Queen's Economics Department Working Paper No. 1491

TESTING FACTORS IN CCE

Nicholas Brown
Queen's University

Joakim Westerlund
Lund University and Deakin University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

10-2022

TESTING FACTORS IN CCE*

Nicholas Brown
Queen's University

Joakim Westerlund†
Lund University
and
Deakin University

October 17, 2022

Abstract

One of the most popular estimators of interactive effects panel data models is the common correlated effects (CCE) approach, which uses the cross-sectional averages of the observables as proxies of the unobserved factors. The present paper proposes a simple test that is suitable for testing hypotheses about the factors in CCE and that is valid provided only that the number of cross-sectional units is large. The new test can be used to test if a subset of the averages is enough to proxy the factors, or if there are observable variables that capture the factors. The test can also be used sequentially to determine the smallest set of averages needed to proxy the factors.

Keywords: Factor model selection; Interactive effects models; CCE estimation.

1 Introduction

Consider the stacked $T \times 1$ vector $\mathbf{y}_i$, observable for $i = 1, ..., N$ cross-sectional units. The model considered for this variable is the same as in Pesaran (2006), and is given by

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

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

*Westerlund would like to thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship.

†Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.

where $\mathbf{X}_i$ is a $T \times k$ matrix of observable regressors, $\boldsymbol{\beta}_i$ is a $k \times 1$ vector of slope coefficients, $\mathbf{F}$ is a $T \times m_0$ vector of unobservable common factors, γ_i and $\boldsymbol{\Gamma}_i$ are $m_0 \times 1$ and $m_0 \times k$ matrices, respectively, of factor loadings, and $\boldsymbol{\varepsilon}_i$ and $\mathbf{V}_i$ are $T \times 1$ and $T \times k$ matrices, respectively, of idiosyncratic errors. The $k \times 1$ vector $\boldsymbol{\beta}_i$ is assumed to be randomly distributed across i with a common mean $\boldsymbol{\beta}$, which is the main object of interest. The products $\mathbf{F}\gamma_i$ and $\mathbf{F}\boldsymbol{\Gamma}_i$ are referred to as "interactive effects", because they interact factors, or time effects, with loadings, or individual effects. Because these effects enter both (1) and (2), $\mathbf{X}_i$ is endogenous, and therefore estimation of (1) is nontrivial.

While there are other approaches that can be used to obtain consistent estimates of $\boldsymbol{\beta}$, in the present paper we focus on the common correlated effects (CCE) approach of Pesaran (2006), which is one of the simplest and most popular approaches around. The idea in this approach is to use the cross-sectional averages of all the observables to estimate (the space spanned by) $\mathbf{F}$ and to estimate $\boldsymbol{\beta}$ by ordinary least squares (OLS) conditional on those averages. However, while certainly simple, the use of all available averages is likely excessive. As in classical model selection problems, removal of redundant factor estimates is expected to reduce noise and increase precision (see Brown et al., 2022, for a formal proof). Moreover, many economic models have a natural role for factors and thus determining which factor estimates that matter is of interest in its own right. Because of these and other reasons, the problem of how to determine which factor estimates to retain in factor models has attracted considerable interest, so much so that it has given rise to a separate strand of literature. Most of this literature supposes that the panel data set is large in the sense that both N and T should be allowed to grow without bound (see, for example, Bai, 2009, Moon and Weidner, 2015, and Pesaran, 2006).[1] Taking advantage of the largeness of the sample, many techniques are available for determining which factors to include. Prominent examples are information criteria, such as those developed by Bai and Ng (2002), and ratios of eigenvalues, such as those of Ahn and Horenstein (2013).

One problem with the approaches just described is that the large-N, large-T condition is rarely met in practice. Take the so-called "wage curve" model (Blanchflower and Oswald,

[1] See Ahn et al. (2013), and Robertson and Sarafidis (2015) for some exceptions in the generalized method of moments (GMM) context.

1994), which relate worker's wages to the rate of unemployment, and is one of the most common motivating example of interactive effects models in the literature (see, for example, Ahn et al., 2013, and Robertson and Sarafidis, 2015). Here the factor loadings may represent workers' unobservable skills and the factors would represent the price of these skills. The problem with this example, as with many other common empirical examples in the literature (see, for example, Ahn et al., 2013), is that the data employed are almost always of the micro type. That is, while N is large, T is not, which means that many existing approaches are not really applicable.

