Estimation and Testing in Panel Data with Cross-Section Dependence

Reese, Simon

2017

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Estimation and Testing in Panel Data with Cross-Section Dependence

Simon Reese

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the School of Economics and Management,
Lund University, Sweden.
To be defended at Holger Crafoord EC3:210, Lund
on April 7, 2017 at 10:15.

Faculty opponent
Takashi Yamagata, University of York
This thesis makes a contribution to the econometrics of panel data with cross-section dependence (CSD). It consists of five self-contained papers. The most popular approach to account for CSD is to attribute the co-movements of observed variables across entities to unobserved common factors and to estimate these factors via simple cross-section averages of the data. The chapters in this thesis investigate the properties of regression estimators that are based on this approach and suggest new tests that rely on cross-section averages to capture the impact of latent common factors.

Chapter two points out a problem with the Common Correlated Effects (CCE) estimator of Pesaran (2006) that appears in the empirically relevant case when the number of factors is strictly less than the number of observables used in their estimation. Specifically, the use of too many observables causes the second moment matrix of the estimated factors to become asymptotically singular, an issue that has not yet been appropriately accounted for. We show that the dominating method of proving the asymptotic properties of the CCE estimator breaks down in this case and suggest a more general method of proof.

Chapter three develops PANICCA, an approach to panel unit root testing that combines the merits of the two most popular approaches for testing for a unit root in panel data with CSD. We take over the separate treatment of idiosyncratic and common components inherent in the PANIC framework which allows to determine the source of nonstationarity specifically. In order to decompose the observed data into its two components, we use cross-section averages of all available variables, a simple, yet powerful approach with good small sample properties.

Chapter four considers models where the impact of latent factors converges to zero as the number of individuals tends to infinity. These so-called weak factors constitute a setting for which the two dominating regression estimators for panel data with CSD are not explicitly conceived for. We investigate the asymptotic properties of these two estimators and set up minimal conditions for the factor strength under which asymptotic normality and consistency can be proven.

Chapter five suggests a Hausman test statistic based on the difference between Pesaran’s (2006) CCE estimator and the OLS estimator. Under the null hypothesis of no cross-section dependence, this difference has a higher rate of convergence than the estimators themselves. As an immediate consequence, the test statistic diverges at a higher rate than the popular CD test of Pesaran (2004) when the data are cross-sectionally correlated, thus leading to a test with higher power.

Chapter six provides a simple and user-friendly way of measuring the contribution of different markets on developments in the fundamental price of an asset. We note that standard models for price discovery can be represented in the form of a common factor model and use a decomposition based on cross-section averages to separate idiosyncratic and common components of the observed price series.

Keywords
Factor-augmented panel regression, CCE estimation, cross-section dependence, common factor models.

Classification system and/or index terms (if any)
JEL Classification: C12, C13, C33, C36

Supplementary bibliographical information

ISSN and key title
0460-0029 Lund Economic Studies no. 203

ISBN
978-91-7753-198-2 (print)
978-91-7753-199-9 (pdf)

Recipient’s notes
Number of pages
227

Price

Security classification

Distributed by Department of Economics, P.O. Box 7082, S-220 07 Lund, Sweden

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature ______________________ Date 2017-03-07
Estimation and Testing in Panel Data with Cross-Section Dependence

Simon Reese

LUND UNIVERSITY

LUND ECONOMIC STUDIES NUMBER 203
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgment</td>
<td>v</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Paper I</td>
<td>11</td>
</tr>
<tr>
<td>Paper II</td>
<td>49</td>
</tr>
<tr>
<td>Paper III</td>
<td>79</td>
</tr>
<tr>
<td>Paper IV</td>
<td>147</td>
</tr>
<tr>
<td>Paper V</td>
<td>177</td>
</tr>
</tbody>
</table>
Acknowledgment

It would hardly be an exaggeration to say that doing a Ph.D. in econometrics has undoubtedly been my number one priority for the last six years of my life. Still, despite all the effort that has been devoted to my thesis, it seems like I merely managed to get a big picture of the vast number of issues I need to know more about in order to become a good econometrician. This being said, I am more motivated than ever to continue doing research! Econometrics offers a unique intellectual challenge, requiring both the discipline to work out proofs and to maintain formal rigour as well as a solid idea about what a certain estimator or test is good for in practice. While doing research in econometrics is a constant reminder of being only half as experienced in theory as a mathematician and being only half as knowledgeable about the real world as an applied economist, it allows working on tremendously exiting issues in empirical economics in a highly analytical way. For this reason, I am extremely happy about having been given the opportunity to spend the last four and a half years, including countless evenings and weekends, on learning more about econometrics. Of course, the work on my thesis would have been hardly as much fun and hardly as successful as it turned out to be without both the company and the help of a number of persons. This section is dedicated to the expression of my gratitude to them.

First and foremost, I am deeply indebted to my supervisor Joakim Westerlund. Joakim, without your endless support and advice I would only have learned a fraction of the things I know now. Thanks for inspiring me with your excitement about research, your incredible amount of creativity and your unique skill in producing research output. Thanks for always encouraging me to carry on when I lost confidence in my qualities as researcher and for putting enormous confidence in my capacity to deliver. Your impact as a supervisor was essential for my professional development during the last years and I could not have wished for a better senior researcher to learn from. I would feel privileged about getting the chance to continue working with you.

My deepest gratitude goes also to my assistant supervisor David Edgerton. David, you have been the key factor behind my motivation to excel during my first years in Lund and you have had a crucial impact on my decision to join the Ph.D. program at the Department of Economics. Thanks for always being genuinely interested in how things are going for me and for providing tons of support and advice on general research issues, be it literature, conferences or funding. Thanks for having shown amazing leadership skills as head of department and for tolerating my attempts to turn the monthly department board meetings into personal Swedish conversa-
tion classes (… which is a gratitude I would like to express to all remaining board members as well).

Fredrik NG Andersson is the third person who deserves appreciation for helping me at important stages of my academic career in Lund. Fredrik, thank you for giving honest and critical feedback on my work and for stepping in whenever your help is needed, irrespective of whether this is as a master thesis supervisor or in order to write a last minute reference letter.

My development as teacher has benefited enormously from being Peter Jochumzen’s teaching assistant in a number econometrics courses. I am very grateful, Peter, for the chance to learn from you how to make econometrics as a subject both intuitively understandable and exciting to students. Thanks for inspiring me with your ideas on how to use new technology in class and for giving me all the freedom I wanted to develop my part of the courses that we had together. Of course, the experience of working with Peter would not have realized without Pontus Hansson’s confidence in my Swedish skills and his willingness to give me teaching responsibilities in Swedish speaking courses.

A number of individuals has made life at the department much more enjoyable, even without having an essential impact on my research. This concerns obviously Lina Ahlin, Aron Berg, Viroj Jienwatcharamongkhol, Karl McShane, Yana Pryymachenko, Margaret Samahita and Anna Welander who have been keeping me company during the entire length of my Ph.D. studies. Thanks for lovely fancy dinners, election nights, BBQ evenings and board game nights. And thanks for obeying my lunch location directives. I also enjoyed to welcome my fellow coursemates Sara Moricz, Jörgen Kratz and Jim Ingebretsen Carlsson from the master program in economics as new colleagues in the year following my own admission to the Ph.D. program. Additionally, other junior staff members, both research group members, former office mates and fellow colleagues, have been very pleasant company during the last years. This concerns Emre Aylar, Alessandro Martinello, Natalia Montinari, Alex Rigos, Martin Strieborny, Danial Ali Akbari, Emmanuel Alfranseder, Elvira Andersson, Björn Thor Arnarson, Katarzyna Burzyńska, Guiyun Cheng, Daniel Ekblom, Albin Erlanson, Caglar Kaya, Gustav Kjellsson, Kristoffer Persson, Yana Petrova, Hilda Ralsmark, Hassan Sabzevari and Hampus Sporre.

It should not go unnoticed that the wider community of people connected to the econometrics group in Lund has exerted its influence on my academic work during the last years: Yushu Li has has provided me with early and indispensable experience in writing and publishing scientific papers through the two projects on which we worked together. Hande Karabiyik and Arturas Juodis have been valuable discussion partners on issues related to our common research interests, both at conferences and during
their respective stays in Lund. Lastly, Mehdi Hosseinkouchack did a truly amazing job as discussant at my final seminar.

During the academic year 2015/2016 I had the pleasure of working at the Department of Quantitative Economics at Maastricht University. I am grateful to Tom Hedelius Stiftelse for financing this stay abroad. The time I had in Maastricht was both highly stimulating in terms of research skills and very pleasant on a personal level. I am immensely thankful to Rasmus Lönn for sharing his office, his thoughts and plenty of coffee coins with me and for backing my fake Swedish identity for the entirety of my stay. Additionally, Aida Abiad, Anne Balter, Marina Friedrich, Verena Jung, Laura Kasper, Vincent Kreuzen, Yicong Lin, Yuliya Shapovalova, Sean Telg and Andrej Winokurow have contributed a lot to the good memories I have from my ten months in Maastricht.

Two persons from the Department of Quantitative Economics at Maastricht University deserve a specific mention. First, Eric Beutner has had an enormous impact on my increased awareness of formal rigour. Eric, thank you for investing so much of your time into our joint paper and for giving me plenty of opportunities to learn how mathematicians think. Lastly, my visiting Ph.D. position in Maastricht would not have been made possible without the commitment of Jean-Pierre Urbain. JP, you have been outstandingly generous with the amount of time and resources you devoted to me, far more than I would have imagined to get as a visiting student. Our regular meetings have provided me with important insights on research, academia and life and I am heartbroken about the fact that you are not longer with us.

It might sound surprising to my close colleagues, but I actually know a handful of people outside of academia. I am thankful to Håkan and Marianne for always making be feel welcome at their house and for running a fantastic raspberry delivery service. Additionally, my beloved Sofia has been a constant source of joy during the past years. Thanks, Sofia, for being the most charming and funny person on earth. Thanks for constantly thinking outside the box and for enriching my view on the world with a passionate humanistic perspective. Thanks for always convincing me that the effort I make is sufficient and for supporting me whatever I do and wherever I want to go. And thanks for the countless hours in which I could forget about everything in the world apart from us two. Finally, I am fundamentally indebted to my parents Wolfgang and Olga whose support is second to none. Thanks for wasting thousands of Euros on my education throughout all these years - I hope I managed to ensure a decent return on investment. Thanks for being right on so many issues in life.\footnote{I disregard from fashion advice in making this statement.} Thanks for opening your door twice or thrice a year when I am in dire need of a
rest. And thanks showing so much enthusiasm for my work even though it is far from the things you tend to consider interesting. There is no doubt that you have been the central driving force behind the completion of this thesis.

Simon Reese
Lund, March 2017
Introduction

Co-movements of economic and financial indicators across entities - be it firms, industries, states or countries - are a well-documented fact; see e.g. Kose et al. (2003) for macroeconomic indicators and Hamao et al. (1990) for stock prices. These co-movements, usually referred to as cross-section dependence (CSD), are commonly attributed to the exposure of economic agents to common shocks as well as the propagation of individual-specific shocks to other economic agents. During the last two decades, CSD has attracted substantial research interest in the panel data econometrics literature which is concerned with the analysis of datasets containing economic indicators of a number of different entities that are observed over a certain time span.

This thesis contains theoretical contributions to the literature on panel data with cross-section dependence. More specifically, I contribute to the literature on common factor models which assume that cross-section dependence arises because most, or even all, variables observed by the researchers are driven by a small number of unobservable common factors.\footnote{The factor approach to CSD is distinct from a spatial approach that considers co-movements to arise from dependence of one entity on other entities in a certain pre-specified geographical or economic vicinity. See Hsiao, 2014, Ch. 9, for a survey of both approaches.} It has been argued that this assumption is very reasonable in macroeconomics since macroeconomic models assume key economic indicators to depend on the same common shocks (Pesaran et al., 2013, p.95).

Common factor models and the pitfalls involved

The common factor model postulates that all observed variables can be decomposed into a component containing individual-specific variation and a common component. The common component is assumed to be due to the effect of a small number of unobservable factors that are common to all individuals and all variables and that have individual-specific slope coefficients. In this regard, the common factor model is similar to the standard model structure in factor analysis. In the context of a regression model, it is the error term rather than the regressand that is assumed to be decomposed into common and idiosyncratic components whereas the decomposition of all covariates into these two parts continues to hold. This structure implies that the common factor model can be seen as a generalisation of two-way fixed effects models that are standard in panel data econometrics: Instead of having one series of period-specific effect that affects all individuals equally and individual-specific effects that are constant over time,
one allows for several series of period-specific effects whose impact differs across individuals. The econometric challenges of the two-way fixed effects model continue to hold for regressions estimated on common factor models: Failure to account for common factors in the error term leads to omitted variable bias if these common factors are correlated with the covariates. Given that the covariates are by assumption driven by the same factors as those found in the error term, estimated regression coefficients will be inconsistent unless highly restrictive assumptions on the nature of the factors and their impact on the observed data are met. Practically, this means that under realistic conditions the estimated impact on, say, economic output of the stock of investments in research and development (see e.g. Eberhardt et al., 2013) will not yield a consistent estimator of the slope coefficients unless the impact of unobservable factors, such as the business cycle or technology shocks, that affect both variables is adequately accounted for. If the time dimension of a panel data set is large enough, the estimation of a regression model is usually preceded by unit root tests to determine the integration order of the observed variables and followed by a cointegration test in case the variables in question have been found to be difference stationary. However, common variation across individuals in the variables under study has been shown to invalidate the results of panel unit root and panel cointegration tests that are developed under the assumption of cross-section independence (Banerjee et al., 2004, 2005). Consequently, results from these tests may lead to an incorrect decision on where the non-stationarity in both economic output and the R&D stock stems from and whether a cointegration relation between the two variables can and does exist.

Properties of the existing estimation approaches

Contributions on panel regression estimators and unit root tests in common factor models, which by now constitute an entire branch in the panel data econometrics literature, are dominated by two closely related ideas on how to account for common variation in the variables under study: The most popular approach is to control for the common factors by augmenting the model that is chosen by the researcher, either to investigate the effect of determinants on the variable of interest (Pesaran, 2006; Bai, 2009) or to obtain a unit root test statistic (Moon and Perron, 2004; Pesaran, 2007). An alternative method is to decompose the observed data into variation that is due to common shocks and purely entity-specific variation. These two components are then treated separately when testing for a unit root (Bai

---

3For a survey, see Pesaran (2015, chapters 29 & 31)
and Ng, 2004, 2010) and/or a cointegration relationship (Gengenbach et al., 2006).

Common to both approaches is that no measurements exist for the common factors that are to be taken account of. Hence, an estimator has to be found. One strand of the literature draws on methods from factor analysis and uses a fixed number of principle components of the data as factor estimates. A different method consists of taking simple cross-section averages of the variables available to the researcher. As shown by Pesaran (2006, 2007), this yields an estimator capable of accounting for variation due to the unobservable factors in a common factor model. Conceptually both methods are closely related: Principle components are by construction merely weighted averages of the individual time series taken over all variables and all individuals. The weights are chosen to ensure that the principle components are orthogonal and that, if the principle components are to be ordered, each principle component captures the variation not explained by prior principle components to the greatest possible extent. This concept is fairly close to that of cross-section averages which are plain averages of each observed variable taken over all individuals in the dataset. However, cross-section averages excel through their simplicity and the relative ease with which the asymptotic properties of estimators and tests based on this approach can be obtained. In fact, estimators for linear regressions based on cross-section averages that are designed for common factor models have been shown to work in a number of different settings. By contrast, the theory underlying principal components-based models is not as widely developed. This is due to additional complications that arise from the computation of the weights that constitute the principle components. And in some instances, it is required to use results from the applied mathematics literature which currently exist only under rather restrictive assumptions (see Moon and Weidner, 2015). For this reason, most of the chapters in this thesis consider cross-section averages as a means of accounting for the effect of common factors.

It can easily be seen that neither principle components nor cross-section averages provide a consistent estimator of the true unobservable common factors. For example, the orthogonal factor estimates obtained from taking principal components of the data can hardly be seen to represent previously mentioned latent factors such as the business cycle and technology shocks, which are very likely to be correlated. In fact, both principal components and cross-section averages are only consistent estimators of certain linear combinations of the true unobservable factors. In the context of panel data regressions, this is not a problem: Latent factors constitute a nuisance term which needs to be accounted for, not a covariate whose slope coefficient the researcher is interested in. The consequences of estimating
some linear combination of the true factors on their associated slope coefficients can hence be disregarded. Furthermore, the estimator of the slope coefficients of all observed variables depends only on the space spanned by the latent factors. Since the space spanned by a certain number of vectors is unaffected by any invertible linear combinations that are applied to these vectors, having a consistent estimator of a linear combination of the true latent factors is as good as being able to consistently estimate the factors themselves. A similar result holds for for a unit root in the common and individual-specific components of the data. The decomposition into these two parts is again only dependent on the space spanned by the factor estimates. And unit root tests applied to the estimated factors are pivotal in that they are unaffected by which (invertible) linear combination of the data one estimates.

Overview of the chapters

Chapter two is concerned with the Common Correlated Effects (CCE) estimator of Pesaran (2006). The CCE estimator consists of augmenting an Ordinary Least Squares (OLS) regression with cross-section averages of all observables, which yields a consistent estimator for data that are characterized by a factor structure. As put forward by Pesaran (2006), a crucial advantage of the CCE estimator is that the researcher is not required to specify the number of latent factors that are assumed to affect the data. The main argument is that the use of cross-section averages mops up all common variation in the data, irrespectively of how many factors it is due to. We show that this reasoning is problematic in the empirically relevant case where the number of observables is strictly greater than the number of latent factors. It turns out that the covariance matrix of the factor estimates is asymptotically singular, causing the inverse of this matrix to diverge. However, this behaviour invalidates the predominant method of proving the asymptotic properties of the CCE estimator, implying that many of the results in the related literature are yet to be proven. We show that an alternative method of proof is derived from the fact that the CCE estimator solely depends on the space spanned by the factor estimates. We construct a regularized factor estimator from the cross-section averages of all observables which is asymptotically non-singular and which spans the same vector space as the factor estimator based on plain cross-section averages. Writing the CCE estimator as a function of the regularized factor estimator allows proving its asymptotic properties straightforwardly without having to handle a diverging inverse matrix.

Chapter three suggests a simple approach to unit root testing in panels with cross-section dependence. We consider the PANIC framework of
Bai and Ng (2004) which consists of decomposing the data into common and idiosyncratic components and testing for a unit root in the two components separately. However, instead of obtaining the factor estimates via taking principle components of a single series, we estimate them by taking cross-section averages of all the data at hand. This parallels the approach of Pesaran (2007) and Pesaran et al. (2013) whose tests lack the possibility of investigating the integration order of common and idiosyncratic components separately. Our approach, called PANICCA naturally matches with the prevalent use of unit root testing as pre-testing step in an analysis involving several variables. It uses the availability of additional data to incorporate more information into the decomposition into common and idiosyncratic components of the variable of interest. Moreover, simulation results suggest that unit root tests based on PANICCA have better small sample properties than their counterparts based on PANIC. This can be explained by noting that the use of cross-section averages circumvents the need to estimate the contribution of individual series to the factor estimates, a step that is inherent in a decomposition based on principal components. We demonstrate the usefulness of PANICCA by applying it to spot and forward exchange rates of 15 OECD countries in order to test the efficient market hypothesis and the forward rate unbiasedness hypothesis in a unifying framework.

Chapter four considers models where the impact of latent factors converges to zero as the number of individuals tends to infinity. This modeling approach is relevant in cases where a common shock that has an impact on the behaviour of a small number of individuals loses its effect as the number becomes large, e.g. because of coordination problems or due to the decreasing relevance of individual behaviour on the aggregate outcome. Our modeling choice is equivalent to the notion of weak factors which denotes factors that affect only an asymptotically negligible fraction of the individuals in the dataset. The two most popular estimators for factor-augmented panel regressions, the CCE estimator of Pesaran (2006) and the interactive fixed effects estimator of Bai (2009), require latent factors to be strong, implying that the fraction of individuals affected by a specific factor grows at the same rate as the total number of individuals in the sample. We investigate the asymptotic properties of these two estimators in the case of weak factors and set up minimal conditions for the factor strength under which asymptotic normality and consistency can be proven. The minimal conditions are related to the relative rate of expansion of cross-sections versus time periods. This allows assessing the required strength of latent factors for the asymptotic properties of CCE and PC estimators to hold in panel datasets of specific dimensions. Our main results are presented for CCE and PC estimators that assume homogeneous slope coefficients for the ob-
served covariates. We supplement these with findings for heterogeneous slopes in the case of the CCE estimator.

Chapter five suggests an alternative route to testing for cross-section dependence for data that is suspected to be driven by unobservable common factors. The existing literature consists predominantly of tests based on the sum of all (squared) pairwise cross-section correlations between the regression residuals of any two individuals in the sample. Instead of constructing a test statistic from model residuals, we use Hausman’s (1978) concept of testing for a significant difference between two alternative regression estimators. The difference considered here is that between Pesaran’s (2006) CCE estimator and the OLS estimator. We show that the CCE estimator is consistent in the absence of cross-section dependence. Together with the well-known inconsistency of the OLS estimator in the presence of CSD, this allows to construct a test statistic in the spirit of Hausman. Furthermore, using asymptotics that assume both the number of time periods and the number of cross-sections to go to infinity, we can show that in the absence of cross-section dependence the CCE estimator converges to the OLS estimator. Given that both estimators converge to the true slope coefficients at rate $\sqrt{NT}$, the difference between them converges to zero at an even faster rate. This requires us to scale up the Hausman test statistic beyond the scaling usually used in the literature, a treatment that also amplifies the rate of divergence of our test statistic in the presence of CSD. In fact, the divergence rate that we achieve is higher than that of the popular CD-test of Pesaran (2004), implying that our test is likely to have more power. Simulation results confirm this presumption in a number of cases, especially in datasets where the number of time periods is as large as the number of cross-sections or even slightly higher.

Chapter six provides a simple and user-friendly way of measuring the contribution of different markets on developments in the fundamental price of an asset. Widely known as price discovery, this issue has been mostly analyzed in the framework of a vector error correction model (VECM) estimated on the series of prices that are observed for the financial product of interest in different markets. While being a convenient tool for obtaining a measure of price discovery when the number of markets is small, the VECM becomes problematic when a large number of markets is considered. One pitfall is that the number of parameters to be estimated increases noticeably with the number of markets involved in the analysis. Furthermore, additional restrictions have to be made in order to identify the fundamental shocks to prices in different markets. These are usually imposed by Cholesky factorization which provides a number of alternative orderings that again markedly increases in the number of markets involved. Since it is usually unknown which ordering it is that applies to the
dataset at hand, a researcher can only obtain upper and lower bounds for the individual contribution of a specific market by considering the resulting contribution for all possible Cholesky factorizations. However, with the number of orderings increasing in the number of markets, including more markets will result in more uncertainty about what the specific contribution of an individual market is. We provide an alternative approach to price discovery by noting that the VECM which is usually employed to analyze price discovery can be rewritten as a common factor model. We then use methods from the literature on panel data with cross-section dependence to estimate the unobservable common factor and suggest a measure of price discovery based on a decomposition of the data into variation due to the common factor and market-specific variation. The practical usefulness of our method is then demonstrated with three empirical examples.

References


Paper I
ON THE ROLE OF THE RANK CONDITION IN CCE
ESTIMATION OF FACTOR-AUGMENTED PANEL
REGRESSIONS *

Hande Karabiyik
VU University Amsterdam

Simon Reese
Lund University

Joakim Westerlund†
Lund University
and
Centre for Financial Econometrics
Deakin University

March 2, 2017

Abstract

A popular approach to factor-augmented panel regressions is the common correlated
effects (CCE) estimator of Pesaran (Estimation and inference in large heterogeneous pan-
els with a multifactor error structure. *Econometrica* 74, 967–1012, 2006). This paper points
to a problem with the CCE approach that appears in the empirically relevant case when
the number of factors is strictly less than the number of observables used in their estima-
tion. Specifically, the use of too many observables causes the second moment matrix of the
estimated factors to become asymptotically singular, an issue that has not yet been appro-
priately accounted for. The purpose of the present paper is to fill this gap in the literature.

**JEL Classification:** C12; C13; C33; C36.

**Keywords:** Factor-augmented panel regression; CCE estimation; Moore–Penrose inverse.

1 Introduction

Consider the scalar \( y_{i,t} \) and the \( k \times 1 \) vector \( x_{i,t} \), where \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \) index the
cross-sectional and time series dimensions, respectively. Except for some simplifications that

---

*Previous versions of the paper were presented at a seminar at Maastricht University, at the 2016 UvA-
Econometrics Panel Data Workshop, at the 9th International Conference on Computational and Financial Eco-
nometrics, and at the IAAE 2016 Annual Conference. The authors would like to thank seminar and workshop partic-
ipants, and in particular Jianqing Fan (Co-Editor), Arturas Juodis, Maurice Bun, Joerg Breitung, Mehdi Hosseink-
ouchack, one Associate Editor, and two anonymous referees for many valuable comments and suggestions. Thank
you also to the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fel-
lowship, and the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number

†Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Tele-
phone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.
are irrelevant for the purpose of the paper, such as the absence of deterministic terms, the model is the same as in Pesaran (2006), and is given by

\[
\begin{align*}
y_i &= X_i \beta + e_i, \\
e_i &= F \gamma_i + \varepsilon_i, \\
X_i &= FF_i + V_i,
\end{align*}
\]

where \( y_i = (y_{i,t}, \ldots, y_{i,T})' \) and \( X_i = (x_{i,1}, \ldots, x_{i,T})' \) are \( T \times 1 \) and \( T \times k \) respectively, \( \beta \) is a \( k \times 1 \) vector of coefficients, \( F = (f_1, \ldots, f_T)' \) is a \( T \times m \) matrix of unobserved common factors with \( \gamma_i \) and \( \Gamma_i \) being the associated vectors of factor loadings, and \( \varepsilon_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})' \) and \( \varepsilon_v = (v_{i,1}, \ldots, v_{i,T})' \) are \( T \times 1 \) and \( T \times k \) matrices, respectively, of idiosyncratic errors. Except for the requirements that \( \varepsilon_{i,t} \) and \( v_{i,t} \) are serially uncorrelated and homoskedastic and \( \beta_i \) are assumed to be the same for all \( i = 1, \ldots, N \), Assumption 1 is the same as Assumptions 1–4 in Pesaran (2006).

**Assumption 1.**

(i) \( \varepsilon_{i,t} \) is independently and identically distributed (iid) across both \( i \) and \( t \) with \( E(\varepsilon_{i,t}) = 0 \), \( E(\varepsilon_{i,t}^2) = \sigma^2 \) and \( E(\varepsilon_{i,t}^4) < \infty \);

(ii) \( v_{i,t} \) is iid across both \( i \) and \( t \) with \( E(v_{i,t}) = 0_{k \times 1} \), \( E(v_{i,t}v_{i,t}') = \Sigma \) positive definite and \( E(||v_{i,t}||^4) < \infty \), where \( ||A|| = \sqrt{\text{tr}(A'A)} \) is the Frobenius norm of the matrix \( A \);

(iii) \( f_t \) is covariance stationary such that \( E(||f_t||^4) < \infty \) and \( E(f_t f_t') = \Sigma_f \) is positive definite;

(iv) \( \gamma_i \) and \( \Gamma_i \) are iid across \( i \), independent of \( \varepsilon_{j,t} \), \( v_{j,t} \) and \( f_t \) for all \( i \) and \( j \), have fixed means \( \gamma \) and \( \Gamma \), respectively, and finite variances;

(v) \( \varepsilon_{i,t}, v_{i,s} \) and \( f_{t} \) are mutually independent for all \( i, j, t, s \) and \( \ell \).

Because of the way that \( F \) enters both (2) and (3) the estimation of \( \beta \) is nontrivial. However, by combining (1) and (3), we have

\[
Z_i = FC_i + U_i,
\]

where \( Z_i = (y_i, X_i)' = (z_{i,1}, \ldots, z_{i,T})' \) is \( T \times (k + 1) \), \( z_{i,t} = (y_{i,t}, x_{i,t}')' \) is \( (k + 1) \times 1 \), \( C_i = (\Gamma_i \beta + \gamma_i, \Gamma_i) \) is \( m \times (k + 1) \), and \( U_i = (u_{i,1}, \ldots, u_{i,T})' = (V_i \varepsilon_i, V_i) \) is \( T \times (k + 1) \). Thus, (1)–(3) can be rewritten equivalently as a static factor model for \( Z_i \), which means that \( F \) can be estimated.
using existing approaches for such models. In the common correlated effects (CCE) approach of Pesaran (2006), the estimator of $\mathbf{F}$ is particularly simple, and is given by

$$\hat{\mathbf{F}} = \mathbf{Z} = \mathbf{F}\mathbf{C} + \mathbf{U},$$

(5)

where $\mathbf{A} = N^{-1} \sum_{i=1}^{N} \mathbf{A}_i$ for any $\mathbf{A}_i$. It is important to note here that under Assumption 1, $\mathbf{U} \overset{p}{\to} 0_{T \times (k+1)}$ as $N \to \infty$, where $\overset{p}{\to}$ signifies convergence in probability. This means that $\hat{\mathbf{F}}$ is consistent for the space spanned by $\mathbf{F}$. The pooled CCE estimator of $\mathbf{b}$ is the conventional pooled OLS estimator with $\hat{\mathbf{F}}$ in place of $\mathbf{F}$;

$$\hat{\mathbf{b}}_p = \left( \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\mathbf{F} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\mathbf{F} \mathbf{y}_i,$$

(6)

where $\mathbf{M}_\mathbf{A} = \mathbf{I}_T - \mathbf{P}_\mathbf{A} = \mathbf{I}_T - \mathbf{A}(\mathbf{A}'\mathbf{A})^+\mathbf{A}'$ for any $T$-rowed matrix $\mathbf{A}$ with $(\mathbf{A}'\mathbf{A})^+$ being the Moore–Penrose (MP) inverse of $\mathbf{A}'\mathbf{A}$. Because of its simplicity and generality, the CCE approach has attracted considerable attention, so much so that there is by now a separate CCE branch of the literature. This literature makes extensive use of the asymptotic distribution of $\sqrt{NT}(\hat{\mathbf{b}}_p - \mathbf{b})$, which has been shown to be normal under a wide variety of circumstances (see, for example, Chudik et al., 2011; Kapetanios et al., 2010; Pesaran et al., 2013; Westerlund and Urbain, 2015).

The current paper is about the way in which the asymptotic distribution of $\sqrt{NT}(\hat{\mathbf{b}}_p - \mathbf{b})$ has been established. A critical first step in all existing proofs is to show that the effect of estimating $\mathbf{F}$ is negligible, such that $\mathbf{F}$ can be treated as known in the rest of the proof. This is done by showing that $(NT)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^\prime (\mathbf{M}_\mathbf{F}\mathbf{C} - \mathbf{M}_\hat{\mathbf{F}}) \mathbf{X}_i$, $(NT)^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i^\prime (\mathbf{M}_\mathbf{F}\mathbf{C} - \mathbf{M}_\hat{\mathbf{F}}) \mathbf{\epsilon}_i$ and $(NT)^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i^\prime (\mathbf{M}_\mathbf{F}\mathbf{C} - \mathbf{M}_\hat{\mathbf{F}}) \mathbf{F}\mathbf{\gamma}_i$, are negligible, such that

$$\sqrt{NT}(\hat{\mathbf{b}}_p - \mathbf{b}) = \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\mathbf{F} \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\hat{\mathbf{F}} (\mathbf{F}\mathbf{\gamma}_i + \mathbf{\epsilon}_i)$$

$$= \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\mathbf{F}\mathbf{C} \mathbf{X}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i^\prime \mathbf{M}_\mathbf{F}\mathbf{C} \mathbf{\epsilon}_i + o_p(1).$$

(7)

Let us therefore consider $\mathbf{M}_\mathbf{F}\mathbf{C} - \mathbf{M}_\hat{\mathbf{F}}$, which can be expanded in the following way:

$$\mathbf{M}_\mathbf{F}\mathbf{C} - \mathbf{M}_\hat{\mathbf{F}} = \mathbf{U} (\hat{\mathbf{F}}\hat{\mathbf{F}})^+ \mathbf{U} + \mathbf{U} (\hat{\mathbf{F}}\hat{\mathbf{F}})^+ \mathbf{C}^\prime \mathbf{F}^\prime + \mathbf{F}\mathbf{C}(\hat{\mathbf{F}}\hat{\mathbf{F}})^+ \mathbf{U}^\prime$$

$$+ \mathbf{F}\mathbf{C}[(\hat{\mathbf{F}}\hat{\mathbf{F}})^+ - (\mathbf{C}^\prime \mathbf{F}\mathbf{C})^+] \mathbf{C}^\prime \mathbf{F}^\prime.$$

(8)

Pesaran (2006) shows that $T^{-1}\hat{\mathbf{F}}\hat{\mathbf{F}} - T^{-1}\mathbf{C}^\prime \mathbf{F}\mathbf{C} \overset{p}{\to} 0_{(k+1) \times (k+1)}$ as $N, T \to \infty$, which is taken to imply that the same result holds for the difference of the inverses. This reasoning has been
used in numerous other studies (see, for example, Chudik et al., 2011; Kapetanios et al., 2010; Pesaran et al., 2013; Westerlund and Urbain, 2015). However, this is not always true. Suppose therefore that $A_n - A_0 \xrightarrow{p} 0_{r \times r}$ as $n \to \infty$ for any real $r \times r$ matrix (matrix sequence) $A_0$ ($A_n$). Let $r_n = \text{rank}(A_n)$ and $r_0 = \text{rank}(A_0)$. Then, according to Andrews (1987), $A_n^+ - A_0^+ \xrightarrow{p} 0_{r \times r}$ if and only if

$$r_n - r_0 \xrightarrow{a.s.} 0$$

as $n \to \infty$, where $\xrightarrow{a.s.}$ signifies almost sure convergence. Therefore, only if $\text{rank}(T^{-1}\hat{F}'\hat{F}) - \text{rank}(T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}) \xrightarrow{a.s.} 0_{(k+1) \times (k+1)}$ as $N, T \to \infty$ does $T^{-1}\hat{F}'\hat{F} - T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C} \xrightarrow{p} 0_{(k+1) \times (k+1)}$ imply $(T^{-1}\hat{F}'\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C})^+ \xrightarrow{p} 0_{(k+1) \times (k+1)}$. While this fact has not gone unnoticed in the CCE literature, it has been thought that provided that

$$\text{rank}(\hat{C}) = m \leq k + 1,$$

the rank of $T^{-1}\hat{F}'\hat{F}$ should converge to the rank of $T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}$, and that this should in turn ensure that $(T^{-1}\hat{F}'\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C})^+ \xrightarrow{p} 0_{(k+1) \times (k+1)}$ (see Chudik et al., 2011; Kapetanios et al., 2010; Pesaran, 2006; Pesaran et al., 2013). However, the rank is a discrete function and for $m < k + 1$ the probability that a rank change occurs before $T^{-1}\hat{F}'\hat{F}$ reaches its asymptotic limit is zero. That is, $P[\text{rank}(T^{-1}\hat{F}'\hat{F}) = \text{rank}(T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C})] \to 0$. The only way to ensure that $\text{rank}(T^{-1}\hat{F}'\hat{F}) - \text{rank}(T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}) \xrightarrow{a.s.} 0$ is therefore to assume that $m = k + 1$.

The purpose of the present paper is in part to make the above discussion of the consequences of improper use of the MP inverse a little more precise, in part to propose a solution. This requires new tools. Section 2 therefore provides some general theory for perturbed matrices. This is a natural starting point, because $T^{-1}\hat{F}'\hat{F} - T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}$ may be regarded as negligible perturbation of $T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}$. In Section 3 the new theory is applied to the problem at hand. The results show that unless (10) is satisfied with equality, $(T^{-1}\hat{F}'\hat{F})^+$ does not converge to $(T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C})^+$, but that $(T^{-1}\hat{F}'\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C})^+$ actually diverges. The method described in the last paragraph is therefore not suitable for evaluating $M_{\hat{F}\hat{C}} - M_{\hat{F}}$, and hence also not suitable for studying the asymptotic distribution of $\sqrt{NT}(\hat{b}_p - \beta)$. An alternative method is therefore needed. This is the subject of Section 4. Section 5 concludes. All proofs are provided in the supplemental material, which also contains some useful illustrations.
2 Some matrix perturbation theory

Consider two positive semi-definite \( r \times r \) matrices \( A_0 \) and \( A_n \). The following assumption is made:

\[
A_n = A_0 + E_n,
\]

where \( E_n \) is a Hermitian perturbation of \( A_0 \). Most of the existing matrix perturbation theory is based on the assumption that \( E_n \) and \( A_0 \) are non-random, that \( n \) is fixed, and that \( \|A_0^+ E_n\| < 1 \), or even \( \|A_0^+ E_n\| < 1 \) (see, for example, Stewart, 1977; Wedin, 1973). The more general conditions that we will be working under are given in Assumption 2.

Assumption 2.

(i) \( E(\|A_0\|) < \infty \);

(ii) \( \|E_n\| \xrightarrow{P} 0 \) as \( n \to \infty \).

Replacing \( A_n, A_0 \) and \( E_n \) with \( T^{-1} \hat{F} \hat{F}, T^{-1} \hat{C} \hat{C} F \hat{F} \) and \( T^{-1} \hat{F} \hat{F} - T^{-1} \hat{C} \hat{C} F \hat{F} \), respectively, it is clear that Assumption 2 provides a natural starting point for our analysis. The questions are: What are the conditions under which \( A_n - A_0 \xrightarrow{\text{as}} 0_{r \times r} \), and what are the consequences if those conditions are not met? We begin by noting that since \( \|A_n - A_0\| = \|E_n\| = o_p(1) \), we have that \( P(r_n \geq r_0) \to 1 \) as \( n \to \infty \) (see, for example, Andrews, 1987, Note 1).

Theorem 1. Suppose that Assumption 2 is met, and that \( r_n - r_0 \xrightarrow{\text{as}} 0 \) as \( n \to \infty \). Then,

\[
\|A_n^+ - A_0^+\| = O_p(\|E_n\|) = o_p(1).
\]

Theorem 1 is similar to Theorem 2 of Andrews (1987). Unfortunately, the results of Andrews (1987) cannot be used to study the consequences of a violation of \( r_n - r_0 \xrightarrow{\text{as}} 0 \). Theorem 2 provides the missing piece.

Theorem 2. Suppose that Assumption 2 is met, and that \( P(r_n > r_0) \to 1 \) as \( n \to \infty \). Then,

\[
\|A_n^+ - A_0^+\| = \Omega_p(\|E_n^{-1}\|),
\]

where \( \Omega_p(\cdot) \) denotes stochastic boundedness from below in the spirit of the definition of Hardy and Littlewood (1914), implying \( \|A_n^+ - A_0^+\| \neq o_p(\|E_n^{-1}\|) \).
Theorem 2 implies that if $A_n$ is “near” $A_0$, but $r_n > r_0$, then its MP inverse can be larger and completely different from $A_0^+$, and the smaller is $E_n$, the worse the problem can be. Hence, if we want $A_n^+$ to be well-behaved in the sense that $\|A_n^+ - A_0^+\| = o_p(1)$, a necessary and sufficient condition is given by (9), that is, $r_n - r_0 \xrightarrow{a.s.} 0$. Unfortunately, this condition is not easily verified.

Chudik et al. (2011), Kapetanios et al. (2010), and Pesaran et al. (2013) all recognize the importance of (9). They claim that if $E_n \xrightarrow{P} 0_{r \times r}$ and rank$(A_0) = r_0$, then $r_n - r_0 \xrightarrow{a.s.} 0$, which is not correct, since $E_n \xrightarrow{P} 0_{r \times r}$ only implies $P(r_n \geq r_0) \to 1$, and not $P(r_n = r_0) \to 1$ (see Andrews, 1987, Note 1). A key result in this regard is that if $A_n$ is a continuous random matrix of full rank, then $r_n - r \xrightarrow{a.s.} 0$ (11) as $n \to \infty$ (see, for example, Feng and Zhang, 2007). This means that for (9) to hold, we require

$$ r_n = r_0 = r $$

for all $n$.

An important consequence of Theorem 1 is that under (9),

$$ \|A_n^+\| = O_p(\|A_0^+\|) = O_p(1) $$

If, on the other hand, (9) is not met, then, by Theorem 2,

$$ \|A_n^+\| = O_p(\|E_n^{-1}\|). $$

### 3 Implications for CCE

We now apply the theory developed in Section 2 to the CCE problem described in Section 1. We begin by noting how $E(||T^{-1}\hat{C}'\hat{F}\hat{F}'\hat{C}||) < \infty$, and by (36) in Pesaran (2006), we also have

$$ ||T^{-1}\hat{F}\hat{F} - T^{-1}\hat{C}'\hat{F}'\hat{F}'\hat{C}|| = O_p(N^{-1}) + O_p((NT)^{-1/2}) $$

showing that Assumption 2 is applicable with $A_n$, $A_0$ and $E_n$ replaced by $T^{-1}\hat{F}\hat{F}$, $T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}$ and $(T^{-1}\hat{F}\hat{F} - T^{-1}\hat{C}'\hat{F}'\hat{F}'\hat{C})$, respectively. As for the rank of $T^{-1}\hat{F}\hat{F}$, in analogy to (11),

$$ \text{rank}(T^{-1}\hat{F}\hat{F}) \xrightarrow{a.s.} k + 1 $$

as $N, T \to \infty$. Moreover, by Assumption 1 (iii), $\text{rank}(T^{-1}\hat{F}\hat{F}) = m$ for all $T$, including $T \to \infty$. The rank of $T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}$ is therefore determined by the rank of $\hat{C}$. Hence, in view of (10),

$$ \text{rank}(T^{-1}\hat{C}'\hat{F}'\hat{F}\hat{C}) = m, $$
for all $N$ and $T$, including $T, N \to \infty$. Suppose now that $m = k + 1$. Since in this case $\text{rank}(T^{-1}\tilde{\Phi}) \overset{a.s.}{\to} \text{rank}(T^{-1}\tilde{C}'F\tilde{C})$, by Theorem 1, we have the following:

$$
\|(T^{-1}\tilde{\Phi})^+ - (T^{-1}\tilde{C}'F\tilde{C})^+\| = O_p(N^{-1}) + O_p((NT)^{-1/2}), \quad (17)
$$

$$
\|(T^{-1}\tilde{\Phi})^+\| = O_p(1). \quad (18)
$$

If, however, $m < k + 1$, then (15) and (16) imply that the rank of $T^{-1}\tilde{\Phi}$ will not converge to that of $T^{-1}\tilde{C}'F\tilde{C}$, and so, by Theorem 2,

$$
\|(T^{-1}\tilde{\Phi})^+ - (T^{-1}\tilde{C}'F\tilde{C})^+\| = \Omega_p(N) + \Omega_p(\sqrt{NT}). \quad (19)
$$

Further use of (14) shows that $\|(T^{-1}\tilde{\Phi})^+\|$ is of the same order.

As an illustration of the implications of the above findings, let us consider $T^{-1}X_i'(M_{ FC} - M_{ F})X_i$. From Lemma 3 of Pesaran (2006), we have $\|T^{-1}X'\tilde{U}\| = O_p(N^{-1}) + O_p((NT)^{-1/2})$. In view of this, $\|\tilde{C}\| = O_p(1), \|T^{-1}F'X_i\| = O_p(1)$ and (8), we have

$$
\|T^{-1}X_i'(M_{ FC} - M_{ F})X_i\|
\leq \|T^{-1}X'\tilde{U}\|\|(T^{-1}\tilde{\Phi})^+\|\|T^{-1}\tilde{U}'X_i\| + 2\|T^{-1}X'\tilde{U}\|\|(T^{-1}\tilde{\Phi})^+\|\|\tilde{C}'\|\|T^{-1}F'X_i\|
+ \|T^{-1}X_i'(F\tilde{C})^2\|\|(T^{-1}\tilde{\Phi})^+ - (T^{-1}\tilde{C}'F\tilde{C})^+\|\|T^{-1}F'X_i\|
= [O_p(N^{-1}) + O_p((NT)^{-1/2})]O_p(\|(T^{-1}\tilde{\Phi})^+\|)
+ O_p(\|(T^{-1}\tilde{\Phi})^+ - (T^{-1}\tilde{C}'F\tilde{C})^+\|), \quad (20)
$$

which is $O_p(N^{-1}) + O_p((NT)^{-1/2})$ if $m = k + 1$ and $O_p(N) + O_p(\sqrt{NT})$ if $m < k + 1$, where the latter contradicts the results reported by Chudik et al. (2011), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). Similar results are obtained when the steps are applied to $\|(NT)^{-1/2}\sum_{i=1}^NX_i'(M_{ FC} - M_{ F})\varepsilon_i\|$ and $\|(NT)^{-1/2}\sum_{i=1}^NX_i'(M_{ FC} - M_{ F})\gamma_i\|$. This means that if $m < k + 1$ the predominating method of proof fails to show that the effect of the estimation of $F$ is negligible.

4 An alternative method of proof

The above results imply that many of the statements in the literature are actually yet to be proven. In what follows, we therefore propose a new method of proof that is appropriate in general, provided that (10) is met. The idea is the same as when analyzing sample second
moment matrices where the elements are of different orders of magnitude (see, for example, Chang and Phillips, 1995), that is, we normalize to ensure convergence to a positive definite matrix. Let us therefore assume without loss of generality that $\mathbf{C} = [\mathbf{C}_m, \mathbf{C}_{-m}]$, where $\mathbf{C}_m$ is an $m \times m$ full rank matrix and $\mathbf{C}_{-m}$ is $m \times (k + 1 - m)$. By similarly partitioning $\mathbf{U} = [\mathbf{U}_m, \mathbf{U}_{-m}]$, we obtain

$$\hat{\mathbf{F}} = [\mathbf{F}\mathbf{C}_m, \mathbf{F}\mathbf{C}_{-m}] + [\mathbf{U}_m, \mathbf{U}_{-m}].$$

Define

$$\mathbf{B} = [\mathbf{B}_m, \mathbf{B}_{-m}] = \begin{bmatrix} \mathbf{C}_m^{-1} & -\mathbf{C}_m^{-1}\mathbf{C}_{-m} \\ 0_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix},$$

which is of full rank under (10). Post-multiplying (21) by $\mathbf{B}$ yields

$$\hat{\mathbf{F}}\mathbf{B} = \mathbf{F}\mathbf{C}_m\mathbf{B} + \mathbf{U}\mathbf{B} = [\mathbf{F}_0\mathbf{I}_{m\times(m+1)}] + [\mathbf{U}_m\mathbf{C}_m^{-1}, \mathbf{U}_{-m} - \mathbf{U}_m\mathbf{C}_m^{-1}\mathbf{C}_{-m}].$$

We now look for a conformable normalization matrix $\mathbf{D}_N$ such that $T^{-1}\mathbf{D}_N\mathbf{B}^\dagger\hat{\mathbf{F}}\mathbf{B}\mathbf{D}_N$ converges to a positive matrix. Here we make use of Lemma 2 of Pesaran (2006), which states that $\|T^{-1}\mathbf{U}'\mathbf{U}\| = O_p(N^{-1})$. Hence, while the upper left $m \times m$ block of $T^{-1}\mathbf{B}^\dagger\hat{\mathbf{F}}\mathbf{B}$ converges to $\mathbf{\Sigma}_t$, the lower right $(k + 1 - m) \times (k + 1 - m)$ block is $O_p(N^{-1})$. This means that the required normalization matrix is given by $\mathbf{D}_N = \text{diag}(\mathbf{I}_m, \sqrt{N}\mathbf{I}_{k+1-m})$. Hence, letting $\hat{\mathbf{F}}^0 = [\mathbf{F}_0\mathbf{I}_{m\times(m+1)}]$ and $\hat{\mathbf{U}}^0 = \mathbf{UBD}_N = [\hat{\mathbf{U}}_m^0, \hat{\mathbf{U}}_{-m}^0] = [\mathbf{U}_m\mathbf{C}_m^{-1}, \sqrt{N}(\mathbf{U}_{-m} - \mathbf{U}_m\mathbf{C}_m^{-1}\mathbf{C}_{-m})]$, the resulting normalized version of $\hat{\mathbf{F}}\mathbf{B}$ is given by

$$\hat{\mathbf{F}}^0 = \mathbf{F}\mathbf{B}\mathbf{D}_N = \mathbf{F}^0 + \hat{\mathbf{U}}^0.$$
Also, by using the decomposition of $\bar{U}^0$ into $\bar{U}^0_m$ and $\bar{U}^0_{-m}$,

$$T^{-1}(\bar{U}^0_m)'\bar{U}^0_m = T^{-1} \begin{bmatrix} (\bar{U}^0_m)'\bar{U}^0_m & (\bar{U}^0_m)'\bar{U}^0_{-m} \\ (\bar{U}^0_{-m})'\bar{U}^0_m & (\bar{U}^0_{-m})'\bar{U}^0_{-m} \end{bmatrix}. $$

By (A.10) in Lemma 2 of Pesaran (2006), we have $\|T^{-1}\bar{U}^0_m\bar{U}^0_m\| = O_p(N^{-1})$, and by further use of Assumption 1 (iv), $\|\Sigma_m\|$ and $\|\Sigma_{-m}\|$ are $O_p(1)$. It follows that

$$\|T^{-1}(\bar{U}^0_m)'\bar{U}^0_m\| \leq N^{-1/2}\|\Sigma_m^{-1}\| \|NT^{-1}\bar{U}^0_m\bar{U}^0_m\| + N^{-1/2}\|\Sigma_{-m}^{-1}\| \|NT^{-1}\bar{U}^0_{-m}\bar{U}^0_{-m}\| \|\Sigma_{-m}\|$$

$$= O_p(N^{-1/2}). \quad (27)$$

It remains to evaluate $T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m}$. Use of $\bar{U}^0_{-m} = \sqrt{N}\mathbf{U}[-\bar{C}^{-1}_{-m}(\bar{C}^{-1}_m)', I_{k+1-m}]' = \sqrt{N}\bar{U}\bar{B}_{-m}$ leads to

$$T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m} = \mathbf{B}^{-1}_{-m}NT^{-1}\bar{U}\bar{U}\bar{B}_{-m}. \quad (28)$$

Let $\Sigma_u = E(u_{i,t}'u_{i,t})$. A straightforward calculation reveals that

$$NT^{-1}\bar{U}\bar{U} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} u_{i,t}'u_{i,t}' = \Sigma_u + O_p(T^{-1/2}), \quad (29)$$

which in turn implies

$$T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m} = \Sigma_u\phi_{-m} + O_p(T^{-1/2}) \quad (30)$$

where $\Sigma_u\phi_{-m} = \mathbf{B}^{-1}_{-m}\Sigma_u\mathbf{B}_{-m}$, a $(k+1-m) \times (k+1-m)$ matrix. It is important to note that this matrix is positive definite, which implies that $T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m}$ is positive definite, too. Hence, $\text{rank}[T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m}] = k+1-m$. By substituting (24)–(28) into (23), we obtain the following:

$$T^{-1}(\bar{F}^0)'\bar{F}^0 = \Sigma_\phi + O_p(N^{-1/2}) + O_p(T^{-1/2}), \quad (31)$$

where

$$\Sigma_\phi = \begin{bmatrix} T^{-1}\bar{F}'\bar{F} & 0_{m \times (k+1-m)} \\ 0_{(k+1-m) \times m} & T^{-1}(\bar{U}^0_{-m})'\bar{U}^0_{-m} \end{bmatrix}.$$
m as \( T \to \infty \), which, together with the positive definiteness of \( T^{-1}(\mathbf{U}_{-m})'\mathbf{U}_{-m} \), in turn implies
\[
\text{rank}(\Sigma_f) \overset{a.s.}{\to} k + 1.
\]
But we also have \( \text{rank}[T^{-1}(\hat{F})'\hat{F}] \overset{a.s.}{\to} k + 1 \) as \( N, T \to \infty \), and so we obtain
\[
\text{rank}[T^{-1}(\hat{F})'\hat{F}] \overset{a.s.}{\to} \text{rank}(\Sigma_f). \tag{32}
\]
By using this result, (31) and Theorem 1, we obtain the following key result:
\[
[T^{-1}(\hat{F})'\hat{F}]^+ = \Sigma_f^+ + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{33}
\]
An important implication of (13) and (33) is that
\[
\|T^{-1}(\hat{F})'\hat{F}\|^+ = O_p(1). \tag{34}
\]
Hence, in contrast to the divergence result obtained for \( \|(T^{-1}\hat{F}\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}\hat{F}\hat{C})^+\| \) and \( \|(T^{-1}\hat{F}\hat{F})^+\|, \|(T^{-1}(\hat{F})'\hat{F})^+\| - \Sigma_f^+\| \) and \( \|(T^{-1}(\hat{F})'\hat{F})^+\| \) do behave “nicely”. From this point on the analysis is similar to the analyses of Chudik et al. (2011), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). The key difference is that here we make use of the fact that \( \mathbf{M}_\hat{F} = \mathbf{M}_{\hat{F}^r} \), which allows us to proceed in roughly the same way as when using \( \hat{F} \) and assuming that \( m = k + 1 \). We therefore put the details in the supplemental material, and just provide here the final result.

