ARTICLE IN PRESS

Economic Modelling xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Economic Modelling

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

Asymptotic collinearity in CCE estimation of interactive effects models☆

Joakim Westerlund[a,*], Yana Petrova[b]

[a] Lund University and Centre for Financial Econometrics, Deakin University, Australia
[b] Lund University, Sweden

ARTICLE INFO

JEL classification:
C12
C13
C33
C36

Keywords:
CCE estimation
Interactive effects
Asymptotic collinearity
Singular signal matrix

ABSTRACT

Researchers sometime fall into the dummy variable trap. A typical scenario in panel data is when wanting to estimate the effect of a regressor that is time invariant, such as sex or race, and accidentally including cross-section specific fixed effects. The problem here is that the fixed effects and the regressor are collinear, which causes the resulting pooled least squares estimator to break down. In interactive effects models such breakdowns can occur even if the regressors are not time invariant. The reason is that the interactive effects are flexible enough to generate a wide range of behaviours that are likely to be shared by the regressors. The current paper considers the challenging case when some of the regressors are asymptotically collinear with the interactive effects. The relevant asymptotic theory is developed and tested in small samples using both simulated and real data.

1. Introduction

Consider the scalar and $k \times 1$ vector of panel data variables $y_{i,t}$ and $\mathbf{x}_{i,t}$, respectively, observable across $t = 1, \ldots, T$ time periods and $i = 1, \ldots, N$ cross-section units. The use of such panel data variables has attracted considerable attention in the empirical economic literature. A major reason for this is the ability to deal with the presence of unobserved heterogeneity in $y_{i,t}$, and the bias that this causes when said heterogeneity is correlated with the regressors in $\mathbf{x}_{i,t}$. The standard approach in the literature is to assume that the unobserved heterogeneity is made up of additive cross-section and time specific constants, or fixed effects (FE), whose effect can be eliminated through demeaning. Because of the demeaning, the resulting pooled ordinary least squares (POLS) is unable to estimate the effect of regressors that are cross-section and/or time invariant. Formally, the FE are collinear with the invariant regressors, which means that after demeaning said regressors are zero. The signal matrix is therefore singular, causing the POLS estimator to break down. This is a well known problem that is typically referred to as the "dummy variable trap".

The most common way to circumvent the dummy variable trap is to simply remove either the problematic regressors or the FE. Which one to remove depends on the purpose of the study. If the purpose is to estimate the effect of the invariant regressors, the FE are removed, whereas if the invariant regressors are not the main focus of the study, then the opposite is generally preferable. The ability to choose is very useful and is facilitated by the known additive structure of the FE.

Of course, in most scenarios of empirical relevance, standard FE are unlikely to be enough, and it is not difficult to find empirical evidence that confirms this. Observations like this have recently led to the consideration of interactive effects (IE) models, in which both cross-section and time specific FE enter in an multiplicative way. The multiplicative form captures the unobserved heterogeneity more flexibly than additive FE, since it allows multiple time specific factors with cross-section specific loadings. This flexibility is the main attraction of IE models.

A common estimation approach to IE models is to first estimate the unknown factors, and then to estimate the effect of the regressors by applying POLS to the defactored variables. Two approaches based on this idea can be identified, which differ mainly in how the factors are estimated in the first step. These are the common correlated effects (CCE) approach of Pesaran (2006), and the principal components (PC) approach of Bai (2009). Both approaches have attracted considerable interest, so much that there is by now a separate literature devoted to them (see Chudik and Pesaran, 2015, for a recent survey). The bulk of the evidence suggests the CCE approach tends to work best in small samples, while at the same time being computationally relatively more convenient. The approach is also very robust, and can be employed under very general conditions on the IE. In fact, save for some mild regulatory conditions, the IE are essentially unrestricted. CCE is therefore very appealing from an applied point of view; hence, our

☆ Westerlund would like to thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship, and the Jan Wallander and Tom Hedelius Foundation for financial support under research grant numbers P2014–0112:1 and W2016–0467:1.

* Correspondence to: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden.
E-mail address: joakim.westerlund@nek.lu.se (J. Westerlund).

http://dx.doi.org/10.1016/j.econmod.2017.07.023
Received 29 July 2017; Accepted 31 July 2017
0264-9993/ © 2017 Elsevier B.V. All rights reserved.

interest in it.