Another problem is that while there are many studies that propose consistent factor selection procedures, in practice there is often considerable disagreement between different procedures, and the resulting estimators of $\boldsymbol{\beta}$ can be very sensitive in this regard (see, for example, Ahn and Horenstein, 2013, and Moon and Weidner, 2015, for discussions).[2] Such disagreements indicate that there is much uncertainty over the selected factors. As far as we are aware, however, there are no formal statistical tests available that can be used in the CCE context that accounts for this uncertainty.

The present paper can be seen as a reaction to the above discussion. The purpose is to develop a test that is suitable for testing hypotheses about the factors in CCE when T is fixed and only N is large. The test can be used to test if a subset of the cross-sectional averages is enough to capture the factors, or if there are observable time series variables that can be used for this purpose. The test can also be used to sequentially determine the smallest set of averages needed to estimate the factors. The asymptotic properties of the test when used in these three ways are investigated in Section 2 and in Section 3 we provide an empirical illustration using as an example the US wage curve. Section 4 concludes. All assumptions, proofs and results of secondary nature are provided in the online appendix, which also contains a small-scale Monte Carlo study.

[2]Brown et al. (2022) show that the asymptotic distribution of the CCE estimator depends on which factors are used for estimation.

2 The test and its asymptotic properties

2.1 Testing estimated factors

Denote by $\mathbf{Z}_i = [\mathbf{y}_i, \mathbf{X}_i]$ the $T \times (k+1)$ matrix of observables. The usual CCE estimator of the space spanned by $\mathbf{F}$ is given simply by $\overline{\mathbf{Z}}$, where $\overline{\mathbf{A}} = N^{-1}\sum_{i=1}^N \mathbf{A}_i$ for any generic matrix $\mathbf{A}_i$.[3] In this paper, however, we want to entertain the possibility that not all $k+1$ averages are needed in the estimation. Let us therefore denote by m the number of averages actually included. Let $M \subseteq \{1, ..., k+1\}$ be the index set corresponding to the selected averages, and denote by $\mathbf{S}_M$ the $(k+1) \times m$ matrix of zeros and ones that selects these averages from $\overline{\mathbf{Z}}$. We also define $\overline{M} = \{1, ..., k+1\}$ as the set of all averages, such that $\mathbf{S}_{\overline{M}} = \mathbf{I}_{k+1}$. The rank of $\mathbf{S}_M$ is given by m and the number of selected averages is the cardinality $|M| = m$. The estimator of the space of $\mathbf{F}$ that we will be considering in this paper is given by

$$\widehat{\mathbf{F}}_M = \overline{\mathbf{Z}}\mathbf{S}_M. \tag{3}$$

The model to be estimated is (1) with $\widehat{\mathbf{F}}_M$ in place of $\mathbf{F}$ and ignoring the heterogeneity of $\boldsymbol{\beta}_i$;

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \widehat{\mathbf{F}}_M\mathbf{g}_{i,M} + \mathbf{e}_i, \tag{4}$$

where $\mathbf{g}_{i,M}$ is a $m \times 1$ vector of factor loadings and the $T \times 1$ vector $\mathbf{e}_i$ is a composite error term capturing not only the original regression error $\boldsymbol{\varepsilon}_i$ but also the heterogeneity of $\boldsymbol{\beta}_i$ and the error coming from the estimation of the factors. One of the challenges we face is how to construct a test that accounts for everything that goes into $\mathbf{e}_i$. The CCE estimators of $\boldsymbol{\beta}$ and $\mathbf{g}_{i,M}$ in (4) are given by

$$\widehat{\boldsymbol{\beta}}_M = \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{X}_i\right)^{-1} \sum_{i=1}^N \mathbf{X}_i'\mathbf{M}_{\widehat{F}_M}\mathbf{y}_i, \tag{5}$$

$$\widehat{\mathbf{g}}_{i,M} = (\widehat{\mathbf{F}}_M'\widehat{\mathbf{F}}_M)^{+}\widehat{\mathbf{F}}_M'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}_M), \tag{6}$$

where $\mathbf{M}_A = \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^{+}\mathbf{A}' = \mathbf{I}_T - \mathbf{P}_A$ for any T-rowed matrix $\mathbf{A}$ and $(\mathbf{A}'\mathbf{A})^{+}$ is the Moore–Penrose inverse of any matrix $\mathbf{A}'\mathbf{A}$ (if $\mathbf{A}$ is not of full column rank).