**Theorem 3.** Under Assumption 1 and condition (10), as \( N, T \to \infty \) with \( T/N \to \tau < \infty \),
\[
\sqrt{NT}(\hat{b}_p - \beta) \overset{d}{\to} N(0_{m \times 1}, \sigma^2 \Sigma^{-1}) + \sqrt{\tau} \Sigma^{-1}(b - d),
\]
where \( b = b_2 - b_1 - b_3 \) and \( d = d_2 - d_1 \) with
\[
\begin{align*}
\mathbf{d}_1 &= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \Sigma[\beta, I_k] - \Gamma'(\bar{\zeta})'\Sigma_u\mathbf{P}_m\Sigma_u\bar{\zeta}'_i, \\
\mathbf{d}_2 &= \lim_{N \to \infty} \Sigma[\beta, I_k] - \bar{\Gamma}'(\bar{\zeta})'\Sigma_u\mathbf{P}_m\sigma^2[1, 0_{1 \times k}]', \\
\mathbf{b}_1 &= \lim_{N \to \infty} \Sigma[\beta, I_k] \bar{\zeta}_i, \\
\mathbf{b}_2 &= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \Gamma'(\bar{\zeta})'\Sigma_u\bar{\zeta}'_i, \\
\mathbf{b}_3 &= \lim_{N \to \infty} \bar{\Gamma}'(\bar{\zeta})' + \sigma^2[1, 0_{1 \times k}]',
\end{align*}
\]
and \( \mathbf{P}_m = \mathbf{B}_m(\mathbf{B}_m'\Sigma_u\mathbf{B}_m) + \mathbf{B}_m' \).

As far as we are aware, Theorem 3 is the first to establish the asymptotic distribution of \( \sqrt{NT}(\hat{b}_p - \beta) \) while properly accounting for the problematic \( m < k + 1 \) case. Interestingly,
while far from obvious from the results of Section 3, the degeneracy that occurs when $m < k + 1$ does not interfere with asymptotic normality. In this sense, Theorem 3 is consistent with the previous literature (see, for example, Pesaran, 2006; Chudik et al., 2011; Kapetanios et al., 2010; Pesaran et al., 2013). We also see that the estimator is biased, which is in agreement with the results of Westerlund and Urbain (2015, Theorem 1). Interestingly, the bias expression given in Theorem 3 is not the same as the one reported in this other paper. The difference is $d$. This term depends on $P_m$, capturing the effect of the degenerate regressors, and is only present under $m < k + 1$. In fact, it is not difficult to show that if $m = k + 1$, such that (10) is satisfied with equality, then

$$
\sqrt{NT}(\hat{b}_P - \beta) \xrightarrow{d} N(0_{m \times 1}, \sigma^2 \Sigma^{-1}) + \sqrt{\tau} \Sigma^{-1}b.
$$

If, in addition, $T/N \to 0$, then

$$
\sqrt{NT}(\hat{b}_P - \beta) \xrightarrow{d} N(0_{m \times 1}, \sigma^2 \Sigma^{-1}),
$$

which under Assumption 1 is the same as in Theorem 4 of Pesaran (2006). Hence, while important under $T/N \to \tau < \infty$, failure to account for the degenerate regressors does not have an effect under the more restrictive assumption that $T/N \to 0$.

## 5 Conclusion

The CCE approach of Pesaran (2006) has attracted considerable interest in the literature on factor-augmented panel regressions. In the present paper we point to a problem with the CCE approach that seems to have gone largely unnoticed in this literature. The problem occurs in the empirically relevant case when $m < k + 1$. Specifically, the use of too many observables causes the second moment matrix of the estimated factors to become asymptotically singular, which in turn invalidates some of the arguments commonly used to establish asymptotic theory. Hence, the bulk of existing theories is actually yet to be proven. A new method of proof is therefore proposed that is shown to alleviate the singularity problem, leading to a straightforward asymptotic analysis.
References


SUPPLEMENT TO “ON THE ROLE OF THE RANK CONDITION IN CCE ESTIMATION OF FACTOR-AUGMENTED PANEL REGRESSIONS”: ILLUSTRATIONS AND PROOFS

Hande Karabiyik  Simon Reese  Joakim Westerlund*
VU University Amsterdam  Lund University  Lund University
and
Centre for Financial Econometrics
Deakin University

March 2, 2017

Abstract

In this supplement, we (i) provide illustrations of some of the results provided in the main text, and (ii) provide the proofs of Theorems 1–3.

1 Illustrations

Illustration 1 (The reduced rank problem). Recall that

\[ A_n = A_0 + E_n, \]

where \( A_0 \) (\( A_n \)) is an \( r \times r \) constant matrix (matrix sequence), and \( E_n \) is a negligible perturbation of \( A_0 \). Theorems 1 and 2 imply that a necessary and sufficient condition for \( A_n^+ - A_0^+ \overset{p}{\to} 0_{r \times r} \) is given by

\[ r_n - r_0 \overset{a.s.}{\to} 0 \]  \hspace{1cm} (S1)

as \( n \to \infty \). We now provide two simple examples that illustrate this point. Suppose first that

\[ A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_n = \begin{bmatrix} 0 & 0 \\ 0 & n^{-1} \end{bmatrix}. \]

Clearly,

\[ A_n - A_0 = E_n = \text{diag}(0, n^{-1}) \to 0_{2 \times 2}. \]

*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.
In spite of this, however, we have
\[ A_n^+ - A_0^+ = \text{diag}(0, n) \nrightarrow 0_{2 \times 2}. \]
If, on the other hand,
\[ E_n = \begin{bmatrix} 0 & n^{-1} \\ 0 & 0 \end{bmatrix}, \]
such that again \( A_n \rightarrow A_0 \), then
\[ A_n^+ - A_0^+ = \frac{1}{1 + n^{-2}} \begin{bmatrix} n^{-2} & 0 \\ n^{-1} & 0 \end{bmatrix} \rightarrow 0_{2 \times 2}. \]
The reason for this difference in the results is that while in the second example \( r_n = r_0 \), in the first example \( r_n > r_0 \). This illustrates that problems can arise when working with the Moore–Penrose (MP) inverse and the rank of \( A_n \) is different from that of \( A_0 \).

**Illustration 2** (The empirical relevance of the \( m < k + 1 \) case). As we explain in the main text, for the standard method of proof to work we need \( m = k + 1 \). The question therefore arises as to how likely this is in practice. The number of regressors, \( k \), is usually a small number that is given by economic theory (and/or previous empirical evidence). Economic theory is, on the other hand, not very informative regarding the number of factors, \( m \) (see, for example, Eberhardt et al., 2013). Therefore, the theoretically implied value of \( k \) has typically little or nothing to do with \( m \). This is important because within CCE choosing \( k \) also means restricting \( m \), and in many applications there is little or no reason to believe that this number should be exactly equal to \( k + 1 \). This is confirmed by the existing empirical evidence. For example, Gaibulloev et al. (2014) regress the GDP growth of a country on its own lagged value, a measure of terrorism, and the investment and population growth rates. They apply an information criterion with which they find evidence of between one and two factors. Another example is provided by Reese and Westerlund (2016), who regress the spot US dollar exchange rate of a currency on the one lagged forward rate to test the forward rate unbiasedness hypothesis. They apply a similar information criterion as Gaibulloev et al. (2014), and report evidence of one factor. There are many examples like these.

**Illustration 3** (The interpretation of the rank condition). Chudik et al. (2011), Kapetanios et al. (2010), and Pesaran et al. (2013) all recognize the importance of the condition in (S1). They claim that if \( E_n \xrightarrow{p} 0_{r \times r} \) and \( \text{rank}(A_0) = r_0 \), then \( r_n - r_0 \xrightarrow{a.s.} 0 \), which is not correct, since
\(E_n \xrightarrow{p} 0_{r \times r}\) only implies \(P(r_n \geq r_0) \to 1\), and not \(P(r_n = r_0) \to 1\) (see Andrews, 1987, Note 1). To see this, consider the following counter-example. Let

\[
A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_n = \begin{bmatrix} 0 & 0 \\ 0 & E_n \end{bmatrix},
\]

where \(E_n\) is a continuous random variable such that \(E_n \xrightarrow{p} 0\). Clearly,

\[
A_n - A_0 = E_n = \text{diag}(0, E_n) \xrightarrow{p} 0_{2 \times 2}.
\]

Furthermore, \(r_n > r_0\) for all \(n < \infty\). In fact, since \(E_n\) is a continuous random variable, we have that \(P(r_n = 1) = P(E_n = 0) = 0\). This shows that \(A_n - A_0 \xrightarrow{p} 0_{r \times r}\) does not imply (S1). It also illustrates how the randomness of the perturbation plays an important role.

**Illustration 4** (The breakdown of the standard method of proof when \(m < k + 1\)). In Section 3 of the main text we show that

\[
\begin{align*}
\text{rank}(T^{-1}\hat{F}\hat{F}) & \xrightarrow{a.s.} k + 1, \\
\text{rank}(T^{-1}\hat{C}'\hat{F}\hat{F}\hat{C}) & = m,
\end{align*}
\]

(S2) (S3)

where the first result holds as \(T, N \to \infty\), while the second holds for all \(N\) and \(T\), including \(T, N \to \infty\). Since

\[
\|T^{-1}\hat{F}\hat{F} - T^{-1}\hat{C}'\hat{F}\hat{F}\hat{C}\| = O_p(N^{-1}) + O_p((NT)^{-1/2}),
\]

according to Theorem 2, this means that if \(m < k + 1\), then

\[
\|(T^{-1}\hat{F}\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}\hat{F}\hat{C})^+\| = O_p(N) + O_p(\sqrt{NT}),
\]

(S4)

with \(\|(T^{-1}\hat{F}\hat{F})^+\|\) being of the same order. This is illustrated in Figure 1, which plots the average norm of \((T^{-1}\hat{F}\hat{F})^+\), \((T^{-1}\hat{F}\hat{F})^+ - (T^{-1}\hat{C}'\hat{F}\hat{F}\hat{C})^+\). The process used to generate the figure is a restricted version of (1)–(3) that sets \(m = 2 < k + 1 = 3\), \(N = T\), \(\hat{f}, v_i, \varepsilon_i \sim N(0, I_5), \beta = [1, 1]'\), \(\gamma = [1, u_1]'\) with \(u_1 \sim U(1, 2)\), and

\[
\Gamma_i = \Gamma = \begin{bmatrix} u_2 \\ u_3 \\ u_4 \end{bmatrix},
\]

where \(u_2, u_3\) and \(u_4\) are all drawn from \(U(1.5, 2.5)\). The number of replications is given by 1,000. The two lines almost coincide and the divergence rate is \(N = T\), just as expected. As mentioned in Section 1, the problem here is that when \(m < k + 1, k + 1 - m\) of the regressors
in \( \hat{F} \) are degenerate. This causes the second moment matrix of \( \hat{F} \) to become asymptotically singular with unbounded inverses as a result.

**Illustration 5** (The alternative method of proof). In Section 4 of the main text we show that in contrast to the divergence result obtained for \( \| (T - 1 \hat{F}^\prime \hat{F})^+ - (T - 1 \hat{C}^\prime \hat{F}^\prime \hat{F} \hat{C})^+ \| \) and \( \| (T - 1 \hat{F}^\prime \hat{F})^+ \| \) when \( m < k + 1 \), \( \| (T - 1 \hat{F}^0 (T - 1 \hat{F}^0)^+) + \Sigma^+_p \| \) and \( \| (T - 1 \hat{F}^0 \hat{F}^0)^+ \| \) do behave “nicely”. This is illustrated in Figure 2, which plots the average of these norms over 1,000 replications of the same data generating process used to generate Figure 1. As expected, while \( \| (T - 1 \hat{F}^0 \hat{F}^0)^+ - \Sigma^+_p \| \) converges to zero, \( \| (T - 1 \hat{F}^0 \hat{F}^0)^+ \| \) converges to a constant. Note also how the properties of \( \| (T - 1 \hat{F}^0 \hat{F}^0)^+ \| \) and \( \| (T - 1 \hat{F}^0 \hat{F}^0)^+ - \Sigma^+_p \| \) are roughly the same as those of \( \| (T - 1 \hat{F}^\prime \hat{F})^+ \| \) and \( \| (T - 1 \hat{F}^\prime \hat{F})^+ - (T - 1 \hat{C}^\prime \hat{F}^\prime \hat{F} \hat{C})^+ \| \) when \( \operatorname{rank}(\hat{C}) = m = k + 1 \).

## 2 Proofs of Theorems 1–3

Before we come to the proofs of Theorems 1 and 2, it is useful to first introduce some notation. Let \( A \) denote a generic square matrix of dimension \( r \times r \). Let \( \lambda_1(A) \geq \ldots \geq \lambda_r(A) \) be the ordered eigenvalues of \( A \), and let \( \rho(A) = \max_{j=1,...,r} |\lambda_j(A)| = \lambda_1(A) \) be the spectral radius of
A. Denote by $\sigma_j(A) = \sqrt{\lambda_j(A'A)}$ the $j$-th singular value of $A$. The spectral norm of $A$ is given by $\|A\|_2 = \sqrt{\rho(A)} = \sqrt{\lambda_1(A)} = \sigma_1(A)$. Note how $\|A\|_2 \leq \|A\|$ (see Horn and Johnson, 1985, Corollary 5.6.35) and $\|A\| \leq \sqrt{r} \|A\|_2$.

Proof of Theorem 1.

Note that $A_n$ that has a dimension of $r \times r$. Now, suppose without loss of generality, that $\text{rank}(A_n) = r_n = r$ for all $n$. Hence, $r_n = r = r_0$. We begin by noting that by Theorem 4.1 of Noble (1976),

$$A_n^+ - A_0^+ = -A_n^+ E_n A_n^+ + A_n^+(A_n^+)'E_n'(I_r - A_0^+ A_n^+) + (I_r - A_n^+ A_n)E_n'(A_n^+)'A_n^+.$$

Since $r_n = r_0 = r$, we have $A_n^+ = A_n^{-1}$, which in turn implies $(I_r - A_n^+ A_n) = 0_{r \times r}$. The same result holds for $I_r - A_0^+ A_0^+$, so that we obtain

$$\|A_n^+ - A_0^+\|_2 \leq \|E_n\|_2 \|A_n^+\|_2 \|A_0^+\|_2. \quad (S6)$$

1Without loss of generality holds because one can decompose any sequence $A_n$ into a finite number of different subsequences, each of which has the property that all its terms have the same rank. Once the proof of the theorem has been established for any such subsequence, it is easy to see that it holds for the sequence as a whole (see Andrews, 1987, page 354).
Consider $A_n$. By the singular value decomposition (SVD), $A_n = V_n \Sigma_n U_n'$, where the $r \times r$ matrix $\Sigma_n$ is a diagonal matrix that contains the singular values of $A_n$ on its diagonal, such that $\Sigma_n = \text{diag}[\sigma_1(A_n), ..., \sigma_r(A_n)]$ with $\sigma_1(A_n) \geq ... \geq \sigma_r(A_n)$, and where $V_n$ and $U_n$ are both $r \times r$ matrices that contain the left and right singular vectors of $A_n$, respectively. Note that $V_n$ and $U_n$ are orthogonal matrices, that is, $V_n V_n' = V_n V_n' = I$, and $U_n' U_n = U_n U_n' = I_r$.

By the properties of the MP inverse (see Abadir and Magnus, 2005, Exercise 10.29), $A_n^+ = (V_n \Sigma_n U_n')^+ = U_n \Sigma_n^+ V_n'$, where $\Sigma_n^+ = \text{diag}[\sigma_1(A_n)^{-1}, ..., \sigma_r(A_n)^{-1}]$. Next, we apply Weyl’s theorem (see Horn and Johnson, 1985, Theorem 4.3.1) to $A_0 = A_n - E_n$, obtaining $\lambda_p(A_0) \leq \lambda_p(A_n) + \lambda_1(-E_n)$ for all $p = 1, ..., r$. This result can be expressed in terms of singular values due to the properties of the matrices involved. Since, $A_0$ ($A_n$) is p.s.d., it follows that $\sigma_p(A_0) = \lambda_p(A_0)$ ($\sigma_p(A_n) = \lambda_p(A_n)$). Furthermore, for the Hermitian matrix $E_n$, it holds that $\sigma_p(-E_n) = |\lambda_p(-E_n)|$. Using these two identities and letting $p = r$, we have

$$\sigma_r(A_0) \leq \sigma_r(A_n) + \sigma_1(-E_n). \quad (S7)$$

But $\sigma_1(-E_n) = \sigma_1(E_n) = \|E_n\|_2$, $\|A_n^+\|_2 = \sigma_r(A_n)^{-1}$ and $\|A_0^+\|_2 = \sigma_r(A_0)^{-1}$, giving

$$\|A_n^+\|_2^{-1} \geq \|A_0^+\|_2^{-1} + \|E_n\|_2 = \|A_0^+\|_2^{-1}(1 + \|A_0^+\|_2\|E_n\|_2),$$

or

$$\|A_n^+\|_2 \leq \|A_0^+\|_2(1 + \|A_0^+\|_2\|E_n\|_2)^{-1}, \quad (S8)$$

where the denominator converges to one as $\|E_n\|_2$ goes to zero since $\|A_0^+\|_2 = \sigma_r(A_0)^{-1} < \infty$.

Direct insertion into (S6) now yields

$$\|A_n^+ - A_0^+\|_2 \leq \|E_n\|_2 \frac{\|A_0^+\|_2^2}{1 + \|A_0^+\|_2\|E_n\|_2} = o_p(1). \quad (S9)$$

The proof is completed by noting that $\|A\|_2 \leq \|A\| \leq \sqrt{r}\|A\|_2$ for square matrix $A$, which means that $O_p(\|A\|_2) = O_p(\|A\|)$.

**Proof of Theorem 2.**

Similarly to the proof of Theorem 1, assuming $r_n = r$ for all $n$, we have $r_n = r > r_0$. According to the SVD, $A_n = V_n \Sigma_n U_n'$, where $V_n$, $\Sigma_n$ and $U_n$ are as in Proof of Theorem 1. Consider the following “regularized” version of $A_n$: $\tilde{A}_n = V_n \tilde{\Sigma}_n U_n'$, where

$$\tilde{\Sigma}_n = \begin{bmatrix} \Delta_{r_0} & 0_{r_0 \times (r-r_0)} \\ 0_{(r-r_0) \times r_0} & 0_{(r-r_0) \times (r-r_0)} \end{bmatrix},$$
and $\Delta_n = \text{diag}[\sigma_1(A_n), ..., \sigma_{r_0}(A_n)]$. In this notation,

$$A_n^+ - A_0^+ = A_n^+ - \tilde{A}_n^+ - (A_0^+ - \tilde{A}_n^+), \quad (S10)$$

Consider $A_0^+ - \tilde{A}_n^+$. Note that rank$(\tilde{A}_n) = r_0$. According to Theorem 2 (b) of Andrews (1987), if (i) $\|\tilde{A}_n - A_0\| = o_p(1)$ as $n \to \infty$, and (ii) $\|\tilde{A}_n^+\| = O_p(1)$, then $\|\tilde{A}_n^+ - A_0^+\| = o_p(1)$. Condition (i) is easy to verify. Indeed, since $A_n - A_0 = E_n$ and $\|E_n\|_2 = o_p(1)$ under Assumption 2 and rank$(\tilde{A}_n) = r_0$, we have that $\|\tilde{A}_n - A_0\| = o_p(1)$ (see Andrews, 1987, page 355). Consider (ii).

By the definition of rank$A$ reduces to

$$\|\tilde{A}_n^+ - A_0^+\| \leq \|E_n^{-1}\|_2. \quad (S14)$$

as required for (ii). It follows that $\|\tilde{A}_n^+ - A_0^+\| = o_p(1)$. From (S10), $\|A_n^+ - A_0^+\|_2 = \|A_n^+ - \tilde{A}_n^+ - (A_0^+ - \tilde{A}_n^+)\|_2$, implying

$$\|A_n^+ - \tilde{A}_n^+\|_2 - \|A_0^+ - \tilde{A}_n^+\|_2 \leq \|A_n^+ - A_0^+\|_2 \leq \|A_n^+ - \tilde{A}_n^+\|_2 + \|A_0^+ - \tilde{A}_n^+\|_2.$$

But $\|A_0^+ - \tilde{A}_n^+\|_2 = o_p(1)$, and therefore

$$\|A_n^+ - \tilde{A}_n^+\|_2 = \|A_n^+ - \tilde{A}_n^+\|_2 + o_p(1). \quad (S12)$$

In what remains we evaluate $\|A_n^+ - \tilde{A}_n^+\|_2$. We begin by noting that

$$A_n^+ - \tilde{A}_n^+ = U_n(\Sigma_n^+ - \tilde{\Sigma}_n^+)V_n^\prime = U_n\begin{bmatrix} 0_{r_0 \times r_0} & 0_{r_0 \times (r - r_0)} \\ 0_{(r - r_0) \times r_0} & \nabla_{r-r_0}^{-1} \end{bmatrix}V_n^\prime,$$

where $\nabla_{r-r_0}^{-1} = \text{diag}[\sigma_{r_0+1}(A_n)^{-1}, ..., \sigma_r(A_n)^{-1}]$. Since $\sigma_{r_0+1}(A_n)^{-1} \leq ... \leq \sigma_r(A_n)^{-1}$, we obtain

$$\|A_n^+ - \tilde{A}_n^+\|_2 = \sigma_r(A_n)^{-1}. \quad (S13)$$

By Weyl’s theorem, $\sigma_p(A_0) - \sigma_r(E_n) \leq \sigma_p(A_n) \leq \sigma_p(A_0) + \sigma_1(E_n)$ for $p = r_0 + 1, ..., r$, which reduces to $\sigma_r(E_n) \leq \sigma_p(A_n) \leq \sigma_1(E_n)$, as $\sigma_{r_0+1}(A_0) = ... = \sigma_r(A_0) = 0$ and further to $\sigma_p(A_n) \in [0; \sigma_1(E_n)]$ since $\sigma_p(A_n)$ is by construction non-negative. Taking the inverse, we obtain $\sigma_1(E_n)^{-1} \leq \sigma_p(A_n)^{-1}$ so that

$$\|A_n^+ - \tilde{A}_n^+\|_2 \leq \sigma_r(E_n)^{-1} \leq \sigma_1(E_n)^{-1}.$$

But $\sigma_1(E_n) = \|E_n\|_2$, which implies

$$\|A_n^+ - \tilde{A}_n^+\|_2 \geq \|E_n^{-1}\|_2. \quad (S14)$$
Consequently,
\[
\|A_n^+ - \tilde{A}_n^+\|_2 = \Omega_p(\|E_n^{-1}\|_2),
\]  
(S15)
where \(\Omega_p(\cdot)\) denoted stochastic boundedness from below and is equivalent to \(\|A_n^+ - \tilde{A}_n^+\|_2 \neq o_p(\|E_n^{-1}\|_2)\). By adding our results,
\[
\|A_n^+ - A_0^\dagger\|_2 = \|A_n^+ - \tilde{A}_n^+\|_2 + o_p(1) = \Omega_p(\|E_n^{-1}\|_2) + o_p(1).
\]  
(S16)
The proof is made complete by noting that the order of \(\|A_n^+ - A_0^\dagger\|_2\) is equal to that of \(\|A_n^+ - A_0^+\|_2\).
\[\Box\]

**Proof of Theorem 3.**

We begin by noting that since rank(\(\mathcal{C}\)) = \(m\) (\(\mathcal{C}\) has full row rank), we have \(\mathcal{C}^+ = \mathcal{C}'(\mathcal{C}\mathcal{C}')^{-1}\), such that \(\mathcal{C}\mathcal{C}^+ = (\mathcal{C}^+)'\mathcal{C}' = \mathbf{I}_m\) (see Abadir and Magnus, 2005, Exercise 10.31). This implies
\[
y_i = X_i\beta + \hat{\mathcal{F}}\mathcal{C}^+\gamma_i - \mathcal{U}\mathcal{C}^+\gamma_i + \varepsilon_i,
\]  
(S17)
which can be substituted into the expression for \(\mathbf{b}_p\) given in (6) in the main text, giving
\[
\sqrt{NT}(\mathbf{b}_p - \beta) = \left(\frac{1}{NT} \sum_{i=1}^{N} X_i'\mathbf{M}_0 X_i\right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'\mathbf{M}_0 (\varepsilon_i - \mathcal{U}\mathcal{C}^+\gamma_i).
\]  
(S18)
Consider the numerator. From \(\mathbf{M}_0 = \mathbf{F}_0\), we have
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'\mathbf{M}_0 (\varepsilon_i - \mathcal{U}\mathcal{C}^+\gamma_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'\mathbf{F}_0\varepsilon_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'\mathbf{M}_0 \mathcal{U}\mathcal{C}^+\gamma_i
\]  
(S19)
with obvious definitions of \(\mathbf{Q}_{NT}\) and \(\mathbf{R}_{NT}\). Consider \(X_i'\mathbf{M}_0\). By using the definition of \(X_i\) and \(\hat{\mathcal{F}}' = (\mathbf{B}')^{-1}\mathbf{D}_N^{-1}(\mathbf{F}_0)'\), we obtain the following very useful expression for \(X_i\):
\[
X_i = F\Gamma_i + V_i = \mathcal{F}\mathcal{C}^+\Gamma_i + V_i = \hat{\mathcal{F}}\mathcal{C}^+\Gamma_i - (\hat{\mathcal{F}} - \mathcal{F}\mathcal{C})\mathcal{C}^+\Gamma_i + V_i
\]  
(S20)
Note in particular how \((\hat{\mathcal{F}})'\mathbf{F}_0 = 0_{(k+1)\times T}\), and therefore \(X_i'\mathbf{M}_0 = [V_i' - \Gamma_i'(\mathcal{C}^+)'\mathcal{U}']\mathbf{M}_0\), which in turn implies
\[
\mathbf{R}_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'\mathbf{M}_0 \mathcal{U}\mathcal{C}^+\gamma_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V_i' - \Gamma_i'(\mathcal{C}^+)'\mathcal{U}']\mathbf{M}_0 \mathcal{U}\mathcal{C}^+\gamma_i
\]  
(S21)
\[
= \mathbf{R}_{0NT} - \mathbf{R}_{1NT} - \mathbf{R}_{2NT},
\]  
(S21)
where

\[
R_{0NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] \mathbf{U} \mathbf{C}^\top \gamma_i,
\]

\[
R_{1NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] p_{F0} \mathbf{U} \mathbf{C}^\top \gamma_i,
\]

\[
R_{2NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] (M_{F0} - M_{F0}) \mathbf{U} \mathbf{C}^\top \gamma_i.
\]

The same trick can be used to show that

\[
Q_{NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{\hat{F}} \varepsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{\hat{F}0} \varepsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] M_{\hat{F}0} \varepsilon_i
\]

\[
= Q_{0NT} - Q_{1NT} - Q_{2NT},
\]

(S22)

where

\[
Q_{0NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] \varepsilon_i,
\]

\[
Q_{1NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] p_{F0} \varepsilon_i,
\]

\[
Q_{2NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [V'_i - \Gamma'_i(\mathbf{C})'] (M_{F0} - M_{F0}) \varepsilon_i.
\]

From this point on the analysis is similar to the analyses of Chudik et al. (2011), Kapetanios et al. (2010), Pesaran et al. (2013), and Westerlund and Urbain (2015). The key difference is that here we make use of the fact that \( M_{\hat{F}} = M_{\hat{F}0} \), which allows us to proceed in roughly the same way as when using \( \hat{F} \) and assuming that \( \text{rank}(\mathbf{C}) = m = k + 1 \). We therefore put the details Section 3, and just provide here the final result.

**Lemma S.1.** Under Assumption 1 and \( m < k + 1 \), as \( N, T \to \infty \) with \( T/N \to \tau < \infty \),

(i) \( R_{2NT} = \sqrt{T} N^{-1/2} d_1 + O_p(N^{-1/2}); \)

(ii) \( Q_{2NT} = \sqrt{T} N^{-1/2} d_2 + O_p(N^{-1/2}); \)

(iii) \( \|R_{1NT}\| = O_p(T^{-1/2}); \)

(iv) \( R_{0NT} = \sqrt{T} N^{-1/2} (b_1 - b_2) + O_p(N^{-1/2}) + O_p(T^{-1/2}); \)

(v) \( \|Q_{1NT}\| = O_p(T^{-1/2}); \)

(vi) \( Q_{0NT} = (NT)^{-1/2} \sum_{i=1}^{N} V_i' \varepsilon_i - \sqrt{T} N^{-1/2} b_3 + O_p(N^{-1/2}) + O_p(T^{-1/2}), \)

31
where

\[
d_1 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} (\Sigma[\beta, I_k] - \Gamma'(C')^+ \Sigma u) P_{-m} \Sigma u C^+ \gamma_i,
\]

\[
d_2 = \lim_{N \to \infty} (\Sigma[\beta, I_k] - \Gamma'(C')^+ \Sigma u) P_{-m} \sigma^2 [1, 0_{1 \times k}]',
\]

\[
b_1 = \lim_{N \to \infty} \Sigma[\beta, I_k] C^+ \gamma,
\]

\[
b_2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \Gamma_i'(C')^+ \Sigma u C^+ \gamma_i,
\]

\[
b_3 = \lim_{N \to \infty} \Gamma'(C')^+ \sigma^2 [1, 0_{1 \times k}]',
\]

with \( P_{-m} = B_{-m}(B_{-m}' \Sigma u B_{-m})^+ B_{-m} \).

In view of (S21), Lemma S.1 implies that under \( T/N \to \tau < \infty \),

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'M_F (\varepsilon_i - \mathbf{UC}^+ \gamma_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i \varepsilon_i + \sqrt{T} N^{-1/2} (b - d) + o_p(1),
\]

where \( b = b_2 - b_1 - b_3 \) and \( d = d_2 - d_1 \). Moreover, since the fourth order moments of \( v_{i,t} \)

and \( \varepsilon_{i,t} \) are bounded by Assumption 1, by a central limit law for iid processes,

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i \varepsilon_i \xrightarrow{d} N(0_{m \times 1}, \sigma^2 \Sigma)
\]

(S23)

as \( N, T \to \infty \). This completes the analysis of the numerator of \( \sqrt{NT}(\hat{b}_p - \beta) \). What is missing

now is the denominator.

**Lemma S.2.** Under Assumption 1 and \( m \leq k + 1 \), as \( N, T \to \infty \),

\[
\frac{1}{NT} \sum_{i=1}^{N} X_i'M_F X_i = \Sigma + O_p(N^{-1}) + O_p(T^{-1/2}).
\]

The results in Lemmas S.1 and S.2 can be inserted together with (S23) into (S18), leading to

the following asymptotic distribution for \( \sqrt{NT}(\hat{b}_p - \beta) \):

\[
\sqrt{NT}(\hat{b}_p - \beta) = \left( \frac{1}{NT} \sum_{i=1}^{N} X_i'M_F X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i'M_F (\varepsilon_i - \mathbf{UC}^+ \gamma_i)
\]

\[
= \Sigma^{-1} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i \varepsilon_i + \sqrt{T} N^{-1/2} (b - d) \right) + o_p(1)
\]

\[
\xrightarrow{d} N(0_{m \times 1}, \sigma^2 \Sigma^{-1}) + \sqrt{T} \Sigma^{-1} (b - d)
\]

as \( N, T \to \infty \) with \( T/N \to \tau < \infty \). This completes the proof of the theorem. ☐
Remark S.1. The fact that the CCE estimator is biased when $T/N \rightarrow \tau < \infty$ is in agreement with the results reported by Westerlund and Urbain (2015), in which CCE is compared with the principal components estimator of Bai (2009). However, as alluded to in Section 4 of the main text, the reported bias expression is only valid in the special case when $m = k + 1$. Theorem 3 can therefore be seen as a generalization of the results of Westerlund and Urbain (2015) to the case when $m \leq k + 1$. The theorem can also be seen as an extension of the results of Pesaran (2006) who assumes that $T/N \rightarrow 0$, which puts a limit on the applicability of the estimator. This is reflected in the Monte Carlo results reported by Westerlund and Urbain (2015), suggesting that bias correction can lead to substantial improvements in practice when $N$ and $T$ are roughly equal.

3 Proofs of Lemmas S.1 and S.2

Proof of Lemma S.1.

Consider (i). We begin by evaluating $(NT)^{-1/2} \sum_{t=1}^{N} \Gamma'(\overline{C}) + \overline{U} (M_{F_0} - M_{F_0}) \overline{U} \gamma_i$. This one of the terms in $R_{2NT}$. Similarly to (8) in the main text, we have

$$M_{F_0} - M_{F_0} = \overline{U}^0 ([\hat{F}^0][\hat{F}^0] + \overline{U}^{0'}) + \overline{U}^0 ([\hat{F}^0][\hat{F}^0] + (F^0)' + F^0([\hat{F}^0][\hat{F}^0]) + \overline{U}^{0'}.$$  \hspace{1cm} (S25)

Consider the first two terms on the right-hand side of this equation. Making use of (31) and the definition of $\overline{U}^0$, we have

$$T\overline{U}^0([\hat{F}^0][\hat{F}^0] + \overline{U}^{0'}) = \begin{bmatrix} \overline{U}^0_m \overline{U}^0_{m-m} \end{bmatrix} \begin{bmatrix} T^{-1}F'F & 0_{m \times (k+1-m)} \\ 0_{(k+1-m) \times m} & T^{-1}(\overline{U}^0_m)\overline{U}^0_{m-m} \end{bmatrix} + \begin{bmatrix} (\overline{U}^0_m)' \\ (\overline{U}^0_{m-m})' \end{bmatrix}.$$ \hspace{1cm} (S26)

The third term is just the transpose of the second. For the forth and last term, we use

$$\Sigma^+_F = [T^{-1}(F^0)'(F^0)] + \begin{bmatrix} 0_{m \times m} & 0_{m \times (k+1-m)} \\ 0_{(k+1-m) \times m} & [T^{-1}(\overline{U}^0_m)\overline{U}^0_{m-m}] \end{bmatrix}.$$ \hspace{1cm} (S27)

33
implying that, since the last \(k + 1 - m\) columns of \(F^0\) are zero,
\[
F^0 ([T^{-1}(\hat{F}^0)'\hat{F}^0]^+ - [T^{-1}(F^0)'F^0]^+) (F^0)'
= \cdots \text{side of (S30)).
\]
By (A.10) and (A.11) of Lemma 2 of Pesaran (2006),
\[
\|T^{-1}U'\hat{F}^0\| \leq \|T^{-1}U'F^0\| + \|T^{-1}U'UBDN\| = O_p(N^{-1/2}).
\]
\[\text{(S28)}\]

By substituting (S26)–(S28) into (S25) and using the definition of \(\hat{F}^0\), we obtain the following expression for \(T(M_{F^0} - M_{\hat{F}^0})\):
\[
T(M_{F^0} - M_{\hat{F}^0}) = \hat{U}^0_m [T^{-1}(U_{m}^{-1})'(U_{m}^{-1})] + (U_{m}^{-1})' + \hat{U}^0_m (T^{-1}F'F)^+ (U_{m}^{-1})'
+ \hat{U}^0_m (T^{-1}F'F)^+ + F'(T^{-1}F'F)^+ (U_{m}^{-1})'
+ \hat{F}^0 ([T^{-1}(\hat{F}^0)'\hat{F}^0]^+ - \Sigma_{\hat{F}^0}^+) (F^0)'.
\]
\[\text{(S29)}\]
We now make use of this to evaluate \(T^{-1}U' (M_{F^0} - M_{\hat{F}^0}) U\). Direct insertion yields
\[
T^{-1}U' (M_{F^0} - M_{\hat{F}^0}) U = T^{-1}U\hat{U}^0_m [T^{-1}(U_{m}^{-1})'(U_{m}^{-1})] + T^{-1}(U_{m}^{-1})'\hat{U}
+ T^{-1}U\hat{U}^0_m (T^{-1}F'F)^+ T^{-1}(U_{m}^{-1})'\hat{U}
+ T^{-1}U\hat{U}^0_m (T^{-1}F'F)^+ T^{-1}F\hat{U} + T^{-1}U'(T^{-1}F'F)^+ T^{-1}(U_{m}^{-1})'\hat{U}
+ T^{-1}U\hat{F}^0 ([T^{-1}(\hat{F}^0)'\hat{F}^0]^+ - \Sigma_{\hat{F}^0}^+) T^{-1}(\hat{F}^0)'\hat{U}.
\]
Consider the first two terms on the right. Note how \(\hat{U}_{m}^{-1} = \sqrt{N}UB_{m}^{-1}\) and \(\hat{U}_{m}^{-1} = UB_{m}\), and that by (29) (or (A.10) in Lemma 2 of Pesaran, 2006), \(\|T^{-1}U'U\| = O_p(N^{-1})\). It follows that
\[
\|T^{-1}U\hat{U}^0_m [T^{-1}(U_{m}^{-1})'(U_{m}^{-1})] + T^{-1}(U_{m}^{-1})'\hat{U}\|
\leq N\|T^{-1}U'U\|^2 \|B_{m}^{-1}\|^2 \|[T^{-1}(U_{m}^{-1})'(U_{m}^{-1})] + T^{-1}(U_{m}^{-1})'\hat{U}\| = O_p(N^{-1}),
\]
and
\[
\|T^{-1}U\hat{U}^0_m (T^{-1}F'F)^+ T^{-1}(U_{m}^{-1})'\hat{U}\| \leq \|T^{-1}U'U\|^2 \|B_{m}^{-1}\|^2 \||(T^{-1}F'F)^+\| = O_p(N^{-2}).
\]
Further use of \(\|T^{-1}U\hat{F}^0\| = O_p((NT)^{-1/2})\) ((A.11) in Lemma 2 of Pesaran, 2006) yields
\[
\|T^{-1}U\hat{F}^0\| \leq \|T^{-1}U'U\|^2 \|B_{m}^{-1}\|^2 \||(T^{-1}F'F)^+\| = O_p(N^{-3/2}T^{-1/2}),
\]
with \(T^{-1}U'(T^{-1}F'F)^+ T^{-1}(\hat{U}_{m})'U\) having the same order. It remains to consider the last term on the right-hand side of (S30). By (A.10) and (A.11) of Lemma 2 of Pesaran (2006),
\[
\|T^{-1}U\hat{F}^0\| \leq \|T^{-1}U'F^0\| + \|T^{-1}U'UBDN\| = O_p(N^{-1/2}).
\]
\[34\]
By using this result and (33) we can show that
\[ T^{-1} \hat{U}' \hat{F}_0 \left(\left[ T^{-1} (\hat{F}_0)' \hat{F}_0 \right] + - \Sigma^+_m \right) T^{-1} (\hat{F}_0)' \hat{U} \]
\[ \leq \| T^{-1} \hat{U}' \hat{F}_0 \|^2 \left( \| T^{-1} (\hat{F}_0)' \hat{F}_0 \| + - \Sigma^+_m \right) = O_p(N^{-3/2}) + O_p(N^{-1} T^{-1/2}). \]

In view of the above orders, it is clear that (S30) is dominated by the first term on the right-hand side, which can be written in a compact notation as \( T^{-1} \hat{U}' \hat{P}_{\hat{U}_{-m}} \hat{U} \). From \( \hat{U}_{-m}^0 = \sqrt{N} \hat{U}_{-m} \),
\[ \sqrt{NT^{-1} \hat{U}' \hat{U}_{-m}^0} = NT^{-1} \hat{U}' \hat{P}_{\hat{U}_{-m}} \hat{U} = \Sigma_u \hat{B}_{-m} + O_p(T^{-1/2}), \quad \text{(S30)} \]
which, via \( \Sigma_{u_{-m}} = B_{-m} \Sigma_u B_{-m} \), implies
\[ NT^{-1} \hat{U}' \hat{P}_{\hat{U}_{-m}} \hat{U} = \sqrt{NT^{-1} \hat{U}' \hat{U}_{-m}^0} \left( T^{-1} (\hat{U}_{-m}^0)' \hat{U}_{-m}^0 + \sqrt{N} T^{-1} (\hat{U}_{-m}^0)' \hat{U} \right) \]
\[ = \Sigma_u B_{-m} (B_{-m}' \Sigma_u B_{-m}) + B_{-m}' \Sigma_u + O_p(T^{-1/2}) \]
\[ = \Sigma_u P_{-m} \Sigma_u + O_p(T^{-1/2}), \quad \text{(S31)} \]

where \( P_{-m} = B_{-m} (B_{-m}' \Sigma_u B_{-m}) + B_{-m}' \Sigma_u \).

By inserting these results into (S30), we obtain
\[ T^{-1} \hat{U}' (M_{F\hat{0}} - M_{F\hat{0}}) \hat{U} = T^{-1} \hat{U}' \hat{P}_{\hat{U}_{-m}} \hat{U} + O_p(N^{-3/2}) + O_p(N^{-1} T^{-1/2}), \quad \text{(S32)} \]
giving
\[ \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Gamma_i (\overline{C})' + \hat{U}' (M_{F\hat{0}} - M_{F\hat{0}} - \hat{P}_{\hat{U}_{-m}}) \overline{U} \overline{C}^+ \gamma_i \right\| \]
\[ = \sqrt{NT} \left\| \frac{1}{N} \sum_{i=1}^N \Gamma_i (\overline{C})' + T^{-1} [\hat{U}' (M_{F\hat{0}} - M_{F\hat{0}}) \hat{U} - \hat{U}' \hat{P}_{\hat{U}_{-m}} \hat{U} ] \overline{C}^+ \gamma_i \right\| \]
\[ = O_p(\sqrt{TN^{-1}}), \quad \text{(S33)} \]
and so
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^N \Gamma_i (\overline{C})' + \hat{U}' (M_{F\hat{0}} - M_{F\hat{0}}) \overline{U} \overline{C}^+ \gamma_i \]
\[ = \sqrt{TN^{-1/2}} \frac{1}{N} \sum_{i=1}^N \Gamma_i (\overline{C})' + NT^{-1} \hat{U}' \hat{P}_{\hat{U}_{-m}} \overline{U} \overline{C}^+ \gamma_i + O_p(\sqrt{TN^{-1}}) \]
\[ = \sqrt{TN^{-1/2}} \frac{1}{N} \sum_{i=1}^N \Gamma_i (\overline{C})' + \Sigma_u P_{-m} \Sigma_u \overline{C}^+ \gamma_i + O_p(\sqrt{TN^{-1}}). \quad \text{(S34)} \]

Let us now consider \( (NT)^{-1/2} \sum_{i=1}^N V_i (M_{F\hat{0}} - M_{F\hat{0}}) \overline{U} \overline{C}^+ \gamma_i \), which is the remaining term in \( R_{2NT} \). The steps used in the analysis of this term are very similar to those used for evaluating
\( (NT)^{-1/2} \sum_{i=1}^{N} \Gamma_i' (C_i')^+ \vec{U}' (M_{F_0} - M_{F_0}) \vec{U} \). Therefore, only essential details are given. We begin by noting that, in view of (S29),

\[
T^{-1} V_i' (M_{F_0} - M_{F_0}) \vec{U} = T^{-1} V_i' \vec{U}_{m,0} [T^{-1} (\vec{U}_{m,0}') \vec{U}_{m,0}] + T^{-1} (\vec{U}_{m,0}') \vec{U} \\
+ T^{-1} V_i' \vec{U}_{m,0} (T^{-1} F'F) + T^{-1} (\vec{U}_{m,0}') \vec{U} \\
+ T^{-1} V_i' \vec{U}_{m,0} (T^{-1} F'F) + T^{-1} F' \vec{U} + T^{-1} V_i' (T^{-1} F'F) + T^{-1} (\vec{U}_{m,0}') \vec{U} \\
+ T^{-1} V_i' \vec{F}^0 (T^{-1} (\vec{F}^0)' \vec{F}^0) + \Sigma_\rho^+ T^{-1} (\vec{F}^0)' \vec{U}.
\]

According to (A.12) and (A.13) Lemma 2 of Pesaran (2006), we have \( \| T^{-1} V_i' \vec{U} \| = O_p(N^{-1}) + O_p((NT)^{-1/2}) \) and \( \| T^{-1} V_i' F \| = O_p(T^{-1/2}) \). Making use of these results, and the fact that

\[
\| T^{-1} V_i' \vec{F}^0 \| \leq \| T^{-1} V_i' \vec{F}^0 \| + \| T^{-1} V_i' \vec{U} B \vec{D} \| = O_p(T^{-1/2}) + O_p(N^{-1/2}).
\]

we can show the following:

\[
\| T^{-1} V_i' \vec{U}_{m,0} [T^{-1} (\vec{U}_{m,0}') \vec{U}_{m,0}] + T^{-1} (\vec{U}_{m,0}') \vec{U} \| \\
\leq N \| T^{-1} V_i' \vec{U} \| \| B_{m,0} \| \| [T^{-1} (\vec{U}_{m,0}') \vec{U}_{m,0}] \| \| T^{-1} \vec{U} \vec{U} \| = O_p(N^{-1}) + O_p((NT)^{-1/2}),
\]

\[
\| T^{-1} V_i' \vec{U}_{m,0} (T^{-1} F'F) + T^{-1} (\vec{U}_{m,0}') \vec{U} \| \leq \| T^{-1} V_i' \vec{U} \| \| T^{-1} \vec{U} \vec{U} \| \| B_m \| \| (T^{-1} F'F) \| = O_p((NT)^{-1}) + O_p(N^{-3/2} T^{-1/2})
\]

\[
\| T^{-1} V_i' \vec{U}_{m,0} (T^{-1} F'F) + T^{-1} F' \vec{U} \| \leq \| T^{-1} V_i' \vec{U} \| \| T^{-1} F' \vec{U} \| \| B_m \| \| (T^{-1} F'F) \| = O_p((NT)^{-1}) + O_p(N^{-3/2} T^{-1/2})
\]

\[
\| T^{-1} V_i' \vec{F} (T^{-1} F'F) + T^{-1} (\vec{U}_{m,0}') \vec{U} \| \leq \| T^{-1} \vec{U} \vec{U} \| \| T^{-1} F' \vec{U} \| \| B_m \| \| (T^{-1} F'F) \| = O_p(N^{-1} T^{-1/2}),
\]

and

\[
\| T^{-1} V_i' \vec{F}^0 (T^{-1} (\vec{F}^0)' \vec{F}^0) + \Sigma_\rho^+ T^{-1} (\vec{F}^0)' \vec{U} \| \leq \| T^{-1} V_i' \vec{F}^0 \| \| [T^{-1} (\vec{F}^0)' \vec{F}^0] + \Sigma_\rho^+ \| = O_p(N^{-3/2}) + O_p(N^{-1} T^{-1/2}) + O_p(T^{-3/2}).
\]

Direct insertion into (S35) yields

\[
T^{-1} V_i' (M_{F_0} - M_{F_0}) \vec{U} = T^{-1} V_i' \vec{P}_{m,0} \vec{U} + O_p(N^{-3/2}) + O_p(N^{-1} T^{-1/2}) + O_p(T^{-3/2}).
\]

(S35)
For the remaining term, we have
\[ \sqrt{NT}^{-1/2} V'_i P_{\mathcal{U}m} \mathcal{U} = T^{-1/2} V'_i \mathcal{U}^0 \left[ T^{-1} (\mathcal{U}^0)_{-m} \mathcal{U}^0_{-m} \right] + \sqrt{NT}^{-1/2} (\mathcal{U}^0_{-m}) \mathcal{U} \]
\[ = \sqrt{NT}^{-1/2} V'_i \mathcal{U} B_{-m} \left[ T^{-1} \mathcal{U} \mathcal{U} B_{-m} \right] + B_{-m} N T^{-1} \mathcal{U} \mathcal{U} \]
\[ = \sqrt{NT}^{-1/2} V'_i \mathcal{U} P_{-m} \Sigma_u + O_p(T^{-1/2}). \]  
(S36)

Under $T/N = O(1)$, this implies
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i (M_{f0} - M_{f0}) \mathcal{U} \mathcal{C}^{+} \gamma_i \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \sqrt{NT}^{-1/2} V'_i \mathcal{U} P_{\mathcal{U}m} \mathcal{U} \mathcal{C}^{+} \gamma_i + O_p(N^{-1} T^{-1/2}) + O_p(N^{-1/2} T^{-1}) + O_p(\sqrt{NT}^{-2}) \]
\[ = \frac{1}{N} \sum_{i=1}^{N} \sqrt{NT}^{-1/2} V'_i \mathcal{U} P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i + O_p(T^{-1/2}). \]  
(S37)

where the first term is
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i \mathcal{U} P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i \]
\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{j=1}^{N} V'_i V'_j \beta + \frac{1}{N} \sum_{j=1}^{N} V'_i \varepsilon_j, \frac{1}{N} \sum_{j=1}^{N} V'_i V'_j \right) P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i \]
\[ = \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \left[ T^{-1} V'_i V'_i \beta + T^{-1} V'_i \varepsilon_i, T^{-1} V'_i V'_i \right] P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i \]
\[ + \frac{1}{N^{3/2} \sqrt{T}} \sum_{i=1}^{N} \left| V'_i \varepsilon_i \right| + V'_i V'_i \gamma_i \right| P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i. \]  
(S38)

From $E(\|v_{il}\|^4) < \infty$ and $E(\|\varepsilon_{il}\|^4) < \infty$, we get $E(\|T^{-1/2} V'_i \varepsilon_i\|^2) = O(1)$. By using this,
$T^{-1} V'_i V'_i = \Sigma + O_p(T^{-1/2})$ and $T/N = O(1)$, we obtain
\[ \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \left[ T^{-1} V'_i V'_i \beta + T^{-1} V'_i \varepsilon_i, T^{-1} V'_i V'_i \right] P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i \]
\[ = \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \left[ \Sigma_{\beta, I_k} P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i + O_p(T^{-1/2}) \right] \]
\[ = \sqrt{T} N^{-1/2} \Sigma_{\beta, I_k} P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i + O_p(T^{-1/2}). \]  
(S39)

As for the second term on the right-hand side of (S38), tedious but straightforward calculations reveal that $E(\|N^{-1} T^{-1/2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} V'_i \varepsilon_j P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i\|^2)$, $E(\|N^{-1} T^{-1/2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} V'_i V'_i P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i\|^2)$ and $E(\|N^{-1} T^{-1/2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} V'_i V'_i P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i\|^2)$ are all $O(1)$. Hence,
\[ \frac{1}{N^{3/2} \sqrt{T}} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left| V'_i \varepsilon_i + V'_i V'_i \gamma_i \right| P_{-m} \Sigma_u \mathcal{C}^{+} \gamma_i \right| = O_p(N^{-1/2}), \]  
(S40)
implying that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i' \Sigma_{\bar{m}} \bar{C}^+ \gamma_i \quad = \quad \sqrt{T} N^{-1/2} \sum_{i=1}^{N} \left[ T^{-1} V_{i} I_{\bar{m}} \beta + T^{-1} V_{i} I_{\bar{m}} \varepsilon_i \right] P_{\bar{m}} \Sigma_{\bar{m}} \bar{C}^+ \gamma_i + O_p(N^{-1/2}) \\
= \quad \sqrt{T} N^{-1/2} \sum_{i=1}^{N} \left[ T^{-1} V_{i} I_{\bar{m}} \gamma_i \right] + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{S41}
\]

We consequently have
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i' (M_{F0} - M_{\bar{F}0}) \bar{U} \bar{C}^+ \gamma_i \\
= \quad \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i' \Sigma_{\bar{m}} \bar{C}^+ \gamma_i + O_p(T^{-1/2}) \\
= \quad \sqrt{T} N^{-1/2} \sum_{i=1}^{N} \left[ T^{-1} V_{i} \gamma_i \right] P_{\bar{m}} \Sigma_{\bar{m}} \bar{C}^+ \gamma_i + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{S42}
\]

This establishes (i).