The purpose of the present paper is to point to a problem with IE models that seems to have gone largely unnoticed in the CCE branch of the literature (see Bai, 2009; Moon and Weidner, 2017for discussions in the PC case). In particular, while an advantage in many regards, the generality of the IE also increases the probability of falling into the dummy variable trap. This is intuitive; as the flexibility of the IE increases, so does their ability to mimic the behaviour of the regressors. In analogy to the above discussion of the problem of demeaning in the FE case, the defactoring exhaust too much variation, causing the signal matrix to become singular. What is more, unlike with FE, the IE are estimated unrestrictedly from the data, which means that the possibility to selectively choose which effects to include is lost. Hence, in IE models it is generally not possible to estimate the effect of regressors that are cross-section and/or time invariant.

In Section 2, we present the model that we will be considering, which can be seen as a collinear regressor extension of the model of Pesaran (2006). The extension amounts to allowing the idiosyncratic component of some of the regressors to go to zero at rate κ. In the limit as $\kappa \to 0$, the variation of the regressors is driven entirely by the IE, which makes the signal matrix of the CCE estimator singular. Section 3 reports our asymptotic results, which are based on letting $N,\ T \to \infty$ and $\kappa \to 0$. What we find is that provided that the rate of shrinking of κ is not too fast relative to the rate at which N and T expand, the CCE estimator is consistent, although generally not asymptotically normal. In spite of this, however, the CCE-based Wald and t-ratio statistics are still asymptotically chi-squared and normal, respectively. This last result is very useful because it means that empirical researchers may proceed as if the estimator is in fact both consistent and asymptotically normal. In Sections 4 and 5, we use both simulated and real data to investigate the accuracy of our asymptotic theory in small samples. Section 6 concludes.

2. Model and assumptions

The IE model considered in the present paper is very similar to the one in Pesaran (2006), and is given by

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\Gamma}_i + \mathbf{V}_i, \tag{2}$$

where $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,T}]'$ is $T \times 1$, $\mathbf{X}_i = [\mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,T}]'$ is $T \times k$, βis $k \times 1$, $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_T]'$ is a $T \times m$ matrix of unobservable common factors with γ_i and Γ_i being the associated $m \times 1$ and $m{\times}k$ matrices of factor loadings, and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i,1}, \dots, \varepsilon_{i,T}]'$ and $\mathbf{V}_i = [\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,T}]'$ are $T \times 1$ and $T{\times}k$ matrices of idiosyncratic errors. The IE are here given by $\mathbf{F}\boldsymbol{\gamma}_i$ and $\mathbf{F}\boldsymbol{\Gamma}_i$. The main difference when compared to Pesaran (2006) is how $\mathbf{V}_i$ in (2) is modeled. Let us introduce a full rank $k{\times}k$ matrix $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2]$, where $\mathbf{G}_1$ and $\mathbf{G}_2$ are $k \times k_1$ and $k \times k_2$, respectively, with $k_2 = k - k_1$. The matrix $\mathbf{G}$rotates the columns in $\mathbf{V}_i$ into k_2 columns that are shrinking to zero and k_1 columns that are not shrinking. Specifically,

$$\mathbf{V}_i\mathbf{G} = [\mathbf{V}_i\mathbf{G}_1, \mathbf{V}_i\mathbf{G}_2] = [\mathbf{E}_{1,i}, \kappa\mathbf{E}_{2,i}], \tag{3}$$

where $\mathbf{E}_{1,i}$ and $\mathbf{E}_{2,i}$ are $T \times k_1$ and $T \times k_2$, respectively. Let $\mathbf{E}_i = [\mathbf{E}_{1,i}, \mathbf{E}_{2,i}] = [\mathbf{e}_{i,1}, \dots, \mathbf{e}_{i,T}]'$, where $\mathbf{e}_{i,t}$ is a $k \times 1$ vector of idiosyncratic errors. Assumption ERR below, which is similar to Assumption 2 in Pesaran (2006), puts restrictions on $\varepsilon_{i,t}$ and $\mathbf{e}_{i,t}$.

Assumption ERR.

(i) $\varepsilon_{i,t}$ and $\mathbf{e}_{i,t}$ are linear stationary processes with absolutely summable autocovariances, zero mean, finite fourth-order moments, $E(\varepsilon_{i,t}^2) = \sigma_{\varepsilon_i}^2$ and $E(\mathbf{e}_{i,t}\mathbf{e}'_{i,t}) = \boldsymbol{\Sigma}_{\mathbf{e}_i}$ for all t, $\sigma_{\varepsilon}^2 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \sigma_{\varepsilon_i}^2 > 0$ and $\boldsymbol{\Sigma}_{\mathbf{e}} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \boldsymbol{\Sigma}_{\mathbf{e}_i}$ positive definite.