[3] An implicit assumption here is that $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$ are the only data available and hence that the only factor candidates available are given by $\overline{\mathbf{Z}}$. This is not necessary. A common scenario in practice, which we discuss at length in the empirical illustration of Section 3, is that in addition to $\overline{\mathbf{Z}}$ there are time series, including common deterministic terms such as a constant and linear trend, that can be viewed as observed common factors. If this is the case, we simply append the observed factors to $\widehat{\mathbf{F}}_M$.

It is standardly assumed in model selection problems (and elsewhere) that the largest model considered is not under-specified (see Bai and Ng, 2002, for an example in the pure factor model context). We do the same.[4] This means that asymptotically the factor estimator based on all $k+1$ averages, $\widehat{\mathbf{F}}_{\overline{M}} = \overline{\mathbf{Z}}$, spans the space of $\mathbf{F}$.

We want to test the null hypothesis that asymptotically $\widehat{\mathbf{F}}_M$, a subset of all $k+1$ averages, spans the space of $\mathbf{F}$. If the null hypothesis is correct, then the excluded averages are redundant and can hence be removed, which is expected to reduce noise and increase precision. If, on the other hand, the null hypothesis is false, then $\widehat{\mathbf{F}}_M$ do not span the space of $\mathbf{F}$. This means that some of the excluded averages are not redundant and that they are in fact necessary for consistent estimation of the factor space. They should therefore be retained.

Easy as it may sound, testing if $\widehat{\mathbf{F}}_M$ spans the space of $\mathbf{F}$ is a nontrivial task. An obvious problem with the test approach described in the last paragraph is that it is not really feasible since $\mathbf{F}$ is unobserved. However, we have assumed that asymptotically $\widehat{\mathbf{F}}_{\overline{M}}$ spans the space of $\mathbf{F}$. In this sense, observing $\widehat{\mathbf{F}}_{\overline{M}}$ is as good as observing $\mathbf{F}$. This motivates implementing the test of whether $\widehat{\mathbf{F}}_M$ spans the space of $\mathbf{F}$ as a test of whether $\widehat{\mathbf{F}}_M$ spans the same space as $\widehat{\mathbf{F}}_{\overline{M}}$. Unfortunately, the fact that $\mathbf{F}$ is unobserved is not the only challenge we face. Another problem that arises is that the dimensionality of $\widehat{\mathbf{F}}_M$ is not the same as that of $\widehat{\mathbf{F}}_{\overline{M}}$. Moreover, even if one compares two estimates of the same dimension that span the same factor space, they will not be equal asymptotically but only equal up to a certain rotation. In order to overcome these problems, in this paper we test if $\widehat{\mathbf{F}}_M \mathbf{g}_{i,M}$ is asymptotically equal to $\widehat{\mathbf{F}}_{\overline{M}} \mathbf{g}_{i,\overline{M}}$. The use of the estimated common components as opposed to the estimated factors not only resolves the difference in dimensionality and rotational indeterminacy issues, but also reduces the dimension of the object being tested. The change of object therefore has many advantages. However, there is also a cost in that while $\widehat{\mathbf{F}}_M$ is consistent provided only that N is large, $\mathbf{g}_{i,M}$ is not. Interestingly enough, this inconsistency is not detrimental in any way, as we will now show.

In order to describe the proposed test statistic, it is useful to define the $T \times 1$ vector of

[4] Our formal assumptions are laid out in full in Section A of the online appendix.

common component differences

$$\mathbf{\Delta}_{i,M} = \widehat{\mathbf{F}}_{\overline{M}}\widehat{\mathbf{g}}_{i,\overline{M}} - \widehat{\mathbf{F}}_{M}\widehat{\mathbf{g}}_{i,M}. \tag{7}$$

A consistent non-parametric estimator of the covariance matrix of this vector is given by

$$\widehat{\mathbf{\Sigma}}_M = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{\Delta}_{i,M} - \overline{\mathbf{\Delta}}_M)(\mathbf{\Delta}_{i,M} - \overline{\mathbf{\Delta}}_M)'. \tag{8}$$