Consider (ii). We have
\[
Q_{2NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i' (M_{F0} - M_{\bar{F}0}) \varepsilon_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma_i (\bar{C}^+) \bar{U} (M_{F0} - M_{\bar{F}0}) \varepsilon_i,
\]
where, in view of (S29),
\[
T^{-1} \bar{U}' (M_{F0} - M_{\bar{F}0}) \varepsilon_i = T^{-1} \bar{U}' \bar{U}_m^{0} [T^{-1} (\bar{U}_m^{0})' \bar{U}_m^{0}] + T^{-1} (\bar{U}_m^{0})' \varepsilon_i \\
+ T^{-1} \bar{U}' \bar{U}_m^{0} (T^{-1} \bar{F} \bar{F}) + T^{-1} (\bar{U}_m^{0})' \varepsilon_i \\
+ T^{-1} \bar{U}' \bar{U}_m^{0} (T^{-1} \bar{F} \bar{F}) + T^{-1} \bar{U}' \bar{F} (T^{-1} \bar{F} \bar{F}) + T^{-1} (\bar{U}_m^{0})' \varepsilon_i \\
+ T^{-1} \bar{U}' \bar{F} (T^{-1} (\bar{F} \bar{F})' - \Sigma_{\bar{F}0}) T^{-1} (\bar{F} \bar{F})' \varepsilon_i.
\]

According to (A.13) in Lemma 2 of Pesaran (2006), \( T^{-1} \bar{U}' \varepsilon_i = O_p(N^{-1}) + O_p((NT)^{-1/2}) \), and in analogy to (A.12) of the same lemma, we can show that \( \| T^{-1} \bar{F} \varepsilon_i \| = O_p(T^{-1/2}) \). This implies
\[
\| T^{-1} \bar{U}' \bar{U}_m^{0} [T^{-1} (\bar{U}_m^{0})' \bar{U}_m^{0}] + T^{-1} (\bar{U}_m^{0})' \varepsilon_i \| \\
\leq N \| T^{-1} \bar{U}' \bar{U} \| \| B_{\bar{m}} \|^2 \| T^{-1} (\bar{U}_m^{0})' \bar{U}_m^{0} \| \| T^{-1} \bar{U}' \varepsilon_i \| = O_p(N^{-1}) + O_p((NT)^{-1/2}),
\]

38
\[
T^{-1}U_0^m(T^{-1}F^+T^{-1}(U_0^m)'\varepsilon_i) \leq \|T^{-1}U'U\| \|T^{-1}U'_\varepsilon_i\| \|B_m\|^2 \|(T^{-1}F^+)\| \\
= O_p(N^{-2}) + O_p(N^{-3/2}T^{-1/2}) ,
\]
\[
T^{-1}U_0^m(T^{-1}F^+T^{-1}F')\varepsilon_i \leq \|T^{-1}U'U\| \|T^{-1}F'_\varepsilon_i\| \|B_m\| \|(T^{-1}F^+)\| \\
= O_p(N^{-1}T^{-1/2}) ,
\]
\[
T^{-1}U_0^m(F(T^{-1}F^+)+T^{-1}U_0^m)\varepsilon_i \leq \|T^{-1}U'U\| \|T^{-1}U'_\varepsilon_i\| \|B_m\| \|(T^{-1}F^+)\| \\
= O_p((NT)^{-1}) + O_p(N^{-3/2}T^{-1/2}) ,
\]
and since
\[
\|T^{-1}(F^0)\varepsilon_i\| \leq \|T^{-1}(F^0)'\varepsilon_i\| + \|T^{-1}D'_mB'U'_\varepsilon_i\| = O_p(T^{-1/2}) + O_p(N^{-1/2}) ,
\]
we also have
\[
\|T^{-1}U_0^m \hat{F}^0([T^{-1}(F^0)'\varepsilon_i] + \Sigma_p^B)T^{-1}(F^0)'\varepsilon_i\| \\
\leq \|T^{-1}U_0^m \hat{F}^0\| \|T^{-1}(F^0)'\varepsilon_i\| \|T^{-1}(F^0)'\varepsilon_i\| + \|\Sigma_p^B\| \\
= O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) ,
\]
It follows that
\[
T^{-1}U_0'(M_{F0} - M_{F0})\varepsilon_i = T^{-1}U_0'P_{U_0}^m\varepsilon_i + O_p(N^{-3/2}) + O_p(N^{-1}T^{-1/2}) \\
+ O_p(T^{-3/2}) ,
\]
and so
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + U'(M_{F0} - M_{F0})\varepsilon_i \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + U'P_{U_0}^m\varepsilon_i + O_p(\sqrt{T}N^{-3/2}) + O_p(N^{-1/2}) \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + \Sigma_uP_{-m}\varepsilon_i + O_p(T^{-1/2}) + O_p(N^{-1/2}) ,
\]
where the last equality follows from imposing \(T/N = O(1)\), and then using the same calculations as when evaluating \((NT)^{-1/2}\sum_{i=1}^{N} V_j\hat{U}P_{-m}\Sigma_u\hat{C}^+\gamma_i\). We also used the fact that \(P_{-m} = P'_{-m}\). For the remaining term, note how
\[
\frac{1}{T} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + \Sigma_uP_{-m}\varepsilon_i \\
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + \Sigma_uP_{-m}\varepsilon_i \\
= \left[ (NT)^{-1} \sum_{i=1}^{N} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + \Sigma_uP_{-m}\varepsilon_i + (NT)^{-1} \sum_{i=1}^{N} \sum_{i=1}^{N} \Gamma_i'(\hat{C}') + \Sigma_uP_{-m}\varepsilon_i \right] .
\]
By applying the same argument used earlier for showing $\|N^{-1}T^{-1/2} \Sigma_{i=1}^{N} \Sigma_{j \neq i}^{N} \Sigma_{t=1}^{T} u_{i,t} u_{j,t}'\| = O_{p}(1)$, we obtain $(NT)^{-1} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \varepsilon \varepsilon_i = O_{p}(T^{-1/2})$. This implies
$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \varepsilon \varepsilon_i = \frac{1}{NT} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \varepsilon \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \varepsilon \varepsilon_i$$
$$= \frac{1}{N} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} T^{-1} \varepsilon \varepsilon_i + O_{p}(T^{-1/2})$$
$$= \sigma^2 \frac{1}{N} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} + O_{p}(T^{-1/2}),$$
where the last result is due to $T^{-1} \varepsilon \varepsilon_i = \sigma^2 + O_{p}(T^{-1/2})$. But $\| (NT)^{-1} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} V \varepsilon_i \|$ and $\| (NT)^{-1} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \beta \varepsilon \varepsilon_i \|$ are $O_{p}(T^{-1/2})$ too, and therefore
$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} V \varepsilon_i \right\| \leq \left\| \frac{1}{NT} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} V \varepsilon_i \right\| + \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \varepsilon \varepsilon_i \right\|$$
$$= O_{p}(\sqrt{NT}) + O_{p}(T^{-1/2}) = O_{p}(T^{-1/2}),$$
with $\| (NT)^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \beta \varepsilon \varepsilon_i \|$ being of the same order. This implies
$$\frac{1}{T} \sum_{i=1}^{N} \Gamma_i'\tilde{C}) + \Sigma_{u} P_{-m} U \varepsilon_i = \frac{1}{N} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} \sigma^2 [1, 0_{1 \times k}]' + O_{p}(T^{-1/2})$$
$$= \Gamma'\tilde{C}) + \Sigma_{u} P_{-m} \sigma^2 [1, 0_{1 \times k}]' + O_{p}(T^{-1/2}), \quad (S45)$$
which can be substituted back into (S44), giving
$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \tilde{U}'(M_{F_0} - M_{\hat{F}_0}) \varepsilon_i$$
$$= \sqrt{T} N^{-1/2} \frac{1}{T} \sum_{i=1}^{N} \Gamma_i'(\tilde{C}) + \Sigma_{u} P_{-m} U \varepsilon_i + O_{p}(T^{-1/2}) + O_{p}(N^{-1/2})$$
$$= \sqrt{T} N^{-1/2} \tilde{U}'(\tilde{C}) + \Sigma_{u} P_{-m} \sigma^2 [1, 0_{1 \times k}]' + O_{p}(T^{-1/2}) + O_{p}(N^{-1/2}). \quad (S46)$$

The arguments used for evaluating $T^{-1} \tilde{U}'(M_{F_0} - M_{\hat{F}_0}) \varepsilon_i$ can be applied also to $T^{-1} V'(M_{F_0} -
\[ M_{F0} \varepsilon_i, \text{ giving} \]
\[ T^{-1}V_i'(M_{F0} - M_{F0})\varepsilon_i \]
\[ = T^{-1}V_i'[U_{m}'(T^{-1}(\hat{U}_{m}' - U_{m}'))\varepsilon_i + T^{-1}(\hat{U}_{m}')\varepsilon_i] \quad (S47) \]
\[ + T^{-1}V_i'[U_{m}'(T^{-1}F'F) + T^{-1}(\hat{U}_{m}')\varepsilon_i + T^{-1}V_i'F(T^{-1}F'F) + T^{-1}(\hat{U}_{m}')\varepsilon_i] \]
\[ + T^{-1}V_i'[\hat{F}^0((T^{-1}(\hat{F}^0)'\hat{F}^0) - \Sigma_{F0}^0)T^{-1}(\hat{F}^0)']\varepsilon_i \]
\[ = T^{-1}V_i'[P_{U_{m}'} \varepsilon_i + O_p(T^{-3/2}) + O_p(T^{-1}N^{-1/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-3/2})]. \quad (S48) \]

This implies
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i'(M_{F0} - M_{F0})\varepsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i'[P_{U_{m}'} \varepsilon_i + O_p(\sqrt{T}N^{-3/2}) + O_p(N^{-1/2})], \quad (S49) \]

where, by using the same steps as before,
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i'[P_{U_{m}'} \varepsilon_i \]
\[ = \sqrt{\frac{N}{T^{3/2}}} \sum_{i=1}^{N} V_i'[U_{m}'B_{m}(B_{m}'\Sigma_u B_{m})'B_{m}' + B_{m}'U_{m}' \varepsilon_i + O_p(T^{-1/2}) + O_p(N^{-1/2}) \]
\[ = \sqrt{\frac{T}{N}}N^{-1/2} \Sigma [\beta, I_k]B_{m}(B_{m}'\Sigma_u B_{m})'B_{m}' + B_{m}'\varepsilon_i[1, 0_{1 \times k}] + O_p(T^{-1/2}) + O_p(N^{-1/2}) \]
\[ = \sqrt{\frac{T}{N}}N^{-1/2} \Sigma [\beta, I_k]P_{m} \sigma^2[1, 0_{1 \times k}] + O_p(T^{-1/2}) + O_p(N^{-1/2}). \quad (S50) \]

The result in (ii) is implied by this.

Next up is (iii). Clearly,
\[ R_{1NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i'[P_{F0} U' \Sigma] + \gamma_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V_i'[F_i'(\Sigma')] + U'P_{F0} U' \Sigma + \gamma_{i \varepsilon}. \]

From (13), we know that \( \|T^{-1}(F'F)^{\gamma} \| = O_p(1) \), and by (A.11) and (A.12) in Lemma 2 of Pesaran (2006), we also have \( \|T^{-1}V_i'F^0\| = O_p(T^{-1/2}) \) and \( \|T^{-1}U_i'F^0\| = O_p((NT)^{-1/2}) \). These results imply
\[ \|T^{-1}V_i'[P_{F0} U\| \leq \|T^{-1}V_i'[F^0\| \|T^{-1}(F'F)^{\gamma}\| \|T^{-1}(F'F)^{\gamma}U\| = O_p(N^{-1/2}T^{-1}), \]
\[ \|T^{-1}U_i'[P_{F0} U\| \leq \|T^{-1}U_i'[F^0\| \|T^{-1}(F'F)^{\gamma}\| = O_p((NT)^{-1}), \]

41
which can in turn be used to show that

\[
\| R_{1NT} \| \leq \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i P_{F_0} \Sigma_{F_0}^{+} \gamma_i \right\| + \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma'_i (\Sigma')^{+} U' P_{F_0} \Sigma_{F_0}^{+} \gamma_i \right\|
\]

\[
\leq \sqrt{NT} \frac{1}{N} \sum_{i=1}^{N} \| T^{-1} V'_i P_{F_0} \Sigma_{F_0}^{+} \gamma_i \| + \sqrt{NT} \frac{1}{N} \sum_{i=1}^{N} \| T^{-1} \Gamma'_i (\Sigma')^{+} U' P_{F_0} \Sigma_{F_0}^{+} \gamma_i \|
\]

\[
\leq \| \Sigma^{+} \| \sqrt{NT} \frac{1}{N} \| T^{-1} V'_i P_{F_0} \Sigma_{F_0} \| \| \gamma_i \|
\]

\[
+ \sqrt{NT} \| (\Sigma')^{+} \| \sum_{i=1}^{N} \| T^{-1} U' P_{F_0} \Sigma_{F_0} \| \| \Sigma^{+} \| \frac{1}{N} \sum_{i=1}^{N} \| \Gamma'_i \| \| \gamma_i \|
\]

\[
= O_p(T^{-1/2}),
\]

(S51)

as required for (iii).

Part (iv) follows from the same arguments used in the proof (i). Specifically, we have

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma'_i (\Sigma')^{+} U' \Sigma_{F_0}^{+} \gamma_i = \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \Gamma'_i (\Sigma')^{+} + \Sigma_{F_0}^{+} \gamma_i + O_p(T^{-1/2}),
\]

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i \Sigma_{F_0}^{+} \gamma_i = \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \Sigma [\beta, I_k] \Sigma_{F_0}^{+} \gamma_i + O_p(N^{-1/2}) + O_p(T^{-1/2}),
\]

implying

\[
R_{0NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i \Sigma_{F_0}^{+} \gamma_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma'_i (\Sigma')^{+} U' \Sigma_{F_0}^{+} \gamma_i
\]

\[
= \sqrt{T} N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} (\Sigma [\beta, I_k] \Sigma_{F_0}^{+} \gamma_i - \Gamma'_i (\Sigma')^{+} \Sigma_{F_0}^{+} \gamma_i) + O_p(N^{-1/2})
\]

\[
+ O_p(T^{-1/2}),
\]

(S52)

as was to be shown.

Next, consider (v). We begin by noting how

\[
\| T^{-1} U' P_{F_0} \epsilon_i \| \leq \| T^{-1} U' F_0' \| \| T^{-1} (F_0' F_0)^{-1} \| \| T^{-1} F_0' \epsilon_i \| = O_p(N^{-1/2} T^{-1}),
\]

implying

\[
\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma'_i (\Sigma')^{+} U' P_{F_0} \epsilon_i \right\| \leq \sqrt{NT} \frac{1}{N} \sum_{i=1}^{N} \| T^{-1} \Gamma'_i (\Sigma')^{+} U' P_{F_0} \epsilon_i \| = O_p(T^{-1/2}).
\]

(S53)

Since \( E(\epsilon; \epsilon'_i) = \sigma^2 I_T \) and

\[
\| T^{-1} V'_i P_{F_0} V_i \| \leq \| T^{-1} V'_i F_0' \| \| T^{-1} (F_0' F_0)^{-1} \| = O_p(T^{-1}),
\]

42
we also have
\[
E \left( \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i P_{F0} \varepsilon_i \right\|^2 \right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{tr}[E(V'_i P_{F0} \varepsilon_i P_{F0} V_j)] = \frac{1}{N} \sum_{i=1}^{N} \text{tr}[E(T^{-1} V'_i P_{F0} V_j)] = O(T^{-1}).
\] (S54)

We can therefore show that
\[
\|Q_{1NT}\| \leq \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i P_{F0} \varepsilon_i \right\| + \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Gamma'_i (\mathcal{C}')^+ U' \varepsilon_i \right\| = O_p(T^{-1/2}),
\] (S55)
as required for (v).

Finally, consider (vi). By using the same steps as in the proof of (ii), we obtain
\[
\frac{1}{T} \sum_{i=1}^{N} \Gamma'_i (\mathcal{C}')^+ U' \varepsilon_i = \frac{1}{N} \sum_{i=1}^{N} \Gamma'_i (\mathcal{C}')^+ \sigma^2 [1, 0_{1 \times k}]' + O_p(T^{-1/2}),
\] (S56)
and so
\[
Q_{0NT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i \varepsilon_i - \sqrt{TN^{-1/2}} \frac{1}{T} \sum_{i=1}^{N} \Gamma'_i (\mathcal{C}')^+ U' \varepsilon_i
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} V'_i \varepsilon_i - \sqrt{TN^{-1/2}} \frac{1}{N} \sum_{i=1}^{N} \Gamma'_i (\mathcal{C}')^+ \sigma^2 [1, 0_{1 \times k}]' + O_p(N^{-1/2})
\] + \(O_p(T^{-1/2}).
\] (S57)

This establishes (vi), and hence the proof of Lemma S.1 is complete.  

**Proof of Lemma S.2.**

Direct substitution from (S20) gives
\[
T^{-1} X'_i M_{F0} X_i = T^{-1} X'_i M_{F0} X_i
\]
\[
= T^{-1} \Gamma'_i (\mathcal{C}')^+ U' M_{F0} \mathcal{U}^+ \Gamma_i - T^{-1} \Gamma'_i (\mathcal{C}')^+ U' M_{F0} V_i
\]
\[
- T^{-1} V'_i M_{F0} \mathcal{U}^+ \Gamma_i + T^{-1} V'_i M_{F0} V_i.
\] (S58)

Consider
\[
\|T^{-1} X'_i (M_{F0} - M_{F0}) X_i\|
\]
\[
= \|T^{-1} \Gamma'_i (\mathcal{C}')^+ U'(M_{F0} - M_{F0}) \mathcal{U}^+ \Gamma_i\| + 2 \|T^{-1} \Gamma'_i (\mathcal{C}')^+ U' (M_{F0} - M_{F0}) V_i\|
\]
\[
+ \|T^{-1} V'_i (M_{F0} - M_{F0}) V_i\|.
\] (S59)
All terms here are known from before, except $\|T^{-1}V'_i(M_{F_0} - M_{\hat{F}_0})V_i\|$, which is of the same order as $\|T^{-1}V'_i(M_{F_0} - M_{\hat{F}_0})\|$. Insertion and simplification yields

$$\|T^{-1}X'_i(M_{F_0} - M_{\hat{F}_0})X_i\| = O_p(N^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}). \quad (S60)$$

By further use of the known orders of $\|T^{-1}\bar{U}'F^0\|$, $\|T^{-1}(F^0)'V_i\|$ and $\|[(T^{-1}(F^0)'F^0)]^{-1}\|$, we can show that $\|T^{-1}\bar{U}'P_{F_0}\bar{U}\| = O_p((NT)^{-1})$, $\|T^{-1}\bar{U}'(_C\bar{I}^*)'\bar{U}'P_{F_0}V_i\| = O_p(N^{-1/2}T^{-1})$ and $\|T^{-1}V'_iP_{F_0}V_i\| = O_p(T^{-1})$, which in turn implies

$$\|T^{-1}X'_iP_{F_0}X_i\| = \|T^{-1}\bar{U}'(_C\bar{I}^*)'\bar{U}'P_{F_0}(_C\bar{I}^*)_{T_i}\| + 2\|T^{-1}\bar{U}'(_C\bar{I}^*)'\bar{U}'P_{F_0}V_i\| + \|T^{-1}V'_iP_{F_0}V_i\| = O_p(T^{-1}). \quad (S61)$$

By using these results and

$$T^{-1}X_i = T^{-1}\bar{U}'(_C\bar{I}^*)'\bar{U}'(_C\bar{I}^*)_{T_i} - T^{-1}\bar{U}'(_C\bar{I}^*)'\bar{U}'V_i - T^{-1}V'_i(_C\bar{I}^*)_{T_i} + T^{-1}V'_iV_i$$

$$= T^{-1}V'_iV_i + O_p(N^{-1}) + O_p((NT)^{-1/2})$$

$$= \Sigma + O_p(N^{-1}) + O_p(T^{-1/2}), \quad (S62)$$

we obtain

$$\frac{1}{NT} \sum_{i=1}^{N} X'_iM_{F_i}X_i = \frac{1}{N} \sum_{i=1}^{N} T^{-1}X'_iX_i - \frac{1}{N} \sum_{i=1}^{N} T^{-1}X'_iP_{F_0}X_i - \frac{1}{N} \sum_{i=1}^{N} T^{-1}X'_i(M_{F_0} - M_{\hat{F}_0})X_i$$

$$= \Sigma + O_p(N^{-1}) + O_p(T^{-1/2}), \quad (S63)$$

as was to be shown. ■
References


Paper II
PANICCA: PANIC ON CROSS-SECTION AVERAGES

SIMON REESEA AND JOAKIM WESTERLUNDB

a Department of Economics, Lund University, Sweden
b Centre for Economic and Financial Econometric Research, Deakin University, Melbourne, Australia

SUMMARY

The cross-section average (CA) augmentation approach of Pesaran (A simple panel unit root test in presence of cross-section dependence. Journal of Applied Econometrics 2007; 22: 265–312) and Pesaran et al. (Panel unit root test in the presence of a multifactor error structure. Journal of Econometrics 2013; 175: 94–115), and the principal components-based panel analysis of non-stationarity in idiosyncratic and common components (PANIC) of Bai and Ng (A PANIC attack on unit roots and cointegration. Econometrica 2004; 72: 1127–1177; Panel unit root tests with cross-section dependence: a further investigation. Econometric Theory 2010; 26: 1088–1114) are among the most popular ‘second-generation’ approaches for cross-section correlated panels. One feature of these approaches is that they have different strengths and weaknesses. The purpose of the current paper is to develop PANICCA, a combined approach that exploits the strengths of both CA and PANIC. Copyright © 2015 John Wiley & Sons, Ltd.

Received 21 October 2014; Revised 27 July 2015

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Consider the panel data variable \( Y_{it} \), observable for \( t = 1, \ldots, T \) time periods and \( i = 1, \ldots, N \) cross-section units. It is well known that unattended cross-section dependence can lead to deceptive inference when testing the null hypothesis of a unit root in such variables. This is certainly true for panel unit root tests devised to test the hypothesis that \( Y_{1,t}, \ldots, Y_{N,t} \) are jointly unit root non-stationary, but the problem is there also when applying univariate unit root tests to each cross-section unit. This finding has led to the development of factor-based ‘second-generation’ test procedures that are robust to cross-section dependence (see Breitung and Pesaran, 2008; Baltagi, 2013, ch. 12, for surveys of the literature). Two of the most popular second-generation tests are the cross-section augmented Im–Pesaran–Shin (CIPS) and Sargan–Bhargava (CSB) tests of Pesaran (2007) and Pesaran et al. (2013). In fact, these tests have in a short period of time become two of the industry’s workhorses, with a large number of applications and also several theoretical extensions (see, for example, Westerlund, 2015a; Westerlund et al., 2015).

As the name suggests, the idea underlying the cross-section average (CA) augmentation approach, originally put forth by Pesaran (2006) in the context of factor-augmented panel regressions, is to use the cross-section average \( \bar{Y}_t \) of \( Y_{it} \) as a proxy for the common component of the data, which is then included in the regression as an additional regressor. But if \( Y_{it} \) is unit root non-stationary then so is \( \bar{Y}_t \), suggesting that, in analogy to the spurious regression phenomenon, the asymptotic distributions of the resulting CIPS and CSB statistics will depend on the Brownian motion generated by \( \bar{Y}_t \). They will therefore be highly non-standard, which in turn makes for complicated implementation. In particular, not only is it necessary to tabulate critical values for each constellation of \( (N, T) \), but there is also a need to truncate the test statistics in order to ensure finite moments. As Pesaran et al. (2013)
show, however, when properly implemented, the CIPS and CSB tests do seem to enjoy relatively good small-sample performance, which is partly expected given the relatively good performance of the CA components estimator (see, for example, Chudik et al., 2011; Kapetanios and Pesaran, 2005; Westerlund and Urbain, 2015). Another feature of the CIPS and CSB statistics is that they assume that the common and idiosyncratic components of the data have the same order of integration, which of course need not be the case in practice.

An alternative test approach that supports asymptotically normal inference is the panel analysis of non-stationarity in idiosyncratic and common components (PANIC) of Bai and Ng (2004, 2010). This approach, which, in contrast to CA augmentation, does not require the common and idiosyncratic components to be integrated of the same order, is arguably the most popular approach in the literature, with even more applications and extensions than CA (see, for example, Bai and Carrion-i-Silvestre, 2009, 2013; Gengenbach et al., 2006; Westerlund, 2014; Westerlund and Hess, 2011; Westerlund and Larsson, 2012). The basic idea in PANIC is to first transform $Y_{it}$ by taking first differences. The method of principal components (PC) is then applied to estimate the first-differenced common and idiosyncratic components, which can be cumulated up to levels. The fact that the components are estimated from a regression in first differences means that the spurious regression problem is avoided, thereby enabling standard normal inference. As the bulk of the existing Monte Carlo evidence show (see, for example, Gengenbach et al., 2006, 2010; Pesaran et al., 2013; Westerlund and Larsson, 2009; Westerlund and Urbain, 2015), however, the use of PC can render PANIC small-sample distorted, especially when $N$ is ‘small’.

The purpose of the present paper is to propose a test procedure that is both general and simple, yet with good small-sample performance. In view of the above discussion, a natural suggestion towards this end is to use PANIC, but to apply it to the estimated CA components rather than to the estimated PC components. As far as we are aware, this is the first attempt to exploit the advantages of both CA and PANIC. The properties of the resulting PANICCA procedure are studied under the condition that the number of panel data variables is at least as large as the number of common factors. Our key findings can be summarized as follows. First, PANICCA inherits the generality of PANIC and enables inference regarding the unit root and cointegration properties of both the common and idiosyncratic components of the data. PANICCA can therefore be seen as a complete panel unit root toolbox. Second, being based on simple CA, PANICCA is very user friendly. In fact, in view of its generality, it is surprisingly simple, requiring nothing but basic averaging and least squares (LS) operations. To make the life of practitioners even simpler, a full suite of GAUSS programs has been made available at http://sites.google.com/site/perjoakimwesterlund/. Third, PANICCA leads to the same asymptotic theory as PANIC. Appropriate critical values can therefore be taken directly from Bai and Ng (2004, 2010). Fourth, the use of CA rather than PC leads to much improved small-sample performance, especially in the type of small- to medium-$N$ panels often encountered in applied work (for a non-exhaustive list, see Lanzafame, 2010; Schmidt and Vosen, 2013; Martín, ; Joseph et al., 2012, 2013; Örsal and Dilan, ; Blomquist and Westerlund, 2014).

In our empirical application we consider an old empirical puzzle within financial economics, namely the failure of the efficient market hypothesis (EMH). Here we demonstrate the usefulness of the generality of PANICCA, as a platform for testing for cointegration both within and between cross-section units. According to EMH, not only should the current forward rate be cointegrated with the future spot rate, but there should also not be any cointegration running across currencies. Interestingly, while separately these cointegrating restrictions have been subject to countless tests (for early contributions see, for example, Hakkio and Rush, 1989; Bailey and Bollerslev, 1989; Crowder, 1994), as far as we are aware, the current paper is the first to consider a joint test of both restrictions.

The balance of the paper is organized as follows. In Section 2 we lay out the assumptions that we will be working under and explain how these compare to the assumptions of PANIC. Section 3 provides an account of the PANICCA procedure and its asymptotic properties, whose accuracy in small samples is...
studied by means of Monte Carlo simulation in Section 4. Section 5 contains the empirical application and Section 6 concludes.

2. MODEL AND ASSUMPTIONS

The data-generating process (DGP) of \( Y_{i,t} \) is assumed to be given by the following common factor model:

\[
Y_{i,t} = \alpha_i' \mathbf{D}_{t,p} + \lambda_i' \mathbf{F}_t + \epsilon_{i,t}
\]

(1)

where \( \epsilon_{i,t} \) is a scalar idiosyncratic error, \( \mathbf{F}_t \) is an \( r \times 1 \) vector of common factors with \( \lambda_i \) being the associated \( (r \times 1) \) vector of loading coefficients, and \( \mathbf{D}_{t,p} = (1, \ldots, t^p)' \) is a \( (p + 1) \times 1 \) vector of trends, for which we consider two specifications: (i) a constant \( (p = 0) \) and (ii) a constant and trend \( (p = 1) \). In this paper, \( Y_{i,t} \) is considered as the variable of interest. However, we do allow for the presence of an \( m \times 1 \) vector of additional variables, henceforth denoted \( X_{i,t} \), whose DGP is given by

\[
X_{i,t} = \beta_i' \mathbf{D}_{t,p} + \mathbf{A}_i' \mathbf{F}_t + \mathbf{u}_{i,t}
\]

(2)

where \( \mathbf{u}_{i,t} \) is an \( m \times 1 \) vector of idiosyncratic errors. Thus, as in Pesaran et al. (2013), we assume the existence of an additional \( m \) variables that are permitted (but not required; see Remark 1) to share the common factors of the variable of interest. This seems very plausible, especially in macroeconomics and finance, where most variables are highly co-moving (see Section 5 for an empirical illustration).

Define \( \mathbf{Z}_{i,t} = \left( Y_{i,t}, X_{i,t}' \right)' \). In view of equations (1) and (2), the DGP of this variable is easily seen to be given by

\[
\mathbf{Z}_{i,t} = \mathbf{B}_i' \mathbf{D}_{t,p} + \mathbf{C}_i' \mathbf{F}_t + \mathbf{V}_{i,t}
\]

(3)

where \( \mathbf{B}_i = (\alpha_i, \beta_i), \mathbf{C}_i = (\lambda_i, \mathbf{A}_i) \) and \( \mathbf{V}_{i,t} = \left( \epsilon_{i,t}, \mathbf{u}_{i,t}' \right)' \). Note that the dimension of \( \mathbf{C}_i \) is \( r \times (m + 1) \). The conditions under which we will be working are summarized below. Here and throughout this paper \( \text{tr}(\mathbf{A}), \text{rk}(\mathbf{A}) \) and \( \|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})} \) denote the trace, rank and Frobenius (Euclidean) norm, respectively, of the matrix \( \mathbf{A}, \mathbf{A} = N^{-1} \sum_{i=1}^N \mathbf{A}_i \) for any \( \mathbf{A}_i \), and \( M < \infty \) is a generic positive number.

Assumption 1. \((1 - \rho_1) \epsilon_{i,t} = \phi_i(L) \epsilon_{i,t} \), where \( \phi_i(L) = \sum_{n=0}^\infty \phi_{i,n} L^n \) with \( \sum_{n=0}^\infty n|\phi_{i,n}| \leq M \) and \( \phi_i(1) > 0 \), and \( \epsilon_{i,t} \) is independently and identically distributed (i.i.d.) across both \( i \) and \( t \) with \( E(\epsilon_{i,t}) = 0, E\left( \epsilon_{i,t}^2 \right) = 1 \) and \( E(\|\epsilon_{i,t}\|) \leq M \).

Assumption 2. \( \Delta \mathbf{u}_{i,t} = \Psi_i(L) \mathbf{e}_{i,t} \), where \( \Psi_i(L) = \sum_{n=0}^\infty \psi_{i,n} L^n \) with \( \sum_{n=0}^\infty n|\psi_{i,n}| \leq M \) and \( \text{rk}[\Psi_i(1)] = m_1 \in [0, m], \text{var}(\Delta \mathbf{u}_{i,t}) = \sum_{n=0}^\infty \psi_n \psi_n' \) is positive definite, and \( \mathbf{e}_{i,t} \) is i.i.d. across both \( i \) and \( t \) with \( E(\mathbf{e}_{i,t}) = 0_{m \times 1}, E\left( \mathbf{e}_{i,t} \mathbf{e}_{i,t}' \right) = \text{Im} \) and \( E(\|\mathbf{e}_{i,t}\|) \leq M \).

Assumption 3. \( \Delta \mathbf{F}_t = \Phi(L) \eta_t \), where \( \Phi(L) = \sum_{n=0}^\infty \Phi_n L^n \) with \( \sum_{n=0}^\infty n|\Phi_n| \leq M \) and \( \text{rk}[\Phi(1)] = r_1 \in [0, r], \text{var}(\Delta \mathbf{F}_t) = \sum_{n=0}^\infty \Phi_n \Sigma_n \Phi_n' \) is positive definite, and \( \eta_t \) is i.i.d. across \( t \) with \( E(\eta_t) = 0_{(m+1) \times 1}, E\left( \eta_t \eta_t' \right) = \Sigma_n \) and \( E(\|\eta_t\|) \leq M \).

Assumption 4. \( \mathbf{C}_i \) is a non-random vector satisfying \( \|\mathbf{C}_i\| \leq M \) and \( \text{rk}(\mathbf{C}) = r \leq m + 1 \) for any \( N \), including \( N \to \infty \).

---

Copyright © 2015 John Wiley & Sons, Ltd.  
DOI: 10.1002/jae
Assumption 5. \( \varepsilon_{i,t} \) and \( \eta_i \) are mutually independent.

Assumption 6. \( E(\|\mathbf{F}_0\|) \leq M \) and \( E(|\varepsilon_{i,0}|) \leq M \) for all \( i \).

Denote by \( \sigma^2_{\varepsilon, i} = \sum_{n=0}^{\infty} \phi^2_{i,n} \omega^2_{\varepsilon, i} = \phi_i(1)^2 \) and \( \tau_{\varepsilon, i} = \left( \omega^2_{\varepsilon, i} - \sigma^2_{\varepsilon, i} \right) / 2 \) the contemporaneous, long-run and one-sided long-run variance of \( \varepsilon_{i,t} \), respectively. Let us further denote by \( \bar{\sigma}^2_{\varepsilon}, \bar{\omega}^2_{\varepsilon} \bar{\phi}_e^4 \) and \( \bar{\lambda}_e \) the cross-sectional averages of \( \sigma^2_{\varepsilon, i}, \omega^2_{\varepsilon, i}, \phi^4_{\varepsilon, i} \) and \( \tau_{\varepsilon, i} \), respectively, where \( \phi^4_{\varepsilon, i} = \omega^4_{\varepsilon, i} \) (to avoid confusion between \( (\bar{\omega}^2)^2 \) and \( N^{-1} \sum_{i=1}^{N} \omega^4_{\varepsilon, i} \)).

Assumption 7. \( \bar{\sigma}^2_{\varepsilon} \rightarrow \sigma^2_{\varepsilon}, \bar{\omega}^2_{\varepsilon} \rightarrow \omega^2_{\varepsilon} \bar{\phi}_e^4 \rightarrow \phi^4_{\varepsilon} \) and \( \bar{\lambda}_e \rightarrow \lambda_e \) as \( N \rightarrow \infty \), where \( \sigma^2_{\varepsilon}, \omega^2_{\varepsilon}, \phi^4_{\varepsilon} \in (0, M) \) and \( |\lambda_e| \leq M \).

The above conditions are very similar to those employed by Bai and Ng (2004, 2010), and we therefore refer to these previous works for a detailed discussion. The main differences are: (i) the assumed presence of the \( m \times 1 \) vector \( \mathbf{X}_{i,t} \), (ii) the requirement that \( rk(\mathbf{C}) = r \leq m + 1 \), (iii) the requirement that \( \varepsilon_{i,t} \) and \( \varepsilon_{i,t} \) are i.i.d. across \( i \), and (iv) the assumed non-randomness of \( \mathbf{C}_i \). Assumptions (i)–(iii) ensure that \( \mathbf{F}_i \) can be estimated using nothing but the simple cross-section average of \( \mathbf{Z}_{i,t} \). The PC equivalent of (ii) is that \( rk(N^{-1} \sum_{i=1}^{N} \mathbf{C}_i \mathbf{C}_i^\prime) = r \leq k \), where \( k \) is the assumed number of common factors, which can be larger or smaller than \( m + 1 \). Hence the usual problem in PC analysis of finding a suitable upper bound on the true number of factors, \( r \), is in CA tantamount to finding an appropriate number of extra variables. The additional observations required in CA is the ‘price’ paid for the relative simplicity with which the factors are estimated. Of course, in many situations, the model of ultimate interest is a multivariate one, as when wanting to regress \( Y_{i,t} \) onto \( \mathbf{X}_{i,t} \), and the unit root testing is just a pre-test step. In situations like this, joint CA estimation of the factors of all the variables of the model is expected to lead to reduced estimation uncertainty when compared to variable-by-variable PC, as in original PANIC (see Westerlund and Urban, 2015). As pointed out in Westerlund and Urban (2015), the requirement that \( rk(\mathbf{C}) = r \) is not testable. It can be relaxed, but then at the cost of additional restrictions on \( \mathbf{C}_i \). Indeed, as Westerlund and Urban (2013) show, if Assumption 4 is violated, then \( \lambda_i \) and \( \mathbf{A}_i \) have to be random and uncorrelated.

As in Bai and Ng (2010), (iii) is not really necessary and can be relaxed to allow for weak cross-section correlation in the ‘idiocynratic’ component, here represented by \( \varepsilon_{i,t} \) and \( \varepsilon_{i,t} \) (see Bai and Ng, 2004). In the terminology of Chudik et al. (2011), \( \varepsilon_{i,t} \) and \( \varepsilon_{i,t} \) may be semi-random correlated without affecting the results derived in the Appendix. The intuition is simple. Suppose for sake of argument that \( \phi_i(L) = 1, \mathbf{I}_m \) \( = \rho_1 = \ldots = \rho_N = 1 \), implying \( \mathbf{A}_{i,t} = \mathbf{A}_i^\prime \Delta \mathbf{F}_i + \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \).

A key requirement for the consistency of the estimated CA factors is that \( ||\mathbf{V}_i|| \leq O_p(N^{-1/2}) \) (see Remark 2 of Section 3), which will be the case if \( \varepsilon_{i,t} \) and \( \varepsilon_{i,t} \) are i.i.d. However, while sufficient, i.i.d.-ness is clearly not a necessary condition. Suppose, for example, that \( \mathbf{A}_{i,t} = \mathbf{A}_i^\prime \Delta \mathbf{F}_i + \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \), where \( \mathbf{A}_i = N^{-\alpha} \mathbf{C}_i \) and \( \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \) is i.i.d. across both \( i \) and \( t \) with mean zero and four finite moments. If \( \alpha \in [1/2, 1] \), such that \( \mathbf{A}_{i,t} \) is semi-weakly cross-correlated, then \( ||\mathbf{V}_i|| \leq N^{-\alpha} ||\mathbf{C}_i|| ||\Delta \mathbf{F}_i|| + ||\mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime \mathbf{A}_i^\prime || = O_p(max\{N^{-\alpha}, N^{-1/2}\}) = O_p(N^{-1/2}) \). The CA factors are therefore still consistent.

As with (iii), assumption (iv) is only for simplicity, and can be relaxed, provided that \( \mathbf{C}_i \) is independent of all the other random elements of the DGP and \( E(||\mathbf{C}_i||^4) \leq M \) (see Bai and Ng, 2004, 2010). Alternatively, we may assume that \( \mathbf{C}_i \) satisfies Assumption 3 of Pesaran (2006).

Remark 1. The assumption that \( Y_{i,t} \) and \( \mathbf{X}_{i,t} \) depend on the same set of factors is not a restriction. Suppose, for example, that the factors to \( Y_{i,t} \) and \( \mathbf{X}_{i,t} \) do not have any elements in common. In order to capture this we introduce the \( r \times r \) orthogonal matrix \( \mathbf{J} = (\mathbf{J}_1, \mathbf{J}_2) \), which is such that \( \mathbf{J}^\prime \mathbf{J} = \mathbf{I}_r \).

The component matrices $J_1$ and $J_2$, which are of dimension $r \times r_1$ and $r \times (r - r_1)$, respectively, are such that $J_1J_1^T = 0_{(r-r_1) \times (r-r_1)}$, $J_1^TA_i = 0_{r_1 \times (m+1)}$ and $J_2^TA_i = 0_{(r-r_1) \times 1}$. The matrix $J$ allows us to rotate $F_t$ as $JF_t = (J_1^T F_1, J_2^T F_2)^T$. Thus, defining $J_1^\top \lambda_i = \lambda_{1,t}$ and $J_2^\top A_i = A_{2,t}$, we have $Y_{i,t} = \alpha_i^\top D_{1,t} + \lambda_i^T F_1 + e_{1,t}, \quad \tilde{Y}_{i,t} = \alpha_i^\top D_{1,t} + \lambda_i^T J_1^T F_1 + e_{1,t}, \quad X_{i,t} = \beta_i^\top D_{2,t} + A_{2,t} F_2 + u_{i,t}$. Therefore, while $Y_{i,t}$ and $X_{i,t}$ may depend on the same set of factors, this is not necessary.

3. PANICCA

The idea behind PANIC is to first transform $Z_{i,t}$ by taking first differences. Since the transformed variable is stationary by assumption, the uncertainty regarding the order of integration of $Z_{i,t}$ is gone, which means that the common and idiosyncratic components can be estimated using existing methods for common factor models. While Bai and Ng (2004, 2010) use PC, in the present paper we use CA. The estimated level components are obtained by simply taking partial sums of the estimated first-differenced components. The unit root and cointegration properties of these components can then be tested using existing tests.

The purpose of the rest of this section is to make the above discussion a little more precise. Let us begin by defining $z_{i,t} = \Delta Z_{i,t}, \quad d_{i,p} = \Delta D_{i,p}, \quad f_t = \Delta F_t$ and $v_{i,t} = \Delta V_{i,t}$. Denote by $G$ the $p \times (p + 1)$ selection matrix of zeros and ones removing the first element of $d_{i,p}$, which is zero, that is, $G_{i,p} = D_{i,p-1}$. Since $p k (G^\top G) = p$, we may further define the $p \times (m + 1)$ matrix $b_i = G (G^\top G)^{-1} B_i$. In this notation, the first-differenced version of equation (3) may be written as

$$z_{i,t} = b_i^\top D_{i,p-1} + C_i^\top f_t + v_{i,t}$$

or, in matrix form,

$$z_i = D_{p-1} b_i + f C_i + v_i$$

where $z_i = (z_{i,2}, \ldots, z_{i,T})^\top$ and $v_i = (v_{i,2}, \ldots, v_{i,T})^\top$ are $(T - 1) \times (m + 1)$, $f = (f_2, \ldots, f_T)^\top$ is $(T - 1) \times r$, and $D_{p-1} = (D_{2,p-1}, \ldots, D_{T,p-1})^\top$ is $(T - 1) \times (p - 1)$. Since $f$ and $C_i$ are not separately identifiable, the best that we can do is to estimate the space spanned by these matrices. Define $M_p = I_{T-1} - D_{p-1} (D_{p-1}^\top D_{p-1})^{-1} D_{p-1}^\top$ for $p = 1$ and $M_p = I_{T-1}$ for $p = 0$. Let $z^{p}_i = (z_{i,2}^p, \ldots, z_{i,T}^p)^\top = M_p z_i$ with similar definitions of $f^p$ and $v^p_i$. In this notation, equation (5) can be written alternatively as

$$z^p_i = f^\top C + v^p_i$$

(6)

The CA estimator of (the space spanned by) $f^p$ is given simply by $\hat{f}^p = M_p \bar{z} = \bar{z}^p = N^{-1} \sum_{i=1}^N M_p z_i$. Let $C_i = [(f^p)^\top (f^p)^{-1} (f^p)^\top]^{-1} (f^p)^\top$ be the LS estimator of $C_i$ in (6) with $fp$ replaced by $\hat{f}^p$. The CA estimator of $v^p_i$ is given by $\hat{v}_i^p = z_i^p - \hat{f}^p \hat{C}_i$. Note that $\hat{f}_i^p$ ($\hat{v}_i^p$) is an estimator of the first-differenced (and detrended) version of $F_t$ ($V_{i,t}$). As an estimator of (the detrended version of) $F_t$ ($V_{i,t}$) we use $\hat{P} = \sum_{n=2}^{T} \hat{P}_n$ ($\hat{V}_{i,t}^p = \sum_{n=2}^{T} \hat{V}_{i,n}^p$).

**Remark 2.** A conceptual difference when compared to Pesaran (2006, 2007) and Pesaran et al. (2013) is that, while in these other works the averages are referred to as ‘factor proxies’, in the present study $\hat{f}^p$ is treated as an estimator of the space spanned by $\hat{f}_i^p$. The reason is simple. We begin by noting that $\tilde{z}_i^p = \tilde{C}_i^p \hat{f}_i^p + \hat{v}_i^p$. This implies

$$\hat{f}_i^p = \tilde{z}_i^p = \tilde{C}_i^p \hat{f}_i^p + \hat{v}_i^p = \tilde{C}_i^p \hat{f}_i^p + O_p(N^{-1/2})$$
where the order of the remainder follows from the fact that $v_{p,t}^p$ is mean zero and independent across $i$. Hence $\hat{e}_{p,t}^p$ is consistent, but not for $\tilde{e}_{p,t}^p$; only for $\tilde{C}^p\tilde{e}_{p,t}^p$, which is enough for our purposes. The rotation by $\tilde{C}$ here illustrates the need for the rank condition in Assumption 4. Suppose, for example, that $r = 1$ but $\tilde{C} = \Theta_{1 \times (m + 1)}$. In this case there is a single common factor present. However, since $\tilde{C}^p\tilde{e}_{p,t}^p = 0$, $\hat{e}_{p,t}^p$ will be unable to capture it.