(ii) $\varepsilon_{i,t}$ and $\mathbf{e}_{j,s}$ are mutually independent for all i, j, t and s.

The scalar κ is key and measures the contribution of the idiosyncratic errors to $\mathbf{X}_i\mathbf{G}_2$, as seen by writing

$$\mathbf{X}_i\mathbf{G} = [\mathbf{X}_i\mathbf{G}_1, \mathbf{X}_i\mathbf{G}_2] = [\mathbf{X}_{1,i}, \mathbf{X}_{2,i}] = [\mathbf{F}\boldsymbol{\Gamma}_i\mathbf{G}_1 + \mathbf{E}_{1,i}, \mathbf{F}\boldsymbol{\Gamma}_i\mathbf{G}_2 + \kappa\mathbf{E}_{2,i}]. \tag{4}$$

In order to appreciate the effect of κ, it is useful to decompose $\mathbf{G}^{-1}$ as $\mathbf{G}^{-1} = [(\mathbf{G}^1)', (\mathbf{G}^2)']'$, where $\mathbf{G}^1$ and $\mathbf{G}^2$ are $k_1 \times k$ and $k_2 \times k$ matrices, respectively. Therefore, letting $\mathbf{G}^{-1}\boldsymbol{\beta} = [(\mathbf{G}^1\boldsymbol{\beta})', (\mathbf{G}^2\boldsymbol{\beta})']' = [\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2']' = \boldsymbol{\alpha}$, we have

$$\mathbf{X}_i\boldsymbol{\beta} = \mathbf{X}_i\mathbf{G}\mathbf{G}^{-1}\boldsymbol{\beta} = \mathbf{X}_{1,i}\boldsymbol{\alpha}_1 + \mathbf{X}_{2,i}\boldsymbol{\alpha}_2, \tag{5}$$

such that $\mathbf{M_F}\mathbf{X}_i\boldsymbol{\beta} = \mathbf{M_F}\mathbf{E}_{1,i}\boldsymbol{\alpha}_1 + \kappa\mathbf{M_F}\mathbf{E}_{2,i}\boldsymbol{\alpha}_2$, which in turn implies that

$$\mathbf{M_F}\mathbf{y}_i = \mathbf{M_F}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{M_F}\boldsymbol{\varepsilon}_i = \mathbf{M_F}\mathbf{E}_{1,i}\boldsymbol{\alpha}_1 + \kappa\mathbf{M_F}\mathbf{E}_{2,i}\boldsymbol{\alpha}_2 + \mathbf{M_F}\boldsymbol{\varepsilon}_i, \tag{6}$$

where $\mathbf{M_A} = \mathbf{I}_T - \mathbf{P_A} = \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'$ for any T-rowed matrix $\mathbf{A}$with $(\mathbf{A}'\mathbf{A})^{+}$ being the Moore–Penrose inverse of $\mathbf{A}'\mathbf{A}$. Suppose that $\mathbf{F}$is known. If $\kappa = 1$, then $\mathbf{E}_{1,i}\boldsymbol{\alpha}_1 + \kappa\mathbf{E}_{2,i}\boldsymbol{\alpha}_2 = \mathbf{E}_i\boldsymbol{\alpha} = \mathbf{V}_i\boldsymbol{\beta}$. Hence, provided that $(NT)^{-1}\sum_{i=1}^N \mathbf{G}'\mathbf{V}_i'\mathbf{M_F}\mathbf{V}_i\mathbf{G} = (NT)^{-1}\sum_{i=1}^N \mathbf{E}_i'\mathbf{M_F}\mathbf{E}_i$ converges to a positive definite matrix, βcan be estimated from a POLS regression of $\mathbf{M_F}\mathbf{y}_i$ onto $\mathbf{M_F}\mathbf{X}_i$. If, however, $\kappa \to 0$, then $\kappa\mathbf{M_F}\mathbf{E}_{2,i}\boldsymbol{\alpha}_2$ vanishes, which means that βis no longer estimable. Moreover, $(NT)^{-1}\sum_{i=1}^N \mathbf{G}'\mathbf{V}_i'\mathbf{M_F}\mathbf{V}_i\mathbf{G}$ is asymptotically singular. This scenario is very different from the "low-rank regressor" case considered by Bai (2009), and Moon and Weidner (2017) in the PC case, in which $(NT)^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{M_F}\mathbf{X}$ is assumed to be positive definite (with probability one).