One possibility for a test statistic is given by $\widehat{\mathbf{\Sigma}}_M^{-1/2}\sqrt{N}\overline{\mathbf{\Delta}}_M$, a $T \times 1$ vector.[5] This test statistic is suitable when wanting to test the equality of the estimated common components at each point in time, and is feasible when T is very small. As T increases, however, so does the number of tests that can possibly be computed.[6] In order to circumvent this problem, we propose combining the point-wise test statistics. Specifically, the following combination statistic based on summing across the individual point-wise test statistics is proposed[7]:

$$T_M = \frac{\sqrt{N}1'_{T\times 1}\overline{\mathbf{\Delta}}_M}{\widehat{\sigma}_M}, \tag{9}$$

where $1_{T\times 1} = [1, ..., 1]'$ is a $T \times 1$ vector of ones, and

$$\widehat{\sigma}_M^2 = 1'_{T\times 1}\widehat{\mathbf{\Sigma}}_M 1_{T\times 1}. \tag{10}$$

The object being tested is the span of $\widehat{\mathbf{F}}_M$. At times, it will be useful to be able to make this dependence on $\widehat{\mathbf{F}}_M$ explicit. We will then write $T(\widehat{\mathbf{F}}_M)$ for T_M

The assumptions that we will be working under in this section are largely the same as in Westerlund and Kaddoura (2022), and are laid out in Section A of the online appendix. Theorem 1 summarizes the large-N properties of the T_M statistic under these assumptions. In order to formally define the null and alternative hypotheses of this test, it is useful to denote by M_0 the smallest set of m_0 averages that are rotationally consistent for $\mathbf{F}$.[8] We can think of M_0 as

[5]Here $\widehat{\mathbf{\Sigma}}_M = \widehat{\mathbf{\Sigma}}_M^{1/2}\widehat{\mathbf{\Sigma}}_M^{1/2\prime}$ with $\widehat{\mathbf{\Sigma}}_M^{1/2}$ being the lower triangular Choleski factor.

[6]A large T is also likely to lead to poor small-sample performance, because of the need to compute and invert the $T \times T$ matrix $\widehat{\mathbf{\Sigma}}_M$.

[7]Our choice of combination statistic is motivated mainly by the fact that the use of the sum enables a straightforward asymptotic analysis. Alternatively, one could consider order statistics (or indeed any other summary measure). The asymptotic theory of such statistics is, however, notoriously difficult, especially in cases such as this when the point-wise statistics are not independent.

[8]The formal definition of M_0 can be found in Section A of the online appendix.

the index set corresponding to the "correct" averages. In this notation, the null and alternative hypotheses being tested here are given by $H_0 : M \supseteq M_0$ and $H_1 : M \subset M_0$, respectively.

Theorem 1. *Suppose that Assumptions 1–7 are met.*

(a) *If H_0 holds, then $T_M \to N(0,1)$ as $N \to \infty$.*

(b) *If H_1 holds, then $T_M = O_p(\sqrt{N})$.*

Theorem 1 (a) implies that T_M is asymptotically correctly sized under H_0;

$$\lim_{N\to\infty} \mathbb{P}(|T_M| > z_{\alpha/2}) = \alpha, \tag{11}$$

where $z_\alpha = \Psi^{-1}(1-\alpha)$ is the $(1-\alpha)$-th quantile of the standard normal cumulative distribution function, here denoted $\Psi(x)$. If H_1 is true, so that $\widehat{\mathbf{F}}_M \mathbf{g}_{i,M}$ is different from $\widehat{\mathbf{F}}_{\overline{M}} \mathbf{g}_{i,\overline{M}}$ even asymptotically, then $T_M = O_p(\sqrt{N})$ by Theorem 1 (b). It follows that

$$\lim_{N\to\infty} \mathbb{P}(|T_M| > z_{\alpha/2}) = 1. \tag{12}$$

Power therefore tends to one.

2.2 Estimating M_0

The T_M test can be used to determine M_0. In principal components analysis the process of determining which factors to include is simplified by the fact that the estimated factors are ordered by the fraction of the total variation in the regression errors they explain. The cross-sectional averages considered here are by contrast unordered. This is a problem in the sense that selection calls for the evaluation of $2^{k+1}-1$ models, which is likely a big number. If $k+1 = 10$, the number of candidate models is $1,023$, and when $k+1 = 20$, this number increases to $1,048,575$. Even with very fast computers, evaluating $2^{k+1}-1$ models may not be practical. And even if it is, the precision of the selection procedure is likely to decrease as the number of candidate models increases.