### 3.1. Testing $e_{i,t}$

Note that $\tilde{C}_t$ can be decomposed as $\tilde{C}_t = (\tilde{\lambda}_t, \tilde{A}_t)$, where $\tilde{\lambda}_t$ is $r \times 1$ and $\tilde{A}_t$ is $r \times m$. We also have $\tilde{y}_{i,t}^p = [\tilde{e}_{i,t}^p, (\tilde{a}_{i,t}^p)^\prime]^\prime$, where $\tilde{e}_{i,t}^p$ is a scalar and $\tilde{a}_{i,t}^p$ is $m \times 1$. Denote by $\tilde{\rho}_p$ the LS slope estimator in a pooled panel regression of $\tilde{e}_{i,t}^p$ onto $\tilde{e}_{i,t-1}^p$. The test statistics considered herein are all taken from Bai and Ng (2010), and are designed to test the null hypothesis that $\rho_1 = \ldots = \rho_N = 1$. The first two test statistics, denoted $P_{a,p}$ and $P_{b,p}$, are defined as follows:

\[
P_{a,0} = \frac{\sqrt{NT} (\hat{\rho}^+_0 - 1)}{\sqrt{2\hat{\phi}^2 / \hat{\omega}^2}},
\]

\[
P_{b,0} = \frac{\sqrt{NT} (\hat{\rho}^+_0 - 1)}{\sqrt{\hat{\phi}^2 / [\hat{\omega}^2 NT^{-1} N^{-1} \sum_{j=1}^N (\hat{e}_{i,-1}^0)' \hat{e}_{i,-1}^0]}},
\]

\[
P_{a,1} = \frac{\sqrt{NT} (\hat{\rho}^+_1 - 1)}{\sqrt{36 \hat{\phi}^4 \hat{\omega}^4 / 5 \hat{\omega}^2}},
\]

\[
P_{b,1} = \frac{\sqrt{NT} (\hat{\rho}^+_1 - 1)}{\sqrt{6 \hat{\phi}^4 \hat{\omega}^4 / [5 \hat{\omega}^2 NT^{-1} N^{-1} \sum_{j=1}^N (\hat{e}_{i,-1}^1)' \hat{e}_{i,-1}^1]}}.
\]

where $\hat{e}_{i,-1}^p = (\hat{e}_{i,2}^p, \ldots, \hat{e}_{i,T-1}^p)'$ and

\[
\hat{\rho}_0^+ = \hat{\rho}_0 + \frac{\hat{\rho}^+_0}{(NT)^{-1} \sum_{j=1}^N (\hat{e}_{i,-1}^0)' \hat{e}_{i,-1}^0},
\]

\[
\hat{\rho}_1^+ = \hat{\rho}_1 + \frac{3 \hat{\phi}^2}{T \hat{\omega}^2},
\]

Here $\hat{\phi}^2_\epsilon$, $\hat{\omega}^2_\epsilon \hat{\phi}^4_\epsilon$ and $\hat{\epsilon}_\epsilon$ are given by the cross-sectional averages of $\hat{\phi}^2_{\epsilon,t}$, $\hat{\omega}^2_{\epsilon,t} \hat{\phi}^4_{\epsilon,t}$ and $\hat{\epsilon}_{\epsilon,t}$, respectively. The first two of these estimated variances are constructed as follows:

\[
\hat{\phi}^2_{\epsilon,t} = \frac{1}{T} \sum_{i=3}^T \hat{e}_{i,t}^2,
\]

\[
\hat{\omega}^2_{\epsilon,t} = \sum_{j=J+1}^{J^{-1}} K(j) \frac{1}{T} \sum_{i+j+3}^T \hat{e}_{i,t}^2 \hat{e}_{i,t-j}^2
\]

where $\hat{e}_{i,t} = \hat{e}_{i,t} - \hat{\rho}_p \hat{e}_{i,t-1}$, $K(j) = 1 - j/(J + 1)$ is the Bartlett kernel and $J$ is the associated kernel bandwidth parameter, which is assumed to satisfy Assumption 8 below. The estimators of $\phi^4_{\epsilon,t}$ and $\lambda_{\epsilon,t}$ are given by $\tilde{\epsilon}_{\epsilon,t} = (\hat{\phi}^2_{\epsilon,t} - \hat{\phi}^2_{\epsilon,t})/2$ and $\hat{\phi}^4_{\epsilon,t} = \hat{\phi}^4_{\epsilon,t}$, respectively.
Assumption 8. $J/\min(\sqrt{N}, \sqrt{T}) \to 0$ as $J, N, T \to \infty$.

The third test statistic that we consider, denoted $P_{MSB},$ is the panel-modified Sargan–Bhargava (PMSB) test statistic of Bai and Ng (2010), as given by

$$P_{MSB} = \frac{\sqrt{N}(N^{-1}T^{-2}\sum_{i=1}^{N}(\hat{\epsilon}_{i,-1}^{0})^{2} - \hat{\omega}^{2}/2)}{\sqrt{\hat{\delta}_{e}^{6}/3}}$$

$$P_{MSB1} = \frac{\sqrt{N}(N^{-1}T^{-2}\sum_{i=1}^{N}(\hat{\epsilon}_{i,-1}^{1})^{2} - \hat{\omega}^{2}/6)}{\sqrt{\hat{\delta}_{e}^{4}/45}}$$

Theorem 1 reports the asymptotic null distributions of $P_{a,p}, P_{b,p}$ and $P_{MSB}$. Its proof, as those of all other formal results, can be found in the online Appendix to this paper (supporting information).

**Theorem 1.** Under Assumptions 1–8 and the null hypothesis that $\rho_1 = \ldots = \rho_N = 1$, as $N, T \to \infty$ with $N/T \to 0$

$$P_{a,p} \to_d N(0, 1),$$

$$P_{b,p} \to_d N(0, 1),$$

$$P_{MSB} \to_d N(0, 1)$$

where $\to_d$ signifies convergence in distribution.

According to Theorem 1 all three test statistics converge to $N(0, 1)$ under the unit root null. In fact, it is not difficult to show that $P_{a,p}$ and $P_{b,p}$ are asymptotically equivalent, which means that for these statistics the convergence is to the same normal variate. Note also that while $P_{a,p}$ and $P_{b,p}$ are (bias-adjusted) $t$-statistics of the largest autoregressive root, $P_{MSB}$ is a ratio of variances. In spite of this difference, however, provided that the alternative formulated as that $|\rho_i| < 1$ for some $i$, all three statistics are left-tailed. The appropriate 5% critical value is therefore given by $-1.645$.

**Remark 3.** The fact that the PANICCA-based statistics are asymptotically $N(0, 1)$ stands in sharp contrast to the results reported by Pesaran (2007) and Pesaran et al. (2013), who use $\tilde{Z}_t$ (and $\tilde{z}_t$) as a ‘proxy’ for $F_t$. This means that if $r_1 > 0$ the asymptotic distributions of their CIPS and CSB test statistics depend on the Brownian motion associated with $F_t$. As alluded to in Section 1, this difference is due to the fact that here the estimation is done using only $\tilde{z}_t$, which is stationary.

### 3.2. Testing $F_t$

The rate of consistency of the CA estimator $\hat{F}_p^p$ of (the space spanned by) $F_p^p$ is the same as that of the PC estimator when $\sqrt{N}/T \to c \leq M$ and it is faster when $\sqrt{N}/T \to \infty$. The results reported by Bai and Ng (2004, Theorems 1 and 3) for the tests of the estimated PC factors therefore go through also in the case of CA estimation.

Testing is conducted in the following fashion. If $r = m + 1 = 1$, such that $\hat{F}_p^p$ (and $F_p^p$) is a scalar, then the testing can be carried out using any existing unit root test. Bai and Ng (2004) focus on the augmented Dickey–Fuller (ADF) test, henceforth denoted $ADF_p$, and hence so do we. Let us therefore define $\Delta \hat{F}_p^p = (\Delta \hat{F}_{p,3+q}^p, \ldots, \Delta \hat{F}_{p,T}^p), \hat{F}_{p,1} = (\hat{F}_{p,2+q}^p, \ldots, \hat{F}_{p,T-1}^p), W = (W_{2+q}, \ldots, W_{T}^p),$ where $W_t = (\Delta \hat{F}_{t-1}^p, \ldots, \Delta \hat{F}_{t-q}^p).$ The ADF statistic is given by
ADF_p = \frac{(\hat{F}_p^{\frac{1}{2}})'M_{p+1}W M_{p+1}\Delta \hat{F}_p}{\hat{\sigma}_q \sqrt{(\hat{F}_p^{\frac{1}{2}})'M_{p+1}W M_{p+1}\hat{F}_p}}

where \( \hat{\sigma}_q^2 = T^{-1}(\Delta \hat{F}_p)'M_{p+1}W M_{p+1}\Delta \hat{F}_p \), \( M_{p+1} = I_{T-q-1} - W(W'W)^{-1}W' \) and the last \( q \) rows of \( M_p \) have been removed to make it conformable with \( W \). Note how the dependence on \( q \) has been suppressed in \( ADF_p \).

If \( r = m + 1 > 1 \), then the following sequential test procedure can be used to determine \( r_1 \), the number of unit roots in \( \hat{F}_p^p \):

1. Set \( k = r \).
2. Compute \( \hat{Y}_k^p = (\hat{Y}_{2,k}, \ldots, \hat{Y}_{T,k})' = M_{p+1} \hat{F}_p \hat{\beta}_k \), where \( \hat{\beta}_k \) is the \( (m + 1) \times k \) matrix of eigenvectors associated with the \( k \) largest eigenvalues of \( T^{-1}(\hat{F}_p^p)'M_{p+1} \hat{F}_p^p \).
3. The test statistic is given by
   \[ MQ_p(k) = T[\hat{w}(k) - 1] \]
   where \( \hat{w}_p(k) \) is the smallest eigenvalue of
   \[ \frac{1}{2} (\hat{Y}_{k-1}^p)'\hat{Y}_k^p + (\hat{Y}_k^p)'\hat{Y}_{k-1}^p - T(\hat{\Sigma}_k + \hat{\Sigma}_k') \]
   \[ (\hat{Y}_{k-1}^p)'\hat{Y}_{k-1}^p \]
   with \( \hat{\Sigma}_k = \sum_{j=1}^{T-1} K(j)T^{-1} \sum_{t=j+1}^T \hat{U}_{t-j,k} \hat{U}_{t-j,k}' \), \( \hat{U}_{t,k} \) is the residual from an LS fit of \( \hat{Y}_{t,k} \) onto \( \hat{Y}_{t-1,k} \), and where \( J \) is assumed to satisfy Assumption 8.
4. If the null hypothesis that \( r_1 = k \) is rejected using \( MQ_p(k) \), set \( k = k - 1 \) and return to step 2. Otherwise, set \( r_1 = k \) and stop.

Remark 4. \( MQ_p \) is the MQc test statistic of Bai and Ng (2004) applied to the estimated CA factors. Bai and Ng (2004) also consider another statistic, denoted MQf. However, since MQc is more general, in this paper we only consider the CA version of this statistic.

Assumption 9. \( q^3/\min(N, T) \rightarrow 0 \) as \( q, N, T \rightarrow \infty \).

In Theorem 2 we report the asymptotic null distributions of \( ADF_p \) and \( MQ_p(k) \). In so doing it is useful to introduce the following detrended Brownian motion:

\[ W_r^p(s) = W_r(s) - \int_0^1 W_r(v)d_p(v)'dv \left( \int_0^1 d_p(v)d_p(v)'dv \right)^{-1} d_p(s) \]

where \( d_p(s) = (1, \ldots, s^p)' \) is the limiting trend function and \( W_r(s) \) is an \( r \times 1 \) vector standard Brownian motion.

Theorem 2. Under Assumptions 1–9 the following results hold as \( N, T \rightarrow \infty \):

(i) Suppose that \( r = m + 1 = 1 \). Under the null hypothesis that \( F_r \) has a unit root

\[ ADF_p \overset{w}{\rightarrow} \frac{\int_0^1 W_{11}^p(s)dW_1(s)}{\sqrt{\int_0^1 [W_{11}^p(s)]^2ds}} \]

where \( \overset{w}{\rightarrow} \) signifies weak convergence.
(ii) Suppose that \( r = m + 1 > 1 \). Under the null hypothesis that \( F_t \) has \( k \leq m + 1 \) unit roots

\[
MQ_p(k) \to w.T[w_p(k) - 1]
\]

where \( w_p(k) \) is the smallest eigenvalue of

\[
\frac{1}{2}[W_p^r(1)W_p^r(1)' - I_k]\left(\int_0^1 W_p^r(s)W_p^r(s)'
ds\right)^{-1}
\]

The asymptotic distribution of \( ADF_p \) is identically the \( ADF \) test distribution, for which critical values are readily available (see, for example, MacKinnon, 1996). Appropriate 1%, 5% and 10% critical values for \( MQ_p(k) \) \((k = 1, \ldots, 6)\) can be found in Table I of Bai and Ng (2004).

The results reported so far make use of Assumption 4, which only requires that the true number of factors, \( r \), is less than or equal to \( m + 1 \). If one would like to pinpoint \( r \), one possibility is to employ an information criterion. This approach has been shown to work in the context of PC estimation (see Bai and Ng, 2002), and, as pointed out by Pesaran et al. (2013, Section 4.1), it is expected to work

Table I. 5% size and size-corrected power for the intercept-only case \((p = 0)\)

| \( N \) | \( T \) | \( P_{a,0}^{PC} \) | \( P_{u,0} \) | \( P_{h,0}^{PC} \) | \( P_{h,0} \) | PMSB0^{|PC} | PMSB0 |
|---|---|---|---|---|---|---|---|
| **Size** | | | | | | | |
| 10 | 10 | 0.2252 | 0.2088 | 0.1304 | 0.1162 | 0.0094 | 0.0094 |
| 10 | 20 | 0.1756 | 0.1478 | 0.098 | 0.0734 | 0.0138 | 0.0138 |
| 10 | 35 | 0.1834 | 0.1574 | 0.1098 | 0.088 | 0.0226 | 0.0138 |
| 10 | 50 | 0.1776 | 0.1504 | 0.1074 | 0.085 | 0.0248 | 0.0136 |
| 20 | 10 | 0.213 | 0.2104 | 0.134 | 0.131 | 0.0696 | 0.0732 |
| 20 | 20 | 0.145 | 0.1452 | 0.0888 | 0.0842 | 0.0392 | 0.0372 |
| 20 | 35 | 0.123 | 0.1174 | 0.0782 | 0.0718 | 0.0284 | 0.0308 |
| 20 | 50 | 0.1234 | 0.1116 | 0.0746 | 0.0664 | 0.0316 | 0.027 |
| 35 | 10 | 0.237 | 0.2314 | 0.1726 | 0.1676 | 0.1382 | 0.1468 |
| 35 | 20 | 0.1298 | 0.1392 | 0.0844 | 0.0892 | 0.06 | 0.0654 |
| 35 | 35 | 0.1086 | 0.1014 | 0.072 | 0.0664 | 0.0408 | 0.0422 |
| 35 | 50 | 0.1044 | 0.104 | 0.0706 | 0.0698 | 0.0392 | 0.0378 |
| 50 | 10 | 0.2526 | 0.2564 | 0.193 | 0.1988 | 0.193 | 0.1976 |
| 50 | 20 | 0.1362 | 0.1422 | 0.095 | 0.0994 | 0.0874 | 0.0838 |
| 50 | 35 | 0.0992 | 0.1008 | 0.065 | 0.0666 | 0.0526 | 0.0528 |
| 50 | 50 | 0.0976 | 0.0994 | 0.0692 | 0.0698 | 0.051 | 0.051 |
| **Size-corrected power** | | | | | | | |
| 10 | 10 | 0.0548 | 0.0708 | 0.0542 | 0.0744 | 0.057 | 0.0642 |
| 10 | 20 | 0.1064 | 0.1572 | 0.1094 | 0.158 | 0.0994 | 0.138 |
| 10 | 35 | 0.2296 | 0.3246 | 0.2384 | 0.3298 | 0.2 | 0.2738 |
| 10 | 50 | 0.395 | 0.5482 | 0.401 | 0.5496 | 0.3343 | 0.4588 |
| 20 | 10 | 0.04 | 0.0926 | 0.0402 | 0.095 | 0.0436 | 0.0836 |
| 20 | 20 | 0.1864 | 0.264 | 0.1874 | 0.2684 | 0.1544 | 0.215 |
| 20 | 35 | 0.5046 | 0.6094 | 0.5142 | 0.6116 | 0.416 | 0.5014 |
| 20 | 50 | 0.7996 | 0.8804 | 0.797 | 0.8782 | 0.6894 | 0.7794 |
| 35 | 10 | 0.0282 | 0.0972 | 0.0278 | 0.0976 | 0.0304 | 0.0886 |
| 35 | 20 | 0.2832 | 0.455 | 0.2872 | 0.4536 | 0.2278 | 0.3622 |
| 35 | 35 | 0.778 | 0.8638 | 0.7798 | 0.8632 | 0.668 | 0.7574 |
| 35 | 50 | 0.9778 | 0.9906 | 0.9772 | 0.9888 | 0.9334 | 0.9604 |
| 50 | 10 | 0.0206 | 0.1172 | 0.0214 | 0.1164 | 0.0256 | 0.1014 |
| 50 | 20 | 0.3658 | 0.5666 | 0.371 | 0.5688 | 0.2906 | 0.4624 |
| 50 | 35 | 0.924 | 0.9642 | 0.9256 | 0.962 | 0.8378 | 0.9094 |
| 50 | 50 | 0.9984 | 0.9998 | 0.9984 | 0.9996 | 0.988 | 0.9924 |

Note: A ‘PC’ superscript signifies that the test is based on original PANIC.
well also for CA. Let $\hat{\psi}_{s}^{p}(s) = x_{t}^{p} - \hat{\psi}_{s}^{p}(s)\hat{c}_{t}$, where $\hat{\psi}_{s}^{p}(s)$ is $\hat{\psi}_{s}$ based on a subset of $s \leq m + 1$ cross-section averages. The information criterion considered in this paper, which can be seen as a multivariate version of the PC-specific IC$_{p3}$ criterion of Bai and Ng (2002), takes the form

$$\text{IC}(s) = \ln(\det(\hat{\Sigma}_{s}(s))) + s \cdot (m + 1)N^{-1}\ln(N)$$

where $\hat{\Sigma}_{s}(s) = (NT)^{-1}\sum_{i=1}^{N}[\hat{\psi}_{s}^{p}(s)]^{\prime}\hat{\psi}_{s}^{p}(s)$ is the sum of squared residuals obtained by fitting equation (6) based on $s \leq m + 1$ cross-section averages. The penalty, $s \cdot (m + 1)N^{-1}\ln(N)$, is the same as in IC$_{p3}$ with $T = 0$ (and $m + 1 = 1$). The reason for this difference is that while the rate of consistency of the PC estimator depends on both $N$ and $T$, as we explain in Remark 2 of Section 3, the rate of consistency of $\hat{\psi}$ only depends on $N$ (see Bai and Ng, 2002, p. 219, for a discussion). The estimator $\hat{\gamma}$ of $\gamma$ is given by

$$\hat{\gamma} = \arg \min_{s = 0, \ldots, m + 1} \text{IC}(s)$$

The consistency of $\hat{\gamma}$ is a direct consequence of Corollary 2 of Bai and Ng (2002), which is not PC specific but applies to any estimator of $\psi$. 

4. MONTE CARLO SIMULATIONS

4.1. Testing $e_{i,t}$

In this section, we use Monte Carlo simulation to assess the relative performance of PANICCA when compared to original PANIC. The DGP used for this purpose is given by a simplified version of equations (1)–(3) that sets $r = 3$, $m = 2$, $\alpha_{i} \sim U(0, 1)$, $\beta_{i} \sim U(0, 1)$, $\lambda'_{i} = (1, l_{i}, l_{i})$ and

$$\Lambda'_{i} = \begin{bmatrix} l_{i} & 1 & l_{i} \\ l_{i} & l_{i} & 1 \end{bmatrix}$$

where $l_{i} = 1.5 \cdot l_{i} - 0.5$ and $1(A)$ is the indicator function for the event $A$. This parametrization of $\lambda_{i}$ and $\Lambda_{i}$ ensures that $\Lambda$ has ones on the main diagonal and 0.25 elsewhere, which means that Assumption 4 is met. Also, $e_{i,t} = \rho e_{i,t-1} + e_{i,t}$, $u_{i,t} = \mu u_{i,t-1} + \varepsilon_{i,t}$ and $F_{t} = \delta F_{t-1} + \eta_{t}$, where $(e_{i,t}, e_{i,t}', \eta_{t}')' \sim N(0_{6 \times 1}, I)$. We begin by considering the 5% size and size-corrected power of $P_{\alpha,p}$, $P_{\beta,p}$ and PMSB$_{a,p}$. In the size experiments, $\rho = \delta = 1$, while in the power experiments $\rho = 0.95$ and $\delta = 0.5$. All results are based on making 5000 draws of panels where $N$ and $T$ are chosen so as to illustrate the main difference between PANIC and PANICCA, which occurs naturally when the sample size is relatively small. We chose $N, T \in \{10, 20, 35, 50\}$, which is consistent with the bulk of empirical work based on PANIC (see, for example, Lanzafame, 2010; Schmidt and Vosen, 2013; Martín, ; Joseph et al., 2012, 2013; Örsal and Dilan, ; Blomquist and Westerlund, 2014).

The tests are implemented as explained in Section 3 with $J$ selected as in Bai and Ng (2010), using the automatic procedure of Newey and West (1994). While the above DGP treats $r$ as known, this is not necessary. One possibility towards this end is to employ the information criterion-based selection approach discussed in the previous section. A practical question that arises when trying to implement this approach is how to select the appropriate subsets of cross-section averages. Since sequential search methods can be very sensitive to small variations in the data, we experimented with grid-search over all possible subsets. This led to very good results with correct selection frequencies that were close 100% in all cases considered, except when $T = 10$ (the results are available upon request). Unfortunately, the grid-search slowed down the procedure quite considerably, making it infeasible for large-scale
Monte Carlo simulations. In this section, we therefore focus on the case when the number of cross-section averages is set equal to \( m + 1 \) (as is done, for example, in Pesaran, 2006, 2007; Pesaran et al., 2013). In our empirical application in Section 5, we report some results based on estimating \( r \).

Because of the high precision in the estimated number of factors, the results reported here for the case when \( m + 1 = r \) are very similar to those obtained based on using the information criterion of Section 3. An alternative approach when \( r \) is unknown is follow the bulk of the existing CA-based works (see, for example, Pesaran, 2006; Pesaran et al., 2013; Chudik et al., 2011; Kapetanios and Pesaran, 2005) and set \( m + 1 \) equal to some sufficiently large value (to ensure that \( m + 1 \geq r \)). A practically relevant scenario with such a strategy is that \( m + 1 > r \), such that the number of cross-section averages is over-specified. In Section 4.3 we elaborate on this.

We begin by considering the size results for the intercept-only case when \( p = 0 \), which are reported in Table I. The information content of this table may be summarized as follows:

- While there are some noticeable distortions, these are mainly among the smaller values of \( N \) and \( T \). In fact, size accuracy is quite good already with \( T = 50 \) and \( N = 35 \), and it increases with increasing values of \( T \) and to a lesser extent with increasing values of \( N \). That the effect of increasing \( T \) is relatively more pronounced is in agreement with the condition that \( N/T \rightarrow 0 \).
- The distortions are generally somewhat smaller for PANICCA than for PANIC. This corroborates the findings of Westerlund and Urbain (2015) in the (stationary) factor-augmented regression case, suggesting that CA tend to be more accurate than PC.
- Looking across the three types of tests, the best size accuracy is generally obtained by using the PSMB tests.
- The PANICCA-based tests are uniformly more powerful than their PANIC-based counterparts. The difference in power is large enough not to be ignored and can in fact be quite substantial.
- In agreement with their relatively high rejection frequencies under the null, the best power is generally obtain by using the \( P_a \) and \( P_b \)-type tests. This finding is consistent with the results of Westerlund (2015b), showing how the local asymptotic power of the PANIC versions of these tests is higher than that of the PANIC-based PSMB test.
- The fact that the difference in size and power is decreasing in \( N \) and \( T \) is consistent with the asymptotic equivalence of the PANICCA- and PANIC-based \( P_a \) and \( P_b \)-tests.

The results reported in Table II for the case with an intercept and trend (\( p = 1 \)) are very similar to those reported in Table I, and we therefore just briefly describe them. The first thing to note is that the size distortions are actually reduced as the linear trend is added, which is somewhat unexpected, because usually the distortions are increasing in \( p \). Another difference worth noting is the power, which is much lower in Table II than in Table I. In fact, the power in the linear trend case only rarely raises above the nominal 5% level. That the power is reduced by the inclusion of the linear trend is a reflection of the well-known ‘incidental trends problem’ (see Westerlund, 2015b), and is therefore expected.

4.2. Testing \( F_t \)

In this subsection, we investigate the performance of the sequential procedure to determine \( r_1 \), the number of unit root factors. The DGP is the same as before. The only difference is the common factors, which are now generated according to \( F_t = \text{diag}(\theta_0 \cdot I_{(r-r_1) \times 1}, I_{r_1 \times 1}) F_{t-1} + \eta_t \), where \( \eta_t \) is as before, \( I_{r \times 1} = (1, \ldots, 1)' \) is an \( r \times 1 \) vector of ones, and \( |\theta_0| < 1 \) is the autoregressive coefficient of the stationary factors (see Bai and Ng, 2004, for a similar parametrization). In the interest of space, we focus on the results for the case when \( N = 20 \) and \( T = 50 \).
### Table II. 5\% size and size-corrected power for the trend case \((p = 1)\)

<table>
<thead>
<tr>
<th>Size</th>
<th>(N)</th>
<th>(T)</th>
<th>(P_{PC}^{a;0})</th>
<th>(P_{a;0})</th>
<th>(P_{PC}^{b;0})</th>
<th>(P_{b;0})</th>
<th>(PMSB_{PC}^{0})</th>
<th>(PMSB_{0})</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>0.1688</td>
<td>0.1908</td>
<td>0.1692</td>
<td>0.166</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>0.1716</td>
<td>0.151</td>
<td>0.1438</td>
<td>0.1214</td>
<td>0.0082</td>
<td>0.0048</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.1646</td>
<td>0.1422</td>
<td>0.1394</td>
<td>0.1146</td>
<td>0.0236</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>0.2264</td>
<td>0.2274</td>
<td>0.2168</td>
<td>0.2198</td>
<td>0.017</td>
<td>0.0198</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.1498</td>
<td>0.1398</td>
<td>0.1342</td>
<td>0.1278</td>
<td>0.0202</td>
<td>0.0186</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>35</td>
<td>0.1308</td>
<td>0.1192</td>
<td>0.1152</td>
<td>0.1048</td>
<td>0.0232</td>
<td>0.0224</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.1226</td>
<td>0.1158</td>
<td>0.1082</td>
<td>0.0998</td>
<td>0.029</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>10</td>
<td>0.2668</td>
<td>0.2694</td>
<td>0.2656</td>
<td>0.271</td>
<td>0.065</td>
<td>0.0674</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>20</td>
<td>0.1696</td>
<td>0.1632</td>
<td>0.1654</td>
<td>0.159</td>
<td>0.0356</td>
<td>0.0396</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>0.1218</td>
<td>0.1198</td>
<td>0.1184</td>
<td>0.1162</td>
<td>0.0342</td>
<td>0.0352</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>50</td>
<td>0.1096</td>
<td>0.1186</td>
<td>0.1054</td>
<td>0.113</td>
<td>0.0362</td>
<td>0.0356</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.2978</td>
<td>0.301</td>
<td>0.3034</td>
<td>0.306</td>
<td>0.1072</td>
<td>0.1078</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>0.19</td>
<td>0.1896</td>
<td>0.192</td>
<td>0.1908</td>
<td>0.0582</td>
<td>0.0524</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>35</td>
<td>0.13</td>
<td>0.1296</td>
<td>0.13</td>
<td>0.1302</td>
<td>0.0398</td>
<td>0.0444</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.1068</td>
<td>0.1094</td>
<td>0.1066</td>
<td>0.1094</td>
<td>0.0392</td>
<td>0.0388</td>
<td></td>
</tr>
</tbody>
</table>

**Size-corrected power**

<table>
<thead>
<tr>
<th>Size</th>
<th>(N)</th>
<th>(T)</th>
<th>(P_{PC}^{a;0})</th>
<th>(P_{a;0})</th>
<th>(P_{PC}^{b;0})</th>
<th>(P_{b;0})</th>
<th>(PMSB_{PC}^{0})</th>
<th>(PMSB_{0})</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>0.0448</td>
<td>0.0494</td>
<td>0.0468</td>
<td>0.0494</td>
<td>0.0442</td>
<td>0.0468</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>0.0554</td>
<td>0.0558</td>
<td>0.0554</td>
<td>0.0568</td>
<td>0.0546</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>35</td>
<td>0.0744</td>
<td>0.0902</td>
<td>0.0748</td>
<td>0.0902</td>
<td>0.0736</td>
<td>0.0886</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.1158</td>
<td>0.1302</td>
<td>0.1156</td>
<td>0.1302</td>
<td>0.1162</td>
<td>0.1316</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>0.0464</td>
<td>0.0474</td>
<td>0.0472</td>
<td>0.0474</td>
<td>0.0456</td>
<td>0.0454</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.0578</td>
<td>0.065</td>
<td>0.058</td>
<td>0.0662</td>
<td>0.057</td>
<td>0.0664</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>35</td>
<td>0.0866</td>
<td>0.1112</td>
<td>0.0982</td>
<td>0.1122</td>
<td>0.0982</td>
<td>0.1108</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.1766</td>
<td>0.2076</td>
<td>0.1756</td>
<td>0.2074</td>
<td>0.1746</td>
<td>0.207</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>10</td>
<td>0.0466</td>
<td>0.0506</td>
<td>0.046</td>
<td>0.0512</td>
<td>0.047</td>
<td>0.0544</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>20</td>
<td>0.0556</td>
<td>0.0756</td>
<td>0.0554</td>
<td>0.0756</td>
<td>0.0566</td>
<td>0.0736</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>0.1236</td>
<td>0.1506</td>
<td>0.1228</td>
<td>0.1494</td>
<td>0.1228</td>
<td>0.1474</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>50</td>
<td>0.248</td>
<td>0.2976</td>
<td>0.2484</td>
<td>0.297</td>
<td>0.2478</td>
<td>0.2942</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.0434</td>
<td>0.0418</td>
<td>0.0438</td>
<td>0.042</td>
<td>0.045</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>0.0622</td>
<td>0.0798</td>
<td>0.063</td>
<td>0.0802</td>
<td>0.0614</td>
<td>0.0762</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>35</td>
<td>0.1526</td>
<td>0.197</td>
<td>0.1536</td>
<td>0.198</td>
<td>0.1508</td>
<td>0.194</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.331</td>
<td>0.39</td>
<td>0.3314</td>
<td>0.3892</td>
<td>0.3308</td>
<td>0.3882</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** A ‘PC’ superscript signifies that the test is based on original PANIC.

The most striking observation that can be made from Table III is that the proposed CA-based estimator \(\hat{r}_1\) of \(r_1\) is much more robust to variations in \(r\) than the corresponding (PANIC) estimator based on PC. In fact, in a majority of cases the PC bias is twice as large as the corresponding CA bias. According to the results reported by Bai and Ng (2004), the PC-based estimator of \(r_1\) is more robust than both the trace test-based procedure of Johansen (1995) and the information criterion of Aznar and Salvador (2002). Being more accurate than PC, the CA-based estimator is expected also to outperform these other estimation approaches.

### 4.3. Robustness Checks

The results reported in Tables I–III are just a small fraction of the total set of results produced (available upon request). The overall impression considering all the results is that the above conclusions seem very robust to changes to the DGP. In this section, we illustrate this using two alternative specifications of \(C_i\); (i) \(r = 2 < m + 1 = 3\) and \(C_i\) generated as before, and (ii) \(C_i = \theta_{r}r_{(m+1)}\). Remember that the number of cross-section averages is set to \(m + 1\). The purpose of (i) is to investigate the effect of an over-specification of the number of cross-section averages. By contrast, in (ii), \(rk(C) = 0 < r = 2\). The purpose here is therefore to illustrate the effect of a violation of Assumption 4. To keep the number of tables manageable we focus on the intercept-only case.
Table III. Correct selection frequency of the estimated number of unit root factors

<table>
<thead>
<tr>
<th>$r$</th>
<th>$r_1$</th>
<th>$\rho$</th>
<th>$\delta_0$</th>
<th>$\hat{r}_1$</th>
<th>$\hat{r}_1^{PC}$</th>
<th>$\hat{r}_1$</th>
<th>$\hat{r}_1^{PC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>—</td>
<td>0.9576</td>
<td>0</td>
<td>0.9798</td>
<td>0.0002</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.5</td>
<td>—</td>
<td>0.9777</td>
<td>0.2502</td>
<td>0.991</td>
<td>0.6368</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0.8</td>
<td>—</td>
<td>0.9764</td>
<td>0.9266</td>
<td>0.9896</td>
<td>0.9794</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>—</td>
<td>0.9740</td>
<td>0.9788</td>
<td>0.9922</td>
<td>0.9898</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.77</td>
<td>0.0188</td>
<td>0.603</td>
<td>0.2532</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>0.77</td>
<td>0.7128</td>
<td>0.6074</td>
<td>0.8276</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.8</td>
<td>0</td>
<td>0.77</td>
<td>0.3634</td>
<td>0.1234</td>
<td>0.3402</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.77</td>
<td>0.9176</td>
<td>0.5946</td>
<td>0.8532</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7284</td>
<td>0.9018</td>
<td>0.5724</td>
<td>0.8162</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.8</td>
<td>0.5</td>
<td>0.7284</td>
<td>0.9176</td>
<td>0.5946</td>
<td>0.8532</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.5</td>
<td>0.7284</td>
<td>0.9176</td>
<td>0.5946</td>
<td>0.8532</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.8</td>
<td>0.5</td>
<td>0.361</td>
<td>0.3066</td>
<td>0.1084</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0.8</td>
<td>0.5</td>
<td>0.3738</td>
<td>0.315</td>
<td>0.1168</td>
<td>0.1066</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.5</td>
<td>0.2326</td>
<td>0.2162</td>
<td>0.085</td>
<td>0.0734</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.6988</td>
<td>0.6384</td>
<td>0.5042</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0.7070</td>
<td>0.6586</td>
<td>0.5026</td>
<td>0.401</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>0</td>
<td>0.7120</td>
<td>0.244</td>
<td>0.5174</td>
<td>0.1986</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.9</td>
<td>0</td>
<td>0.7128</td>
<td>0.2</td>
<td>0.5156</td>
<td>0.2056</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.7014</td>
<td>0.204</td>
<td>0.5124</td>
<td>0.1738</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5522</td>
<td>0.7056</td>
<td>0.169</td>
<td>0.447</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5428</td>
<td>0.5442</td>
<td>0.1182</td>
<td>0.165</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5402</td>
<td>0.0824</td>
<td>0.1158</td>
<td>0.0892</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.9</td>
<td>0.5</td>
<td>0.44</td>
<td>0.035</td>
<td>0.1194</td>
<td>0.0068</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.2334</td>
<td>0.0158</td>
<td>0.0766</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.989</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>0.9998</td>
<td>1</td>
<td>0.9888</td>
<td>0.984</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>1</td>
<td>0.7762</td>
<td>0.9856</td>
<td>0.3128</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>1</td>
<td>0.2952</td>
<td>0.988</td>
<td>0.0994</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.9744</td>
<td>0.0382</td>
<td>0.9778</td>
<td>0.024</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8622</td>
<td>0.9726</td>
<td>0.4022</td>
<td>0.656</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7777</td>
<td>0.8654</td>
<td>0.3172</td>
<td>0.2818</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.8</td>
<td>0.5</td>
<td>0.7356</td>
<td>0.3024</td>
<td>0.2864</td>
<td>0.0242</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.9</td>
<td>0.5</td>
<td>0.7320</td>
<td>0.1006</td>
<td>0.29</td>
<td>0.0848</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0.4054</td>
<td>0.0042</td>
<td>0.184</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

Note: $r$, $r_1$, $\rho$ and $\delta_0$ refer to the true number of factors, number of unit root factors, autoregressive coefficient of the idiosyncratic component, and the autoregressive coefficient of stationary factors, respectively. The ‘PC’ superscript signifies that the estimated number of unit root factors is based on original PANIC.

The size and power results for the tests of $e_{i,t}$ contained in Tables IV and V for cases (i) and (ii), respectively, are strikingly similar to the results reported in Table I. The conclusions drawn in the previous section are therefore unaffected by an over-specification of the number of cross-section averages, which is to be expected, because Assumption 4 is still satisfied, although not with equality as in Table I. The good performance when $C_t = 0_{r \times (m+1)}$ is more unexpected. Thus, while theoretically a restriction, in practice violations of Assumption 4 need not be detrimental for the performance of PANICCA-based unit root tests on the idiosyncratic components, which is consistent with the findings of Pesaran (2004, Section 7) and Kapetanios et al. (2011, Section 4).

Table VI reports some correct selection frequencies for the number of unit root factors for case (i). The main difference when compared to Table III occurs when $r_1 = r$, in which case $\hat{r}_1$ is more sensitive to variations in $\rho$. However, PANICCA still outperforms PANIC. For all other values of $r_1$, the results are very similar to those reported in Table III.
Table IV. 5% size and size-corrected power for the intercept-only case \((p = 0)\) when \(m + 1 = 3 > r = 2\)

<table>
<thead>
<tr>
<th>Size</th>
<th>Size-corrected power</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N)</td>
<td>(T)</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>35</td>
<td>10</td>
</tr>
<tr>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Note: A ‘PC’ superscript signifies that the test is based on original PANIC.

5. AN APPLICATION TO THE EMH

A financial market is said to be efficient if prices fully reflect all available information and no profit opportunities are left unexploited. Economic agents form their expectations rationally and rapidly arbitrage away any deviations of the expected returns consistent with supernormal profits. Therefore, if currency markets are efficient, the spot (forward) exchange rate should embody all relevant information, and it should not be possible to forecast one spot (forward) rate as a function of another. In what follows we refer to this proposition of the EMH as the ‘efficient cross-market hypothesis’ (ECMH).

Also, provided that agents are risk neutral and that the risk premium is stationary, the current forward rate should be an unbiased predictor of the future spot rate. This is the forward rate unbiasedness hypothesis (FRUH).

The validity of the above propositions of the EMH has been, and still is, one of the most heavily researched areas in the financial literature. However, a lot of controversy still exists about the method that must be applied to test for its existence. In particular, the use of cointegration techniques has become very popular, and is by now the workhorse of the industry (see Zivot, 2000, for a survey of the cointegration-based literature). Indeed, since the seminal work of Hakkio and Rush (1989), it is well recognized that the FRUH requires that the future spot and current forward rates are cointegrated and one-to-one. Also, if the ECMH holds, then spot and forward rates cannot be cointegrated across markets.
Interestingly, while each of these propositions of the EMH occupies a huge literature (for early contributions see, for example, Hakkio and Rush, 1989; Baillie and Bollerslev, 1989; Crowder, 1994), we know of no previous study that has tried to formalize the connection between the two. In particular, since both spot and forward rates from across a variety of markets exhibit unit root-like behavior, a natural question concerns the source of the non-stationarity. To formalize matters slightly, let us denote $s_{i,t}$ the log spot (forward) rate of market $i$ at time $t$. In terms of the model of Section 2, $Y_{i,t} = s_{i,t+1}$ and $X_{i,t} = f_{i,t}$. Consider $s_{i,t}$. According to the ECMH, this variables must not be cointegrated across markets. In order to appreciate the implications of this, it is useful to note that

$$s_{i,t+1} - \theta s_{j,t+1} = (\alpha_i - \theta \alpha_j) D_{i,p} + (\lambda_i - \theta \lambda_j) F_t + e_{i,t} - e_{j,t}$$

Obviously, being idiosyncratic, $e_{i,t}$ and $e_{j,t}$ cannot be cointegrated for $i \neq j$. Hence, for the ECMH to hold it must be that $(\lambda_i - \theta \lambda_j) F_t$ and/or $e_{1,t}, \ldots, e_{N,t}$ are unit root non-stationary, such that $s_{i,t+1} - \theta s_{j,t+1}$ is unit root non-stationary too. Similarly, for $f_{i,t} - \theta f_{j,t}$ to be non-stationary, we require that $(A_j - \theta A_j) F_t$ and/or $u_{1,t}, \ldots, u_{N,t}$ are unit root non-stationary. Of course, only one of the conditions have to be met for the EMH not to fail, and in the current paper we therefore test whether $e_{1,t}, \ldots, e_{N,t}$ and $u_{1,t}, \ldots, u_{N,t}$ are unit root non-stationary. But we also have

$$f_{i,t} = (\alpha_i - \beta_i) D_{i,p} + (\lambda_i - \Lambda_i) F_t + e_{i,t} - u_{i,t}$$

Table V. 5% size and size-corrected power for the intercept-only case ($p = 0$) when $C_i = 0_{r \times (m+1)}$

<table>
<thead>
<tr>
<th>Size</th>
<th>$N$</th>
<th>$T$</th>
<th>$P_{0.00}$</th>
<th>$P_{0.01}$</th>
<th>$P_{0.02}$</th>
<th>$P_{0.03}$</th>
<th>$P_{0.04}$</th>
<th>$P_{0.05}$</th>
<th>PMSB$_0^{PC}$</th>
<th>PMSB$_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>0.2148</td>
<td>0.2006</td>
<td>0.1154</td>
<td>0.1056</td>
<td>0.0378</td>
<td>0.0376</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>0.1478</td>
<td>0.1214</td>
<td>0.0844</td>
<td>0.062</td>
<td>0.0328</td>
<td>0.0326</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>35</td>
<td>0.1428</td>
<td>0.1194</td>
<td>0.0868</td>
<td>0.0672</td>
<td>0.0438</td>
<td>0.0338</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.1438</td>
<td>0.1116</td>
<td>0.096</td>
<td>0.0674</td>
<td>0.0484</td>
<td>0.0364</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>0.2108</td>
<td>0.2228</td>
<td>0.1328</td>
<td>0.1382</td>
<td>0.0504</td>
<td>0.0492</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.1072</td>
<td>0.1090</td>
<td>0.06</td>
<td>0.0692</td>
<td>0.0218</td>
<td>0.0252</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>35</td>
<td>0.0974</td>
<td>0.0882</td>
<td>0.0696</td>
<td>0.0556</td>
<td>0.0318</td>
<td>0.034</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>50</td>
<td>0.0892</td>
<td>0.0794</td>
<td>0.0628</td>
<td>0.0526</td>
<td>0.04</td>
<td>0.036</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>10</td>
<td>0.2516</td>
<td>0.2752</td>
<td>0.1778</td>
<td>0.1924</td>
<td>0.108</td>
<td>0.1162</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>20</td>
<td>0.1064</td>
<td>0.116</td>
<td>0.066</td>
<td>0.0746</td>
<td>0.0358</td>
<td>0.0378</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>0.08</td>
<td>0.085</td>
<td>0.0582</td>
<td>0.0584</td>
<td>0.0328</td>
<td>0.0356</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>50</td>
<td>0.0692</td>
<td>0.0766</td>
<td>0.051</td>
<td>0.0548</td>
<td>0.0358</td>
<td>0.0374</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.3072</td>
<td>0.3104</td>
<td>0.232</td>
<td>0.2348</td>
<td>0.1594</td>
<td>0.173</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>0.1220</td>
<td>0.134</td>
<td>0.0848</td>
<td>0.0972</td>
<td>0.05</td>
<td>0.0590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>35</td>
<td>0.0696</td>
<td>0.0856</td>
<td>0.051</td>
<td>0.0592</td>
<td>0.0328</td>
<td>0.0398</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.071</td>
<td>0.074</td>
<td>0.055</td>
<td>0.0554</td>
<td>0.0382</td>
<td>0.0404</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: A ‘PC’ superscript signifies that the test is based on original PANIC.
Table VI. Correct selection frequency of the estimated number of unit root factors for the intercept-only case \( (p = 0) \) when \( r < m + 1 \)

<table>
<thead>
<tr>
<th>( r )</th>
<th>( r_1 )</th>
<th>( \rho )</th>
<th>( \delta_0 )</th>
<th>( r_1 )</th>
<th>( \hat{r}_1^{PC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>—</td>
<td>0.9292</td>
<td>0.9144</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>—</td>
<td>0.5996</td>
<td>0.7252</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.8</td>
<td>—</td>
<td>0.3330</td>
<td>0.0678</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.9</td>
<td>—</td>
<td>0.4042</td>
<td>0.0262</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>—</td>
<td>0.0922</td>
<td>0.0120</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.8172</td>
<td>0.9042</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0.8272</td>
<td>0.9258</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.8</td>
<td>0</td>
<td>0.8164</td>
<td>0.2116</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.9</td>
<td>0</td>
<td>0.8226</td>
<td>0.0628</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.7826</td>
<td>0.0204</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.7638</td>
<td>0.7878</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6192</td>
<td>0.6054</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.8</td>
<td>0.5</td>
<td>0.5164</td>
<td>0.0772</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5332</td>
<td>0.0242</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>0.2104</td>
<td>0.0062</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9998</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>1.0000</td>
<td>0.7816</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>1.0000</td>
<td>0.2944</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.9466</td>
<td>0.0358</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.9716</td>
<td>0.9886</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8796</td>
<td>0.8804</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.8</td>
<td>0.5</td>
<td>0.7858</td>
<td>0.3016</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.9</td>
<td>0.5</td>
<td>0.7856</td>
<td>0.0974</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0.3504</td>
<td>0.0042</td>
</tr>
</tbody>
</table>

Note: A ‘PC’ superscript signifies that the test is based on original PANIC.

which means that for \( s_{i,t+1} \) and \( f_{i,t} \) to be cointegrated and one-to-one, as dictated by the FRUH, the following additional conditions must be satisfied: \( \alpha_i = \beta_i \), \( (\lambda_i - A_i)F_i \) is stationary, and \( e_{i,t} \) and \( u_{i,t} \) must be either stationary, or cointegrated and one-to-one.

As the above discussion makes clear, the ECMH and FRUH arise naturally as restrictions on the general factor model considered here. All in all, we have the following four restrictions:

- **R1.** \( e_{1,t}, \ldots, e_{N,t} \) and \( u_{1,t}, \ldots, u_{N,t} \) are unit root non-stationary.
- **R2.** \( \alpha_i = \beta_i \).
- **R3.** \( (\lambda_i - A_i)F_i \) is stationary.
- **R4.** \( e_{i,t} \) and \( u_{i,t} \) are either stationary, or cointegrated and one-to-one.

While R1 is a test of the ECMH, R2–R4 test the FRUH. In what remains we test each of restrictions. The test machinery developed in the present paper is ideally suited for this task, as it does not place any restrictions on the source of the (potential) non-stationarity of the data. Also, unlike original PANIC, in which the common component of each variable would be estimated separately, in PANICCA the common component of both variables is estimated jointly, leading to increased efficiency (see Westerlund and Urbain, 2015).\(^1\) The dataset that we use is the same as in Westerlund (2007), and consists of monthly spot and forward exchange rates relative to the US dollar. The sample covers 15 OECD countries between February 1997 and July 2006. Hence \( N = 15 \) and \( T = 115 \). The choice of dataset is motivated in part by comparability, and in part by the preference of Westerlund (2007) to treat the factors as stationary, a restriction that is never tested.

\(^1\) As a referee of this journal pointed out, one could also consider a multivariate extension of original PANIC.
Table VII. Unit root and cointegration test results.

<table>
<thead>
<tr>
<th>Common component</th>
<th>k</th>
<th>MQ₀(k)</th>
<th>Reject?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>−60.9015</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>−2.2407</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
<td>—</td>
</tr>
</tbody>
</table>

Idiosyncratic component

<table>
<thead>
<tr>
<th>Test</th>
<th>Forward</th>
<th>Spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₀</td>
<td>−3.1485</td>
<td>−3.0226</td>
</tr>
<tr>
<td></td>
<td>(0.0016)</td>
<td>(0.0025)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel-t</td>
<td>−3.2713</td>
<td>0.0005</td>
</tr>
<tr>
<td>Group-t</td>
<td>−3.3109</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

We begin by testing R1; that is, we test whether the estimated idiosyncratic components of both the spot and forward rates can be characterized as unit root non-stationary. The tests are implemented as described in Sections 3 and 4. Also, since both variables do not appear to be trending, we focus on the constant-only specification. The results reported in Table VII are mixed. In particular, while according to \( P₀ \), the unit root null should be rejected, according to PMSB₀ it should not. The evidence based on \( P₀ \) is more ambiguous, favoring a rejection at the 10% level but not at the 5% level. Of course, given the broad formulation of the alternative hypothesis (that there is at least one country for which the idiosyncratic component is stationary), ideally the results should be overwhelmingly against the null in the case of a rejection. However, this is not what we observe. In view of this, and the tendency of \( P₀ \) and \( P₀ \) to over-reject in small-\( N \) panels (see Section 4), in what follows we treat the idiosyncratic components of both \( sᵢ;ₜ \) and \( fᵢ;ₜ \) as unit root non-stationary, a conclusion that is supported by some (unreported) unit-by-unit ADF test results. The implication of this result is that \( sᵢ;ₜ \) and \( fᵢ;ₜ \) cannot be cointegrated across countries, which is consistent with R1 and hence with the ECMH.

Having tested R1 we now continue to R2–R4, which represent the FRUH part of the EMH. We begin by examining the the common component. Application of the IC information criterion described in Section 3 yields \( r = 1 \), and by further application of the MQ₀-based sequential test procedure, we obtain \( r = 1 \). Hence the common component \( Fᵣ \) of \( Zᵢ;ₜ \) consists of one unit root factor. One implication of this result is that \( (λᵣ − Λᵣ)Fᵣ \) can only be stationary if \( λᵣ = Λᵣ \) (such that \( (λᵣ − Λᵣ)Fᵣ = 0 \)). Unfortunately, since \( Fᵣ \) is only estimated up to a matrix rotation, this restriction cannot be tested. However, knowing that \( \hat{F}₀ = (\hat{F}₀,₁, \hat{F}₀,₂) \)' = \( \sum_{n=2}^{t}(\hat{X}₀,ₙ,₁, \hat{X}₀,ₙ,₂) \)' estimates \( \text{Cov} \hat{F}_₀ \), it is possible to check whether \( \lambdaᵣ \) and \( Λᵣ \) are equal on average by looking at the difference \( \hat{F}₀,₁ − \hat{F}₀,₂ \), which should be zero under equality. In order to get a feeling for the appropriateness of this restriction, we look at Figure 1, which plots \( \hat{F}₀,₁ \) and \( \hat{F}₀,₂ \). Both factors exhibit clear and strikingly similar unit root-like behavior. In fact, the lines representing the two factors almost coincide, an observation that is supported by a formal \( t \)-test of equality in mean, which we take as evidence in favor of R3. The non-stationarity of the factors further implies that previous results based on assuming either that the factors are stationary or indeed absent altogether (as in Westerlund, 2007) should be reconsidered.
A test of R4 involves testing for cointegration between $\tilde{e}_{i,t}$ and $\tilde{u}_{i,t}$, which is carried out in a very straightforward fashion. Note in particular that since $\tilde{e}_{i,t}$ and $\tilde{u}_{i,t}$ are (asymptotically) cross-section independent, we may apply any first-generation test statistic designed for such cross-section independent panels. We choose the panel-$t$ and group-$t$ statistics of Pedroni (2004), which are two of the most popular (and scrutinized) test statistics in the literature. The main difference between the two is that while the panel-$t$ statistic is based on within pooling, the group-$t$ statistic is based on between pooling. The results reported in Table VII suggest that the no cointegration null is strongly rejected even at the 1% level, which we take as evidence in favor of cointegration.