The problem that arises when $\kappa \to 0$ is that $\mathbf{X}_{2,i}$ becomes collinear with $\mathbf{F}$in (1). The last k_2 columns of $\mathbf{M_F}\mathbf{X}_i\mathbf{G}$ are therefore asymptotically zero, which means that there is no variation left for the estimation of α_2. This causes the POLS estimator to break down. The probability of break-down depends to a large extent on the assumptions placed on $\mathbf{F}$. The more general is $\mathbf{F}$, the more likely it is that the projection onto space spanned by $\mathbf{F}$will exhaust the variation in $\mathbf{X}_i$. A standard assumption in the literature on factor-augmented regression models is that $T^{-1}\mathbf{F}'\mathbf{F}$ converges to a positive definite matrix (see, for example, Bai, 2009; Greenaway-McGrevy et al., 2012; Moon and Weidner, 2017; Pesaran, 2006), which rules out many empirically relevant cases, such as when $\mathbf{F}$is trending. However, as pointed out by Westerlund (2017), the CCE estimator is actually valid under much more general conditions. In fact, the only thing we need is that there exists an $m{\times}m$ diagonal normalization matrix $\mathbf{D}_T$ such that $\mathbf{D}_T^{-1}\mathbf{f}_t$ has certain moments.

Assumption F. Consider the $m{\times}m$ matrix $\mathbf{D}_T = \mathrm{diag}(T^{p_1}, \dots, T^{p_m})$ with $p_j \geq 1/2$ for all j. The following is assumed:

(i) $\lim_{T\to\infty} E(\| \mathbf{D}_T^{-1}\mathbf{F}'\mathbf{F}\mathbf{D}_T^{-1} - \boldsymbol{\Sigma}_\mathbf{F} \|^2) = 0$, where $\mathrm{rank}(\boldsymbol{\Sigma}_\mathbf{F}) \overset{a.s.}{=} m$ and $E(\| \boldsymbol{\Sigma}_\mathbf{F} \|^2) < \infty$, where $\mathrm{rank}\,\mathbf{A}$ and $\| \mathbf{A} \| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$ denote the trace and the Frobenius (Euclidean) norm of the matrix $\mathbf{A}$, respectively.

(ii) $\lim_{N,\,T\to\infty} E(\| \sqrt{N}\mathbf{D}_T^{-1}\mathbf{F}'\bar{\mathbf{a}} \|^2) < \infty$ and $\lim_{T\to\infty} E(\| \mathbf{D}_T^{-1}\mathbf{F}'\mathbf{a}_i \|^2) < \infty$ for all i, where $\mathbf{a}_i \in \{\mathbf{E}_i, \boldsymbol{\varepsilon}_i\}$.

Because of the asymptotic collinearity, the estimation of βis clearly a nontrivial issue. The problem becomes even more interesting if we in addition assume that $\mathbf{F}$is unknown. The reason is that if $\mathbf{F}$is unknown, then there is a problem of how to control the endogeneity caused by the presence of $\mathbf{F}$in both (1) and (2). However, by combining the two equations, we have

$$\mathbf{Z}_i = \mathbf{F}\mathbf{C}_i + \mathbf{U}_i, \tag{7}$$

where $\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i] = [\mathbf{z}_{i,1}, \dots, \mathbf{z}_{i,T}]'$ is $T \times (k+1)$, $\mathbf{z}_{i,t} = [y_{i,t}, \mathbf{x}_{i,t}']'$ is $(k+1) \times 1$, $\mathbf{C}_i = [\boldsymbol{\Gamma}_i\boldsymbol{\beta} + \boldsymbol{\gamma}_i, \boldsymbol{\Gamma}_i]$ is $m \times (k+1)$, and $\mathbf{U}_i = [\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,T}]' = [\mathbf{V}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \mathbf{V}_i]$ is $T \times (k+1)$. Thus, the model in (1) and (2) can be rewritten equivalently as a static factor model for $\mathbf{Z}_i$, which is convenient because it means that $\mathbf{F}$can be estimated using existing approaches for such models. In CCE the estimator of $\mathbf{F}$is particularly simple, and is given by

$$\hat{\mathbf{F}} = \overline{\mathbf{Z}}, \tag{8}$$

Download English Version:

https://daneshyari.com/en/article/7347252

Download Persian Version:

https://daneshyari.com/article/7347252

Daneshyari.com