We therefore look for a test-based procedure to determine M_0 that does not require evaluation of all $2^{k+1}-1$ models. As a source of inspiration, we look at the statistical literature on variable selection in high-dimensional models (see, for example, Fujikoshi, 2022, and the references provided therein). Kick-one-out (KOO) methods, originally proposed for use with information criteria, reduce the number of computational statistics from $2^{k+1}-1$ to $k+1$. The general idea is that if the criterion value for the subset with one variable removed from the full set is greater than that of the full set, the removed variable is selected. In the present context, this means that we should exclude one average at a time and only retain those averages for which there is no evidence against H_0. The below test procedure is similar in spirit to the one considered by Fujikoshi and Sakurai (2019).

Sequential estimation of M_0:

1. Set $j=1$, $\widehat{M}=\overline{M}$ and $M_j=\overline{M}\setminus\{j\}$.
2. Compute $T_{M_j}=T(\widehat{\mathbf{F}}_{M_j})$.
3. Denote by c_N a "critical value" such that $c_N\to\infty$ and $c_N/\sqrt{N}\to 0$. If $|T_{M_j}|\le c_N$, set $\widehat{M}=\widehat{M}\setminus\{j\}$, whereas if $|T_{M_j}|>c_N$, $\widehat{M}$ is untouched.
4. If $j<k+1$, set $j=j+1$ and go back to step 2, whereas if $j=k+1$, the procedure is stopped.

A word about c_N. Assume for a moment that the step-2 test was carried out using the conventional critical value based on the standard normal distribution, $z_{\alpha/2}$, instead of c_N. The limiting behaviour of T_{M_j} in the above sequential setup is dictated by Theorem 1. Hence, since T_{M_j} is divergent if truly important averages are kicked out, with probability approaching one the set of averages retained in the sequential procedure will not be smaller than M_0. However, since the probability of rejecting a true null hypothesis is given by α, the probability of selecting too many averages will not be zero, and therefore the set of retained averages will tend to be larger than M_0 even asymptotically. The use of a diverging critical value eliminates the risk of over-specification, which means that consistency is possible, provided that $c_N/\sqrt{N}\to 0$. Theorem 2 formalizes this.

Theorem 2. *Suppose that Assumptions 1–7 are met. Then, as* $N \to \infty$,

$$\mathbb{P}(\widehat{M} = M_0) \to 1.$$

As far as we are aware, the sequential test-based procedure considered here is the first that enables consistent estimation of M_0 in the CCE context when T is fixed and only N is large. A naturally consistent estimator of the number of factors, $m_0 = |M_0|$, is given $\widehat{m} = |\widehat{M}|$.

2.3 Testing known factors

As Bai and Ng (2006) point out, it may be appealing to associate estimated factors with observed time series variables as this enables economic interpretation, although admittedly in CCE the estimated factors are already quite easy to interpret (see Stauskas and Westerlund, 2022, for a discussion). We therefore want to be able to test if the observed variables span the space of the factors. This is similar to the testing problem considered in Section 2.2 and the solution is similar, too.

Suppose we observe $\mathbf{G}$, a $T \times m$ matrix of economic time series variables. Analogously to the situation faced in Section 2.2, while we are ultimately interested in the relationship between $\mathbf{G}$ and $\mathbf{F}$, we do not observe $\mathbf{F}$ and so we replace it with $\widehat{\mathbf{F}}_{\overline{M}}$, which we know asymptotically spans the same space. The hypothesis to be tested is therefore that $\mathbf{G}$ spans the space of $\mathbf{F}$, which we implement as a test of whether asymptotically $\mathbf{G}$ and $\widehat{\mathbf{F}}_{\overline{M}}$ span the same space. The test statistic is the same as in Section 2.2, except that we replace $\widehat{\mathbf{F}}_M$ with $\mathbf{G}$, that is, we use $T_G = T(\mathbf{G})$.

Theorem 3. *Suppose that Assumptions 1–7 are met. Then* T_G *has the same asymptotic properties as* T_M *in Theorem 1.*

Which variables include in $\mathbf{G}$ is not always obvious. One implication of Theorem 3 is that T_G can be applied in the same sequential fashion as T_M in Section 2.3 to determine the variables that span the space of $\mathbf{F}$.