The next step in the test of R4 involves testing if the cointegrating slope on $f_{i,t}$ is indeed unity, as postulated by theory. The estimated cointegrating slopes of both the common and idiosyncratic components are reported in Table VIII. Again, given the consistency of the component estimates, the estimation of the cointegrating relationship can be carried out as if the components are in fact observed. We therefore follow the usual practice and apply fully modified LS (FMLS) and dynamic LS (DLS) techniques (see, for example, Pedroni, 2001). These are robust to endogeneity, but in the panel

Table VIII. Estimation results of the cointegrating slope

<table>
<thead>
<tr>
<th>Common component</th>
<th>MFLS</th>
<th>p-value</th>
<th>DLS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>1.0013</td>
<td>0.4853</td>
<td>0.9989</td>
<td>0.5126</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idiosyncratic component</th>
<th>MFLS</th>
<th>p-value</th>
<th>DLS</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group mean estimation</td>
<td>Slope</td>
<td>0.9363</td>
<td>0.0246</td>
<td>0.9354</td>
</tr>
<tr>
<td>Panel estimation</td>
<td>Slope</td>
<td>1.0221</td>
<td>0.0000</td>
<td>1.0235</td>
</tr>
</tbody>
</table>

Note: ‘FMLS’ and ‘DLS’ refer to the fully modified LS and dynamic LS estimator, respectively. While the group mean estimator is based on between pooling, the panel estimator is based on within pooling. The reported $p$-values correspond to a test of whether the slope is equal to one.
case not to cross-section dependence, which is also not necessary since the idiosyncratic components are cross-section independent by assumption. Analogous to the cointegration testing, we consider both a group estimator and a panel estimator, which both allow for country-specific fixed effects. The first thing to note is that the slope estimates are very close to one. In the case of the common factors, the evidence against the null hypothesis of a unit slope is insignificant, as is the evidence for the idiosyncratic component based on the group mean estimator. However, according to the panel estimator, the slope is significantly different from one. But since the point estimate is very close to one, our overall interpretation of the results is still in support of the unit slope hypothesis. The bulk of the evidence is therefore in favor of R4.

Since the deterministic component is eliminated prior to estimating the components of the data, unlike R1, R3 and R4, in PANICCA there is no natural test of R2. Westerlund and Blomquist (2013) develop a (PANIC-based) test for the presence of a linear trend in equation (1), which is based on testing whether the average of the first-differenced data is zero. We test whether \( (\bar{s}_{i,+1} - \bar{f}_i) \) is zero on average, which can be done using a simple \( t \)-test. The logic behind this test stems from the fact that under R3 and R4 \( (\bar{s}_{i,+1} - \bar{f}) = N^{-1} \sum_{i=1}^{N} (\bar{s}_{i,+1} - \bar{f}_i) \) is a consistent estimator of \( (\bar{\alpha} - \bar{\theta}) \), which is zero under R2. Applying this test to the data, we find that \( (\bar{s}_{i,+1} - \bar{f}) \approx 0.0003 \) and the associated \( t \)-statistic is 0.25, leading to a clear non-rejection of the zero intercept null. Hence this test does not provide any evidence against R2.

The results reported in this section suggest that the evidence against the EMH is weak, at best. In fact, most of the restrictions of the hypothesis seem to be satisfied in our sample. This is noteworthy because, despite the wide acceptance of the EMH in theory, most cointegration-based studies tend to reject the EMH (see Zivot, 2000, for a review of the literature). One explanation of this difference in the results is the generality of the DGP considered here, which does not impose any assumptions on the nature of the non-stationarity of the data. In fact, PANICCA seems to provide a natural platform for testing the unit root and cointegration implications of the EMH.

6. CONCLUSIONS

The CA approach of Pesaran (2006) is one of the most convenient approaches around for dealing with the effects of cross-section dependence. However, the way that this approach is implemented when testing for unit roots has resulted in test statistics with non-standard asymptotic distributions and, as a result, complicated implementation. The current paper can be seen as a reaction to this. The purpose is to develop CA-based tests that support asymptotically normal inference. As a starting point we take the PANIC approach of Bai and Ng (2004, 2010), which is one of the most general panel unit root test approaches around. Original PANIC uses PC to estimate the common and idiosyncratic components of the data. CA is more convenient and has been shown to perform relatively well in small samples. These considerations lead naturally to PANICCA: PANIC based on CA rather than PC.

ACKNOWLEDGEMENTS

A previous version of this paper was presented at the 6th Italian Congress of Econometrics and Empirical Economics in Salerno, and at seminars at Lund University and at the Norwegian School of Economics in Bergen. The authors would like to thank conference and seminar participants, and in particular Jonas Andersson, Herman van Dijk (co-editor), Yushu Li, and two anonymous referees for many valuable comments and suggestions. Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship. Both authors thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014-0112:1.
REFERENCES


Define $e_{it} = \sum_{n=2}^{t} (\Delta e_{it})^p$. Let $s_t = \hat{f}_t^p - C' f_t^p$, $S_t = \sum_{n=2}^{t} s_n$ and $d_i = \hat{\lambda}_i - C' \lambda_i$.

Lemma A.1. Under Assumptions 1–7, uniformly in $i = 1, \ldots, N$ and $t = 2, \ldots, T$,

\begin{align*}
|s_t| &= O_p(N^{-1/2}), \quad (i) \\
|S_t| &= O_p(\sqrt{T} N^{-1/2}), \quad (ii) \\
|d_i| &= O_p(T^{-1/2}) + O_p(N^{-1}). \quad (iii)
\end{align*}

Proof of Lemma A.1.

From $\hat{z}_t^p = z_t^p = C' f_t^p + v_t^p$, we obtain $s_t = \hat{z}_t^p - C' f_t^p = v_t^p = O_p(N^{-1/2})$, as required for (i). The result in (ii) is a direct consequence of this;

\begin{equation}
||T^{-1/2}S_t|| = N^{-1/2} \left\| \frac{1}{\sqrt{T}} \sum_{n=2}^{t} \sqrt{N} v_n^p \right\| = O_p(N^{-1/2}).
\end{equation}
Lemma A.2. Under the conditions of Lemma A.1,

\[ \frac{1}{NT^2} \sum_{i=1}^{T} \sum_{t=1}^{N} (\hat{\theta}_i^{p})^2 = \frac{1}{NT^2} \sum_{i=1}^{T} \sum_{t=1}^{N} (\epsilon_i^{p})^2 + O_p(N^{-1}) + O_p(T^{-1}). \]

Proof of Lemma A.2.
This proof is analogous to Proof of Lemma 1 in Bai and Ng (2010). Let us denote by $A^-$ the Moore–Penrose inverse of the matrix $A$. Note in particular that if $A$ has full row rank, then $A^- = A'(AA')^{-1}$, whereas if $A$ has column row rank, then $A^- = (A'A)^{-1}A'$. Thus, since $\bar{C}$ has full row rank, we have $\bar{C}^{-1} = (\bar{C}C)^{-1}$, such that $\bar{C}C^T = I$. Making use of this result,

$$y'_{i,t} = \lambda'_i t'_i + (\Delta e_{i,t})^p = \lambda'_i(\bar{C}^-)'t'_i + (\Delta e_{i,t})^p. \quad (4)$$

Moreover,

$$y'_{i,t} = \lambda'_i t'_i + (\bar{\Delta}e_{i,t})^p. \quad (5)$$

Subtracting (4) from (5), we obtain

$$\Delta e_{i,t}^p = (\Delta e_{i,t})^p + \lambda'_i(\bar{C}^-)'t'_i - \hat{C}_i t'_i = (\Delta e_{i,t})^p - \lambda'_i(\bar{C}^-)'(t'_i - \hat{C}_i t'_i) - (\lambda_i - \bar{C}^- \lambda_i)\hat{t'_i} = (\Delta e_{i,t})^p - \lambda'_i(\bar{C}^-)'s_i - d'_i \hat{t'_i}. \quad (6)$$

Insertion into the definition of $\hat{e}_{i,t}^p$ now yields

$$\hat{e}_{i,t}^p = \frac{1}{N T^2} \sum_{t=1}^{N} \sum_{t=2}^{T} (\Delta e_{i,t})^2 = \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t}^p + \frac{2}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \hat{a}_{i,t} + \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} a_{i,t}^2$$

with implicit definitions of $I$ and $II$. By Lemma A.1,

$$II = \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} a_{i,t}^2 = \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} [-\lambda'_i(\bar{C}^-)'s_i - d'_i \hat{t'_i}]^2$$

$$\leq \frac{2}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} [(\lambda'_i(\bar{C}^-)'s_i)^2 + (d'_i \hat{t'_i})^2]$$

$$\leq 2||\bar{C}^-||^2 \frac{1}{N} \sum_{i=1}^{N} ||\lambda_i||^2 \frac{1}{T} \sum_{t=2}^{T} ||T^{-1/2} s_i||^2 + \frac{2}{N} \sum_{i=1}^{N} ||d_i||^2 \frac{1}{T^2} \sum_{t=2}^{T} ||\bar{t'_i}||^2$$

$$= O_p(N^{-1}) + O_p(T^{-1}). \quad (7)$$

Consider $I$;

$$I = \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \hat{a}_{i,t} = - \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} \lambda'_i(\bar{C}^-)'s_i - \frac{1}{N T^2} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t} d'_i \hat{t'_i}.$$
By the Cauchy–Schwarz inequality,
\[
\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t}^p \lambda_{i} (\mathbf{C}^{-1})' \mathbf{S}_t \right\| \\
\leq \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \lambda_{i} \left\| \mathbf{C}^{-1/2} \right\| \right\| T^{-1/2} \mathbf{S}_t \right\|
\leq N^{-1/2} \left\| \mathbf{C}^{-1/2} \right\| \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \lambda_{i} e_{i,t}^p \right\| ^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| T^{-1/2} \mathbf{S}_t \right\| ^2 \right)^{1/2}
= O_p(N^{-1}),
\]
and
\[
\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} e_{i,t}^p \tilde{d}_i \tilde{F}_t \right\| \leq \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left\| \frac{N}{d_i} \sum_{t=2}^{T} e_{i,t}^p \right\| \left\| \tilde{F}_t \right\|
\leq T^{-1/2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \frac{1}{N} \sum_{i=1}^{N} d_i e_{i,t}^p \right\| ^2 \right)^{1/2} \left( \frac{1}{T^2} \sum_{t=2}^{T} \left\| \tilde{F}_t \right\| ^2 \right)^{1/2}
= O_p(T^{-1}) + O_p(T^{-1/2} N^{-1}),
\]
where the last result holds, because
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} d_i e_{i,t}^p \right\| \leq \left( \frac{1}{N} \sum_{i=1}^{N} \left\| d_i \right\| ^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| e_{i,t}^p \right\| ^2 \right)^{1/2} = O_p(T^{-1/2}) + O_p(N^{-1}).
\]
It follows that
\[
I = O_p(T^{-1}) + O_p(N^{-1}),
\]
which in turn implies
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left( \tilde{d}_i \tilde{F}_t \right)^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left( e_{i,t}^p \right)^2 + O_p(N^{-1}) + O_p(T^{-1}),
\]
as was to be shown.

\[\]

Lemma A.3. Under the conditions of Lemma A.1,
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left( \left( e_{i,1}^p \right)^2 - \left( e_{i,1}^o \right)^2 \right) = O_p(\sqrt{NT}^{-1}),
\]
\[\]
(i)
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left( \left( e_{i,T}^o \right)^2 - \left( e_{i,T}^p \right)^2 \right) = O_p(\sqrt{NT}^{-1}) + O_p(N^{-1/2}),
\]
\[\]
(ii)
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left( \Delta e_{i,t}^p \right)^2 = O_p(\sqrt{NT}^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}).
\]
\[\]
(iii)
Proof of Lemma A.3.

Part (i) is obvious. Consider (ii). From $\hat{e}_{i,t} = e_{i,t}^p + a_{i,t}$,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [(e_{i,t}^p)^2 - (e_{i,T}^p)^2] = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (2e_{i,T}^p a_{i,T} - a_{i,T}^2)$$

$$= \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} e_{i,T}^p a_{i,T} - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} a_{i,T}^2. \quad (10)$$

Making use of the fact that $a_{i,t} = -\lambda_i'(\mathbf{C})' \mathbf{s}_t - \mathbf{d}_t' \hat{f}_t$, we can show that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} a_{i,T}^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} [-\lambda_i'(\mathbf{C})' \mathbf{V}_T - \mathbf{d}_t' \hat{f}_t]^2$$

$$\leq 2\sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} ||\lambda_i||^2 ||\mathbf{C}||^2 ||T^{-1/2} \mathbf{V}_T||^2 + \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{d}_i||^2 ||T^{-1/2} \hat{f}_t||^2 \right)$$

$$= \sqrt{N}[O_p(T^{-1}) + O_p(N^{-1})] = O_p(\sqrt{NT^{-1}}) + O_p(N^{-1/2}),$$

and

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} e_{i,T}^p a_{i,T} = \left|\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} e_{i,T}^p \lambda_i' \right|\right| ||\mathbf{C}||^2 ||T^{-1/2} \mathbf{V}_T|| + \sqrt{NT}^{-1/2} \left|\left| \frac{1}{N} \sum_{i=1}^{N} e_{i,T}^p \mathbf{d}_i' \right|\right| ||T^{-1/2} \hat{f}_t||$$

$$= O_p(N^{-1/2}) + \sqrt{NT}^{-1/2}[O_p(T^{-1/2}) + O_p(N^{-1})] = O_p(N^{-1/2}) + O_p(\sqrt{NT^{-1}}).$$

The result in (ii) is implied by this.

For (iii), note that $\Delta e_{i,t}^p = \Delta e_{i,T}^p + \Delta a_{i,t}$, where $\Delta a_{i,t} = -\lambda_i'(\mathbf{C})' \mathbf{s}_t - \mathbf{d}_t' \hat{f}_t$. Therefore, $(\Delta e_{i,t}^p)^2 = (\Delta e_{i,T}^p)^2 + 2\Delta e_{i,T}^p \Delta a_{i,t} + (\Delta a_{i,t})^2$. By using this result, we obtain

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta e_{i,t}^p)^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} [2\Delta e_{i,T}^p \Delta a_{i,t} + (\Delta a_{i,t})^2]$$

$$\leq \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{i,T}^p \Delta a_{i,t} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta a_{i,t})^2$$

$$= I + II, \quad (11)$$

with implicit definitions of $I$ and $II$. The order of $II$ can be obtained in the following fashion:

$$II = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} [-\lambda_i'(\mathbf{C})' \mathbf{s}_t - \mathbf{d}_t' \hat{f}_t]^2 \leq \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} [||\lambda_i'(\mathbf{C})' \mathbf{s}_t||^2 + ||\mathbf{d}_t' \hat{f}_t||^2]$$

$$\leq 2\sqrt{N} \left( ||\mathbf{C}||^2 \frac{1}{N} \sum_{i=1}^{N} ||\lambda_i||^2 \frac{1}{T} \sum_{t=2}^{T} ||\mathbf{s}_t||^2 + \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{d}_i||^2 \frac{1}{T} \sum_{t=2}^{T} ||\hat{f}_t||^2 \right)$$

$$= \sqrt{N}[O_p(N^{-1}) + O_p(T^{-1})] = O_p(N^{-1/2}) + O_p(\sqrt{NT^{-1}})$$
For $I$,

$$I = -\frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{i,t}^{p} \lambda_{i}^{f} (\mathcal{C}^{-})' s_{i} - \frac{2}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{i,t}^{p} \mathbf{d}_{i}^{f}.$$

By using the fact that $s_{t} = \mathbf{v}_{i}^{p}$, we obtain

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{i,t}^{p} \lambda_{i}^{f} (\mathcal{C}^{-})' s_{i} \right\| \leq \frac{1}{T} \sum_{t=2}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta e_{i,t}^{p} \mathbf{d}_{i}^{f} \right\| \left\| \mathcal{C}^{-} \right\| \left\| \mathbf{v}_{i}^{p} \right\| \leq N^{-1/2} \left\| \mathcal{C}^{-} \right\| \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Delta e_{i,t}^{p} \mathbf{d}_{i}^{f} \right\| \right)^{2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \sqrt{N} \mathbf{v}_{i}^{p} \right\| \right)^{2} = O_{p}(N^{-1/2}),$$

and

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta e_{i,t}^{p} \mathbf{d}_{i}^{f} \mathbf{d}_{i}^{f} \right\| \leq \sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{d}_{i}^{f} \right\|^{2} \right)^{1/2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \Delta e_{i,t}^{p} \mathbf{d}_{i}^{f} \right\| \right)^{2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \sqrt{N} \mathbf{v}_{i}^{p} \right\| \right)^{2} = O_{p}(\sqrt{NT^{-1}}) + O_{p}(T^{-1/2}) + O_{p}(N^{-1/2}),$$

where the last result follows from

$$\left\| \frac{1}{T} \sum_{t=2}^{T} \mathbf{d}_{i}^{f} \Delta e_{i,t}^{p} \right\| \leq \left\| \frac{1}{\sqrt{T}} \sum_{t=2}^{T} \mathbf{d}_{i}^{f} \Delta e_{i,t}^{p} \right\| \leq T^{-1/2} \left\| \mathbf{d}_{i}^{f} \right\| \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \Delta e_{i,t}^{p} \right\| \right)^{1/2} \left( \frac{1}{T} \sum_{t=2}^{T} \left\| \mathbf{d}_{i}^{f} \right\| \right)^{1/2} = O_{p}(T^{-1/2}) + O_{p}(N^{-1/2}).$$

Hence,

$$I = O_{p}(\sqrt{NT^{-1}}) + O_{p}(T^{-1/2}) + O_{p}(N^{-1/2}),$$

and so we obtain

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta e_{i,t}^{p})^{2} = O_{p}((\sqrt{NT^{-1}}) + O_{p}(T^{-1/2}) + O_{p}(N^{-1/2}).$$

(12)
This establishes (iii) and hence the proof of the lemma is complete. ■

**Lemma A.4.** Under the condition of Lemma A.1,

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\epsilon}_{i,t-1}^p \Delta \hat{\epsilon}_{i,t}^p = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \epsilon_{i,t-1}^p (\Delta \epsilon_{i,t})^p + O_p(\sqrt{NT}^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}).
$$

**Proof of Lemma A.4.**

This proof follows from the same steps used in the proof of Lemma 2 in Bai and Ng (2010). We begin by noting that

$$(\hat{\epsilon}_{i,t}^p, t) = (\hat{\epsilon}_{i,t-1}^p + \Delta \hat{\epsilon}_{i,t}^p) = (\epsilon_{i,t}^p, t - 1) + 2\hat{\epsilon}_{i,t-1}^p \Delta \epsilon_{i,t}^p + (\Delta \epsilon_{i,t}^p)^2$$

which implies

$$\frac{1}{T} \sum_{t=2}^{T} \epsilon_{i,t-1}^p \Delta \epsilon_{i,t}^p = \frac{1}{2T} (\epsilon_{i,T}^p)^2 - \frac{1}{2T} (\epsilon_{i,1}^p)^2 - \frac{1}{2T} \sum_{t=2}^{T} (\Delta \epsilon_{i,t}^p)^2.$$ 

A similar result applies to $T^{-1} \sum_{t=2}^{T} \epsilon_{i,t-1}^p \Delta \epsilon_{i,t}^p$. Hence,

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} (\epsilon_{i,t-1}^p \Delta \epsilon_{i,t}^p - \epsilon_{i,t-1}^p \Delta \epsilon_{i,t}^p) \\
\leq \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{i,t}^p)^2 - (\epsilon_{i,T}^p)^2 + \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{i,1}^p)^2 - (\epsilon_{i,t}^p)^2 \\
+ \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\Delta \epsilon_{i,t}^p)^2 - (\Delta \epsilon_{i,t}^p)^2 \\
= O_p(\sqrt{NT}^{-1}) + O_p(T^{-1/2}) + O_p(N^{-1/2}),
$$

(13)

as required. ■

**Proof of Theorem 1.**

When $p = 0$ ($p = 1$) Lemmas A.2 and A.4 correspond to Lemmas 1 and 2 (Lemma 4) in Bai and Ng (2010). The proof of Theorem 1 therefore follows from using the same steps as in the proofs of Theorems 1 and 2 in this other paper. ■

**Proof of Theorem 2.**

As mentioned in Section 3.2, the rate of consistency of $\hat{F}_t^p$ is faster than in the case of PC estimation. In view of this, the proof of Theorem 2 when $p = 0$ ($p = 1$) follows directly from the proof of Theorem 1 (3) in Bai and Ng (2004). ■
Estimation of factor-augmented panel regressions with weakly influential factors

Simon Reese\textsuperscript{a} and Joakim Westerlund\textsuperscript{a,b}

\textsuperscript{a}Department of Economics, Lund University, Lund, Sweden; \textsuperscript{b}Centre for Economics and Financial Econometrics Research, School of Accounting, Economics and Finance, Deakin University, Melbourne, Australia

ABSTRACT

The use of factor-augmented panel regressions has become very popular in recent years. Existing methods for such regressions require that the common factors are strong, an assumption that is likely to be mistaken in practice. Motivated by this, the current article offers an analysis of the effect of weak, semi-weak, and semi-strong factors on two of the most popular estimators for factor-augmented regressions, namely, principal components (PC) and common correlated effects (CCE).

KEYWORDS

Common factor models; factor-augmented panel regressions; non-strong common factors.

JEL CLASSIFICATION

C12; C13; C33.

1. Introduction

Consider the scalar and $m \times 1$ vector of observable panel data variables $y_{i,t}$ and $x_{i,t}$, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ index the cross-sectional and time series dimensions, respectively. The data generating process (DGP) of the $T \times 1$ vector $y_{i} = (y_{i,1}, \ldots, y_{i,T})'$ is similar to the DGP of Bai (2009a), Greenaway-McGrevy et al. (2012), Pesaran (2006), and Westerlund and Urbain (2015), and is given by

$$ y_{i} = X_{i} \beta_{i} + \eta_{i}, $$

$$ \eta_{i} = F \lambda_{i} + \nu_{i}, $$

where $X_{i} = (x_{i,1}, \ldots, x_{i,T})'$ is $T \times m$, $\beta_{i}$ is a $m \times 1$ vector of slope coefficients, $F = (f_{1}, \ldots, f_{T})'$ is a $T \times r$ matrix of unobservable common factors with $\lambda_{i}$ being the associated $r \times 1$ vector of factor loadings, and $\nu_{i} = (\nu_{i,1}, \ldots, \nu_{i,T})'$ is a $T \times 1$ vector of idiosyncratic errors.\textsuperscript{1} In (1), $\beta_{i}$ is allowed to vary across the cross-section. Most existing work (see, for example, Bai, 2009a; Greenaway-McGrevy et al., 2012; Westerlund and Urbain, 2015), however, focuses on the case when $\beta_{1} = \cdots = \beta_{N} = \beta$, and therefore, so shall we. Hence, unless otherwise stated, $\beta_{i}$ is assumed to be homogenous, although we also consider the case when the slopes are not all equal.

The above model is the prototypical pooled panel regression with a factor error structure, in which $\nu_{i}$ is assumed independent of $X_{i}$. If $F$ is also independent of $X_{i}$, then (1) is nothing more than a static panel data regression with exogenous regressors. As is well known, such models can be consistently estimated using least squares (LS), which is true even if $F$ is unobserved and hence omitted in the estimation. If, however, $X_{i}$ is correlated with $F$, then consistency may be lost.\textsuperscript{2} To allow for this possibility, we follow Pesaran (2006) and assume that

$$ X_{i} = FA_{i} + E_{i}, $$

where $A_{i}$ is a $r \times m$ matrix of factor loadings in the equations for $\eta_{i}$ and $\nu_{i}$, and $E_{i}$ is a $T \times 1$ vector of idiosyncratic errors.

CONTACT Joakim Westerlund joakim.westerlund@nek.lu.se Department of Economics, Lund University, Box 7082, S-220 07 Lund, Sweden.

\textsuperscript{1}If the model includes unit-specific fixed effects, then $y_{i}$, $X_{i}$, $\eta_{i}$, $F$, and $\nu_{i}$ are simply the correspondingly (time) demeaned variables.

\textsuperscript{2}One occasion when LS will remain consistent even if $X_{i}$ is correlated with $F$ is if the factor loadings in the equations for $\eta_{i}$ and $X_{i}$ are uncorrelated with zero expected values.

© 2016 Taylor & Francis Group, LLC
where $\Lambda_i$ is a $m \times r$ loading matrix and $E_i = (\epsilon_{i,1}, \ldots, \epsilon_{i,T})'$ is a $T \times m$ matrix of idiosyncratic errors. This specification of $X_i$ makes for a nontrivial estimation problem in the sense that $F$ can no longer be ignored, but has to be estimated and included in (1) as an additional regressor. By combining (1)–(3),

$$Z_i = FC_i + U_i,$$

where $Z_i = (y_i, X_i) = (z_{i,1}, \ldots, z_{i,T})'$ is $T \times (m+1)$, $z_{i,t} = (y_{i,t}, x_{i,t}')'$ is $(m+1) \times 1$, $C_i = (\Lambda_i' \beta + \lambda_i, \Lambda_i')$ is $r \times (m+1)$, and $U_i = (u_{i,1}', \ldots, u_{i,T}')' = (E_i \beta + v_i, E_i)$ is $T \times (m+1)$. Thus, (1)–(3) can be rewritten equivalently as a static factor model for $Z_i$, which is convenient because it means that $F$ can be estimated (up to a matrix rotation) using existing approaches for such models. The two main approaches are principal components (PC) (see Bai, 2009a; Greenaway-McGrevy et al., 2012) and common correlated effects (CCE) (see Pesaran, 2006).

Despite the generated regressor problem caused by the use of estimated rather than known factors, the CCE and PC estimators of $\beta$ are $\sqrt{NT}$-consistent and asymptotically normal. This requires that the factors are strong, however, which of course need not be the case in practice, as when the factors are only weakly influential (see Chudik et al., 2011, p. C58; Chudik and Pesaran, 2013, p. 15, for some motivating examples). This case is discussed to some extent in a survey of Bai and Ng (2008), who use Monte Carlo simulation to show that the PC factors can be severely impaired when the factors are not strong (see also Boivin and Ng, 2006). They therefore reach the conclusion that “although much work has been accomplished in this research, much remains to be done” (p. 155). These findings recently motivated Chudik et al. (2011) to more formally consider the implications of weak, semi-weak, and semi-strong factors on the CCE estimator of $\beta$. The following quote summarizes their findings: “As predicted by the theory the CCE estimator performs well and show very little size distortions, in contrast with the iterated PC approach of Bai (2009a), which exhibit significant size distortions. The latter is partly due to the fact that in the presence of weak or semi-strong factors the PC estimates of the (rotated) factors need not be consistent. This problem does not affect the CCE estimator because it does not aim at consistent estimation of the factors but deals with error cross-section dependence generally by using cross-section averages to mop up such effects” (p. C47).

The present article is motivated by these findings. Our starting point is the preference of Chudik et al. (2011) to assume that the researcher knows with full certainty which factors that are strong and which factors that are not, and also that the non-strong factor component is uncorrelated with $X_i$, which means that, in analogy to the literature on omitted variables, it can be omitted without consequence. A more realistic scenario is that the researcher is unaware of both the strength and correlation of the factors and, therefore, that some of the supposedly strong factors may potentially be non-strong and possibly correlated with $X_i$. Onatski (2012) considers the case when some of the included factors are semi-strong, but only within the context of PC estimation of a pure common factor model. According to his results, the presence of such factors causes the PC estimator to become inconsistent.

Our purpose in the present study is to provide an analysis of the effects of the presence of non-strong factors on the CCE and PC estimators. In doing so, we generalize the present literature in several directions. First, to the best of our knowledge, the present study is the first to consider estimation of factor-augmented regressions when some of the included factors are potentially non-strong and/or correlated with $X_i$. Second, the study is the first to allow for a dependence between the strength of the factors and the relative expansion rate of $N$ and $T$, a feature that is shown to deliver significant insight. Third, the study is also the first to enable a direct comparison between the CCE and PC estimators when the factors are non-strong. Fourth, the study is the first to consider the properties of the PC estimator in case of a violation of the common slope assumption.

---

1. As a reviewer to this journal correctly pointed out, our specification imposes a linear impact of the common factors. The results in this article need not hold if the relation is in fact nonlinear.
2. See Onatski (2012) for a brief review of some articles that have considered the presence of non-strong factors within the context of a pure common factor model.
2. Assumptions

The conditions under which we will be working are summarized in Assumptions HOM, HET, ERR, LAM, RK–CCE, RK–PC, and KAP. Here and throughout this paper \( tr(\mathbf{A}) \), \( rk(\mathbf{A}) \) and \( ||\mathbf{A}|| = \sqrt{tr(\mathbf{A}^T\mathbf{A})} \) denote the trace, rank and Frobenius (Euclidean) norm, respectively, of the matrix \( \mathbf{A} \), and \( M < \infty \) is a generic positive number.

**Assumption HOM.** \( \beta_1 = \cdots = \beta_N = \beta \).

As mentioned in the introduction, the assumption of a common slope coefficient \( \beta \) is standard in the literature. One exception is Pesaran (2006) (see also Chudik et al., 2011), who allows some variation in \( \beta \), by requiring that it satisfies a random coefficient assumption. Because of this, our main analysis in Section 3.1, which is conducted under Assumption HOM, is in Section 3.2 complemented with an analysis under Assumption HET.

**Assumption HET.**

(i) \( \beta_i = \beta + \xi_i \), where \( \xi_i \) is independently and identically distributed (iid) with \( E(\xi_i) = 0 \),
\[ E(\xi_i^2) = \Sigma_\xi \text{ positive semidefinite}, ||\beta|| \leq M \text{ and } ||\Sigma_\xi|| \leq M. \]

(ii) \( \xi_i \) is independent of \( \lambda_i, \Lambda_i, v_{i,t}, E_i, \text{ and } F_i \).

**Assumption ERR.**

(i) \( v_{i,t} \) is iid across both \( i \) and \( t \) with \( E(v_{i,t}) = 0 \), \( E(v_{i,t}^2) = \sigma_{v,i}^2 > 0 \), and \( E(v_{i,t}^4) \leq M \).

(ii) \( \epsilon_{i,t} \) is iid across both \( i \) and \( t \) with \( E(\epsilon_{i,t}) = 0 \), \( E(\epsilon_{i,t}\epsilon_{i,t}') = \Sigma_\epsilon \text{ positive definite} \) and
\[ E(||\epsilon_{i,t}||^4) \leq M. \]

(iii) \( f_t \) is covariance stationary such that \( E(||f_t||^4) \leq M \) and \( E(f_t^2f_t') = \Sigma_f \text{ is positive definite}. \)

(iv) \( v_{i,t}, E_i, \text{ and } F \) are mutually and pairwise independent.

Assumption ERR is quite restrictive, but can be relaxed at the expense of added expositional and technical complexity. However, as argued by (Chudik et al., 2011, p. C53), this seems unnecessary in the present case, because the results regarding the strength of the factors are unlikely to be affected by the removal of nuisance parameters. The assumption that \( v_{i,t} \) is iid over time is, for instance, not necessary and can be relaxed by simply replacing \( \sigma_{v,i}^2 \) by \( \omega_{v,i}^2 = \sum_{s=-\infty}^{\infty} E(v_{i,t}v_{i,t-s}). \) Standard (parametric or kernel-based) estimators can be used if \( \omega_{v,i}^2 \) is unknown. Serial correlation in \( \epsilon_{i,t} \) can be accounted for in the same way.

**Assumption LAM.**

(i) \( \lambda_i = N^{-\alpha}\lambda_i^0 \) and \( \Lambda_i = N^{-\alpha}\Lambda_i^0 \), where \( \alpha \in [0, 1] \).

(ii) \( \lambda_i^0 \) and \( \Lambda_i^0 \) are either random such that \( E(||\lambda_i^0||^4) \leq M \) and \( E(||\Lambda_i^0||^4) \leq M \), or non-random such that \( ||\lambda_i^0|| \leq M \) and \( ||\Lambda_i^0|| \leq M \).

(iii) \( \lambda_i^0 \) and \( \Lambda_i^0 \) are independent of \( v_{i,t}, E_i \) and \( F \).

Assumption LAM allows for very general cross-section dependencies. Indeed, in contrast to Onatski (2012), who only considers the case when \( \alpha = 1/2 \), the factors may be strong (\( \alpha = 0 \)), semi-strong (0 < \( \alpha < 1/2 \)), semi-weak (1/2 \( \leq \alpha < 1 \)), or weak (\( \alpha = 1 \)) (see Chudik et al., 2011, Definition 3.1, and the discussion that follows it). The fact that \( \alpha \) is the same for all loadings in \( \lambda_i \) and \( \Lambda_i \) is not very restrictive. If the loadings shrink to zero at different rates, then the results will be dominated by the slowest shrinking loading, and in such a case our results can be thought of as emanating from an analysis of the effects of the strongest factor. Indeed, as (Chudik and Pesaran, 2013, p. 11) point out, in a multifactor setup with differing rates of shrinking \( \alpha \) can be thought of as representing their maximum.

---

5Consider \( \lim_{N \to \infty} N^{-\gamma} \sum_{i=1}^{N} ||\lambda_i|| < \infty \). According to (Chudik et al., 2011, Condition (3.7)), the conditions for \( \lambda_i \) to be strong, semi-strong, semi-weak, and weak are given by \( \gamma = 1, \gamma = 1/2 \leq \gamma < 1, 0 < \gamma < 1/2, \) and \( \gamma = 0 \). These conditions correspond to our classification via the relation \( \alpha = 1 - \gamma \). For ease of exposition, we deviate from Chudik et al. (2011) by counting \( \alpha = 1/2 \) to semi-weak factors.
In addition to Assumption LAM, we require that the fitted number of factors is equal to the true number, \( r \). Let us therefore denote by \( k \) the number of factors used in PC. Let \( \mathbf{C}^0 = N^{-1} \sum_{i=1}^{N} \mathbf{C}_i^0 \) and \( \mathbf{Q}^0 = N^{-1} \sum_{i=1}^{N} \mathbf{C}_i^0 \mathbf{C}_i^0' = N^{-1} \mathbf{C}^0 \mathbf{C}^0', \) where \( \mathbf{C}_i^0 = [(\mathbf{A}_i^0)' \beta + \lambda_i^0, (\mathbf{A}_i^0)'] \) is such that \( \mathbf{C}_i = N^{-\alpha} \mathbf{C}_i^0 \) and \( \mathbf{C}_i^0 = (\mathbf{c}_i^0_1, \ldots, \mathbf{c}_i^0_N)' \). Also, \( \mathbf{C}^* = E(\mathbf{C}^0) \) and \( \mathbf{Q}^* = E(\mathbf{Q}^0). \) The following assumptions are enough for our purposes.

**Assumption RK–CCE.** \( ||\mathbf{C}^0 - \mathbf{C}^*|| = o_p(1), \) where \( rk(\mathbf{C}^0) = rk(\mathbf{C}^*) = r = m + 1 \) and \( ||\mathbf{C}^*|| \leq M. \)

**Assumption RK–PC.** \( ||\mathbf{Q}^0 - \mathbf{Q}^*|| = o_p(1), \) where \( rk(\mathbf{Q}^0) = rk(\mathbf{Q}^*) = r = k \) and \( ||\mathbf{Q}^*|| \leq M. \)

Assumptions RK–CCE and RK–PC are more restrictive than those employed by (Pesaran, 2006, Equation (21)) and (Bai, 2009b, Section C.3) which are tantamount to requiring that \( rk(\mathbf{C}^0) = rk(\mathbf{C}^*) = r \leq m + 1 \) and \( rk(\mathbf{Q}^0) = rk(\mathbf{Q}^*) = r \leq k, \) respectively. However, since relaxing Assumptions RK–CCE and RK–PC greatly complicates the proofs without affecting the conclusions (see Section 3.1 for a discussion), we choose to work under the more restrictive assumptions. As Assumption RK–CCE makes clear, the number of common factors permitted within the CCE approach should be equal to the number of observables, \( m + 1. \) Specifically, the number of common factors should be equal to the number of observables that depend on \( F. \) In the DGP considered here, \( F \) enters the equations of both \( y_i \) and \( x_i, \) which means that there should be \( m + 1 \) factors. These restrictions make CCE less flexible than PC, in which \( k \) can be set arbitrarily by the researcher. That is, unlike in CCE, in PC the number of common factors is not bounded by the number of observables. Indeed, as Bai (2009a) shows, provided that it enters the equation for \( y_i \) in PC \( F \) need not enter the equation for \( x_i. \) Hence, not even \( rk(\mathbf{Q}^0) = rk(\mathbf{Q}^*) = r \leq k \) is necessary. Such allowances are possible also within the CCE framework (see Pesaran, 2006). However, as Westerlund and Urbain (2013) show, this requires assuming the loadings \( \lambda_i \) and \( A_i \) to be mutually uncorrelated which, if false, renders the CCE estimator inconsistent. Hence, even if RK–CCE can in principle be relaxed, in most situations of practical relevance this is not the case. PC is therefore more general in this regard.

Chudik et al. (2011) assume the existence of \( r \leq m + 1 \) strong factors and \( n \) non-strong factors. Since the researcher is assumed to know the strength of the factors, the strong factors can be treated exactly as in Pesaran (2006). Moreover, by assuming that the non-strong factors are mean zero and independent of all other random elements of the model (including \( x_i \)), in analogy to the classical literature on omitted variables, the non-strong factors can be ignored, provided that \( n \) is fixed or increases at a slower rate than \( N \) (see Chudik and Pesaran, 2013, Example 4, for an illustration of the effects of omitted factors when the factors are correlated with \( x_i \)).\(^6\) Of course, in practice one never knows beforehand which factors are strong and which factors are not, and therefore, the included factors are likely to have different strengths. Indeed, as Onatski (2012) shows, in some circumstances methods originally designed to estimate \( r \) in the strong factor-only case remain valid also in the non-strong factor case. Hence, from a practical point of view the most relevant scenario is when the strength of the included factors is unknown.

The assumption that \( \mathbf{C}^0 (\mathbf{Q}^0) \) has the same rank as \( \mathbf{C}^* (\mathbf{Q}^*) \) is not necessary and can be relaxed at the cost of technical complexity. A minimal requirement in case of \( \mathbf{C}^0 \) is that \( P[rk(\mathbf{C}^0) = rk(\mathbf{C}^*)] \rightarrow 1 \) as \( N, T \rightarrow \infty \) (see Andrews, 1987), which obviously satisfied under Assumption RK–CCE (see also Chudik et al., 2011, Proof of Lemma A.2, for a formal argument).

For the sake of simplicity, in what follows we will frequently refer to Assumption RK–CCE/RK–PC with the understanding that Assumption RK–CCE (RK–PC) only applies to CCE (PC).

\(^6\)In the classical literature the number of omitted variables is usually treated as fixed. A contribution of Chudik et al. (2011) over this literature is therefore the consideration of an increasing number of non-strong factors. However, since these factors are not allowed to be correlated with the explanatory variables, which is highly plausible in practice, this contribution is mainly technical.
The types of factors that can be permitted are related to the relative expansion rate of \( N \) and \( T \). In order to capture relationships of this sort, we make use of the following assumption.

**Assumption KAP.** \( T = N^\kappa \), where \( \kappa > 0 \).

Assumption KAP is less “flexible” than assuming that \( T \) is proportional to \( N^\kappa \). However, since the conclusions are qualitatively the same, and since assuming \( T = N^\kappa \) greatly simplifies both transparency and notation, we opt for the less flexible specification in the present article. Note in particular how Assumption KAP does not put any restrictions on the values taken by \( \kappa \), provided that \( \kappa > 0 \). This is in contrast to Onatski (2012), who assumes that \( N/T \to c > 0 \), which under Assumption KAP is analogous to requiring \( \kappa = 1 \).

3. Results

3.1. Results under Assumption HOM

As is well known, since \( F \) and \( C_i \) are not separately identifiable, \( F \) can only be estimated up to a matrix rotation. Depending on the assumed data generating process (DGP), the CCE and PC estimators of the suitably rotated version of \( F \) can be constructed in different ways. For example, while in Bai (2009a) \( F \) is estimated by applying PC to the residuals from the LS fit of (1), in Greenaway-McGrevy et al. (2012), Kapetanios and Pesaran (2005), and Westerlund and Urbain (2015) the estimation is carried out by applying PC (and CCE) to \( Z_i \). In the current setup, the estimators are applied to \( Z_i \) in order to exploit the common factor structure of \( X_i \). Hence, while the CCE estimator is just \( \hat{F}_{CCE} = (\hat{f}_{CCE}^1, \ldots, \hat{f}_{CCE}^T)' = \hat{Z} = N^{-1} \sum_{i=1}^{N} Z_i \), the PC estimator, denoted \( \hat{F}_{PC} = (\hat{f}_{PC}^1, \ldots, \hat{f}_{PC}^T)' \), is \( \sqrt{T} \) times the matrix consisting of the eigenvectors corresponding to the \( k \) largest eigenvalues of the \( T \times T \) matrix \( ZZ' \), where \( Z = (Z_1, \ldots, Z_N) \) is \( T \times N(m+1) \).

The motivation for using panel estimation has always been, and continues to be, the increased accuracy that becomes available when the slope coefficients can be assumed to be homogenous. Most of the existing literature on factor-augmented regressions is therefore based on assuming that the slopes are indeed homogenous. Moreover, since under homogeneity the most efficient method of combining the data across the cross-section is “pooling,” as opposed to “mean grouping,” most estimators are pooled. In fact, in case of PC the literature has not yet ventured outside the homogeneous slope environment, and to the best of our knowledge all PC estimators in the literature are pooled (see, for example, Bai, 2009a; Greenaway-McGrevy et al., 2012). In this article we therefore focus on the pooled CCE and PC estimators under Assumption HOM, although in Section 3.2 we also consider briefly the properties of these estimators under Assumption HET. In Appendix D we provide some results for both the individual and the group mean CCE estimators considered by Pesaran (2006).

The pooled factor-augmented CCE and PC estimators of \( \beta \) have the following form:

\[
\hat{\beta}_{n}^p = \left( \sum_{i=1}^{N} X_i' M_{\hat{F}_n} X_i \right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}_n} y_i,
\]

where \( n \in \{\text{CCE, PC}\} \), the \( p \) superscript signifies that the estimators are pooled, and \( M_{\hat{F}_n} = I_T - \hat{F}_n' (\hat{F}_n \hat{F}_n')^{-1} \hat{F}_n' \). Hence, \( \hat{\beta}_{n}^p \) is just the LS estimator of \( \beta \) with \( \hat{F}_n \) in place of \( F \). Theorem 1, which is our main result, gives the asymptotic distribution of \( \sqrt{NT} (\hat{\beta}_{n}^p - \beta) \). Before we come to the theorem, however, we need to introduce some notation. In particular, let us define

\[
b_{n,i} = b_{1n,i} - b_{2n,i} - b_{3n,i},
\]

where

\[
b_{1CCE,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0,
\]

and

\[
b_{2CCE,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0,
\]

\[
b_{3CCE,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0,
\]

\[
b_{1PC,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0,
\]

\[
b_{2PC,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0,
\]

\[
b_{3PC,i} = \lambda_i^0 (\bar{C}^0)' \Sigma_{ii} (\bar{C}^0)' \lambda_i^0.
\]
\[
\mathbf{b}_{2\text{CCE},j} = \Sigma_{\epsilon,j}(\beta, \mathbf{I}_m)(\mathbf{C}^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{3\text{CCE},j} = \sigma_{\nu,j}^2(\mathbf{C}_0^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{1\text{PC},j} = \Sigma_{\epsilon,j}(\beta, \mathbf{I}_m)(\mathbf{C}^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{2\text{PC},j} = \Sigma_{\epsilon,j}(\beta, \mathbf{I}_m)(\mathbf{C}^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{3\text{PC},j} = \sigma_{\nu,j}^2(\mathbf{C}_0^0)\Sigma_{\epsilon,j}^0.
\]

for \( n = \text{CCE} \), and

\[
\mathbf{b}_{1\text{PC},j} = \Sigma_{\epsilon,j}(\beta, \mathbf{I}_m)(\mathbf{C}^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{2\text{PC},j} = \Sigma_{\epsilon,j}(\beta, \mathbf{I}_m)(\mathbf{C}^0)\Sigma_{\epsilon,j}^0, \\
\mathbf{b}_{3\text{PC},j} = \sigma_{\nu,j}^2(\mathbf{C}_0^0)\Sigma_{\epsilon,j}^0.
\]

for \( n = \text{PC} \). Theorem 1 is stated in terms of

\[ b_n = N^{-1} \sum_{i=1}^{N} b_{n,i}, \quad \mathbf{W} = N^{-1} \sum_{i=1}^{N} \sigma_{\nu,i}^2 \Sigma_{\epsilon,j}^0, \quad \text{and} \quad \Sigma_{\epsilon} = N^{-1} \sum_{i=1}^{N} \Sigma_{\epsilon,j}. \]

Theorem 1. Suppose that Assumptions HOM, ERR, LAM, RK–CCE/RK–PC, and KAP hold, and that \( \alpha < 1/2 \). Suppose also that \( \kappa \in K_{\text{CCE}} = (2\alpha, 3 - 4\alpha) \) for \( n = \text{CCE} \), and \( \kappa \in K_{\text{PC}} = (\max(1/2 - \alpha, (4\alpha + 1)/3), 2 - \alpha) \) for \( n = \text{PC} \). Then, as \( N, T \to \infty \),

\[
\sqrt{NT}(\hat{\beta}^n - \beta) \to_d N \left( 0, \lim_{N \to \infty} \Sigma_{\epsilon}^{-1} \mathbf{W} \Sigma_{\epsilon}^{-1} \right) + \lim_{N \to \infty} \Sigma_{\epsilon}^{-1} N^{(\kappa - 1)/2} b_n,
\]

where \( \to_d \) signifies convergence in distribution.

The asymptotic distribution given in Theorem 1 is similar to those reported by (Bai, 2009a, Theorem 3) and (Pesaran, 2006, Theorem 4). In particular, we see that the asymptotic distribution consists of two terms: (i) a normal variate with mean zero and covariance matrix \( \lim_{N \to \infty} \Sigma_{\epsilon}^{-1} \mathbf{W} \Sigma_{\epsilon}^{-1} \), which is identical to the asymptotic distribution of the LS estimator of \( \beta \) when \( F \) is known, and (ii) a bias term involving \( N^{(\kappa - 1)/2} b_n \) that depends on both \( \kappa \) and \( n \in \{\text{CCE}, \text{PC}\} \). Since the normal variate does not depend on \( n \), from a distributional point of view, the main difference between the estimators is the bias. Westerlund and Urbain (2015) compare the properties of the CCE and PC estimators under the strong factor assumption (see also Sarafidis and Wansbeek, 2012). They end up with similar bias expressions. Their analysis suggests that, except for some highly specialized cases, it is not possible to work out the relative bias, at least not analytically. The same is true in the present more general case. Of course, in this article we are not interested in the magnitude of the different bias components per se, but rather in how the distributional results are affected by \( \alpha \) (and \( \kappa \)). The following observations regarding the impact of \( \alpha \) can be made:

- Both estimators require \( \alpha < 1/2 \), that is, the factors need to be strong or semi-strong. If this is not the case, the asymptotic distributions of the CCE and PC estimators become dependent on nuisance parameters reflecting, among other things, the value of \( \alpha \geq 1/2 \). This finding is in agreement with the results of Onatski (2012), showing that the PC estimator of \( F \) is inconsistent when \( \alpha = 1/2 \).
- The ranges of values for \( \kappa \) that can be allowed under Theorem 1, \( K_{\text{CCE}} \) and \( K_{\text{PC}} \), depend critically on \( \alpha \). Consider the two extreme cases of \( \kappa = 0 \) and \( \alpha = 1/2 - \nu \), where \( \nu > 0 \) is a small number. When \( \alpha = 0 \), \( K_{\text{CCE}} = (0, 3) \) and \( K_{\text{PC}} = (1/2, 2) \), whereas when \( \alpha = 1/2 - \nu \), then \( K_{\text{CCE}} = (1 - 2\nu, 1 + 4\nu) \) and \( K_{\text{PC}} = (1 - 4\nu/3, 3/2 + \nu) \). This result leads to the following two conclusions. First, \( K_{\text{CCE}} \) and \( K_{\text{PC}} \) get narrower when \( \alpha \) increases. Second, \( \alpha \) determines not only the width of \( K_{\text{CCE}} \) and \( K_{\text{PC}} \) but also their relative width; when \( \alpha = 0 \), \( K_{\text{CCE}} \) is relatively wide, but this relation reverses when \( \alpha = 1/2 - \nu \) with \( \nu \) sufficiently small. Hence, when it comes to the allowable values of \( \kappa \), the choice of which estimator to use depends critically on \( \alpha \).

\[ \text{In case of the PC estimator, the bias also appears in Theorem 3 of Bai (2009a). However, there is no bias in Theorem 4 of Pesaran (2006). However, Pesaran assumes } \frac{T}{N} = \frac{N}{N^2} = o(1), \text{ in which case the bias in our Theorem 1 converges to zero. In this sense, Theorem 1 generalizes Theorem 4 of Pesaran (2006) in two directions: (i) it allows for non-strong factors, and (ii) there is no requirement that } \frac{N}{T} \text{ should go to zero.} \]
When $\alpha \to 1/2$, $K_{CCE}$ and $K_{PC}$ shrink towards the value one, suggesting that the weaker the factor
the more restrictive the conditions placed on $N$ and $T$.

While theoretically a restriction, the assumed relationship between $N$ and $T$ is very relevant from
an applied point of view. Indeed, since $\kappa = \ln(T)/\ln(N)$ for any $N$ and $T$, as the following examples
illustrate, the practical implications of Theorem 1 are very straightforward. Many empirical studies use
data sets where $N \approx T$, suggesting that $\kappa \approx 1$ (see, for example, Chudik et al., 2011; Bertoli and
Fernández-Huertas Moraga, 2012; Calderón et al., 2015; Herzer et al., 2012). As Theorem 1 makes clear in,
terms of the allowable values of $\alpha$, when $\kappa = 1$ there is no difference between the CCE and PC
estimators. The choice of which estimator to use in this case is therefore mainly a matter of personal
preference. However, there are also studies where $N$ and $T$ are quite different, in which case $\kappa \neq 1$.
In Arnold et al. (2011), for example, $N = 21$ and $T = 34$, implying that $\kappa \approx 1.62$. Since in this case
$\kappa < 5/3$, when it comes to the allowable values of $\alpha$, the PC estimator is the preferred choice. By contrast,
in studies such as in Eberhardt et al. (2013) where $N > T$, the CCE estimator is clearly preferred, for in
this case the set of allowable values of $\alpha$ is always larger for CCE than for PC.

