.

* Correspondence to: Department of Statistics, University of Missouri-Columbia, Columbia, MO 65211, United States. Tel.: +1 573 772 1577; fax: +1 573 884 5524.
E-mail address: zhangxianyang@gmail.com (X. Zhang).

0165-1765/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2013.05.026

involves dividing the full sample into several recursive subsamples and constructing a normalization matrix based on them. In the statistical literature, Shao (2010) developed the self-normalized approach to inference for time series data that uses an inconsistent long run variance (LRV) estimator based on recursive subsample estimates. The self-normalized method is an extension of Lobato (2001) from the sample autocovariances to more general approximately linear statistics and it coincides with KV's fixed-b approach in the inference of the mean of a stationary time series by using the Bartlett kernel and letting $b = 1$. Although the above inference procedures are proposed in different settings and for different problems and data structure, they share a common feature in the sense that the underlying smoothing parameters in the asymptotic covariance matrix estimator such as the number of basis functions, the number of cluster groups and the number of recursive subsamples, play a similar role as the bandwidth in the HAC estimator.

The goal of this note is to introduce a general class of estimators for estimating the LRV matrix in the inference of stationary time series models estimated by GMM. Our proposal includes the traditional lag window type (or HAC) covariance estimator, the projection-based covariance estimator, the cluster-based covariance estimator and the blockwise recursive subsampling-based covariance estimator as special cases. The general covariance estimator considered here involves projecting the original data onto a space spanned by a sequence of basis functions (not necessarily orthogonal), where the number of basis functions K plays a key role in determining asymptotic properties of the estimator. Under the fixed-K asymptotics, we show that the Wald statistic based on the general LRV estimator converges to an (approximate) F distribution with a scale constant depending only on K and the number of restrictions being tested. Thus our result provides a unification of the various recently proposed fixed-smoothing inference procedures in the first order sense.

We introduce some notation. Denote by $\lfloor a \rfloor$ the integer part of a real number a. Let $L^2[0, 1]$ be the space of square integrable functions on $[0, 1]$. Denote by $D[0, 1]$ the space of functions on $[0, 1]$ which are right continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Denote by "$\Rightarrow$" weak convergence in the $\mathbb{R}^{q_0}$-valued function space $D^{q_0}[0, 1]$, where $q_0 \in \mathbb{N}$. Define "$\rightarrow^d$" convergence in distribution. We use "$\otimes$" to denote the Kronecker product in matrix algebra. The notation $N(\mu, \Sigma)$ is used to denote the multivariate normal distribution with mean μ and covariance Σ. Let χ^2_k be a random variable following χ^2 distribution with k degrees of freedom.

2. Basic setup and assumptions

In linear and nonlinear models with moment conditions, it is standard to employ GMM to estimate the model parameters. We follow the GMM setup as described in KV. Consider a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$ of interest, where Θ is the parameter space. Denote θ_0 the true parameter of θ which is an interior point of Θ. Let y_t denote a vector of observed data and assume the moment conditions

$$E[f(y_t, \theta)] = 0, \quad t = 1, 2, \ldots, T \tag{1}$$

hold if and only if $\theta = \theta_0$, where $f(\cdot)$ is $m \times 1$ vector of functions with $m \geq d$ and $\text{rank}(E[\partial f(y_t, \theta_0)/\partial\theta']) = d$. When $m > d$, the parameter θ is over-identified with the degree of over-identification $v = m - d$. Define the partial sum $g_t(\theta) = T^{-1}\sum_{j=1}^t f(y_j, \theta)$. Then the GMM estimator of θ_0 is given by

$$\hat{\theta}_T = \underset{\theta\in\Theta}{\text{argmin}}\, g_T(\theta)' W_T g_T(\theta), \tag{2}$$

where W_T is a $m \times m$ semi-positive definite weighting matrix. Further define

$$G_t(\theta) = (G_{t1}(\theta), \ldots, G_{tm}(\theta))' = \frac{\partial g_t(\theta)}{\partial\theta'} = \frac{1}{T}\sum_{j=1}^{t}\frac{\partial f(y_j, \theta)}{\partial\theta'}.$$