3 Empirical illustration

In this section, we illustrate the usefulness of the T_G test using as an example the US wage curve. Westerlund et al. (2019) employ three estimators of this curve; namely, (i) regular OLS, (ii) OLS based on treating US real gross domestic product, consumer price index and industrial production as observable factors, and (iii) CCE. All three estimators include cross-section fixed effects. The main finding is that the evidence of a wage curve gets weaker as the generality of the allowable interactive effects increases, which is taken to imply that CCE is relatively more suitable. There is, however, little in the way of formal statistical evidence to suggest that CCE is superior to the other, simpler estimators considered.

The present illustration is motivated by this last observation. The purpose is to test if fixed effects alone or together with the observed factors asymptotically span the space of the estimated CCE factors. If they do there is no need to use CCE. This possibility suggests that there is another, competing interpretation of the reported variation in the wage curve evidence; namely, that is it a result of increased variability brought about by over-specification of the factor space.

The data set used is the same as in Westerlund et al. (2019), and is taken from the Panel Study of Income Dynamics. It comprises $T = 7$ biannual observations (2002–2014) for $N = 888$ male household heads between 25 and 65 years of age. The dependent variable is log real wages, and the included regressors are unemployment, education, working experience and experience squared (see, for example, Card, 1999, for a similar line-up).[9] Because of high correlation between the averages of experience and experience squared, the CCE estimator is implemented using only the averages of wages, unemployment, education and experience. A vector of ones is also included, which is tantamount to allowing for household fixed effects. The total number of factor estimates is therefore equal to five.

The $|T_G|$ value for testing whether asymptotically fixed effects alone are enough to estimate the factor space is equal to 62.51 and if we also include the observed factors said test value equals 48.21, which are both highly significant even at the conservative 1% significance level. We therefore conclude that the two sets of observed factors are not enough to capture the un-

[9] We refer to Westerlund et al. (2019) for a more detailed description of the data.

known factors, which is consistent with the interpretation of Westerlund et al. (2019) that the variation in the results is due to model under-specification when the observed factors are used.

4 Conclusion

In this paper, we introduced a simple test that can be used to test hypotheses about the factors in CCE. The new test can be used to test if a subset of the averages asymptotically span the space of the unobserved factors, or if there are observable variables that span the same space. The test can also be used sequentially to determine the smallest set of averages needed to span the factor space.

References

Ahn, S. C., Y. H. Lee, and P. Schmidt (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* **101**, 219–255.

Ahn, S. C., and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica* **81**, 1203–1227.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* **91**, 299–323.

Bai, J., and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Bai, J., and S. Ng (2006). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* **131**, 507–537.

Blanchflower, D. G., and A. J. Oswald (1994). *The wage curve.* Cambridge, MA: MIT Press.

Brown, N., P. Schmidt, and J. M. Wooldridge (2022). Simple alternatives to the common correlated effects model. Unpublished manuscript.

Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O., and D. Card (Eds.), *Handbook of Labor Economics* **3A**, Amsterdam: North Holland.

Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295–300.

Hagemann, A. (2012). A simple test for regression specification with non-nested alternatives. *Journal of Econometrics* **166**, 247–254.

Fujikoshi, Y. (2022). High-dimensional consistencies of KOO methods in multivariate regression model and discriminant analysis. *Journal of Multivariate Analysis* **188**, 104860.

Fujikoshi, Y., and T. Sakurai (2019). Consistency of test-based method for selection of variables in high-dimensional two-group discriminant analysis. *Japanese Journal of Statistics and Data Science* **2**, 155–171.

Moon, H. R., and M. Weidner (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* **83**, 1543–1579.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica* **74**, 967–1012.

Pesaran, M. H. (2021). General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics* **60**, 13–50.

Robertson, D., and V. Sarafidis (2015). IV estimation of panels with factor residuals. *Journal of Econometrics* **185**, 526–541.

Stauskas, O., and J. Westerlund (2022). Tests of equal forecasting accuracy for nested models with estimated CCE factors. *Journal of Business & Economic Statistics* **40**, 1745–1758.

Westerlund, J., Y. Petrova and M. Norkute (2019). CCE in fixed-T panels. *Journal of Applied Econometrics* **34**, 746–761.

Westerlund, J., and Y. Kaddoura (2022). CCE in heterogenous fixed-T panels. *Econometrics Journal* **25**, 719–738.