In order to illustrate the implications of Theorem 1 in Table 1, we report some Monte Carlo results for
the simulated means and standard deviations (STDs) of $\sqrt{NT}(\hat{\beta}^p_{PC} - \beta)$ and $\sqrt{NT}(\hat{\beta}^p_{CCE} - \beta)$, and
their theoretical predictions. The DGP is a restricted version of the one given in (4), and sets $m = 1$, $r = 2,
\beta = 0$, and $(f_t, v_t, \epsilon_t)^\prime \sim N(0, \text{diag}(\Sigma_f, \sigma^2_{v, i}, \Sigma_{\epsilon, i}))$, where $\Sigma_f = 4I_2$, $\sigma^2_{v, i} = 4$, and $\Sigma_{\epsilon, i} = I_m$. While
the variation in $\lambda_0^\prime$ and $\Lambda_0^\prime$ is not unimportant, for the purpose of illustration it is enough to consider
a single change. We therefore set $\lambda_0^\prime = \lambda^* \epsilon_1 + [0.5 - 1(i > \lfloor 0.5N \rfloor)]t_2$, where $\epsilon_1 = (1, 0)^\prime$, $t_2 = (1, 1)^\prime$, $1(A)$ is the indicator function for the event $A$, and $|x|$ is the integer part of $x$. Thus, if $\lambda^* = 0$, then $\lambda_0^\prime$ is
$0.5t_2$ for the first $\lfloor 0.5N \rfloor$ units and $-0.5t_2$ for the rest. The corresponding specification of $(\Lambda_0^\prime)^\prime$ is given by $(\Lambda_0^\prime)^\prime = \Lambda^* \epsilon_2 + [0.5 - 1(i > \lfloor 0.5N \rfloor)]t_2$, where $\epsilon_2 = (0, 1)^\prime$. In order to ensure that Assumptions RK–CCE and RK–PC are satisfied, in this experiment we set $\lambda^* = 1$ and $\Lambda^* = 0.4$. All results are
based on 3,000 replications. We see that the asymptotic theory provides very accurate approximations of
actual behavior when $\alpha \in \{0, 1/4\}$, regardless of the value of $T$. However, this is no longer the case when
$\alpha = 1/2$. Specifically, while the variance predictions are still very accurate, the predicted means are way
off target. A similar pattern emerges when looking at the relative performance of the two estimators.
Specifically, while the STDs tend to be very close, there is a difference in bias. In the particular DGP
considered here CCE is more biased than PC, but it is possible to find other DGPs in which it is the
other way around. The results therefore give little or no practical guidance, except that one should be
careful not to assume that all factors are strong when in fact some (or indeed all) of the factors may be
weak.

Consider Assumption RK–CCE. As already mentioned in Section 2, the condition that $r = m + 1$ can be relaxed to allow $r \leq m + 1$ Pesaran (2006). However, this means that the $r \times (m + 1)$ matrix $\overline{C}^0$
will be a singular if $r < m + 1$, as $rk(\overline{C}^0) = r \leq (m + 1)$. This singularity is a source of great technical
complexity. However, the net effect on the results reported in Theorem 1 is very simple and intuitive.
Indeed, all one has to do is to replace all instances of $(\overline{C}^0)^{-1}$ in the definitions of $b_{1CCE,i}$, $b_{2CCE,i}$, and
$b_{3CCE,i}$ by $(\overline{C}^0)^{-1}$, where $A^-$ is the Moore–Penrose inverse of the matrix $A$. The same is true for the PC
estimator. Of course, as already pointed out, the focus here is not really on the exact form of these bias
expressions, but rather on the effects of $\alpha$ and $\kappa$, which are unaffected by allowing $r \leq m + 1$.

Theorem 1 implies that the two estimators are consistent and that the bias disappears at rate
$(NT)^{-1/2}N(\kappa^{-1/2}) = N^{-1}$. However, requiring that $\alpha < 1/2$ and $\kappa \in K_{CCE}$ and/or $\kappa \in K_{PC}$ is only
necessary for deriving the asymptotic distributions, and not for consistency. This is made clear in the
following corollary to Theorem 1, which provides the relevant conditions for consistency.

---

8While the results do depend to some extent on the variation in $\lambda_0^\prime$ and $\Lambda_0^\prime$, in terms of the effect of $\kappa$ and $\alpha$, the conclusions
do not depend on how $\lambda_0^\prime$ and $\Lambda_0^\prime$ are generated.
Table 1. Monte Carlo evaluation of the asymptotic distributions of the pooled PC and CCE estimators.

<table>
<thead>
<tr>
<th>( \kappa )</th>
<th>( \alpha )</th>
<th>( T = 100 )</th>
<th>( T = 200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 3/4 )</td>
<td>0</td>
<td>-1.0957</td>
<td>-1.1231</td>
</tr>
<tr>
<td>( 1 )</td>
<td>0</td>
<td>-0.6389</td>
<td>-0.6589</td>
</tr>
<tr>
<td>( 3/4 )</td>
<td>1/4</td>
<td>-1.1591</td>
<td>-1.1231</td>
</tr>
<tr>
<td>( 1 )</td>
<td>1/4</td>
<td>-0.6850</td>
<td>-0.6589</td>
</tr>
<tr>
<td>( 3/4 )</td>
<td>1/2</td>
<td>-1.4381</td>
<td>-1.1231</td>
</tr>
<tr>
<td>( 1 )</td>
<td>1/2</td>
<td>-1.0813</td>
<td>-0.6589</td>
</tr>
<tr>
<td>( 3/4 )</td>
<td>0</td>
<td>-1.2607</td>
<td>-1.2746</td>
</tr>
<tr>
<td>( 1 )</td>
<td>0</td>
<td>-0.6088</td>
<td>-0.6749</td>
</tr>
<tr>
<td>( 3/4 )</td>
<td>1/4</td>
<td>-1.3176</td>
<td>-1.2746</td>
</tr>
<tr>
<td>( 1 )</td>
<td>1/4</td>
<td>-0.6444</td>
<td>-0.6749</td>
</tr>
<tr>
<td>( 3/4 )</td>
<td>1/2</td>
<td>-1.6379</td>
<td>-1.2746</td>
</tr>
<tr>
<td>( 1 )</td>
<td>1/2</td>
<td>-1.0835</td>
<td>-0.6749</td>
</tr>
</tbody>
</table>

*“Mean” and “STD” refer to the simulated mean and standard deviation of \( \sqrt{NT}(\hat{\beta}_n^P - \beta) \), respectively. The theoretical predictions, reported as “Theory,” are obtained by simulating the asymptotic distributions given in Theorem 1.*

**Corollary 1.** Suppose that Assumptions HOM, ERR, LAM, RK–CCE/RK–PC, and KAP hold. Suppose also that \( \kappa > \max\{6\alpha - 4, 3\alpha - 3/2\} \) with \( \alpha < 1 \) for \( n = \text{CCE} \), and \( \kappa > \max\{2\alpha, 4\alpha - 1\} \) for \( n = \text{PC} \). Then, as \( N, T \to \infty \),

\[
||\hat{\beta}_n^P - \beta|| = o_p(1).
\]

Corollary 1 suggests that there is a trade-off between the allowable values of \( \alpha \) and the restrictions placed on \( \kappa \). On the one hand, since \( \max\{6\alpha - 4, 3\alpha - 3/2\} \leq \max\{2\alpha, 4\alpha - 1\} \) for all \( \alpha \in [0, 1) \), PC is relatively demanding when it comes to the values of \( \kappa \) needed for consistency. On the other hand, if \( \kappa \) is just large enough, PC is consistent even when \( \alpha = 1 \), which is not the case for CCE. Of course, since the relevant condition on \( \kappa \) in this case is given by \( \kappa > 3 \), allowing \( \alpha = 1 \) comes at a high cost in terms of both bias (as \( N^{(\kappa-1)/2} \) is increasing in \( \kappa \)) and the required size of \( T \) (in the sense that \( T >> N \)).

The finding that the PC estimator is consistent even when \( \alpha \geq 1/2 \) would seem to go against the results of (Onatski, 2012, Theorem 1) showing how the PC estimator of \( \beta \) is rendered inconsistent when \( \alpha = 1/2 \). However, this is actually not the case. The difference lies with the current more flexible specification of \( \alpha \) and \( \kappa \). Indeed, if we assume that \( \kappa = 1 \), which under Assumption KAP is equivalent to Onatski’s (2012) requirement that \( N/T \to \epsilon > 0 \), then, as Corollary 1 makes clear, the PC estimator is inconsistent for all \( \alpha \geq 1/2 \) (including \( \alpha = 1/2 \)). The results reported herein are therefore consistent with those of Onatski (2012).

As an illustration of Corollary 1, we collect the mean and STD of \( \hat{\beta}_n^{\text{CCE}} - \beta \) over 3,000 simulated samples and report them in Table 2. We use the same DGP as before but now with the following parameter values: \( \Sigma_f = 10I_2 \), \( \sigma_{v,i}^2 = 10 \), \( \Sigma_{\epsilon,i} = 10^{-4}I_m \), and \( \kappa = 1 \). The factor loadings are obtained by setting \( \lambda_e^* = 1 \), \( \Lambda_e^* = 10 \), and by increasing the variation around \( \lambda_e^*e_1 \) and \( \Lambda_e^*e_2 \) by a factor of 20 relative to the previous DGP. For the purpose of demonstrating the requirements for convergence, it is convenient to look at the STDs as a function of the sample size. For \( \alpha < 1 \), the values decrease constantly, but the closer \( \alpha \) is to one, the slower this decrease becomes. As expected, the STDs stagnate when \( \alpha = 1 \), and they even begin to increase if the value of \( \alpha \) is pushed above one.

### 3.2. Results under Assumption HET

As mentioned in Section 3.1, in the strong factor case, while the CCE strand of the literature has considered both Assumptions HOM and HET (see, for example, Pesaran, 2006; Chudik et al., 2011), the PC strand has only considered Assumption HOM (see, for example, Bai, 2009a; Greenaway-McGrevy et al., 2012; Westerlund and Urbain, 2015). In this section, we study the pooled CCE and PC estimators...
under Assumption HET. Hence, even in the strong factor case, the results reported here represents an extension of the existing PC results under HOM.

**Theorem 2.** Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold. Suppose also that 
\[ \kappa > \max\{\alpha, 6\alpha - 3, 3\alpha - 1\} \] 
with \( \alpha < 3/4 \) for \( n = CCE \), and \( \kappa > \max\{4\alpha - 2, \alpha + 1/4, 1/2 - 2\alpha\} \) with \( \alpha < 7/4 \) for \( n = PC \). Then, as \( N, T \to \infty \),

\[
\sqrt{N}(\hat{\beta}_n^p - \beta) \overset{d}{\to} N\left(0, \lim_{N \to \infty} \Sigma^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \Sigma_{\epsilon,i} \Sigma \right) \Sigma^{-1}\right).
\]

Some remarks are in order.

- According to Theorems 1 and 2, the heterogeneity of \( \beta_1, \ldots, \beta_N \) leads to a slower rate of convergence relative to the case when \( \beta_1 = \ldots = \beta_N \), from \( \sqrt{NT} \) to \( \sqrt{N} \). This is consistent with the results of Pesaran, 2006, Theorems 2 and 3.

- Unlike what happens under Assumption HOM, under HET \( \hat{\beta}_n^p \) is asymptotically unbiased even for \( \kappa > 1 \). Note in particular how in the strong factor case asymptotic unbiasedness holds for all (permissable) \( \kappa > 0 \). This finding is consistent with those of Pesaran, 2006, Theorems 2 and 3 and Chudik et al. (2011, Section 4), where in the latter study all non-strong factors are omitted from the estimation process.

- Interestingly, the set of admissible combinations of \( \alpha \) and \( \kappa \) is actually larger under Assumption HET than under Assumption HOM. That is, relaxing the homogeneity assumption also means relaxing the restrictions placed \( \alpha \) and \( \kappa \). For example, in terms of \( \alpha \), while under HOM the relevant requirement for asymptotic normality is that \( \alpha < 1/2 \), under HET it is \( \alpha < 3/4 \) for \( \hat{\beta}_n^p \) and \( \alpha < 7/4 \) for \( \hat{\beta}_n^p \).

That is, while under HOM the factors have to be at most semi-strong, under HET we can also allow for semi-weak factors. The “price” of the increased generality under HET is of course the reduced rate of consistency mentioned above.

Corollary 2 provides the conditions on \( \alpha \) and \( \kappa \) required to ensure consistency under Assumption HET.

**Corollary 2.** Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold. Suppose also that 
\[ \kappa > \max\{6\alpha - 4, 3\alpha - 3/2\} \] 
with \( \alpha < 1 \) for \( n = CCE \), and \( \kappa > \max\{2\alpha, 4\alpha - 1\} \) for \( n = PC \). Then, as \( N, T \to \infty \),

\[
||\hat{\beta}_n^p - \beta|| = o_p(1).
\]

The conditions placed on \( \kappa \) and \( \alpha \) are therefore the same as in Corollary 1 (under Assumption HOM).

---

**Table 2.** Monte Carlo evaluation of the consistency of the pooled CCE estimator.

<table>
<thead>
<tr>
<th>Mean</th>
<th>STD</th>
<th>Relative STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/α</td>
<td>0.5</td>
<td>0.95</td>
</tr>
<tr>
<td>50</td>
<td>1.0005</td>
<td>1.0004</td>
</tr>
<tr>
<td>100</td>
<td>1.0001</td>
<td>1.0003</td>
</tr>
<tr>
<td>150</td>
<td>1.003</td>
<td>1.0006</td>
</tr>
<tr>
<td>200</td>
<td>1.0002</td>
<td>1.0004</td>
</tr>
<tr>
<td>250</td>
<td>1.0000</td>
<td>0.9993</td>
</tr>
<tr>
<td>300</td>
<td>0.9999</td>
<td>0.9994</td>
</tr>
<tr>
<td>350</td>
<td>0.9999</td>
<td>0.9988</td>
</tr>
<tr>
<td>400</td>
<td>0.9998</td>
<td>0.9993</td>
</tr>
<tr>
<td>450</td>
<td>0.9998</td>
<td>0.9986</td>
</tr>
<tr>
<td>500</td>
<td>0.9999</td>
<td>0.9987</td>
</tr>
<tr>
<td>550</td>
<td>1.0001</td>
<td>0.9986</td>
</tr>
<tr>
<td>600</td>
<td>1.0000</td>
<td>0.9987</td>
</tr>
</tbody>
</table>

*Mean* and “STD” refer to the simulated mean and standard deviation of \( \hat{\beta}_n^p - \beta \), respectively. “Relative STD” refers to the ratio of the standard deviation corresponding to the given value of \( N \) and that of a sample with \( N = 50 \).
4. Conclusion

The present study considers a factor-augmented regression model in which the factor loadings go to zero at the rate $N^{-\alpha}$, where $\alpha \in [0, 1]$ and $N$ is related to $T = N^\kappa$. The purpose is to study the effect of $\alpha$ on two of the most popular estimators for factor-augmented regressions, namely, PC and CCE. A standard assumption in the literature is that the slopes are homogenous. Our findings in this case can be summarized as follows. First, in order to ensure $\sqrt{N T}$-consistency and limiting normal distributions, both estimators require strong or semi-strong factors ($\alpha < 1/2$). The sets of allowable values of $\kappa$ generally differ between the estimators. However, both sets shrink toward the value one as the factors become weaker ($\alpha \rightarrow 0$). Second, unless $\kappa < 1$, both estimators are asymptotically biased. Third, the estimators can be consistent when the factors are semi-weak ($1/2 \leq \alpha < 1$), but only PC allows consistent estimation in the weak factor case ($\alpha = 1$). Additional requirements on $\kappa$ have to be fulfilled for consistency to be possible. CCE is very general in this regard, having binding restrictions on $\kappa$ only for semi-weak factors. By contrast, PC restricts $\kappa$ from below already for semi-strong factors and requires that $\kappa > 3$ when the factors are weak, which is clearly very restrictive.

While the PC strand of the literature on factor-augmented regressions assumes that the slopes are homogenous, in the CCE strand there has been some attempts to relax this assumption. Motivated by this development, we also consider the properties of the CCE and PC estimators when the heterogeneity of the slopes can be given a random coefficient representation. From a qualitative point of view, our results for the heterogeneous slope case are very similar to those obtained under homogenous slopes. The main difference is threefold. First, the estimators are asymptotically unbiased for all (permissible) values of $\kappa$, including $\kappa \geq 1$. Second, quite unexpectedly, the required restrictions on $\kappa$ and $\alpha$ to ensure asymptotic normality (and also consistency in case of the PC estimator) are less restrictive when the slopes are allowed to be heterogeneous than when they are restricted to be homogenous. Third, the increased generality in terms of $\kappa$ and $\alpha$ under heterogenous slopes has a “price” in terms of a relatively slow rate of consistency.

Appendix A: Notation

The model for $Z_{i,t} = (y_{i,t}, X'_{i,t})'$ can be written in matrix notation as

$$Z_i = FC_i + U_i,$$

where $Z_i = (y_i, X_i) = (z_{i,1}, \ldots, z_{i,T})'$ is $T \times (m + 1)$, $F = (f_1, \ldots, f_T)'$ is $T \times r$, $C_i = (\Lambda_i', \beta_i, \lambda_i, \Lambda_i')$ is $r \times (m + 1)$, and $U_i = (u_{i,1}, \ldots, u_{i,T})' = (E_i' \beta + v_i, E_i)$ is $T \times (m + 1)$. Alternatively, the model for $z_{i,t}$ can be written as the following $N$-dimensional system:

$$z_t = C_i f_t + u_t,$$

where $z_t = (z'_{1,t}, \ldots, z'_{N,t})'$ and $u_t = (u'_{1,t}, \ldots, u'_{N,t})'$ are $N(m + 1) \times 1$, and $C = (C_1, \ldots, C_N)'$ is $N(m + 1) \times r$. The matrix notation

$$Z = FC' + U$$

will also be used, where $Z = (Z_1, \ldots, Z_N)$ and $U = (U_1, \ldots, U_N)$ are $T \times N(m + 1)$. In what follows, the representations in (A1)–(A3) will be used interchangeably.

Many of the results can be expressed in terms of $D_{CCE} = \hat{F}CCE - FC$ and $D_{PC} = \hat{F}PC - FH$, where

$$\overline{C} = N^{-1} \sum_{i=1}^{N} C_i = N^{-\alpha} \overline{C}_0, \overline{H} = \overline{Q}(T^{-1} \hat{F} \hat{F}^T)^{-1} \overline{Q} = N^{-2\alpha} \overline{H}_0, \overline{Q} = N^{-1} \sum_{i=1}^{N} C_i C_i' = N^{-1} \overline{C} \overline{C}'$$

and $\overline{H} = \overline{Q}(T^{-1} \hat{F} \hat{F}^T)^{-1} \overline{Q}$. It will therefore be convenient to introduce some special notation to simplify such expressions. Consider the PC estimator. As in (Bai, 2003, p. 158), if we denote by $V_T$ the $k \times k$ diagonal matrix consisting of the first $k$ eigenvalues of $(NT)^{-1}Z Z'$ in descending order, then, by
the definition of eigenvalues and eigenvectors, \( \hat{\mathbf{F}}_{\mathbf{PC}} = (NT)^{-1} \mathbf{Z} \mathbf{Z}' \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1} \). In this notation,

\[
\begin{align*}
\mathbf{D}_{\mathbf{PC}} &= \hat{\mathbf{F}}_{\mathbf{PC}} - \mathbf{F} \mathbf{H} = (NT)^{-1} \mathbf{Z} \mathbf{Z}' \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1} - (NT)^{-1} \mathbf{F} \mathbf{C}' \mathbf{F}' \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1} \\
&= (NT)^{-1} (\mathbf{Z} \mathbf{Z}' - \mathbf{F} \mathbf{C}' \mathbf{F}') \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1} = (NT)^{-1} (\mathbf{U} \mathbf{U}' + \mathbf{U} \mathbf{C}' \mathbf{F}' + \mathbf{F}' \mathbf{U}) \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1} \\
&= (NT)^{-1} \mathbf{G} \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{V}_T^{-1},
\end{align*}
\]

where \( \mathbf{G} = (\mathbf{U} \mathbf{U}' + \mathbf{U} \mathbf{C}' \mathbf{F}' + \mathbf{F}' \mathbf{U}) = (\mathbf{g}_1, \ldots, \mathbf{g}_T)' \) is \( T \times T \) and \( \mathbf{g}_t = (\mathbf{U} \mathbf{u}_t + \mathbf{F}' \mathbf{u}_t + \mathbf{U} \mathbf{c}_t) \) is \( T \times 1 \). Note that the dimension of \( \mathbf{D}_{\mathbf{PC}} \) is \( T \times (m + 1) \). It is further convenient to write \( \mathbf{D}_{\mathbf{PC}} = (\mathbf{d}_1^\prime, \ldots, \mathbf{d}_T^\prime)' \), where

\[
\begin{align*}
\mathbf{d}_t^\prime &= \hat{\mathbf{f}}_{\mathbf{PC}}^\prime - \mathbf{H} \mathbf{f}_t = (NT)^{-1} \mathbf{V}_T^{-1} (\hat{\mathbf{F}}_{\mathbf{PC}})' (\mathbf{U} \mathbf{u}_t + \mathbf{F}' \mathbf{u}_t + \mathbf{U} \mathbf{c}_t) \\
&= (NT)^{-1} \mathbf{V}_T^{-1} (\hat{\mathbf{F}}_{\mathbf{PC}})' \mathbf{g}_t
\end{align*}
\]

is \( (m + 1) \times 1 \). The corresponding quantities in case of CCE are given by

\[
\begin{align*}
\mathbf{D}_{\mathbf{CCE}} &= \hat{\mathbf{f}}_{\mathbf{CCE}}^\prime - \mathbf{F} \mathbf{C} = \mathbf{U}, \\
\mathbf{d}_t^\prime &= \hat{\mathbf{f}}_{\mathbf{CCE}}^\prime - \mathbf{C} \mathbf{f}_t = \mathbf{u}_t.
\end{align*}
\]

For notational simplicity, \( R \) will henceforth be used to indicate the order of lengthy remainder terms. For example, if \( A = O_p(N^n) + O_p(T^b) \), then we write \( A = O_p(R) \), where \( R = N^n + T^b \) (with the dependence on \( N \) and \( T \) in \( R \) omitted).

\section*{Appendix B: Auxiliary lemmas}

In Lemma PC1 we show how Assumption RK–PC can be used to show that \( ||\mathbf{H}_0^0 - \mathbf{H}^*|| = o_p(1) \), where \( \mathbf{H}_0^0 \) is the relevant rotation matrix in case of PC. The corresponding rotation matrix in case of CCE is given by \( \mathbf{C}_0^0 \), for which we already know from Assumption RK–CCE that \( ||\mathbf{C}_0^0 - \mathbf{C}^*|| = o_p(1) \). There is therefore no “Lemma CCE1.” All other lemmas come in pairs; for every CCE lemma, there is a corresponding PC lemma.

**Lemma PC1.** Under Assumptions ERR, LAM, RK–PC, and KAP, with \( \kappa > \max\{2\alpha, 4\alpha - 1\} \),

\[
||\mathbf{H}_0^0 - \mathbf{H}^*|| = o_p(1),
\]

where \( \mathbf{H}^* = (\mathbf{Q}^*)^{1/2} \mathbf{S} \mathbf{V}^{-1/2}, \mathbf{V} = \text{diag}(v_1, \ldots, v_r), v_1 > \cdots > v_r > 0 \) are the eigenvalues of \( \mathbf{S}_0^0 = (\mathbf{Q}^*)^{1/2} \Sigma_0 (\mathbf{Q}^*)^{1/2} \), and \( \mathbf{S} \) is the associated matrix of eigenvectors.

**Proof of Lemma PC1.** The proof of Lemma PC1 is similar to that of Theorem 2 in Stock and Watson (1999). We begin by evaluating \( T^{-1} \hat{\mathbf{F}}_{\mathbf{PC}} \mathbf{F} \). As will soon become clear, this can be done indirectly by evaluating the limit of the PC objective function, which we write as \( tr[(\mathbf{Z} - \mathbf{F}^* \mathbf{C}')(\mathbf{Z} - \mathbf{F}^* \mathbf{C}')]' \), where \( \mathbf{F}^* \) is the chosen value for \( \mathbf{F} \). Using the normalization \( T^{-1} \mathbf{F}^* \mathbf{F}^* = \mathbf{I}_r \), we can concentrate the objective function with respect to \( \mathbf{C} = \mathbf{Z}' \mathbf{F}^* (\mathbf{F}^* \mathbf{F}^*)^{-1} = T^{-1} \mathbf{Z}' \mathbf{F}^* \), leading to the concentrated objective function

\[
tr(\mathbf{Z}' \mathbf{M}_T \mathbf{Z}) = tr(\mathbf{Z}' \mathbf{Z}) - tr(\mathbf{Z}' \mathbf{P}_T \mathbf{Z}),
\]

where \( \mathbf{P}_A = \mathbf{A}(\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' \) for any matrix \( \mathbf{A} \). Hence, minimizing \( (NT)^{-1} tr[(\mathbf{Z} - \mathbf{F}^* \mathbf{C}')(\mathbf{Z} - \mathbf{F}^* \mathbf{C}')]' \) is equivalent to maximizing \( Q(\mathbf{F}^*) = (NT)^{-1} tr(\mathbf{Z}' \mathbf{P}_T \mathbf{Z}) \). Substitution of \( \mathbf{Z} = \mathbf{F}' + \mathbf{U} \) into \( Q(\mathbf{F}^*) \) gives

\[
\begin{align*}
Q(\mathbf{F}^*) &= (NT)^{-1} tr(\mathbf{Z}' \mathbf{P}_T \mathbf{Z}) \\
&= (NT)^{-1} tr(\mathbf{F}' + \mathbf{U}) \mathbf{P}_T \mathbf{F}' (\mathbf{F}' + \mathbf{U}) \\
&= (NT)^{-1} [tr(\mathbf{F}' \mathbf{P}_T \mathbf{F}') + tr(\mathbf{F}' \mathbf{P}_T \mathbf{U}) + tr(\mathbf{U}' \mathbf{P}_T \mathbf{F}') + tr(\mathbf{U}' \mathbf{P}_T \mathbf{U})],
\end{align*}
\]
where

\[(NT)^{-1} \text{tr}(U'PF \cdot FC') = \frac{1}{NT} \sum_{i=1}^{N} \text{tr}(U'_i [PF \cdot FC_i]) = \frac{1}{\sqrt{NT}} \text{tr} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} C_i U'_i F^* (T^{-1} F^* F^*)^{-1} T^{-1} F^* F \right) = O_p((NT)^{-1/2}),\]

with \((NT)^{-1} \text{tr}(CF'P_F \cdot U)\) having the same order. Similarly,

\[(NT)^{-1} \text{tr}(U'PF \cdot U) = \frac{1}{NT} \sum_{i=1}^{N} \text{tr}(U'_i [PF \cdot U_i]) = \frac{1}{NT} \sum_{i=1}^{N} \text{tr} \left( T^{-1/2} U'_i [F^* (T^{-1} F^* F^*)^{-1} T^{-1/2} F^* F] U_i \right) = O_p(T^{-1}).\]

By using this and the fact that \(N^{-1} \text{tr}(CC) = \bar{Q} = N^{-2\alpha} \bar{Q}^0\), where \(||\bar{Q}^0 - Q^\alpha|| = o_p(1)\) (Assumption RK–PC), we obtain

\[Q(F^*) = (NT)^{-1} \text{tr}(CF'P_F \cdot FC') + O_p((NT)^{-1/2}) + O_p(T^{-1}) = T^{-1} \text{tr}(F'P_F \cdot F\bar{Q}) + O_p((NT)^{-1/2}) + O_p(T^{-1}) = N^{-2\alpha} T^{-1} \text{tr}(F'P_F \cdot Q^\alpha) + O_p((NT)^{-1/2}) + O_p(T^{-1}).\]

Moreover, because \(Q^*\) is positive definite, we may write \(Q^* = (Q^*)^{1/2}(Q^*)^{1/2}\), suggesting that

\[Q(F^*) = N^{-2\alpha} Q_F(F^*) + O_p((NT)^{-1/2}) + O_p(T^{-1}), \tag{A8}\]

where \(Q_F(F^*) = T^{-1} \text{tr}(F^0 P_F \cdot F^0)\) and \(F^0 = F(Q^*)^{1/2}\). Consider \(Q_F(F^*)\). Denote by \(S_T\) the eigenvectors of the \(r \times r\) matrix \(T^{-1} F^0 F^0\) and let \(V_T\) be the associated diagonal matrix of eigenvalues. It follows that \(T^{-1} F^0 F^0 S_T = S_T V_T\), and therefore, \(T^{-1} F^0 F^0 (F^0 S_T) = (F^0 S_T) V_T\), where now \(F^0 S_T\) contains the eigenvectors of the \(T \times T\) matrix \(T^{-1} F^0 F^0\). By using this and \(F^0 F^* = I_T\), we obtain

\[Q_F(F^*) = T^{-1} \text{tr}(F^0 P_F \cdot F^0) = T^{-2} \text{tr}(F^0 F^* F^0 F^0) = T^{-1} \text{tr}(F^0 F^* (T^{-1} F^0 F^0) F^0),\]

which means that \(Q_F(F^*)\) is maximized for \(F^* = \sqrt{T} F^0 S_T\). While infeasible, \(F^*\) lends itself to simple asymptotics. Specifically, letting \(S = \lim_{T \to \infty} S_T\) and \(\Sigma^0 = \lim_{T \to \infty} T^{-1} F^0 F^0 = \lim_{T \to \infty} (Q^*)^{1/2} T^{-1} F^0 F^0 (Q^0)^{1/2} = (Q^*)^{1/2} \Sigma_F (Q^*)^{1/2}\), we can show that

\[Q_F(F^*) = T^{-1} \text{tr}(F^0 P_F \cdot F^0) = T^{-1} \text{tr}(F^0 F^* (F^* F^* - F^* F^* F^0)) = \text{tr}[T^{-1} F^0 F^0 S_T (S_T' T^{-1} F^0 F^0 S_T) S_T' T^{-1} F^0 F^0] = \text{tr}[\Sigma^0 S (S' \Sigma^0 S)^{-1} \Sigma^0] + o_p(1). \tag{A9}\]

In order to appreciate the relevance of (A9) for the evaluation of \(T^{-1} \hat{F}^{PC}F\), note that using \(T^{-1} F^{PC} \hat{F}^{PC} = I_r\), we obtain

\[Q_F(F^{PC}) = T^{-1} F^0 P_F F^0 = T^{-2} F^0 F^{PC} F^{PC} F^0 = (Q^*)^{1/2} (T^{-1} F^0 F^{PC}) (T^{-1} F^{PC} F) (Q^*)^{1/2}.\]

Hence, if we can show that \(Q_F(F^{PC}) - Q_F(F^*) = o_p(1)\), then the limit of \(T^{-1} \hat{F}^{PC}F\) can be worked out from that of \(Q_F(F^*)\) above. We start with the following expansion:

\[N^{-2\alpha} [Q_F(F^*) - Q_F(F^{PC})] = N^{-2\alpha} [Q_F(F^*) - Q_F(F)] - N^{-2\alpha} [Q_F(F^{PC}) - Q_F(F)], \tag{A10}\]
where
\[ N^{-2a}[Q_F(\hat{F}^P) - Q_F(F)] = [N^{-2a}Q_F(\hat{F}^P) - Q(\hat{F}^P)] + [Q(\hat{F}^P) - N^{-2a}Q_F(\hat{F}^P)] + N^{-2a}[Q_F(\hat{F}^*P) - Q_F(F)] \]

By (A8), the first term on the right-hand side is \( O_p((NT)^{-1/2}) + O_p(T^{-1}) \). Also, since \( \hat{F}^P \) and \( \hat{F}^* \) are the maximizers of \( Q(F^*) \) and \( Q_F(F^*) \), respectively, and \( |Q(F^*) - N^{-2a}Q_F(F^*)| = o_p(1) \), the second term is of the same order. It remains to consider the third term. Note that \( \hat{F}^* = \sqrt{T}F^0S_T = \sqrt{T}F(Q^*)^{1/2}S_T \).

The eigenvalue interpretation of \( E^0S_T = F(Q^*)^{1/2}S_T \) means that \( S_T(Q^*)^{1/2}(T^{-1}F)F(Q^*)^{1/2}S_T = I_r \) and \( T^{-1}F = [(Q^*)^{1/2}S_TS_T(Q^*)^{1/2}]^{-1} \). Making use of these relationships, we obtain

\[
P_{F^*} = \hat{F}^*(\hat{F}^*S^*)^{-1}\hat{F}^* = T^{-1}F(Q^*)^{1/2}S_T[S_T(Q^*)^{1/2}(T^{-1}F)F(Q^*)^{1/2}S_T]^{-1}S_T(Q^*)^{1/2}F' = T^{-1}F(Q^*)^{1/2}S_TS_T(Q^*)^{1/2}F = T^{-1}F(Q^*)^{1/2}S_TS_T(Q^*)^{1/2}F = T^{-1}F(Q^*)^{1/2}F' = T^{-1}F(F)F^{-1}F' = P_F.
\]

Therefore,
\[
Q_F(\hat{F}^*) - Q_F(F) = T^{-1}tr[F^0(P_{F^*} - P_F)F^0] = 0,
\]
which in turn implies
\[
N^{-2a}[Q_F(\hat{F}^P) - Q_F(F)] = O_p((NT)^{-1/2}) + O_p(T^{-1}).
\]

and
\[
N^{-2a}[Q_F(\hat{F}^*) - Q_F(\hat{F}^P)] = O_p((NT)^{-1/2}) + O_p(T^{-1}),
\]

or, by imposing \( T = N^\kappa \),
\[
Q_F(\hat{F}^*) - Q_F(\hat{F}^P) = O_p(N^{2\alpha - 1/2}T^{-1/2}) + O_p(N^{2\alpha}T^{-1}) = O_p(N^{2\alpha - (1/2)}T^{1/2}) + O_p(N^{2\alpha - \kappa}).
\]

Provided that \( \kappa > \max[2\alpha, 4\alpha - 1] \), this is \( o_p(1) \). By using this and (A9),
\[
T^{-1}F^0(P_{F^P} - P_F)F^0 = (Q^*)^{1/2}(T^{-1}F(\hat{F}^P)F)(Q^*)^{1/2} = \Sigma^0S'S^0S^{-1}S'\Sigma^0 + o_p(1),
\]

and therefore
\[
T^{-1}\hat{F}^PC'F = (S'\Sigma^0S)^{-1/2}S'\Sigma^0(Q^*)^{-1/2} + o_p(1).
\]

Note that \( S \) is the eigenvector matrix of \( \Sigma^0 \) with \( V = \lim_{T \to \infty} V_T \) being the associated diagonal matrix of eigenvalues. It follows that \( \Sigma^0S = SV \), and since \( S = S' = S^{-1} \), we also have that \( S'\Sigma^0S = V \). Hence, since \( V \) is symmetric,
\[
(S'\Sigma^0S)^{-1/2}S'\Sigma^0 = V^{-1/2}VS' = V^{1/2}S',
\]
giving
\[
T^{-1}\hat{F}^PC'F = (S'\Sigma^0S)^{-1/2}S'\Sigma^0(Q^*)^{-1/2} + o_p(1) = V^{1/2}S'(Q^*)^{-1/2} + o_p(1),
\]

and so we obtain
\[
N^{2\alpha}H = N^{2\alpha}Q(T^{-1}F(\hat{F}^P)F)V_T^{-1} = Q^*(V^{1/2}S'(Q^*)^{-1/2})V^{-1} + o_p(1)
\]

\[
H^* = (Q^*)^{1/2}SV^{-1/2}.
\]

**Lemma CCE2.** Under Assumptions ERR and LAM,
\[
\frac{1}{T} \sum_{t=1}^{T} \|d_t^{CEF}\|^2 = O_p(N^{-1}).
\]
Proof of Lemma CCE2. The proof of Lemma CCE2 is a simple consequence of the fact that $||\sqrt{N}\tilde{u}_t|| = O_p(1)$, as seen by writing

$$\frac{N}{T} \sum_{t=1}^{T} ||d_{CCE}^t||^2 \leq \frac{1}{T} \sum_{t=1}^{T} ||\sqrt{N}\tilde{u}_t||^2 = O_p(1).$$

**Lemma PC2.** Under Assumptions ERR and LAM,

$$\frac{1}{T} \sum_{t=1}^{T} ||d_{PC}^t||^2 = O_p(T^{-1}) + O_p(N^{-1}).$$

Proof of Lemma PC2. By the definition of $d_{PC}^t$,

$$\frac{1}{T} \sum_{t=1}^{T} ||d_{PC}^t||^2 \leq ||V_{T^{-1}}||^2 \frac{3}{N^2T^3} \sum_{t=1}^{T} (||\hat{F}_{PC}'\tilde{u}_t||^2 + ||\hat{F}_{PC}'\hat{F}_{C}'u_t||^2 + ||\hat{F}_{PC}'\hat{U}\hat{C}_t||^2).$$

Here

$$\frac{1}{N^2T^3} \sum_{t=1}^{T} (||\hat{F}_{PC}'\tilde{u}_t||^2) = \frac{1}{N^2T^3} \sum_{t=1}^{T} \left(\frac{1}{T} \sum_{s=1}^{T} ||\hat{f}_{PC}^s||^2\right) \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (N^{-1/2}u_t' u_t)^2\right) = O_p(T^{-1}) + O_p(N^{-1}),$$

where the last equality holds because of the normalization $T^{-1}\hat{F}_{PC}'\hat{F}_{PC} = I$, suggesting that $T^{-1}||\hat{F}_{PC}||^2 = T^{-1} \sum_{t=1}^{T} ||\hat{f}_{PC}^t||^2 = r$. Also, since $N^{-1/2}u_t' u_t = O_p(1)$ and $N^{-1/2}u_t' u_t = O_p(1)$ whenever $t \neq s$,

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (N^{-1/2}u_t' u_t)^2 = \frac{1}{T} \sum_{t=1}^{T} (N^{-1/2}u_t' u_t)^2 + \frac{1}{N} \sum_{t=1}^{T} \sum_{s=1}^{T} (N^{-1/2}u_t' u_t)^2 = O_p(1) + O_p(TN^{-1}).$$

Further use of $||N^{-1/2}C' u_t|| = N^{-\alpha}||N^{-1/2}C^0 u_t|| = O_p(N^{-\alpha})$ gives

$$\frac{1}{N^2T^3} \sum_{t=1}^{T} ||(\hat{F}_{PC}'C' u_t)||^2$$

$$\leq \frac{1}{N^2T^3} \sum_{t=1}^{T} \left(\frac{1}{T} \sum_{s=1}^{T} ||f_{PC}^s||^2\right) \left(\frac{1}{T} \sum_{t=1}^{T} ||f_t||^2\right) \left(\frac{1}{T} \sum_{t=1}^{T} ||N^{-1/2}C^0 u_t||^2\right)$$

$$= O_p(N^{-\alpha+1}),$$

with $N^{-2}T^{-3} \sum_{t=1}^{T} ||(\hat{F}_{PC}'\hat{U}\hat{C}_t)||^2$ being of the same order. Hence, since $||V_{T^{-1}}|| = O_p(1)$,

$$\frac{1}{T} \sum_{t=1}^{T} ||d_{PC}^t||^2 \leq ||V_{T^{-1}}||^2 \frac{3}{N^2T^3} \sum_{t=1}^{T} (||\hat{F}_{PC}'\tilde{u}_t||^2 + ||(\hat{F}_{PC}'C' u_t||^2 + ||(\hat{F}_{PC}'\hat{U}\hat{C}_t)||^2)$$

$$= O_p(T^{-1}) + O_p(N^{-1}) + O_p(N^{-(\alpha+1)}) = O_p(T^{-1}) + O_p(N^{-1}),$$

and so the proof is complete.
Lemma CCE3. Under Assumptions ERR and LAM,
\[ ||\sqrt{N}T^{-1/2}F'D_{CCE}|| = O_p(1). \]

Proof of Lemma CCE3. The proof is completed by noting that
\[ \sqrt{N}T^{-1/2}F'D_{CCE} = \sqrt{N} \sum_{t=1}^{T} f_t d_t^{\text{CCE}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \sqrt{N}u_e = O_p(1). \] (A16)

Lemma PC3. Under Assumptions ERR and LAM,
\[ ||\sqrt{N}T^{-1/2}F'D_{PC}|| = O_p(R_1), \]
where
\[ R_1 = N^{-3\alpha} + N^{-1/2} + (1 + N^{1/2-2\alpha})T^{-1/2} + \sqrt{N}T^{-1}. \]

Proof of Lemma PC3. Write
\[ \sqrt{N}T^{-1/2}F'D_{PC} = \sqrt{N} \sum_{t=1}^{T} f_t d_t^{\text{PC}} = \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} \hat{f}_t g_t^{\text{PC}} V_T^{-1} \]
\[ = \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t (u'_t U' + u'_t CF' + f'_t C'U') \hat{F}_{PC} V_T^{-1}. \] (A17)

The first term on the right involves
\[ \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t u'_t U' \hat{F}_{PC} = \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t u'_t U' D_{PC} + \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t u'_t U' FH, \]
where
\[ \left| \left| \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t u'_t U' D_{PC} \right| \right| \leq \left( \frac{1}{T} \sum_{t=1}^{T} \left| \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u'_t u_s \right| \right| \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} ||d_{PC}||^2 \right)^{1/2} \]
\[ = \left[ O_p(\sqrt{N}T^{-1/2}) + O_p(1) \right][O_p(N^{-1/2}) + O_p(T^{-1/2})] \]
\[ = O_p(\sqrt{NT^{-1}}) + O_p(N^{-1/2}) + O_p(T^{-1/2}), \]
as \[ T^{-1} \sum_{s=1}^{T} ||d_{PC}||^2 = O_p(N^{-1}) + O_p(T^{-1}) \] by Lemma PC2 and
\[ \frac{1}{\sqrt{N^T}} \sum_{t=1}^{T} f_t u'_t u_s = \sqrt{NT^{-1/2}}N^{-1}f_t u'_u u_s + \frac{1}{\sqrt{N^T}} \sum_{t \neq s}^{T} f_t u'_u u_s = O_p(\sqrt{NT^{-1/2}}) + O_p(1). \]

Also,
\[ \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} f_t u'_t U' F = \frac{1}{\sqrt{NT^{3/2}}} \sum_{t=1}^{T} \sum_{s=1}^{T} f_t u'_t u_s f'_s. \]
\[
\begin{align*}
\sqrt{N} \frac{1}{\sqrt{T} NT} \sum_{t=1}^{T} f_t u'_t f_t' + \frac{1}{\sqrt{T} \sqrt{NT}} \sum_{s \neq t}^{T} f_t u'_t f'_t
\end{align*}
\]

\[
= O_p(\sqrt{NT}^{-1/2}) + O_p(T^{-1/2}) = O_p(\sqrt{NT}^{-1/2}),
\]
suggesting that, with \(\|H\| = O_p(N^{-2\alpha})\),

\[
\left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t u'_t U' F_{PC} \right|
\]

\[
\leq \left( \frac{1}{T} \sum_{s=1}^{T} \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u'_t C \right|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \|C_{PC}\|^2 \right)^{1/2}
\]

\[
= O_p(N^{-\alpha}) \left[ O_p(N^{-\alpha}) + O_p(T^{-1/2}) \right]
\]

\[
= O_p(N^{-\alpha+1/2}) + O_p(N^{-\alpha} T^{-1/2}),
\]

and

\[
\left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t u'_t C F' F \right| \leq N^{-\alpha} \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u'_t C_0 \right| \|T^{-1} F' F\| = O_p(N^{-\alpha}),
\]
giving

\[
\left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t u'_t C F' PC \right|
\]

\[
\leq \left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t u'_t C F' D_{PC} \right| + \left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t u'_t C F' F \right| \|H\|
\]

\[
= O_p(N^{-\alpha+1/2}) + O_p(N^{-\alpha} T^{-1/2}) + O_p(N^{-3\alpha}),
\]  

(A19)

with \(\|N^{-1/2} T^{-3/2} \sum_{t=1}^{T} f_t f'_t C' U' F_{PC}\|\) being of the same order. Hence, letting

\[
R_1 = N^{-3\alpha} + N^{-1/2} + (1 + N^{1/2-2\alpha}) T^{-1/2} + \sqrt{NT}^{-1},
\]

with \(\|V_T^{-1}\| = O_p(1)\),

\[
\|\sqrt{NT}^{-1/2} F' D_{PC}\|
\]

\[
\leq \left| \frac{1}{\sqrt{NT}^{3/2}} \sum_{t=1}^{T} f_t (u'_t U' + u'_t C F' + f'_t C' U') F_{PC} \right| \|V_T^{-1}\| = O_p(R_1),
\]

(A20)
as required.
Lemma CCE4. Under Assumptions ERR and LAM,
\[ NT^{-1} D^{CCE} D^{CCE} = \Sigma_u + O_p(T^{-1/2}), \]
where
\[ \Sigma_u = \frac{1}{N} \sum_{i=1}^{N} \Sigma_{u,i}, \]
\[ \Sigma_{u,i} = E(u_{i,i} u_{i,i}') = \begin{bmatrix} \beta' \Sigma_{e,i} \beta + \sigma_{\epsilon,i}^2 & \beta' \Sigma_{e,i} \\ \Sigma_{e,i} & \Sigma_{e,i} \end{bmatrix}. \]

Proof of Lemma CCE4. A direct calculation reveals that
\[ NT^{-1} D^{CCE} D^{CCE} = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N} u_{i,t} u_{i,t}', \]
\[ = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} u_{i,t} u_{i,t}' + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{j \neq i}^{N} u_{i,t} u_{j,t}' \]
\[ = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} u_{i,t} u_{i,t}' + O_p(T^{-1/2}) \]
\[ = \Sigma_u + O_p(T^{-1/2}), \]
as was to be shown.

Lemma PC4. Under Assumptions ERR, LAM, and RK–PC,
\[ NT^{-1} D^{PC} D^{PC} = N^{-2\alpha} H^0 (Q^0)^{-1} S^0 (Q^0)^{-1} H^0 + O_p(R_4), \]
where
\[ S^0 = \frac{1}{N} \sum_{i=1}^{N} C_i^0 \Sigma_{u,i} C_i^0, \]
\[ R_4 = N^{-3\alpha-1} + N^{-4\alpha-1/2} + (N^{-2\alpha-1/2} + N^{-4\alpha} + N^{-1}) T^{-1/2} + (N^{-2\alpha} + N^{-1/2}) T^{-1} \]
\[ + (1 + N^{1/2-3\alpha}) T^{-3/2} + (N^{-4\alpha+1} + N^{-2\alpha+1/2}) T^{-2} + N^{-2\alpha} T^{-5/2}. \]

Proof of Lemma PC4. By definition,
\[ NT^{-1} D^{PC} D^{PC} = N^{-1} T^{-3} V^{-1} E^{PC} G' G E^{PC} V_T^{-1}, \]
\[ = N^{-1} T^{-3} V^{-1} [H^0 F' G^t G H + H^0 F' G D^{PC} + D^{PC} G' G D^{PC} + D^{PC} G' G D^{PC}] V_T^{-1}. \]
Consider the second term on the right-hand side. Clearly,
\[ ||N^{-1} T^{-3} F' G D^{PC}|| = \left( \frac{1}{NT^3} \sum_{s=1}^{T} ||F' G g_s D_s^{PC}|| \right)^{1/2} \]
\[ \leq \left( \frac{1}{T} \sum_{s=1}^{T} \left( \frac{1}{T} \sum_{s=1}^{T} ||d_s^{PC}||^2 \right)^{1/2}, \right) \]
where the order of the second term in the product is given by Lemma PC2. As for the first term,
\[ N^{-1}T^{-2}F'G_g, \]
\[ = N^{-1}T^{-2}F'(UU' + FC'U' + UCF')(UU_s + FC'u_s + UCF_s) \]
\[ = N^{-1}T^{-2}(F'UU'UU_s + F'FC'U'Uu_s + F'UCF'Uu_s + F'UU'FC'u_s + F'FC'U'FC'u_s \]
\[ + F'UCF'FC'u_s + F'UU'UCF_s + F'FC'UUCF_s + F'UCF'UCF_s) \]
\[ = j_1 + \cdots + j_9, \] \hspace{1cm} (A25)
where the dependence on \( s \) in \( j_1, \ldots, j_9 \) has been suppressed.

