Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 597–620
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Asymptotic properties of a robust variance matrix estimator for panel data when T is large

Christian B. Hansen

University of Chicago, Graduate School of Business, 5807 South Woodlawn Ave., Chicago, IL 60637, USA

Available online 20 November 2006

Abstract

I consider the asymptotic properties of a commonly advocated covariance matrix estimator for panel data. Under asymptotics where the cross-section dimension, n, grows large with the time dimension, T, fixed, the estimator is consistent while allowing essentially arbitrary correlation within each individual. However, many panel data sets have a non-negligible time dimension. I extend the usual analysis to cases where n and T go to infinity jointly and where $T \to \infty$ with n fixed. I provide conditions under which t and F statistics based on the covariance matrix estimator provide valid inference and illustrate the properties of the estimator in a simulation study.

JEL classification: C12; C13; C23

Keywords: Panel; Heteroskedasticity; Autocorrelation; Robust; Covariance matrix

1. Introduction

The use of heteroskedasticity robust covariance matrix estimators, cf. White (1980), in cross-sectional settings and of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators, cf. Andrews (1991), in time series contexts is extremely common in applied econometrics. The popularity of these robust covariance matrix estimators is due to their consistency under weak functional form assumptions. In particular, their use allows the researcher to form valid confidence regions about a set of parameters from a model of interest without specifying an exact process for the disturbances in the model.

E-mail address: chansen1@chicagoGSB.edu.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.10.009

With the increasing availability of panel data, it is natural that the use of robust covariance matrix estimators for panel data settings that allow for arbitrary within individual correlation are becoming more common. A recent paper by Bertrand et al. (2004) illustrated the pitfalls of ignoring serial correlation in panel data, finding through a simulation study that inference procedures which fail to account for within individual serial correlation may be severely size distorted. As a potential resolution of this problem, Bertrand et al. (2004) suggest the use of a robust covariance matrix estimator proposed by Arellano (1987) and explored in Kezdi (2002) which allows arbitrary within individual correlation and find in a simulation study that tests based on this estimator of the covariance parameters have correct size.

One drawback of the estimator of Arellano (1987), hereafter referred to as the "clustered" covariance matrix (CCM) estimator, is that its properties are only known in conventional panel asymptotics as the cross-section dimension, n, increases with the time dimension, T, fixed. While many panel data sets are indeed characterized by large n and relatively small T, this is not necessarily the case. For example, in many differences-in-differences and policy evaluation studies, the cross-section is composed of states and the time dimension of yearly or quarterly (or occasionally monthly) observations on each state for 20 or more years.

In this paper, I address this issue by exploring the theoretical properties of the CCM estimator in asymptotics that allow n and T to go to infinity jointly and in asymptotics where T goes to infinity with n fixed. I find that the CCM estimator, appropriately normalized, is consistent without imposing any conditions on the rate of growth of T relative to n even when the time series dependence between the observations within each individual is left unrestricted. In this case, both the OLS estimator and the CCM estimator converge at only the $\sqrt{n}$-rate, essentially because the only information is coming from cross-sectional variation. If the time series process is restricted to be strongly mixing, I show that the OLS estimator is $\sqrt{nT}$-consistent but that, because high lags are not down weighted, the robust covariance matrix estimator still converges at only the $\sqrt{n}$-rate. This behavior suggests, as indicated in the simulations found in Kezdi (2002), that it is the n dimension and not the size of n relative to T that matters for determining the properties of the CCM estimator.

It is interesting to note that the limiting behavior of $\widehat{\beta}$ changes "discontinuously" as the amount of dependence is limited. In particular, the rate of convergence of $\widehat{\beta}$ changes from $\sqrt{n}$ in the "no-mixing case" to $\sqrt{nT}$ when mixing is imposed. However, despite the difference in the limiting behavior of $\widehat{\beta}$, there is no difference in the behavior of standard inference procedures based on the CCM estimator between the two cases. In particular, the same t and F statistics will be valid in either case (and in the $n \to \infty$ with T fixed case) without reference to the asymptotics or degree of dependence in the data.

I also derive the behavior of the CCM estimator as $T \to \infty$ with n fixed, where I find the estimator is not consistent but does have a limiting distribution. This result corresponds to asymptotic results for HAC estimators without truncation found in recent work by Kiefer and Vogelsang (2002, 2005), Phillips et al. (2003), and Vogelsang (2003). While the limiting distribution is not proportional to the true covariance matrix in general, it is proportional to the covariance matrix in the important special case of iid data across individuals,[1]

[1]Note that this still allows arbitrary correlation and heteroskedasticity within individuals, but restricts that the pattern is the same across individuals.