Using the mean value theorem for each element of g_T, we have $g_T(\hat{\theta}_T) = g_T(\theta_0) + \tilde{G}_T(\hat{\theta}_T - \theta_0)$, where $\tilde{G}_T = (G_{T1}(\tilde{\theta}_{T1}), \ldots, G_{Tm}(\tilde{\theta}_{Tm}))'$ and $\tilde{\theta}_{Tj}$ is between θ_0 and $\hat{\theta}_T$ for each $1 \leq j \leq m$. Note that $G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0$ by the first order condition, which implies that

$$G_T(\hat{\theta}_T)' W_T g_T(\theta_0) + G_T(\hat{\theta}_T)' W_T \tilde{G}_T(\hat{\theta}_T - \theta_0)$$
$$= G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0.$$

Solving the above equation, we have

$$T^{1/2}(\hat{\theta}_T - \theta_0) = -(G_T(\hat{\theta}_T)' W_T \tilde{G}_T)^{-1} G_T(\hat{\theta}_T)' W_T (T^{1/2} g_T(\theta_0)).$$

To derive the asymptotic distribution of $\hat{\theta}_T$, we make the following high-level assumptions as KV and Sun (2010).

Assumption 2.1. $\hat{\theta}_T \rightarrow^p \theta_0$.

Assumption 2.2. $T^{1/2} g_{\lfloor Tr \rfloor}(\theta_0) \Rightarrow \Delta W_m(r)$ where

$$\Delta\Delta' = \Omega = \sum_{j=-\infty}^{+\infty} E[f(y_t, \theta_0) f(y_{t-j}, \theta_0)'],$$

and $W_m(r)$ is a m—dimensional vector of independent standard Brownian motions.

Assumption 2.3. $\tilde{G}_T \rightarrow^p G_0$ uniformly for all $\tilde{\theta}_{Tj}$ between $\hat{\theta}_T$ and θ_0, where $G_0 = E[\partial f(y_j, \theta_0)/\partial\theta']$ and $1 \leq j \leq m$.

Assumption 2.4. The weighting matrix W_T is symmetric and semi-positive definite such that $W_T \rightarrow^p W_0$ and $G_0' W_0 G_0$ is positive definite.

Under Assumptions 2.1–2.4, it is easy to see that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \rightarrow^d -(G_0' W_0 G_0)^{-1} G_0' W_0 \Delta W_m(1) =^d N(0, V_0),$$

where "$=^d$" denotes "equal in distribution" and the asymptotic covariance matrix $V_0 := (G_0' W_0 G_0)^{-1} G_0' W_0 \Omega W_0 G_0 (G_0' W_0 G_0)^{-1}$. To make inference on θ_0, we have to estimate G_0, W_0 and the LRV matrix Ω. Under the above assumptions, G_0 and W_0 can be consistently estimated by their sample counterparts $G_T(\hat{\theta}_T)$ and W_T respectively. It remains to estimate the LRV matrix Ω. In the next section, we introduce a general class of estimators for Ω and V_0.

3. LRV estimators

To present the idea, we focus on the hypothesis testing problem that $H_0 : r(\theta_0) = 0$ versus the alternative that $H_a : r(\theta_0) \neq 0$, where $r(\theta)$ is a $p \times 1$ continuously differentiable function with the first order derivative matrix $R(\theta) = \partial r(\theta)/\partial\theta'$ and $p \leq d$. Let

$$\hat{V}_T = (G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1}$$
$$\times (G_T(\hat{\theta}_T)' W_T \hat{\Omega}_T W_T G_T(\hat{\theta}_T))(G_T(\hat{\theta}_T)' W_T G_T(\hat{\theta}_T))^{-1},$$

be an estimator of V_0, where $\hat{\Omega}_T$ is the LRV estimate of Ω. The Wald statistic for testing H_0 against H_a is defined as

$$F_T = T r(\hat{\theta}_T)' \hat{D}_T^{-1} r(\hat{\theta}_T)/p, \tag{3}$$

where $\hat{D}_T = R(\hat{\theta}_T)\hat{V}_T R(\hat{\theta}_T)'$. The widely used lag window type LRV estimator is given by

$$\hat{\Omega}_T = \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathcal{K}\left(\frac{i-j}{bT}\right) f(y_i, \hat{\theta}_T) f(y_j, \hat{\theta}_T)', \tag{4}$$