Consider \( j_1 \), which can be expanded as
\[ ||j_1|| = ||N^{-1}T^{-2}F'UU'Uu_s|| \]
\[ = \left| \left| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{k=1}^{T} f_k u_k u_s u_s' \right| \right| \]
\[ \leq \left| \left| \frac{1}{NT^2} \sum_{k=1}^{T} f_k u_k u_s u_s' \right| \right| + \left| \left| \frac{1}{NT} \sum_{t \neq s} \sum_{k=1}^{T} f_k u_k u_s u_s' \right| \right| \]
\[ \leq NT^{-2}||N^{-2}f_s u_s u_s' u_s|| + \frac{\sqrt{N}}{T^{3/2}} \left| \left| \frac{1}{NT^3/2} \sum_{k \neq s} \sum_{t} f_k u_k u_s u_s' \right| \right| \]
\[ + \frac{\sqrt{N}}{T^{3/2}} \left| \left| \frac{1}{NT^3/2} \sum_{t \neq s} \sum_{k \neq t} f_k u_k u_s u_s' \right| \right| \]
\[ = O_p(NT^{-2}) + O_p(\sqrt{NT^{-3/2}}) + O_p(T^{-1/2}), \] \hspace{1cm} (A26)
where last equality follows from direct evaluation of the individual terms. The order of the second term, for example, is obtained as
\[ \left| \left| \frac{1}{NT^2} \sum_{k \neq s}^{T} f_k u_k u_s u_s' u_s \right| \right| \leq \frac{\sqrt{N}}{T^{3/2}} \left| \left| \frac{1}{NT} \sum_{k \neq s}^{T} \sum_{i=1}^{N} f_k u_k u_s u_s' u_s \right| \right| \left| \left| \frac{1}{N} \sum_{j=1}^{N} u_j u_j' u_j u_j' \right| \right| \leq O_p(\sqrt{NT^{-3/2}}). \]

Consider \( j_2 \). Let \( \Sigma_u = \text{diag}(\Sigma_{u,1}, \ldots, \Sigma_{u,N}) \), such that \( E(u_i u_i') = \Sigma_u \), an \( N(m+1) \times N(m+1) \) matrix. Note that
\[ C^0(u_i u_i' - \Sigma_u)u_s = \sum_{i=1}^{N} \sum_{j=1}^{N} C^0_{i} [u_i u_i' - E(u_i u_i')] u_s j_s \]
\[ = \sum_{i=1}^{N} C^0_{i} (u_i u_i' - \Sigma_u) u_s j_s + \sum_{i=1}^{N} \sum_{j \neq i} C^0_{i} u_i u_i' j_s u_s j_s \]
where both terms on the right are mean zero and independent across \( s \neq t \), suggesting
\[ \frac{1}{NT} \sum_{t \neq s}^{T} C^0_{i} u_i u_i' u_s \]
\[ = \frac{1}{NT} \sum_{t \neq s}^{T} C^0_{i} (u_i u_i' - \Sigma_u) u_s + \frac{(T-1)}{NT} C^0_{i} \Sigma_u u_s \]

96
\[ ||j_2|| = ||N^{-1/2} T^{-2} F^\prime FC^\prime U^\prime u_s || \]
\[ \leq ||T^{-1} F^\prime F|| \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} C^\prime u_t u'_t \right| \]
\[ \leq N^{1/2-\alpha} T^{-1} ||T^{-1} F^\prime F|| ||N^{-1/2} C^0 u_s || ||N^{-1} u'_t u_s || \]
\[ + N^{-\alpha} ||T^{-1} F^\prime F|| \left| \frac{1}{\sqrt{NT}} \sum_{t \neq s}^{T} C^0 u_t u'_t u_s \right| \]
\[ = O_p(N^{1/2-\alpha} T^{-1}) + O_p(N^{-\alpha} T^{-1}) + O_p(N^{-\alpha+1/2}). \quad (A27) \]

For \( j_3 \), we use the fact that
\[ \frac{1}{\sqrt{NT}} \sum_{k=1}^{T} f_k u'_k u_s = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{k=1}^{T} f_k u'_{i,k} u_{i,s} \]
\[ = \frac{\sqrt{N}}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f_t u'_i u_{i,s} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{k \neq s}^{T} f_k u'_{i,k} u_{i,s} \]
\[ = O_p(\sqrt{NT}^{-1/2}) + O_p(1), \]
giving
\[ ||j_3|| = ||N^{-1/2} T^{-2} F U C^\prime u_s || \leq \frac{1}{N^\alpha T} \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u'_t C^0 \right| \left| \frac{1}{\sqrt{NT}} \sum_{k=1}^{T} f_k u'_k u_s \right| \]
\[ = N^{-\alpha} T^{-1} [O_p(\sqrt{NT}^{-1/2}) + O_p(1)] = O_p(N^{1/2-\alpha} T^{-3/2}) + O_p(N^{-\alpha} T^{-1}). \quad (A28) \]

Also, since
\[ N^{-1/2} T^{-2} F^\prime U^\prime F = \frac{1}{\sqrt{NT^2}} \sum_{t=1}^{T} \sum_{k=1}^{T} f_t u'_t f'_k \]
\[ = \frac{\sqrt{N}}{T} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f_t u'_i f'_i + \frac{1}{T} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{k \neq t}^{T} f_i u'_i f'_k \]
\[ = O_p(\sqrt{NT}^{-1}) + O_p(T^{-1}) = O_p(\sqrt{NT}^{-1}), \]
we can show that
\[ ||j_4|| = ||N^{-1/2} T^{-2} F^\prime U^\prime F C^\prime u_s || \leq N^{-\alpha} ||N^{-1/2} T^{-2} F^\prime U^\prime F|| ||N^{-1/2} C^0 u_s || \]
\[ = O_p(N^{1/2-\alpha} T^{-1}). \quad (A29) \]
As for \( j_5 \), by using arguments similar to those used in the above,

\[
||j_5|| = ||N^{-1}T^{-2}F'FC'U'FC'u_i|| \\
\leq N^{-2\alpha T^{-1/2}}||T^{-1}F'F||||NT||^{-1/2}C^0U'F|| ||N^{-1/2}C^0u_i|| \\
= O_p(N^{-2\alpha T^{-1/2}}),
\]

with \( ||j_6|| \) being of the same order.

For \( j_7 \),

\[
||j_7|| = ||N^{-1}T^{-2}F'U'U'Cf_i|| \\
= \left|\left| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{k=1}^{T} f_t u'_t u'_k C f_i \right|\right| \\
\leq \frac{1}{N^{\alpha - 1/2}T} \left|\left| \frac{1}{N^{3/2}T} \sum_{t=1}^{T} f_t u'_t u'_k C^0f_i \right|\right| + \frac{1}{N^{\alpha \sqrt{T}}} \left|\left| \frac{1}{NT^{3/2}} \sum_{t=1}^{T} \sum_{k \neq t} f_t u'_t u'_k C f_i \right|\right| \\
= O_p(N^{-\alpha + 1/2}T^{-1}) + O_p(N^{-\alpha}T^{-1/2}),
\]

where the last equality holds because

\[
\left|\left| \frac{1}{N^{3/2}T} \sum_{t=1}^{T} f_t u'_t u'_t C^0f_i \right|\right| \\
\leq \left( \frac{1}{T} \sum_{t=1}^{T} ||N^{-1}f_t u'_t||^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} ||N^{-1/2}u'_t C^0||^2 \right)^{1/2} \|f_i\| = O_p(1),
\]

and

\[
\left|\left| \frac{1}{NT^{3/2}} \sum_{t=1}^{T} \sum_{k \neq t} f_t u'_t u'_k C^0f_i \right|\right| \\
\leq \left( \frac{1}{T} \sum_{k \neq t} \left|\left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u'_t u'_k \right|\right|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{k \neq t} ||N^{-1/2}u'_k C^0||^2 \right)^{1/2} \|f_i\| = O_p(1).
\]

The order of \( j_8 \) can be obtained from

\[
\frac{1}{NT} \sum_{t=1}^{T} C^0' u_t u_t C^0 \\
= \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} C^0'(u_t u_t' - \Sigma_u) C^0 + N^{-1} C^0' \Sigma_u C^0 \\
= \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} C^0_i (u_{t,i} u_{t,i}' - \Sigma_{u,i}) C^0_i + \frac{1}{\sqrt{NT}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} C^0_i u_{i,j} u_{i,j}' C^0_j \\
+ \frac{1}{N} \sum_{i=1}^{N} C_i \Sigma_{u,i} C^0_i \\
= O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(1) = O_p(1),
\]

98
where
\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} C_j^0 u_{i,t} u_{j,t}' C_j^0 = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} C_i^0 u_{i,t} u_{i,t}' C_i^0 + \frac{1}{NT} \sum_{t=1}^{T} \sum_{j \neq i}^{N} C_j^0 u_{i,t} u_{j,t}' C_j^0
\]
= \Omega_p(1) + \Omega_p(T^{-1/2}),
\]
implying
\[
||j_8|| = ||N^{-1} T^{-2} F' FC' U' UCF_i|| \leq N^{-2\alpha} ||T^{-1} F F|| \left( \frac{1}{NT} \sum_{t=1}^{T} C_0^0 u_{i,t} u_{i,t}' \right) ||f_i|| = \Omega_p(N^{-2\alpha}). \tag{A32}
\]

It remains to consider \(j_9\), whose order is given by
\[
||j_9|| = ||N^{-1} T^{-2} F' UCF' UCF_i|| = \frac{1}{N^{2\alpha} T} \left( \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_t u_t' C_0 \right) ||f_t|| = \Omega_p(N^{-2\alpha} T^{-1}). \tag{A33}
\]

Hence, by putting everything together,
\[
||N^{-1} T^{-2} F' G' g_t|| \leq ||j_1|| + \cdots + ||j_9|| = \Omega_p(R_2), \tag{A34}
\]
where
\[
R_2 = N^{-(\alpha + 1/2)} + N^{-2\alpha} + T^{-1/2} + N^{1/2 - \alpha} T^{-1} + \sqrt{NT^{-3/2}} + NT^{-2},
\]
and therefore
\[
||N^{-1} T^{-3} F' G' GD_{PC}|| \leq \left( \frac{1}{T} \sum_{t=1}^{T} ||N^{-1} T^{-2} F' G' g_t||^2 \right)^{1/2} \left( \frac{1}{T} \sum_{t=1}^{T} ||d_{PC}^t||^2 \right)^{1/2} = \Omega_p(N^{-1/2} R_2) + \Omega_p(T^{-1/2} R_2). \tag{A35}
\]
The order of \(||N^{-1} T^{-3} D_{PC} G' G' GF||\) is the same. Thus, since \(||V^{-1}|| = \Omega_p(1)\), \(||H|| = \Omega_p(N^{-2\alpha})\), and the order of \(||N^{-1} T^{-3} D_{PC} G' GD_{PC}||\) is dominated by that of \(||N^{-1} T^{-3} F' G' GD_{PC}||\), we obtain
\[
NT^{-1} D_{PC} D_{PC} = N^{-1} T^{-3} V^{-1} T H' F' G' G F H + H' F' G' GD_{PC} + D_{PC} G' GF H + D_{PC} G' GD_{PC} V^{-1} \]
= \(N^{-1} T^{-3} V^{-1} T H' F' G' GF H V^{-1} + \Omega_p(N^{-4\alpha + 1/2} R_2) + \Omega_p(N^{-2\alpha} T^{-1/2} R_2). \tag{A36}
\]
Consider \(N^{-1} T^{-3} F' G' GF\), which we expand in the following obvious fashion:
\[
N^{-1} T^{-3} F' G' GF
\]
= \(1/NT^3 \sum_{t=1}^{T} \sum_{t=1}^{T} F' g_t g_t' F \]
= \(1/NT^3 \sum_{t=1}^{T} \sum_{t=1}^{T} F' (U u_t + F C' u_t + UCF_i)(u_t' U' + u_t' C' F + f_t' C') F \]
= \(1/NT^3 \sum_{t=1}^{T} \sum_{t=1}^{T} (F' U u_t u_t' U' F + F' FC' u_t u_t' U' F + F' UCF_i u_t' U' F + F' U u_t u_t' C' F + F' FC' u_t u_t' C' U' F) \]
+ \(F' FC' u_t u_t' C' F + F' UCF_i u_t' C' F + F' U u_t l'_t C' U' F + F' FC' u_t l'_t C' U' F + F' UCF_i l'_t C' U' F) \)
= \(J_1 + \cdots + J_9 \). \tag{A37}
For $J_1$,

$$J_1 = \frac{1}{NT^3} \sum_{t=1}^{T} F' u_t^i u_t^f' U F$$

$$= \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{k=1}^{T} f_t u_t^i u_t^j u_t^k f_t'$$

$$= \frac{1}{NT^3} \sum_{t=1}^{T} f_t u_t^i u_t^j u_t^f' + \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{s \neq t}^{T} f_t u_t^i u_t^j u_t^f' + \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{k \neq t}^{T} f_t u_t^i u_t^j u_t^k f_t'$$

$$+ \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \sum_{k \neq t}^{T} f_t u_t^i u_t^j u_t^k f_t'$$

where

$$\left| \frac{1}{NT^3} \sum_{t=1}^{T} f_t u_t^i u_t^j u_t^f' \right| \leq \frac{N}{T^2} \frac{1}{T} \sum_{t=1}^{T} \left| f_t u_t^i u_t^f' \right|^2 = O_p(NT^{-2})$$

$$\left| \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{s \neq t}^{T} f_t u_t^i u_t^j u_t^f' \right| = O_p(NT^{-3/2})$$

and

$$\left| \frac{1}{NT^3} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \sum_{k \neq t}^{T} f_t u_t^i u_t^j u_t^k f_t' \right| = O_p(T^{-1})$$

$$\sum_{t=1}^{T} \sum_{s \neq t}^{T} f_t u_t^i u_t^j u_t^f' / NT^3$$ is of the same order as $\sum_{t=1}^{T} \sum_{s \neq t}^{T} f_t u_t^i u_t^j u_t^f' / NT^3$. Hence,

$$||J_1|| = O_p(NT^{-2}) + O_p(T^{-1}) + O_p(N^{1/2}T^{-3/2})$$

(A38)

For $J_2$, we use

$$\frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} C^0 u_t^i u_t^j u_t^f' = \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} C^0 u_t^i u_t^j u_t^f' + \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s \neq t}^{T} C^0 u_t^i u_t^j u_t^f'$$

where

$$\frac{1}{NT^2} \sum_{t=1}^{T} C^0 u_t^i u_t^j u_t^f'$$

$$= \frac{1}{NT^2} \sum_{t=1}^{T} C^0 (u_t^i u_t^f' - \Sigma_i) u_t^f' + \frac{1}{NT^2} \sum_{t=1}^{T} C^0 u_t^f'$$

$$= \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{i=1}^{N} C^0 (u_t^i u_t^f' - \Sigma_i) u_t^f' + \frac{\sqrt{N}}{T^{3/2}} \frac{1}{N^{3/2}} \sum_{t=1}^{T} \sum_{i=1}^{N} C^0 u_t^i u_t^f'$$

$$+ \frac{1}{\sqrt{NT}^3/2} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} C^0 u_t^i u_t^f'$$

$$= O_p(N^{-1/2}T^{-3/2}) + O_p(\sqrt{NT}^{-3/2}) = O_p(\sqrt{NT}^{-3/2})$$

100
and, by exactly the same argument,

\[
\frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=t+1}^{T} C_{0t}^0 u_t u_s f_s' = \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{s=t+1}^{T} \sum_{i=1}^{N} C_{0i}^0 (u_{i,t} u_{i,s} - \Sigma_{u,i}) u_{i,s} f_s' + \frac{1}{T NT} \sum_{t=1}^{T} \sum_{s=t+1}^{T} \sum_{i=1}^{N} \sum_{j \neq t}^{N} C_{i}^0 u_{i,t} u_{j,t} u_{j,s} f_s'
\]

\[
+ \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{s=t+1}^{T} \sum_{i=1}^{N} C_{i}^0 \Sigma_{u,i} u_{i,s} f_s'
\]

\[= O_p(N^{-1/2}T^{-1}) + O_p(T^{-1}) + O_p(N^{-1/2}T^{-3/2}) = O_p(T^{-1}).\]

Therefore,

\[
\left\| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} C_{0t}^0 u_t u_s f_s' \right\| \leq O_p(T^{-1}) + O_p(\sqrt{NT^{-3/2}}),
\]

giving

\[
||J_2|| \leq N^{-\alpha} ||T^{-1}F'F|| \left\| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} C_{0t}^0 u_t u_s f_s' \right\| = O_p(N^{-\alpha} T^{-1}) + O_p(N^{1/2-\alpha} T^{-3/2}). \quad (A39)
\]

We know from before that \(||N^{-1/2}T^{-2}F'UUF|| = O_p(\sqrt{NT^{-1}})\) and \(||(NT)^{-1/2}F'UC_0|| = O_p(1).\) This implies

\[
||J_3|| \leq \frac{1}{N^\alpha \sqrt{T}} ||(NT)^{-1/2}F'UC_0|| \left\| \frac{1}{\sqrt{NT^2}} \sum_{t=1}^{T} \sum_{s=1}^{T} f_t u_t u_s f_s' \right\| = O_p(N^{1/2-\alpha} T^{-3/2}). \quad (A40)
\]

Also, since \(J_4 = J_2, ||J_4||\) is of the same order of magnitude as \(||J_2||.\)

Consider \(J_5,\) where

\[
\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} C_{t}^0 u_t u_i C_0^0
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{i}^0 u_t u_i C_0^0
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} C_{i}^0 u_t u_i C_0^0 + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \neq t}^{N} C_{i}^0 u_t u_j C_0^0
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} C_{i}^0 u_t u_i C_0^0 + O_p(T^{-1/2})
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} C_{i}^0 \Sigma_{u,i} C_0^0 + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} C_{i}^0 (u_{i,t} u_{i,t} - \Sigma_{u,i}) C_0^0 + O_p(T^{-1/2})
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} C_{i}^0 \Sigma_{u,i} C_0^0 + O_p((NT)^{-1/2}) + O_p(T^{-1/2})
\]

\[= \tilde{S}^0 + O_p(T^{-1/2}),
\]
suggesting that

\[
J_5 = \frac{1}{NT^3} \sum_{t=1}^{T} F'FC'u'u'C'F
\]

\[
= N^{-2\alpha} (T^{-1}F'F) \frac{1}{NT} \sum_{t=1}^{T} C^0u'u'C^0 (T^{-1}F'F)
\]

\[
= N^{-2\alpha} (T^{-1}F'F)S^0 (T^{-1}F'F) + O_p(N^{-2\alpha}T^{-1/2}).
\] (A41)

The order of \(||J_6||\) is

\[
||J_6|| = \left\| \frac{1}{NT^3} \sum_{t=1}^{T} F'UC_f'u'C'F \right\|
\]

\[
\leq \frac{1}{N^{2\alpha}T} \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} f_tu'C^0 \right\| \left\| T^{-1}F'F \right\| = O_p(N^{-2\alpha}T^{-1}),
\] (A42)

while that of \(||J_7||\) is \(O_p(N^{1/2-\alpha}T^{-3/2})\), as follows from the fact that \(J_7 = J_3\). Similarly, since \(J_8 = J_6\), we have \(||J_8|| = O_p(N^{-2\alpha}T^{-1})\).

Finally, consider \(J_9\), which can be evaluated in the following fashion:

\[
||J_9|| = \left\| \frac{1}{NT^3} \sum_{t=1}^{T} F'UC_f'f'C'U'F \right\|
\]

\[
\leq \frac{1}{N^{2\alpha}T} \frac{1}{T} \sum_{t=1}^{T} \left\| (NT)^{-1/2}F'UC^0 \right\| ^2 ||f_t||^2 = O_p(N^{-2\alpha}T^{-1}).
\] (A43)

Hence, by adding the results,

\[
N^{-1}T^{-3}F'G'GF = J_1 + \cdots + J_9
\]

\[
= N^{-2\alpha} (T^{-1}F'F)S^0 (T^{-1}F'F) + O_p(NT^{-2}) + O_p(T^{-1})
\]

\[
+ O_p(N^{1/2}T^{-3/2}) + O_p(N^{-2\alpha}T^{-1/2}),
\] (A44)

which in turn implies

\[
NT^{-1}D'PCD'PC
\]

\[
= N^{-1}T^{-3}V_T^{-1}H'F'G'GFHVV_T^{-1} + O_p(N^{-(4\alpha+1)/2}R_2) + O_p(N^{-2\alpha}T^{-1/2}R_2)
\]

\[
= N^{-2\alpha} V_T^{-1}H' (T^{-1}F'F)S^0 (T^{-1}F'F)HVV_T^{-1} + O_p(R_3),
\] (A45)

where

\[
R_3 = N^{-(4\alpha+1)/2}R_2 + N^{-2\alpha} T^{-1/2}R_2 + N^{-4\alpha} (NT^{-2} + T^{-1} + N^{1/2}T^{-3/2} + N^{-2\alpha} T^{-1/2})
\]

\[
= N^{-3\alpha - 1} + N^{-4\alpha - 1/2} + (N^{-2\alpha - 1/2} + N^{-4\alpha}) T^{-1/2}
\]

\[
+ N^{-2\alpha - 1} T^{-1} + (N^{1/2-3\alpha}) T^{-3/2} + N^{1-2\alpha} T^{-5/2} + (N^{-2\alpha+1/2} + N^{4\alpha+1}) T^{-2}
\]

is the order of \(O_p(N^{-(4\alpha+1)/2}R_2) + O_p(N^{-2\alpha}T^{-1/2}R_2)\) plus \(N^{-4\alpha}\) times the four reminder terms in the above expression for \(N^{-1}T^{-3}F'G'GF\).
Consider $\overline{H}$. According to Lemma PC3, $\|T^{-1}F'D^{PC}\| = O_p((NT)^{-1/2}R_1)$, from which it follows that

\[
\overline{H} = \overline{Q}(T^{-1}F'D^{PC})V_T^{-1} = N^{-2\alpha}\overline{Q}^0(T^{-1}F'F\overline{H})V_T^{-1} + N^{-2\alpha}\overline{Q}^0(T^{-1}F'D^{PC})V_T^{-1} = N^{-2\alpha}\overline{Q}^0T^{-1}F'F\overline{H}V_T^{-1} + O_p(N^{-1}(2\alpha+1/2)T^{-1/2}R_1), \tag{A46}
\]

or, since $\overline{Q}^0$ is invertible,

\[
T^{-1}F'F\overline{H}V_T^{-1} = N^{2\alpha}(\overline{Q}^0)^{-1}\overline{H} + O_p((NT)^{-1/2}R_1) = (\overline{Q}^0)^{-1}\overline{H}^0 + O_p((NT)^{-1/2}R_1),
\]

with which we obtain

\[
NT^{-1}D^{PC}D^{PC} = N^{-2\alpha}V_T^{-1}H(T^{-1}F'F)S^0(T^{-1}F'F)H^0V_T^{-1} + O_p(R_3) = N^{-2\alpha}\overline{H}^0(\overline{Q}^0)^{-1}S^0(\overline{Q}^0)^{-1}\overline{H}^0 + O_p(R_4), \tag{A47}
\]

where

\[
R_4 = R_3 + (NT)^{-1}R_1 = N^{-3\alpha-1} + N^{-4\alpha-1/2} + (N^{-2\alpha-1/2} + N^{-4\alpha} + N^{-1})T^{-1/2} + (N^{-2\alpha} + N^{-1/2})T^{-1} + (1 + N^{1/2-3\alpha})T^{-3/2} + (N^{-4\alpha+1} + N^{-2\alpha+1/2})T^{-2} + N^{-2\alpha}T^{-5/2}
\]

This completes the proof.

**Lemma CCE5.** Under Assumptions ERR,

\[
NT^{-1}E_i'D^{CCE} = \Sigma_{\epsilon,i}(\beta, I_m) + O_p(\sqrt{NT^{-1/2}}),
\]

\[
\frac{1}{T} \sum_{i=1}^N E_i'D^{CCE} = \frac{1}{N} \sum_{i=1}^N \Sigma_{\epsilon,i}(\beta, I_m) + O_p(T^{-1/2}).
\]

**Proof of Lemma CCE5.** For the first result, note that $E(\epsilon_{it}, u'_{it}) = E(\epsilon_{it}(\nu_{it} + \epsilon'_{i,t}\beta, \epsilon'_{i,t})) = (\Sigma_{\epsilon,i}\beta, \Sigma_{\epsilon,i}) = \Sigma_{\epsilon,i}(\beta, I_m)$. This suggests

\[
NT^{-1}E_i'D^{CCE} = \frac{N}{T} \sum_{i=1}^T \epsilon_{it}D^{CCE}_i = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^N \epsilon_{it}u'_{jt} = \Sigma_{\epsilon,i}(\beta, I_m) + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{i=1}^T (\epsilon_{it}u'_{jt} - \Sigma_{\epsilon,i}(\beta, I_m)) + \sqrt{N} \frac{1}{\sqrt{NT}} \sum_{i=1}^T \sum_{j\neq i} \epsilon_{i,t}u'_{j,t} = \Sigma_{\epsilon,i}(\beta, I_m) + O_p(\sqrt{NT^{-1/2}}).
\]

For the second result, we can use the calculations above to obtain

\[
\frac{1}{T} \sum_{i=1}^N E_i'D^{CCE} = \frac{1}{N} \sum_{i=1}^N \Sigma_{\epsilon,i}(\beta, I_m) + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T [\epsilon_{i,t}u'_{it} - \Sigma_{\epsilon,i}(\beta, I_m)] + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{i=1}^N \sum_{t=1}^T \sum_{j\neq i} \epsilon_{i,t}u'_{j,t} = \frac{1}{N} \sum_{i=1}^N \Sigma_{\epsilon,i}(\beta, I_m) + O_p(T^{-1/2}),
\]

and so the proof of the lemma is complete.
Lemma PC5. Under Assumptions ERR, LAM, and RK–PC,

$$\| NT^{-1} E_i^\prime D_{PC}^\prime \| = O_p(N^{-\alpha/2}) + O_p(\sqrt{N}T^{-1/2}) + O_p(NT^{-3/2}),$$

$$\frac{1}{T} \sum_{i=1}^{N} E_i^\prime D_{PC}^\prime = \frac{1}{N^{\alpha+1}} \sum_{i=1}^{N} \Sigma e_i(\beta, I_m)C_i(\beta)^{-1}H_i^0 + O_p(R_5),$$

where

$$R_5 = N^{-(\alpha+1)/2} + T^{-1/2} + \sqrt{NT^{-3/2}}.$$}

Proof of Lemma PC5. Consider the first result. By the definition of $D_{PC}$,

$$NT^{-1} E_i^\prime D_{PC}^\prime = \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t}(u_i^\prime U' + u_i^\prime C^t) \hat{p}_C^t V_{T}^{-1}. \tag{A48}$$

Ignoring $V_{T}^{-1}$, the first term on the right is

$$\frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime U' \hat{p}_C^t = \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime U'D_{PC} + \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime U' \hat{p}_C^t \hat{p}_C^t.$$}

Clearly,

$$\left\| \frac{1}{T} \sum_{t \neq s}^{T} \epsilon_{i,t} u_i^\prime u_s \right\| \leq \left( \frac{1}{T} \sum_{t \neq s}^{T} |\epsilon_{i,t}|^2 \right)^{1/2} \sqrt{N} \left( \frac{1}{T} \sum_{t \neq s}^{T} |N^{-1/2} u_i^\prime u_s|^2 \right)^{1/2} = O_p(\sqrt{N}),$$

giving

$$\left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime U'D_{PC} \right\| = \left\| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \epsilon_{i,t} u_i^\prime u_s d_{PC}^s \right\| \leq \left( \frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime u_s \right\|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \|d_{PC}^s\|^2 \right)^{1/2} = O_p(\sqrt{N})[O_p(N^{-1/2}) + O_p(T^{-1/2})] = O_p(1) + O_p(\sqrt{NT^{-1/2}}).$$

Furthermore,

$$\left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime U'f_i \right\| = \left\| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \epsilon_{i,t} u_i^\prime u_s f_s^t \right\| \leq \left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime u_s f_s^t \right\| + \left\| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s \neq t} \epsilon_{i,t} u_i^\prime u_s f_s^t \right\| \leq \frac{1}{T} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime u_t f_t^t \right\| + \frac{1}{N \sqrt{T}} \left\| \frac{1}{T} \sum_{j \neq i}^{N} \sum_{t=1}^{T} \epsilon_{i,t} u_i^\prime u_j f_j^t \right\| + \frac{\sqrt{N}}{T} \left\| \frac{1}{T} \sum_{j \neq i}^{N} \sum_{t=1}^{T} \sum_{s \neq t} \epsilon_{i,t} u_i^\prime u_j u_s f_s^t \right\| \leq O_p(1) + O_p(\sqrt{NT^{-1/2}}).$$
\[ Q(T^{-1}) + O_p(\sqrt{N}T^{-1}) \]

and so, with \( ||H|| = O_p(N^{-2\alpha}) \),

\[
\left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i'u_i' \hat{U}' \hat{F} \right\| \leq \left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i'u_i' \hat{F} \right\| ||H||
\]

\[
= O_p(1) + O_p(\sqrt{N}T^{-1/2}) + O_p(N^{-1-2\alpha}T^{-3/2}). \quad (A49)
\]

Let us now consider the second term in the expansion of \( NT^{-1}\epsilon_i'D^{PC} \). By definition, \( H = T^{-1}Q^0\hat{F}'\hat{P}C \). \( \hat{H}V_T^{-1} = N^{-2\alpha}T^{-1}Q^0\hat{F}'\hat{P}C \). \( \hat{H}V_T^{-1} \), or, since \( Q^0 \) is invertible, \( T^{-1}F'\hat{P}C \). \( \hat{H} = (Q^0)^{-1}H = (Q^0)^{-1}H^0 \). Hence, \( ||T^{-1}F'\hat{P}C|| = ||Q^0||^{-1}H^0 \). Moreover,

\[
\left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} u_i'u_i' \right\| \leq N^{-\alpha} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} u_i'C^0_{i,t} \right\| + \frac{N^{1/2-\alpha}}{\sqrt{T}} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} \frac{1}{\sqrt{N}} \sum_{j \neq i}^{N} u_i'C^0_{j,t} \right\|
\]

\[
= O_p(N^{-\alpha}) + O_p(N^{1/2-\alpha}T^{-1/2}).
\]

Finally, since

\[
\left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} f_i'C^0u_i'F \right\| \leq \sqrt{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{i,t} f_i' \right\| ||(NT)^{-1/2}C^0u_i'F|| = O_p(\sqrt{N}),
\]

and

\[
\left\| \frac{1}{T^{3/2}} \sum_{t=1}^{T} \epsilon_{i,t} f_i'C^0u_i'D^{PC} \right\|
\]

\[
= \left\| \frac{1}{T^{3/2}} \sum_{t=1}^{T} \sum_{s=1}^{T} \epsilon_{i,t} f_i'C^0u_i'd_{s}^{PC} \right\|
\]

\[
\leq \sqrt{N} \left( \frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{i,t} f_i'(N^{-1/2}C^0u_i) \right\| \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \left\| d_{s}^{PC} \right\| \right)^{1/2}
\]

\[
= \sqrt{N}[O_p(N^{-1/2}) + O_p(T^{-1/2})] = O_p(1) + O_p(\sqrt{N}T^{-1/2}),
\]

the order of the third term becomes

\[
\left\| \frac{1}{T^2} \sum_{t=1}^{T} \epsilon_{i,t} f_i'C'u_i' \hat{F} \right\|
\]

\[
\leq \frac{1}{N^{3\alpha}T} \left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{i,t} f_i'C^0u_i'F \right\| ||H^0|| + \frac{1}{N^{\alpha}T} \left\| \frac{1}{T^{3/2}} \sum_{t=1}^{T} \epsilon_{i,t} f_i'C^0u_i'D^{PC} \right\|
\]

\[
= N^{-3\alpha}T^{-1}O_p(\sqrt{N}) + N^{-\alpha}T^{-1/2} [O_p(1) + O_p(\sqrt{N}T^{-1/2})]
\]

\[
= O_p(N^{1/2-3\alpha}T^{-1}) + O_p(N^{-\alpha}T^{-1/2}) + O_p(N^{1/2-\alpha}T^{-1})
\]

\[
= O_p(N^{-\alpha}T^{-1/2}) + O_p(N^{1/2-\alpha}T^{-1}). \quad (A50)
\]
By adding the above results,
\[
||NT^{-1}E'_iD_{PC}|| \leq \left| \frac{1}{T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}(u'_i U' + u'_i C' F' + f'_i C' U') \hat{E}_{PC} \right| \left| \left| V_T^{-1} \right| \right|
= O_p(1) + O_p(\sqrt{NT^{-1/2}}) + O_p(N^{1-2\alpha} T^{-3/2}). \tag{A51}
\]

The proof of the second result is very similar to that of the first. We begin by noting that
\[
\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i U' \hat{F}_{PC} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i U' D_{PC} + \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i U' F_H.
\]

Clearly,
\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i u_s = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i u_s + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t \neq s}^{T} \epsilon_{i,t}u'_i u_s
= O_p(1) + O_p(T^{-1}) = O_p(1),
\]

and therefore
\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i u_s = \frac{\sqrt{N}}{T} \sum_{i=1}^{N} \epsilon_{i,s}(N^{-1} u'_s u_s) + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t \neq s}^{T} \epsilon_{i,t}u'_i u_s
= O_p(\sqrt{NT^{-1}}) + O_p(1).
\]

Thus,
\[
\left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i U' D_{PC} \right| \right| = \left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \epsilon_{i,t}u'_i u_s \hat{d}_{PC} \right| \right|
\leq \left( \frac{1}{T} \sum_{s=1}^{T} \left| \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i u_s \right| \right| \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \left| |d_{PC}^{s}|| \right| \right)^{1/2}
= [O_p(\sqrt{NT^{-1}}) + O_p(1)][O_p(N^{-1/2}) + O_p(T^{-1/2})]
= O_p(N^{-1/2}) + O_p(\sqrt{NT^{-3/2}}) + O_p(T^{-1/2}).
\]

Moreover,
\[
\left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i U' F \right| \right|
= \left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \epsilon_{i,t}u'_i u_s f'_s \right| \right|
\leq \left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t}u'_i u_t f'_t \right| \right| + \left| \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{s} \epsilon_{i,t}u'_i u_s f'_s \right| \right|
\]
Consider the second term in the expansion of \( O_u \),
\[
\sum_{i=1}^{NT} \sum_{t=1}^{T} \epsilon_{i,t} u_i' T^{-1/2} P \epsilon_{i,t} + \sum_{i=1}^{NT} \sum_{t=1}^{T} \epsilon_{i,t} u_i' T^{-1/2} P \epsilon_{i,t}
\]
and
\[
\sum_{i=1}^{NT} \sum_{t=1}^{T} \epsilon_{i,t} u_i' T^{-1/2} P \epsilon_{i,t}
\]
from which we deduce that, with \( ||\hat{H}|| = O_p(N^{-2\alpha})\),
\[
\frac{1}{NT^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{i,t} u_i' T^{-1/2} P \epsilon_{i,t} \leq \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \epsilon_{i,t} u_i' T^{-1/2} P \epsilon_{i,t} ||\hat{H}||
\]
\[
= O_p(N^{-1/2}) + O_p(\sqrt{NT}^{-3/2}) + O_p((NT)^{-1/2}) + O_p(T^{-1})
\]
where \( E(\epsilon_{i,t} u_i') = E[\epsilon_{i,t}(v_{i,t} + \epsilon_{i,t}' \beta, \epsilon_{i,t})] = (\Sigma_{\epsilon_{i,t}} \beta, \Sigma_{\epsilon_{i,t}}) = \Sigma_{\epsilon_{i,t}}(\beta, I_m) \), suggesting that
\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \epsilon_{i,t} u_i' \Sigma_{\epsilon_{i,t}} C_{ij}^0
\]
\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \epsilon_{i,t} u_i' \Sigma_{\epsilon_{i,t}} C_{ij}^0 + \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} u_i' \Sigma_{\epsilon_{i,t}} C_{ij}^0
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon_{i,t}}(\beta, I_m) C_{ij}^0 + \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{N} (\epsilon_{i,t} u_i' - \Sigma_{\epsilon_{i,t}}(\beta, I_m)) C_{ij}^0
\]
\[
+ \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \epsilon_{i,t} u_i' C_{ij}^0
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon_{i,t}}(\beta, I_m) C_{ij}^0 + O_p((NT)^{-1/2}) + O_p(T^{-1/2})
\]
\[
= \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon_{i,t}}(\beta, I_m) C_{ij}^0 + O_p(T^{-1/2}).
\]
Therefore,
\[
\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} u_j' CF' \hat{F}^{PC} V_T^{-1} = \frac{1}{N^{a+1}T} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{N} \epsilon_{i,t} u_j' C_j^0 (Q^0) -1 H^0
\]
\[
= \frac{1}{N^{a+1}} \sum_{i=1}^{N} \Sigma \epsilon_{i}(\beta, I_m) C_i^0 (Q^0) -1 H^0 + O_p(N^{-a} T^{-1/2}). \quad (A54)
\]

One term remains, the order of which is given by
\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' C^0' U' F \right\| \leq \left\| \frac{1}{N^{a}T} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' C^0' U' D^{PC} \right\|
\]
\[
= N^{-a} T^{-1} O_p(N^{-2a}) + N^{-a} T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})]
\]
\[
= O_p(N^{-(a+1)/2} T^{-1/2}) + O_p(N^{-a} T^{-1}). \quad (A55)
\]
as follows from noting that
\[
\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' C^0' U' F \right\| \leq \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' \right\| \left\| (NT)^{-1/2} C^0' U' F \right\| = O_p(1),
\]
and
\[
\left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' C^0' U' D^{PC} \right\|
\]
\[
= \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{T} \epsilon_{i,t} f_i' C^0' u_j d^{PC}_j \right\|
\]
\[
\leq \left( \frac{1}{T} \sum_{i=1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} f_i' (N^{-1/2} C^0' u_j) \left\|^2 \right\right)^{1/2} \left( \frac{1}{T} \sum_{i=1}^{T} \left\| d^{PC}_j \right\|^2 \right)^{1/2}
\]
\[
= O_p(N^{-1/2}) + O_p(T^{-1/2}).
\]

Direct substitution now yields
\[
\frac{1}{T} \sum_{i=1}^{N} E[D^{PC}] = \frac{1}{N^{a+1}} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{i,t} u_j' (U' + u_j' C' F') \hat{F}^{PC} V_T^{-1}
\]
\[
= \frac{1}{N^{a+1}} \sum_{i=1}^{N} \Sigma \epsilon_{i}(\beta, I_m) C_i^0 (Q^0) -1 H^0 + O_p(R_5), \quad (A56)
\]
where
\[
R_5 = N^{-1/2} + T^{-1/2} + \sqrt{N} T^{-3/2}.
\]
This establishes the second result, and hence the proof of the lemma is complete.
Lemma CCE6. Under Assumptions ERR, LAM, and RK–CCE,

\[ ||NT^{-1}v'_iD^{CCE}|| = O_p(1) + O_p(\sqrt{NT}^{-1/2}), \]

\[ \frac{1}{N} \sum_{i=1}^{N} NT^{-1}v'_iD^{CCE}(\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{v,i}(1,0)(\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' + O_p(T^{-1/2}). \]

Proof of Lemma CCE6. Consider the first result. Note that \( E(v_{i,t}u'_{i,t}) = E[v_{i,t}(v_{i,t} + \epsilon'_{i,t}\beta, \epsilon'_{i,t})] = (\sigma^2_{v,i}, 0) \). Making use of this and

\[ NT^{-1}v'_iD^{CCE} = \frac{N}{T} \sum_{t=1}^{T} v_{i,t} \frac{1}{N} \sum_{j=1}^{N} u'_{i,j} = \frac{1}{T} \sum_{t=1}^{T} v_{i,t}u'_{i,t} + \frac{1}{T} \sum_{t=1}^{T} v_{i,t}u'_{i,t} \]

we obtain

\[ ||NT^{-1}v'_iD^{CCE}|| \leq \sigma^2_{v,i} + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{t=1}^{T} |v_{i,t}u'_{i,t} - \sigma^2_{v,i}(1,0)| + \frac{\sqrt{N}}{\sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{j \neq i}^{N} v_{i,t}u'_{j,t}, \]

as required for the first result.

In order to show the second result, we again make use of the above expansion of \( NT^{-1}v'_iD^{CCE} \), which, together with

\[ \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} |v_{i,t}u'_{i,t} - \sigma^2_{v,i}(1,0)| (\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' \right| = O_p(1), \]

\[ \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} v_{i,t}u'_{j,t} (\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' \right| = O_p(1), \]

can be used to show that

\[ \frac{1}{N} \sum_{i=1}^{N} NT^{-1}v'_iD^{CCE}(\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{v,i}(1,0)(\mathbf{C}^0)^{-1}(\mathbf{A}_i^0)' + O_p(T^{-1/2}). \]

Lemma PC6. Under Assumptions ERR, LAM, and RK–PC,

\[ ||NT^{-1}v'_iD^{PC}|| = O_p(N^{-\alpha}) + O_p(R_6), \]

\[ \frac{1}{N} \sum_{i=1}^{N} NT^{-1}v'_iD^{PC}(\mathbf{H}^0)^{-1}(\mathbf{A}_i^0)' = \frac{1}{N^{1+\alpha}} \sum_{i=1}^{N} \sigma^2_{v,i}(1,0)\mathbf{C}_i^0(\mathbf{Q}^0)^{-1}(\mathbf{A}_i^0)' + O_p(N^{-\alpha}T^{-1/2}) + O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{1/2}T^{-3/2}), \]

where

\[ R_6 = N^{-1/2} + (1 + N^{1/2-\alpha})T^{-1/2} + \sqrt{NT}^{-1} + NT^{-3/2}. \]
Proof of Lemma PC6. This proof is very similar to Proof of Lemma PC5. Consider the first result. We have

\[ NT^{-1} v'_i D^{PC} = \frac{1}{T^2} \sum_{t=1}^{T} \left( v_{i,t} (u'_i U' + u'_i C F' + f'_i C' U') \right) \hat{F}^{PC} V^{-1}_T. \]  

(H57)

Here

\[ \frac{1}{T^2} \sum_{t=1}^{T} v_{i,t} u'_i U'^{PC} = \frac{1}{T^2} \sum_{t=1}^{T} v_{i,t} u'_i D^{PC} + \frac{1}{T^2} \sum_{t=1}^{T} v_{i,t} u'_i F H, \]

with

\[ \left\| \frac{1}{NT^2} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right\| \]

\[ = \left\| \frac{1}{NT^2} \sum_{t=1}^{T} \sum_{s=1}^{T} v_{i,t} u'_i d_s \right\| \]

\[ \leq \left\{ \frac{1}{T} \sum_{s=1}^{T} \right\} \left( \frac{1}{NT} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right)^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} ||d_s^{PC}||^2 \right)^{1/2} \]

\[ = \left[ O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}) \right] \left[ O_p(N^{-1/2}) + O_p(T^{-1/2}) \right] \]

\[ = O_p(N^{-3/2}) + O_p(N^{-1} T^{-1/2}) + O_p(T^{-3/2}) + O_p(N^{-1/2} T^{-1}), \]  

(A58)

where we have made use of the fact that

\[ \left\| \frac{1}{NT} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right\| \]

\[ \leq \left\| \frac{1}{NT} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right\| + \left\| \frac{1}{NT} \sum_{j \neq i}^{N} \sum_{t=1}^{T} v_{i,t} u'_i u_j \right\| \]

\[ \leq \frac{1}{N} \left\| \frac{1}{T} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right\| + \frac{1}{T} \left\| \frac{1}{N} \sum_{j \neq i}^{N} v_{i,j} u'_i u_j \right\| + \frac{1}{\sqrt{NT}} \left\| \sum_{j \neq i}^{N} \sum_{t=1}^{T} v_{i,t} u'_i u_j \right\| \]

\[ = O_p(N^{-1}) + O_p(T^{-1}) + O_p((NT)^{-1/2}). \]

By using the same steps as in the proof of Lemma PC5, we can further show that

\[ \left\| \frac{1}{NT^2} \sum_{t=1}^{T} v_{i,t} u'_i u_i \right\| \]

\[ \leq \frac{1}{NT} \left\| \frac{1}{T} \sum_{t=1}^{T} v_{i,t} u'_i u_i f'_t \right\| + \frac{1}{NT} \left\| \sum_{j \neq i}^{N} \sum_{t=1}^{T} v_{i,t} u'_i u_j f'_t \right\| \]

\[ + \frac{1}{N \sqrt{T}} \left\| \frac{T^{3/2}}{T} \sum_{t=1}^{T} \sum_{s \neq t}^{T} v_{i,t} u'_i u_s f'_t \right\| + \frac{1}{\sqrt{NT}} \left\| \sum_{j \neq i}^{N} \sum_{t=1}^{T} v_{i,t} u'_i u_j f'_t \right\| \]

\[ = O_p(N^{-1} T^{-1}) + O_p(T^{-3/2}) + O_p(N^{-1} T^{-1/2}) + O_p(N^{-1/2} T^{-1}) \]

\[ = O_p(T^{-3/2}) + O_p(N^{-1} T^{-1/2}) + O_p(N^{-1/2} T^{-1}), \]  

110
and so we obtain
\[
\left\| \frac{1}{T^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_{i,t}' U' \right\| \\
\leq N \left( \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_{i,t}' U' D_{PC} \right) + N^{1-2\alpha} \left( \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_{i,t}' F \right) \left\| |H|^0 \right\| \\
= O_p(N^{-1/2}) + O_p(T^{-1/2}) + O_p(N T^{-3/2}) + O_p(\sqrt{N T^{-1}}). \quad (A59)
\]
Again, in analogy to the proof of Lemma PC5, we may write
\[
\frac{1}{T} \sum_{t=1}^{T} T \sum_{i=1}^{N} v_{i,t} u_{i,t} C^0_{i} = (\sigma^2_{v,i}, 0)
\]
where \( E(v_{i,t} u_{i,t}) = E(v_{i,t} (v_{i,t} + \delta_{i,t} \beta, \epsilon_{i,t}) \} = (\sigma^2_{v,i}, 0) \) and
\[
\frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} v_{i,t} u_{i,t} C^0_{i} = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} v_{i,t} u_{i,t} C^0_{j}
\]
yielding
\[
\frac{1}{T^2} \sum_{i=1}^{T} v_{i,t} u_{i,t} C^0_{i} = \frac{1}{T^2} \sum_{i=1}^{T} v_{i,t} u_{i,t} C^0_{j} = \sigma^2_{v,i} (1, 0) C^0_{i} + \sqrt{N} \frac{1}{\sqrt{T}} \sum_{i=1}^{T} v_{i,t} u_{i,t} C^0_{j}
\]
Finally, since
\[
\left\| \frac{1}{NT^2} \sum_{t=1}^{T} v_{i,t} f_i' \right\| = \frac{1}{N^{\alpha+1/2} T} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} v_{i,t} f_i' \right\| \left\| (NT)^{-1/2} C^0 U' F \right\|
\]
and
\[
\left\| \frac{1}{NT^2} \sum_{t=1}^{T} v_{i,t} f_i' C^0 U' D_{PC} \right\| \leq \left( \frac{1}{N^{\alpha+1/2} \sqrt{T}} \left( \frac{1}{T} \sum_{t=1}^{T} \left\| v_{i,t} f_i' (N^{-1/2} C^0 U_i) \right\| \right)^2 \right)^{\frac{1}{2}} \left( \frac{1}{T} \sum_{s=1}^{T} \left| d_s^0 \right|^2 \right)^{\frac{1}{2}}
\]
111
which can be combined with the above results to obtain

\[ \sum_{t=1}^{T} \left| \frac{1}{T^2} \sum_{i=1}^{N} \nu_i \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq O_p(N^{-\alpha+1/2}) + O_p(T^{-1/2}), \]  

(A61)

which can be combined with the above results to obtain

\[ \sum_{t=1}^{T} \left| \frac{1}{T^2} \sum_{i=1}^{N} \nu_i \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq \frac{1}{T^2} \sum_{i=1}^{N} \left| \mathbf{F}_T \right| \leq O_p(N^{-\alpha+1/2}) + O_p(T^{-1/2}), \]  

(A62)

where

\[ R_6 = N^{-(\alpha+1)/2} + (1 + N^{1/2-\alpha}) T^{-1/2} + \sqrt{NT^{-1} + NT^{-3/2}}. \]

Hence,

\[ ||NT^{-1} \nu|D^P|| = O_p(N^{-\alpha}) + O_p(R_6), \]  

(A63)

as required.

As in the proof of CCE6, the second result follows from a simple manipulation of the proof of the first. Therefore, only essential details will be given. Note first that

\[ \frac{1}{N} \sum_{i=1}^{N} NT^{-1} \nu \mathbf{D}^P (\mathbf{H}^0) - 1 (\mathbf{A}^0)^T = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i (\mathbf{H}^0)_t - 1 (\mathbf{A}^0)^T. \]  

(A64)

We begin by considering the first and third terms on the right. These are negligible, and therefore, the analysis is unaffected by the scaling of \( V_T^{-1} (\mathbf{H}^0) - 1 (\mathbf{A}^0)^T \). Ignoring this matrix, the order of the first term is given by

\[ \left| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \mathbf{F}_T \right| = O_p(N^{-\alpha+1/2}) + O_p((NT)^{-1/2}) + O_p(T^{-1}) + [O_p((NT)^{-1/2}) + O_p(T^{-1})] O_p(N^{-2\alpha}) = O_p(N^{-(\alpha+1)/2}) + O_p((NT)^{-1/2}) + O_p(T^{-1}). \]  

(A65)

In order to appreciate that this must be so, note that

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \mathbf{u}_i \mathbf{u}_i^T \leq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \mathbf{u}_i \mathbf{u}_i^T + \frac{1}{\sqrt{T}} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \mathbf{u}_i \mathbf{u}_i^T \leq O_p(N^{-1/2}) + O_p(T^{-1/2}), \]

112
and therefore
\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' D_P^C \right\| = \left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' d_s^C \right\| \\
\leq \left( \frac{1}{T} \sum_{i=1}^{N} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' u_i' \right\|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \| d_s^C \|^2 \right)^{1/2} \\
= [O_p(N^{-1/2}) + O_p(T^{-1/2})] [O_p(N^{-1/2}) + O_p(T^{-1/2})] \\
= O_p(N^{-1}) + O_p((NT)^{-1/2}) + O_p(T^{-1}).
\]

Moreover,
\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' F \right\| \\
\leq \frac{1}{T} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' u_i' f_t' \right\| + \frac{\sqrt{N}}{T^{3/2}} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \sum_{t=1}^{T} v_{i,t} u_i' u_j' f_t' \right\| \\
+ \frac{1}{\sqrt{NT}} \left\| \frac{1}{NT^{3/2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j \neq i}^{N} \sum_{s \neq t}^{T} v_{i,t} u_i' u_j' f_s' \right\| \\
= O_p(T^{-1}) + O_p(N^{1/2} T^{-3/2}) + O_p((NT)^{-1/2}) + O_p(T^{-1}) \\
= O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{1/2} T^{-3/2}).
\]
The second term is
\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' C T^{-1} F' P^C V^{-1} T (\hat{H}^0) - 1 (A_i^0)' \\
= \frac{1}{N^{w+1} T} \sum_{j=1}^{N} \sum_{i=1}^{T} \sum_{j=1}^{N} v_{i,t} u_i' c_j^0 (\hat{Q}^0) - 1 (A_i^0)' \\
= \frac{1}{N^{w+1} T} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{i,t} u_i' c_j^0 (\hat{Q}^0) - 1 (A_i^0)' + \frac{1}{N^{w+1} T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} v_{i,t} u_i' c_j^0 (\hat{Q}^0) - 1 (A_i^0)' \\
= \frac{1}{N^{w+1}} \sum_{i=1} v_{i,t}^2 (1,0) c_j^0 (\hat{Q}^0) - 1 (A_i^0)' \\
+ \frac{1}{N^{w+1/2} \sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} [v_{i,t} u_i' C_j^0 (\hat{Q}^0) - 1 (A_i^0)' \\
+ \frac{1}{N^{w+1/2} \sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j \neq i}^{N} v_{i,t} u_i' c_j^0 (\hat{Q}^0) - 1 (A_i^0)' \\
= \frac{1}{N^{w+1}} \sum_{i=1} v_{i,t}^2 (1,0) c_j^0 (\hat{Q}^0) - 1 (A_i^0)' + O_p(N^{-\alpha} T^{-1/2}).
\]
The third and final term is, again ignoring the scaling by \( V_T^{-1}(\tilde{H}^0)^{-1}(A^0_i)' \),

\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t C' U^t P_C \right\| 
\leq \left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t C' U' F \right\| \|H\| + \left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t C' U' D^P_C \right\| 
= O_p(N^{-(3\alpha+1)/2}T^{-1/2}) + O_p(N^{-\alpha}T^{-1}),
\] (A67)

which uses

\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t C' U' F \right\| = \frac{1}{N^{\alpha}T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} v_{it} f_i^t \right\| \| NT^{-1/2} C^0 U' F \| = O_p(N^{-\alpha}T^{-1}),
\]

and

\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t C' U' D^P_C \right\| 
= \left\| \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} v_{it} f_i^t C' u_s d^P_C \right\| 
\leq \frac{1}{N^{\alpha}T} \left( \frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it} f_i^t \right\| ^{1/2} \left( \frac{1}{T} \sum_{s=1}^{T} \| d^P_C \|^2 \right) ^{1/2} \right) ^{1/2} 
= N^{-\alpha}T^{-1/2} [O_p(N^{-1/2}) + O_p(T^{-1/2})] 
= O_p(N^{-3\alpha-1/2}T^{-1/2}) + O_p(N^{-\alpha}T^{-1}).
\]

The above results imply

\[
\frac{1}{N} \sum_{i=1}^{N} NT^{-1} v_j^P D^P_C (\tilde{H}^0)^{-1}(A^0_j)' = \frac{1}{N^{1+\alpha}} \sum_{i=1}^{N} \sigma_i^2 (1, 0) C_i^0 (Q_i^0)^{-1}(A_i^0)' + O_p(N^{-\alpha}T^{-1/2}) 
\quad + O_p((NT)^{-1/2}) + O_p(T^{-1}) + O_p(N^{1/2}T^{-3/2}).
\] (A68)

This establishes the second result, and hence the proof of the lemma is complete.

**Lemma CCE7.** Under Assumptions LAM and RK–CCE,

\[
T||((\tilde{F}^{\text{CCE}})' \tilde{F}^{\text{CCE}})^{-1} - (C' F' F C)^{-1}|| = O_p(N^{3\alpha-1/2}T^{-1/2}) + O_p(N^{4\alpha-1}),
\]

\[
||T^{-1} (\tilde{F}^{\text{CCE}})' T^{-1} (\tilde{F}^{\text{CCE}})^{-1} - F C E F C E P C E D C E|| = O_p(N^{2\alpha}).
\]

**Proof of Lemma CCE7.** From

\[
\tilde{F}^{\text{CCE}} = (F C + D^{\text{CCE}})' (F C + D^{\text{CCE}}) = C' F F C + D^{\text{CCE}} C + C' F D^{\text{CCE}} + D^{\text{CCE}} P C E D^{\text{CCE}},
\]

we obtain

\[
T^{-1} ||(\tilde{F}^{\text{CCE}})' T^{-1} (\tilde{F}^{\text{CCE}})' - C' F F C|| 
\leq 2N^{-\alpha}T^{-1/2} || \sqrt{N} T^{-1/2} D^{\text{CCE}} F || || C^0 || + N^{-1} || N T^{-1} D^{\text{CCE}} D^{\text{CCE}} || 
= O_p(N^{-\alpha}T^{-1/2}) + O_p(N^{-1}),
\] (A69)
or

\[ N^{2\alpha} T^{-1} ||(\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}} - C′F′FC|| = O_p(N^{\alpha-1/2} T^{-1/2}) + O_p(N^{2\alpha-1}). \]  

(A.70)

This last result is crucial. Note in particular that since \( N^{2\alpha} T^{-1} C′F′FC = T^{-1} (C^0)^′ F′F^0 \) is positive definite, we can use a power series expansion of the type

\[
(N^{2\alpha} T^{-1} (\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}})^{-1} = (T^{-1} (\hat{C}^0)^′ F′F^0)^{-1}
- (T^{-1} (\hat{C}^0)^′ F′F^0)^{-1}N^{2\alpha} T^{-1} ((\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}} - C′F′FC)(T^{-1} (\hat{C}^0)^′ F′F^0)^{-1}
+ (T^{-1} (\hat{C}^0)^′ F′F^0)^{-1}N^{2\alpha} T^{-1} ((\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}} - C′F′FC)(T^{-1} (\hat{C}^0)^′ F′F^0)^{-1}
\times N^{2\alpha} T^{-1} ((\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}} - C′F′FC)(T^{-1} (\hat{C}^0)^′ F′F^0)^{-1} - \ldots
\]

to show that

\[
T ||(\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}})^{-1} - (\hat{C}′F′FC)^{-1}||
\leq N^{2\alpha} ||(\hat{C}^0)^′ F′F^0||^4 ||(T^{-1} F′F)^{-1}||^2 N^{2\alpha} T^{-1} ||(\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}} - C′F′FC||
= O_p(N^{3\alpha-1/2} T^{-1/2}) + O_p(N^{4\alpha-1}). \]  

(A.71)

Also, given the positive definiteness of \( T^{-1} (\hat{C}^0)^′ F′F^0 \), we have

\[ N^{-2\alpha} T ||(\hat{F}^{\text{CCE}})′ \hat{F}^{\text{CCE}})^{-1}|| = O_p(1). \]  

(A.72)

**Lemma PC7.** Under Assumptions LAM and RK–PC,

\[ T ||(\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1} - (\hat{H}′F′FH)^{-1}|| = O_p(N^{4\alpha-1/2} T^{-1/2} R_{10}), \]

where

\[ R_{10} = N^{-1/2} \sqrt{T} + N^{-5\alpha} + \sqrt{NT^{-1/2}}. \]

**Proof of Lemma PC7.** The proof of this lemma is slightly simpler than that of Lemma CCE7 in that there is no need for any expansions of the inverse. The trick is to note that

\[
((\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1} - (\hat{H}′F′FH)^{-1}
= -((\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1}((\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}} - \hat{H}′F′FH)(\hat{H}′F′FH)^{-1}.
\]

By using this, \( T^{-1} (\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}} = I_r, O_p(R_{10}) + O_p(N^{-2\alpha} R_1) = O_p(R_{10}), \) and \( ||\hat{H}|| = O_p(N^{-2\alpha}) \), we obtain

\[
T ||(\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1} - (\hat{H}′F′FH)^{-1}||
\leq (NT)^{-1/2} ||(T^{-1} (\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1}||\sqrt{NT^{-1/2}} ||(\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}} - \hat{H}′F′FH|| ||(T^{-1} \hat{H}′F′FH)^{-1}||
= (NT)^{-1/2} ||(T^{-1} (\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1}||\sqrt{NT^{-1/2}} ||D^{\text{PC}} F′PC + \hat{H}′D′PC|| ||(T^{-1} \hat{H}′F′FH)^{-1}||
\leq (NT)^{-1/2} ||(T^{-1} (\hat{F}^{\text{PC}})′ \hat{F}^{\text{PC}})^{-1}||\sqrt{NT^{-1/2}} ||D^{\text{PC}} F′PC + \hat{H}′D′PC|| ||H^{-1}||^2 ||(T^{-1} \hat{F}′F)^{-1}||
= (NT)^{-1/2}[O_p(R_{10}) + O_p(N^{-2\alpha} R_1)]O_p(N^{4\alpha}) = O_p(N^{4\alpha-1/2} T^{-1/2} R_{10}). \]  

(A.73)

**Lemma CCE8.** Under Assumptions ERR and LAM,

\[ ||NT^{-1}X_d\hat{D}^{\text{CCE}}|| = O_p(1) + O_p(\sqrt{NT^{-1/2}}). \]
Proof of Lemma CCE8. From Lemmas CCE3 and CCE5,
\[ ||NT^{-1}X_i^iD_{CCE}|| = ||NT^{-1}(FA_i^0 + E_i^0)D_{CCE}|| \]
\[ \leq N^{1/2-\alpha}T^{-1/2}||A_0^0||\sqrt{NT^{-1}}D_{CCE}|| + ||NT^{-1}E_i^0D_{CCE}|| \]
\[ = O_p(N^{1/2-\alpha}T^{-1/2}) + O_p(1) + O_p(\sqrt{NT^{-1}}/2) = O_p(1) + O_p(\sqrt{NT^{-1}}/2), \]
as was to be shown.

Lemma PC8. Under Assumptions ERR, LAM, and RK–PC,
\[ ||NT^{-1}X_i^iD_{PC}|| = O_p(N^{1/2-\alpha}T^{-1/2}R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1}}) + O_p(N^{-3/2}). \]

Proof of Lemma PC8. Lemmas PC3 and PC5 imply
\[ ||NT^{-1}X_i^iD_{PC}|| = ||NT^{-1}(FA_i^0 + E_i^0)^0D_{PC}|| \]
\[ \leq N^{1/2-\alpha}T^{-1/2}||A_0^0||\sqrt{NT^{-1}}D_{PC}|| + ||NT^{-1}E_i^0D_{PC}|| \]
\[ = O_p(N^{1/2-\alpha}T^{-1/2}R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1}}) + O_p(N^{-3/2}). \]

Lemma CCE9. Under Assumptions ERR, LAM, RK–CCE, KAP, and \( \kappa > \max\{0, 2\alpha - 1\} \),
\[ T^{-1}X_i^iM_{F_{CCE}}X_i = \Sigma \epsilon_i + o_p(1), \]
\[ \frac{1}{NT} \sum_{i=1}^N X_i^iM_{F_{CCE}}X_i = \Sigma \epsilon + o_p(1). \]

Proof of Lemma CCE9. We can expand the expression as
\[ T^{-1}X_i^iM_{F_{CCE}}X_i = T^{-1}X_i^iM_{F_{FC}}X_i - T^{-1}X_i^i(M_{F_{C}} - M_{F_{CCE}})X_i. \] (A74)

Now, from the definitions of \( M_{F_{CCE}} \) and \( M_{F_{FC}} \), we have
\[ M_{F_{C}} - M_{F_{CCE}} \]
\[ = D_{CCE}((\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1}D_{CCE} + D_{CCE}((\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1}C F \]
\[ + F_{C}(\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1}D_{CCE} + F_{C}((\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1} - (\sqrt{C F} F C)^{-1}(\sqrt{C F} F C)^{-1}||F \]
\[ Note furthermore that \]
\[ ||T^{-1}X_i^iF|| \leq N^{-\alpha}||A_0^0||||T^{-1}F|| + T^{-1/2}||T^{-1/2}E_i^0F|| = O_p(N^{-\alpha}) + O_p(T^{-1/2}). \] (A76)

These two results and Lemmas CCE7 and CCE8 imply
\[ ||T^{-1}X_i^i(M_{F_{C}} - M_{F_{CCE}})X_i|| \]
\[ = N^{-2}||NT^{-1}X_i^iD_{CCE}||^2||(T^{-1}(\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1}|| \]
\[ + 2N^{(\alpha+1)}||NT^{-1}X_i^iD_{CCE}|| ||T^{-1}((\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1}C F \]
\[ + N^{-2\alpha}||T^{-1}X_i^iF||^2||C F ||^2T)||((\hat{\epsilon}_{CCE}/\hat{\epsilon}_{CCE})^{-1} - (\sqrt{C F} F C)^{-1}|| \]
\[ = N^{-2}[O_p(1) + O_p(\sqrt{NT^{-1/2}})^2O_p(N^{2\alpha}) \]
\[ + N^{(\alpha+1)}[O_p(1) + O_p(\sqrt{NT^{-1/2}})]O_p(N^{2\alpha})O_p(N^{-\alpha}) + O_p(T^{-1/2}) \]
\[ + N^{2\alpha}[O_p(N^{-\alpha}) + O_p(T^{-1/2})]^2O_p(N^{3\alpha-1/2}T^{-1/2}) + O_p(N^{4\alpha-1}) \]
\[ = O_p(N^{2\alpha-2}) + O_p(N^{2\alpha-3/2}T^{-1/2}) + O_p(N^{2\alpha-1}T^{-1}) \]
\[ + O_p(N^{-1}) + O_p(N^{-\alpha-1}T^{-1/2}) + O_p((NT)^{-1/2}) + O_p(N^{\alpha-1/2}T^{-1}) \]

116
\[ + O_p(N^{-\alpha-1/2} T^{-1/2}) + O_p(N^{-1}) + O_p(N^{-1/2} T^{-1}) + O_p(N^\alpha T^{-1/2}) \]
\[ + O_p(N^\alpha T^{-3/2}) + O_p(N^{2\alpha-1} T^{-1}) \]
\[ = O_p(N^{2\alpha-2}) + O_p(N^{-1}) + O_p(N^{2\alpha-3/2} T^{-1/2}) + O_p(N^\alpha T^{-1/2}) + O_p((NT)^{-1/2}) \]
\[ + O_p(N^\alpha T^{-1}) + O_p(N^{2\alpha-1} T^{-1}) \]
\[ = O_p(N^{2\alpha-2}) + O_p(N^{-1}) + O_p(N^{2\alpha-3(1+\kappa)/2}) + O_p(N^\alpha T^{-1/2}) + O_p(N^{-1(1+\kappa)/2}) \]
\[ + O_p(N^{\alpha-1/2}) + O_p(N^{\alpha} T^{-1}) \]
\[ \quad \text{(A77)} \]

which is clearly \( o_p(1) \) under \( T = N^{\kappa} \) and \( \kappa > \max\{0, 2\alpha - 1\} \). Also, the first term on the right-hand side of (A74) is given by

\[
T^{-1}X_i'\hat{M}_{\text{ECE}}X_i = T^{-1}E_i'\hat{M}_{\text{ECE}}E_i
\]
\[ = T^{-1}E_i' - T^{-1}(T^{-1/2}E_i'F)\tilde{C}^0[T^{-1}(\tilde{C}^0)'F'\tilde{C}^0]^{-1}(\tilde{C}^0)'(T^{-1/2}F'E_i)
\]
\[ = \Sigma_{\epsilon,i} + o_p(1). \quad (A78) \]

When taken together, these two results imply that if \( \kappa > \max\{0, 2\alpha - 1\} \), we have

\[
T^{-1}X_i'\hat{M}_{\text{ECE}}X_i = \Sigma_{\epsilon,i} + o_p(1). \quad (A79)
\]

The second result of the lemma follows from

\[
\left\| \frac{1}{NT} \sum_{i=1}^{N} X_i'\hat{M}_{\text{ECE}}X_i \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \| T^{-1}X_i'\hat{M}_{\text{ECE}}X_i \|
\]
and so we are done.

**Lemma PC9.** Under Assumptions ERR, LAM, RK-PC, and KAP,

\[
\frac{1}{NT} \sum_{i=1}^{N} X_i'\hat{M}_{\text{PC}}X_i = \Sigma_{\epsilon} + o_p(1).
\]

**Proof of Lemma PC9.** \((NT)^{-1} \sum_{i=1}^{N} X_i'\hat{M}_{\text{PC}}X_i\) can be expanded as follows:

\[
\frac{1}{NT} \sum_{i=1}^{N} X_i'\hat{M}_{\text{PC}}X_i = \frac{1}{NT} \sum_{i=1}^{N} X_i'\hat{M}_{\text{PC}}X_i - \frac{1}{NT} \sum_{i=1}^{N} X_i'(\hat{M}_{\text{PC}} - \hat{M}_{\text{PC}})X_i. \quad (A80)
\]

From (A75) and Lemmas PC7 and PC8, we obtain

\[
\| T^{-1}X_i'(\hat{M}_{\text{PC}} - \hat{M}_{\text{PC}})X_i \|
\]
\[ \leq N^{-2} \| (NT)^{-1}X_i'D_{\text{PC}} \| \| (T^{-1}(\hat{\epsilon}_{\text{PC}})'\hat{\epsilon}_{\text{PC}})^{-1} \|
\]
\[ + 2N^{-2(\alpha+1)} \| (NT)^{-1}X_i'D_{\text{PC}} \| \| (T^{-1}(\hat{\epsilon}_{\text{PC}})'\hat{\epsilon}_{\text{PC}})^{-1} \| \| \hat{\epsilon}_{\text{PC}} \| \| T^{-1}\hat{\epsilon}_{\text{PC}} \|
\]
\[ + N^{-4\alpha} \| (NT)^{-1}X_i'D_{\text{PC}} \| \| \hat{\epsilon}_{\text{PC}} \| \| T^{-1}\hat{\epsilon}_{\text{PC}} \| \| (T^{-1}(\hat{\epsilon}_{\text{PC}})'\hat{\epsilon}_{\text{PC}})^{-1} \| \| (\hat{\epsilon}_{\text{PC}})'\hat{\epsilon}_{\text{PC}} \|
\]
\[ = N^{-2}[O_p(N^{1/2-\alpha}T^{-1/2}R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2})]2O_p(1)
\]
\[ + N^{-2}\alpha[O_p(N^{1/2-\alpha}T^{-1/2}R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2})] \]
\[ \times O_p(1)[O_p(N^{-\alpha}) + O_p(T^{-1/2})]
\]
\[ + N^{-2\alpha}[O_p(N^{-\alpha}) + O_p(T^{-1/2})]2O_p(N^{4\alpha-1-2}T^{-1/2}R_{10})
\]
\[ = O_p(N^{2\alpha-1}) + O_p(N^{-\alpha-2}) + O_p(N^{-(\alpha+3)/2}T^{-1/2}) + O_p(N^{-(\alpha-1)T^{-1/2}} + O_p(N^{-1}T^{-1})
\]
\[ + O_p(N^{-2\alpha T^{-1}}) + O_p(N^{-\alpha} T^{-3/2}) + O_p(T^{-2}), \quad (A81)
\]
which does not impose any restrictions on the admissible values of $\alpha$ and $\kappa$. Hence, by using the fact that $||((NT)^{-1} \sum_{i=1}^{N} X_{i}^{i} M_{FCE} X_{i})|| \leq N^{-1} \sum_{i=1}^{N} ||T^{-1} X_{i}^{i} M_{FCE} X_{i}||$, we can proceed analogously to Proof of Lemma CCE9 to obtain

$$\frac{1}{NT} \sum_{i=1}^{N} X_{i}^{i} M_{FCE} X_{i} = \Sigma_{\epsilon} + o_{p}(1),$$

(A82)

which holds irrespectively of the values of $\alpha$ and $\kappa$.

**Lemma CCE10.** Under Assumptions ERR, LAM, RK–CCE, and KAP,

$$- T^{-1} X_{i}^{i} M_{FCE} D^{CCE} C^{-1} \lambda_{i} = N^{-1} (b_{1CCE, i} - b_{2CCE, i}) + O_{p} (R_{8}),$$

$$- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_{i}^{i} M_{FCE} D^{CCE} C^{-1} \lambda_{i} = \sqrt{TN}^{-1/2} (\bar{b}_{1CCE} - \bar{b}_{2CCE}) + \sqrt{NT} O_{p} (R_{8}),$$

where

$$b_{1CCE, i} = \Lambda_{i}^{-1} (\overline{C}_{i}^{0})^{-1} \sum_{i} (\overline{C}_{i}^{0})^{-1} \lambda_{i}^{0},$$

$$b_{2CCE, i} = \sum_{\epsilon, i}(\beta, I_{\epsilon}, m) (\overline{C}_{i}^{0})^{-1} \lambda_{i}^{0},$$

$$\bar{b}_{1CCE} = \frac{1}{N} \sum_{i=1}^{N} b_{1CCE, i},$$

$$\bar{b}_{2CCE} = \frac{1}{N} \sum_{i=1}^{N} b_{2CCE, i},$$

$$R_{8} = N^{2\alpha -2} + N^{2\alpha -(3+\kappa)/2} + N^{\alpha -(1+\kappa)/2} + N^{-(1+\kappa)}.$$

**Proof of Lemma CCE10.** Clearly, $M_{FCE} D^{CCE} C^{-1} = M_{FCE} (\hat{F}^{CE} - \hat{F}^{C}) C^{-1} = -M_{FCE} F$, and therefore

$$- T^{-1} X_{i}^{i} M_{FCE} D^{CCE} C^{-1} \lambda_{i} = T^{-1} X_{i}^{i} M_{FCE} F \lambda_{i}$$

$$= T^{-1} \Lambda_{i} F M_{FCE} F \lambda_{i} + T^{-1} E_{i}^{i} M_{FCE} F \lambda_{i}$$

$$= k_{1} + k_{2}.$$ (A83)

Consider $k_{1}$. We can write

$$\overline{C} F M_{FCE} F \overline{C} = D^{CCE} M_{FCE} D^{CCE} = D^{CCE} M_{FCE} D^{CCE} - D^{CCE} (M_{FCE} - M_{FCE}) D^{CCE},$$

from which we obtain

$$k_{1} = T^{-1} \Lambda_{i} F M_{FCE} F \lambda_{i}$$

$$= T^{-1} \Lambda_{i} (\overline{C}^{-1} F M_{FCE} F \overline{C}^{-1} \lambda_{i})$$

$$= T^{-1} \Lambda_{i} (\overline{C}^{-1} D^{CCE} M_{FCE} D^{CCE} \lambda_{i}) + T^{-1} \Lambda_{i} (\overline{C}^{-1} D^{CCE} (M_{FCE} - M_{FCE}) D^{CCE} \lambda_{i})$$

$$= k_{11} + k_{12}. (A84)$$

First, consider $k_{12}$. The decomposition in (A75) suggests that

$$D^{CCE} (M_{FCE} - M_{FCE}) D^{CCE}$$

$$= D^{CCE} D^{CCE} ((\hat{F}^{CE})^{i} \hat{F}^{CE})^{-1} D^{CCE} D^{CCE} + D^{CCE} D^{CCE} ((\hat{F}^{CE'})^{i} \hat{F}^{CE'})^{-1} \hat{F}^{CE'} D^{CCE}$$

$$+ D^{CCE} D^{CCE} ((\hat{F}^{CE})^{i} \hat{F}^{CE})^{-1} D^{CCE} D^{CCE}$$

$$+ D^{CCE} D^{CCE} ((\hat{F}^{CE'})^{i} \hat{F}^{CE'})^{-1} \hat{F}^{CE'} D^{CCE}. (A85)$$
Application of Lemmas CCE3, CCE4, and CCE7 to this expression yields

\[ \| T^{-1} D_{CCE} (M_F^{CCE} - M_{FCE}) D_{CCE} \| \]
\[ \leq N^{-2} \| NT^{-1} D_{CCE} D_{CCE} \| ^2 \| (T^{-1} (\hat{F}_{CCE})' \hat{F}_{CCE})^{-1} \| 
\]
\[ + 2N^{-2} (\alpha + 3/2) T^{-1/2} ||C||^2 \| NT^{-1} D_{CCE} D_{CCE} \| \| (T^{-1} (\hat{F}_{CCE})' \hat{F}_{CCE})^{-1} \| \| \sqrt{NT^{-1/2}} F D_{CCE} \| 
\]
\[ + N^{-2} (\alpha + 1) T^{-1} ||C||^2 \| \sqrt{NT^{-1/2}} D_{CCE} F \| \| (T^{-1} (\hat{F}_{CCE})' \hat{F}_{CCE})^{-1} - (C' F' F C)^{-1} \| 
\]
\[ = N^{-2} O_p (N^{2\alpha}) + N^{-2} (\alpha + 3/2) T^{-1/2} O_p (N^{2\alpha}) 
\]
\[ + N^{-2} (\alpha + 1) T^{-1} [O_p (N^{4\alpha - 1}) + O_p (N^{3\alpha - 1} T^{-1/2})] = O_p (R_7), \]

where

\[ R_7 = N^{2\alpha - 2} + N^{\alpha - 3/2} T^{-1/2}. \]

This result can in turn be used to show that

\[ \| k_{12} \| = \| T^{-1} A_i (\hat{C})' D_{CCE} (M_F^{CCE} - M_{FCE}) D_{CCE} C^{-1} \lambda_i \| \]
\[ \leq \| A_i^0 \| \| (\hat{C})^{-1} \| ^2 \| T^{-1} D_{CCE} (M_F^{CCE} - M_{FCE}) D_{CCE} \| \| A_i^0 \| \| \lambda_i^0 \| \]
\[ = O_p (R_7), \] (A86)

Let us now consider \( k_{11} \). Write \( M_F^{CCE} = I_T - P_F^{CCE} \), such that \( D_{CCE} M_F^{CCE} D_{CCE} = D_{CCE} D_{CCE} - D_{CCE} P_F^{CCE} D_{CCE} \). Here, given Lemma CCE3 and the fact that \( F \) and \( F ^{CCE} \) span the same vector space,

\[ \| T^{-1} A_i (\hat{C})' D_{CCE} P_F^{CCE} D_{CCE} C^{-1} \lambda_i \| \]
\[ = \| T^{-1} A_i (\hat{C})' D_{CCE} P_F^{CCE} (\hat{F}^{CCE} F C')^{-1} C' F' D_{CCE} C^{-1} \lambda_i \| \]
\[ = \| T^{-1} A_i (\hat{C})' D_{CCE} F (F' F)^{-1} F' D_{CCE} C^{-1} \lambda_i \| \]
\[ \leq (NT)^{-1} \| A_i^0 \| \| (\hat{C})^{-1} \| ^2 \| \sqrt{NT}^{-1/2} D_{CCE} F \| \| (T^{-1} F' F)^{-1} \| \| \lambda_i^0 \| \]
\[ = O_p (N^{-1} T^{-1}), \]

which implies

\[ k_{11} = T^{-1} A_i (\hat{C})' D_{CCE} M_F^{CCE} D_{CCE} C^{-1} \lambda_i \]
\[ = T^{-1} A_i (\hat{C})' D_{CCE} D_{CCE} C^{-1} \lambda_i + O_p (N^{-1} T^{-1}). \] (A87)

By combining (A86) and (A87),

\[ k_1 = k_{11} + k_{12} \]
\[ = N^{-1} A_i^0 [(\hat{C})^{-1}]' (NT^{-1} D_{CCE} D_{CCE}) (\hat{C})^{-1} \lambda_i^0 + O_p (R_7) + O_p (N^{-1} T^{-1}). \]

Application of Lemma CCE4 now yields

\[ k_1 = \sqrt{T} N^{-1} b_{1CCE,i} + O_p (R_7) + O_p (N^{-1} T^{-1}), \] (A88)

where \( b_{1CCE,i} = A_i^0 [(\hat{C})^{-1}]' \Sigma_i (\hat{C})^{-1} \lambda_i^0 \).

Next, consider \( k_2 \). By using \( M_F^{CCE} F \lambda_i = M_F^{CCE} F C^{-1} \lambda_i = 0 \), and then substitution for \( M_F^{CCE} - M_{FCE} \),

\[ k_2 = T^{-1} E_i M_F^{CCE} F \lambda_i \]
\[ = - T^{-1} E_i (M_F^{CCE} - M_{FCE}) F \lambda_i \]
\[ = - T^{-1} E_i D_{CCE} ((\hat{F}_{CCE})' \hat{F}_{CCE})^{-1} D_{CCE} F \lambda_i - T^{-1} E_i D_{CCE} ((\hat{F}_{CCE})' \hat{F}_{CCE})^{-1} C' F' \lambda_i \]
\[ - T^{-1} E_i (\hat{F}^{CCE})' (\hat{F}^{CCE})^{-1} D_{CCE} F \lambda_i - T^{-1} E_i (\hat{F}^{CCE})' (\hat{F}^{CCE})^{-1} C' F' \lambda_i \]
\[ = - k_{21} - \cdots - k_{24}. \]
Consider \( k_{21}, k_{23}, \) and \( k_{24} \). Here we may make use of Lemmas CCE3, CCE5, and CCE7 to show that

\[
||k_{21}|| = ||T^{-1}E_iD^{CCE}((F^{CCE})^\dagger F^{CCE})^{-1}D^{CCE}F\lambda_i|| \\
\leq N^{-(\alpha+3/2)}T^{-1/2}||NT^{-1/2}E_iD^{CCE}||((T^{-1}(F^{CCE})^\dagger F^{CCE})^{-1}||||\sqrt{N}T^{-1/2}D^{CCE}F||||\lambda_i^0|| \\
= N^{-(\alpha+3/2)}T^{-1/2}[O_p(1) + O_p(\sqrt{NT^{-1/2}})]O_p(N^{2\alpha}) \\
= O_p(N^{\alpha-1}T^{-1}) + O_p(N^{\alpha-3/2}T^{-1/2}),
\]

(A89)

\[
||k_{23}|| = ||T^{-1}E_i\left(\frac{CCE}{F}\right)^\dagger F^{CCE})^{-1}D^{CCE}F\lambda_i|| \\
\leq N^{-2\alpha-1/2}T^{-1/2}||T^{-1/2}E_iF||||\left(\frac{CCE}{F}\right)^\dagger ((T^{-1}(F^{CCE})^\dagger F^{CCE})^{-1}||||\sqrt{N}T^{-1/2}D^{CCE}F||||\lambda_i^0|| \\
= O_p(N^{\alpha-1}T^{-1}) + O_p(N^{\alpha-3/2}T^{-1/2}),
\]

(A90)

and

\[
||k_{24}|| = ||T^{-1}E_i\left(\frac{CCE}{F}\right)^\dagger F^{CCE})^{-1}D^{CCE}F\lambda_i|| \\
\leq N^{-3\alpha}T^{-1/2}||T^{-1/2}E_iF||||\left(\frac{CCE}{F}\right)^\dagger ((T^{-1}(F^{CCE})^\dagger F^{CCE})^{-1}||||\sqrt{N}T^{-1/2}D^{CCE}F||||\lambda_i^0|| \\
= O_p(N^{\alpha-1}T^{-1}) + O_p(N^{\alpha-3/2}T^{-1/2}).
\]

(A91)

\( k_{22} \) can be expanded in the following obvious fashion:

\[
k_{22} = T^{-1}E_iD^{CCE}((F^{CCE})^\dagger F^{CCE})^{-1}\left(\frac{CCE}{F}\right)^\dagger F\lambda_i \\
= T^{-1}E_i\left(\frac{CCE}{F}\right)^\dagger C^{\dagger}F^{-1}C^{\dagger}F^{\dagger}F\lambda_i + T^{-1}E_iD^{CCE}((F^{CCE})^\dagger F^{CCE})^{-1} - \left(\frac{CCE}{F}\right)^\dagger F\lambda_i,
\]

where, via Lemmas CCE5 and CCE7,

\[
||T^{-1}E_i\left(\frac{CCE}{F}\right)^\dagger F^{CCE})^{-1} - \left(\frac{CCE}{F}\right)^\dagger F\lambda_i|| \\
\leq N^{-(2\alpha+1)}||NT^{-1}E_iD^{CCE}||||((F^{CCE})^\dagger F^{CCE})^{-1} - \left(\frac{CCE}{F}\right)^\dagger F\lambda_i||||\sqrt{N}T^{-1/2}D^{CCE}F||||\lambda_i^0|| \\
= O_p(N^{\alpha-1}T^{-1} + O_p(N^{\alpha-3/2}T^{-1/2} + O_p(N^{\alpha-1}T^{-1})),
\]

Moreover,

\[
T^{-1}E_iD^{CCE}C^{\dagger}F^{\dagger}F\lambda_i = N^{-1}(NT^{-1})E_iD^{CCE}C^{\dagger}F^{\dagger}F\lambda_i = N^{-1}\Sigma_{\epsilon,i}(\beta, I_m)(C^{\dagger})^{-1}\lambda_i^0 + O_p(N^{-1}T^{-1/2}).
\]

Hence,

\[
k_{22} = N^{-1}b_{2CCE,i} + O_p(N^{2\alpha-2}) + O_p(N^{2\alpha-3/2}T^{-1/2}) + O_p(N^{\alpha-1}T^{-1} + O_p(N^{\alpha-1}T^{-1}).
\]

(A92)

where \( b_{2CCE,i} = \Sigma_{\epsilon,i}(\beta, I_m)(C^{\dagger})^{-1}\lambda_i^0 \). Taking together (A89)–(A92), it is clear that

\[
k_2 = -k_{21} - \cdots - k_{24} \\
= -N^{-1}b_{2CCE,i} + O_p(N^{2\alpha-2}) + O_p(N^{2\alpha-3/2}T^{-1/2}) + O_p(N^{\alpha-1}T^{-1}) \\
+ O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}).
\]

(A93)

By adding (A88) and (A93),

\[
T^{-1}X_i\hat{M}_{\epsilon CCE}D^{CCE}C^{\dagger}F^{\dagger}F\lambda_i = k_1 + k_2 = N^{-1}(b_{1CCE,i} - b_{2CCE,i}) + O_p(R_8),
\]

(A94)

where, under \( T = N^K \),

\[
R_8 = N^{2\alpha-2} + N^{2\alpha-3/2}T^{-1/2} + N^{\alpha-1}T^{-1/2} + N^{-1/2}T^{-1} \\
= N^{2\alpha-2} + N^{2\alpha-(3+\kappa)/2} + N^{\alpha-(1+\kappa)/2} + N^{-(1/2+\kappa)}.
\]

120
The second result follows directly from the first:

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i^\prime M_{\text{CCE}} D_{\text{CCE}}^{-1} \lambda_i = \sqrt{NT} \frac{1}{N} \sum_{i=1}^{N} T^{-1} X_i^\prime M_{\text{CCE}} D_{\text{CCE}}^{-1} \lambda_i \]

\[ = \sqrt{T N^{-1/2}} (\bar{b}_{1\text{CCE}} - \bar{b}_{2\text{CCE}}) + \sqrt{NT} \alpha_p(R_8), \]

where \( \bar{b}_{1\text{CCE}} \) and \( \bar{b}_{2\text{CCE}} \) are simply the average \( b_{1\text{CCE},i} \) and \( b_{2\text{CCE},i} \), respectively.

**Lemma PC10.** Under Assumptions ERR, LAM, RK–PC, and KAP,

\[ - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i^\prime M_{\text{PC}} D_{\text{PC}} H^{-1} \lambda_i = \sqrt{T N^{-1/2}} (\bar{b}_{1\text{PC}} - \bar{b}_{2\text{PC}}) + O_p(R_{12}), \quad (A95) \]

where

\[ \bar{b}_{1\text{PC}} = \frac{1}{N} \sum_{i=1}^{N} A_i^\prime (Q^0)_{\text{PC}}^{-1} S^0 (Q^0)^{-1} \lambda_i^0, \]

\[ \bar{b}_{2\text{PC}} = \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon_i}(\beta, I_m) C^0_i (Q^0)^{-1} \lambda_i^0, \]

\[ R_{12} = (N^{2\alpha+1/2} R_{11} + N^{2\alpha-1/2} R_4 + N^{\alpha-1/2} R_5) \sqrt{T} + N^{-3/2} R_{10} \]

\[ + (N^{-3\alpha} R_1 + N^{-(\alpha+1/2)} R_1 + N^{-\alpha} R_{10} + N^{-2\alpha-1/2} R_2) T^{-1/2} + N^{-\alpha} R_1 T^{-3/2}, \]

\[ R_{11} = N^{-4} + (N^{-5\alpha+3/2} + N^{-2\alpha+2}) T^{-1/2} + (N^{-3} + N^{-2\alpha+3/2} + N^{-(4\alpha+1)}) T^{-1} \]

\[ + (N^{-5\alpha+1/2} + N^{-2\alpha+1}) T^{-3/2} + (N^{-2} + N^{-4\alpha} + N^{-2\alpha+1/2}) T^{-2} \]

\[ + N^{-2\alpha} T^{-5/2} + N^{-1} T^{-3} + T^{-4}. \]

**Proof of Lemma PC10.** Analogous to Proof of Lemma CCE10, we can write

\[ - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i^\prime M_{\text{PC}} D_{\text{PC}} H^{-1} \lambda_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} A_i^\prime F_{\text{PC}} M_{\text{PC}} F_{\lambda_i} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^\prime M_{\text{PC}} F_{\lambda_i} \]

\[ = k_1 + k_2, \quad (A96) \]

where

\[ k_1 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} A_i^\prime (H^{-1})^\prime D_{\text{PC}} (M_{\text{PC}} - M_{\text{PC}}) D_{\text{PC}} H^{-1} \lambda_i \]

\[ + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} A_i^\prime (H^{-1})^\prime D_{\text{PC}} H^2 D_{\text{PC}} H^{-1} \lambda_i \]

\[ = k_{11} + k_{12}, \quad (A97) \]

and

\[ k_2 = - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^\prime D_{\text{PC}} (\hat{F}_{\text{PC}} - \hat{F}_{\text{PC}}) D_{\text{PC}} F_{\lambda_i} - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^\prime D_{\text{PC}} (\hat{F}_{\text{PC}} - \hat{F}_{\text{PC}}) H^2 F_{\lambda_i} \]

\[ - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^\prime F_{\lambda_i} \]

121
\[-\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E^i F H (((\hat{F}^\text{PC})' \hat{F}^\text{PC})^{-1} - (H' F^\text{PC} H)^{-1}) \hat{H} F^\text{PC} \lambda_i \]

\[= -k_{21} - \cdots - k_{24}.\]

Consider \(k_{12}\). Application of Lemmas PC2 and PC3 yields

\[
||\sqrt{N T^{-1/2}} D^\text{PC} (M F^\text{PC} - M_{\text{PC}}) D^\text{PC}||
\leq \sqrt{N T} ||T^{-1/2} D^\text{PC} D^\text{PC}|| + ||\sqrt{N T^{-1/2}} D^\text{PC} F|| ||H||
= \sqrt{N T} [O_p(T^{-1}) + O_p(N^{-1})] + O_p(R_1) O_p(N^{-2\alpha}) = O_p(R_{10}),
\]

where

\[R_{10} = \sqrt{N T^{-1/2}} + N^{-1/2} \sqrt{T} + N^{-2\alpha} R_1 = N^{-1/2} \sqrt{T} + N^{-5\alpha} + \sqrt{T N^{-1/2}},\]

and therefore

\[
||T^{-1} D^\text{PC} (M F^\text{PC} - M_{\text{PC}}) D^\text{PC}||
\leq N^{-\alpha} ||T^{-1} D^\text{PC} D^\text{PC}||^2 ||(T^{-1} (\hat{F}^\text{PC})' \hat{F}^\text{PC})^{-1}||
+ 2 N^{-\alpha} ||T^{-1/2} ||H|| ||T^{-1} D^\text{PC} D^\text{PC}|| ||(T^{-1} (\hat{F}^\text{PC})' \hat{F}^\text{PC})^{-1}|| ||\sqrt{N T^{-1/2}} F D^\text{PC}||
+ N^{-\alpha} ||\hat{H}||^2 ||\sqrt{N T^{-1/2}} D^\text{PC} F||^2 T ||(\hat{F}^\text{PC})' \hat{F}^\text{PC}|| - (\hat{H} F^\text{PC} \hat{H})||
= N^{-\alpha} [O_p(N^{-1}) + O_p(T^{-1})]^2
+ N^{-2(2\alpha+1/2)} T^{-1/2} [O_p(N^{-1}) + O_p(T^{-1})] O_p(R_1)
+ N^{-4\alpha} T^{-1} O_p(R_1^2) O_p(N^{4\alpha - 1/2} T^{-1/2} R_{10}).
\]

The order of the first term on the right is \(N^{-2}(N^{-1} + T^{-1})^2 \leq N^{-4} + (NT)^{-2}\), while the order of the second is

\[N^{-2(2\alpha+1/2)} T^{-1/2} (N^{-1} + T^{-1}) R_1
= (N^{-5\alpha+3/2} + N^{-2(2\alpha+3/2)} T^{-1/2} + N^{-4(2\alpha+1/2)} T^{-1})
+ (N^{-2(2\alpha+1)} + N^{-5(2\alpha+1/2)} T^{-3/2} + N^{-4\alpha} + N^{-2(2\alpha+1/2)} T^{-2} + N^{-2\alpha} T^{-5/2}).\]

The order of the third term can be written as follows, via the definition of \(R_{10}\):

\[N^{-3/2} T^{-3/2} R_1^2 R_{10} = N^{-3/2} T^{-3/2} R_1^2 [\sqrt{NT^{-1/2}} + \sqrt{TN^{-1/2}}] + N^{-2(2\alpha+3/2)} T^{-3/2} R_1^3,
\]

where, after considerable simplification,

\[N^{-3/2} T^{-3/2} R_1^2 [\sqrt{NT^{-1/2}} + \sqrt{TN^{-1/2}}]
\leq (N^{-3} + N^{-6\alpha+2}) T^{-1} + (N^{-2} + N^{-4\alpha+1}) T^{-2} + (N^{-4\alpha} + N^{-1}) T^{-3} + T^{-4},\]

and

\[N^{-2(2\alpha+3/2)} T^{-3/2} R_1^3 \leq (N^{-2(2\alpha+3)} + N^{-8\alpha}) T^{-3} + N^{-2\alpha} T^{-9/2},\]

implying

\[N^{-3/2} T^{-3/2} R_1^2 R_{10}
\leq (N^{-3} + N^{-6\alpha+2}) T^{-1} + N^{-11\alpha+3/2} T^{-3/2} + (N^{-4\alpha+1} + N^{-2}) T^{-2}
+ (N^{-4\alpha} + N^{-1}) T^{-3} + T^{-4}.
\]
Thus, letting
\[ R_{11} = N^{-4} + (N^{-5\alpha+3/2} + N^{-2(2\alpha+2)})T^{-1/2} + (N^{-3} + N^{-2(2\alpha+3/2) + N^{-2(4\alpha+1)})T^{-1} + (N^{-5\alpha+1/2} + N^{-2(2\alpha+1)})T^{-3/2} + (N^{-2} + N^{-4\alpha} + N^{-2(2\alpha+1/2)})T^{-2} + N^{-2\alpha}T^{-5/2} + N^{-1}T^{-3} + T^{-4}, \]
we have
\[ ||T^{-1}D^{PC}(M_{FH} - M_{F_{PC}})D^{PC}|| = O_p(R_{11}). \]  
(A100)

It follows that, with \((H^0)^{-1} = (H^0)(H^0)^{-1},\)
\[ ||k_{12}|| = \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Lambda_i(H^{-1})'D^{PC}(M_{FH} - M_{F_{PC}})D^{PC}H^{-1}\lambda_i \right| \]
\[ \leq N^{2\alpha+1/2}\sqrt{T} \left( \frac{1}{N} \sum_{i=1}^{N} ||\Lambda_i^0||||H^{-1}||^2||T^{-1}D^{PC}(M_{FH} - M_{F_{PC}})D^{PC}||||\lambda_i^0|| \right) \]
\[ = O_p(N^{2\alpha+1/2}\sqrt{T}R_{11}). \]  
(A101)

\(k_{11}\) can be rewritten as follows:
\[ k_{11} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Lambda_i(H^{-1})'D^{PC}D^{PC}H^{-1}\lambda_i \]
\[ - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Lambda_i(H^{-1})'D^{PC}P_{FH}D^{PC}H^{-1}\lambda_i \]  
(A102)

where, by Lemma PC3,
\[ \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Lambda_i(H^{-1})'D^{PC}P_{FH}(H^0)(H^0)^{-1}H^0F'D^{PC}H^{-1}\lambda_i \right| \]
\[ \leq \frac{1}{N^{1/2-2\alpha}\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} ||\Lambda_i^0||||H^0||^2||\sqrt{NT^{-1/2}D^{PC}F}||2||T^{-1/2}H^0F'(H^0)^{-1}||\lambda_i^0|| \right) \]
\[ = O_p(N^{2\alpha-1/2}T^{-1/2}R_{11}^2). \]

From (A97), (A101), and (A102), we therefore obtain
\[ k_1 = \sqrt{TN^{2\alpha-1/2}} \left( \frac{1}{N} \sum_{i=1}^{N} \Lambda_i^0(H^{-1})'\frac{1}{NT^{-1/2}}D^{PC}(H^0)^{-1}\lambda_i^0 \right) + O_p(N^{2\alpha+1/2}\sqrt{T}R_{11}) + O_p(N^{2\alpha-1/2}T^{-1/2}R_{11}^2), \]
and so, by further use of Lemma PC4,
\[ k_1 = \sqrt{TN^{-1/2}\tilde{b}_{1PC}} + O_p(\sqrt{TN^{2\alpha-1/2}R_3}) + O_p(N^{2\alpha+1/2}\sqrt{T}R_{11}) + O_p(N^{2\alpha-1/2}T^{-1/2}R_{11}^2), \]  
(A103)

where \(\tilde{b}_{1PC} = N^{-1} \sum_{i=1}^{N} \Lambda_i^0(Q^{-1})^{-1}S(Q^{-1})^{-1}\lambda_i^0. \)
Next, consider $k_2$. Making use of Lemmas PC3, PC5, and PC7, it is possible to show that

$$\|k_{21}\| = \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T D^{PC} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} D^{PC} F \lambda_i \right\|$$

$$\leq N^{-(\alpha+1)} \frac{1}{N} \sum_{i=1}^{N} \|NT^{-1/2} E_i^T D^{PC} \| \| (T^{-1} (\hat{F}^{PC})' \hat{F}^{PC})^{-1} \| \sqrt{NT^{-1/2} D^{PC} F} \| \| \lambda_i^0 \|$$

$$= N^{-(\alpha+1)} \left[ O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2}) \right] \| \lambda_i^0 \|$$

$$= O_p(N^{-\alpha} T^{-1/2} R_1), \quad (A104)$$

and

$$\|k_{23}\| = \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T F H ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} D^{PC} F \lambda_i \right\|$$

$$\leq N^{-3\alpha} T^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \|T^{-1/2} E_i^T F \| \| \hat{H}^0 \| \| (T^{-1} (\hat{F}^{PC})' \hat{F}^{PC})^{-1} \| \sqrt{NT^{-1/2} D^{PC} F} \| \| \lambda_i^0 \|$$

$$= O_p(N^{-3\alpha} T^{-1/2} R_1), \quad (A105)$$

For $k_{22}$,

$$k_{22} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T D^{PC} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} \hat{H} \hat{F} F \lambda_i$$

$$= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T D^{PC} \hat{H} \lambda_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T D^{PC} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\hat{H} \hat{F} F \hat{H})^{-1} \hat{H} \hat{F} F \lambda_i,$$

where, by Lemmas PC5 and PC7,

$$\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^T D^{PC} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\hat{H} \hat{F} F \hat{H})^{-1} \hat{H} \hat{F} F \lambda_i \right\|$$

$$\leq \sqrt{T} N^{-(3\alpha+1/2)} \frac{1}{N} \sum_{i=1}^{N} \|NT^{-1} E_i^T D^{PC} \| \|T^{-1} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\hat{H} \hat{F} F \hat{H})^{-1} \| \| \lambda_i^0 \|$$

$$= \sqrt{T} N^{-(3\alpha+1/2)} \left[ O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2}) \right] \| \lambda_i^0 \|$$

$$= O_p(N^{\alpha/2-1} R_1) + O_p(N^{\alpha-1/2} T^{-1/2} R_1) + O_p(N^{\alpha} T^{-3/2} R_1).$$
Another application of Lemma PC5 yields
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i\mathbf{D}^{PC} \mathbf{H}^{-1} \lambda_i
= \sqrt{T}N^{-1/2} \frac{1}{T} \sum_{i=1}^{N} E_i\mathbf{D}^{PC} \mathbf{H}^{-1} \lambda_i
= \sqrt{T}N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \sum_{\epsilon,i} (\beta, I_m) C_i^{0'} (\mathbf{Q})^{-1} \lambda_i^0 + O_p(\sqrt{T}N^{\alpha-1/2}R_5),
\]
and so, with \( \mathbf{b}_{2PC} = N^{-1} \sum_{i=1}^{N} \sum_{\epsilon,i} (\beta, I_m) C_i^{0'} (\mathbf{Q})^{-1} \lambda_i^0 \),
\[
k_{22} = \sqrt{T}N^{-1/2} \mathbf{b}_{2PC} + O_p(\sqrt{T}N^{\alpha-1/2}R_5) + O_p(N^{\alpha-1/2}R_5R_{10}). \tag{A107}
\]
By adding (A104)–(A107), we obtain
\[
k_2 = \sqrt{T}N^{-1/2} \mathbf{b}_{2PC} + O_p(\sqrt{T}N^{\alpha-1/2}R_5) + O_p(N^{\alpha-1/2}R_5R_{10}) + O_p(N^{-3\alpha}T^{1/2}R_1)
+ O_p(N^{-3\alpha+1/2}R_1) + O_p(N^{-\alpha+1/2}T^{1/2}R_1) + O_p(N^{-\alpha}T^{-3/2}R_1)
+ O_p(N^{-\alpha}T^{-1/2}R_{10}). \tag{A108}
\]
The above results for \( k_1 \) and \( k_2 \) imply
\[
- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' \mathbf{M}_{PC} \mathbf{D}^{PC} \mathbf{H}^{-1} \lambda_i = k_1 + k_2 = \sqrt{T}N^{-1/2} (\mathbf{b}_{1PC} - \mathbf{b}_{2PC}) + O_p(R_{12}),
\]
where
\[
R_{12} = (N^{2\alpha+1/2}R_{11} + N^{2\alpha-1/2}R_4 + N^{\alpha-1/2}R_5) \sqrt{T} + N^{-(3\alpha+1/2)}R_1 + N^{\alpha-1/2}R_5R_{10}
+ (N^{-3\alpha+1/2}R_1 + N^{-\alpha+1/2}R_1 + N^{-\alpha}R_{10} + N^{2\alpha-1/2}R_5^2) T^{-1/2} + N^{-\alpha}T^{-3/2}.
\]

**Lemma CCE11.** Under Assumptions ERR, LAM, RK–CCE, and KAP,
\[
T^{-1}X_i' \mathbf{M}_{\bar{F}CCE} \nu_i = T^{-1} \mathbf{E}_i \nu_i - N^{-1} \mathbf{b}_{3CCE,i} + O_p(R_9) + O_p((NT)^{-1/2}),
\]
\[
- \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' \mathbf{M}_{\bar{F}CCE} \nu_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i' \nu_i - \sqrt{T}N^{-1/2} \mathbf{b}_{3CCE} + \sqrt{NT}O_p(R_9) + O_p(N^{-1/2}),
\]
where
\[
\mathbf{b}_{3CCE,i} = \sigma_{\nu,i}^2 \mathbf{A}_i^{0'} (\mathbf{C})^{-1} (1,0)',
\]
\[
\mathbf{b}_{3CCE} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{b}_{3CCE,i},
\]
\[
R_9 = N^{2\alpha-2} + N^{3\alpha-2-\kappa/2} + N^{2\alpha-3/2-\kappa/2} + N^{\alpha-1-\kappa/2}
+ N^{3\alpha-(3/2+\kappa)} + N^{2\alpha-(1+\kappa)} + N^{\alpha-(1/2+\kappa)}.
\]

**Proof of Lemma CCE11.** Let us rewrite \( T^{-1}X_i' \mathbf{M}_{\bar{F}CCE} \nu_i \) as
\[
T^{-1}X_i' \mathbf{M}_{\bar{F}CCE} \nu_i = T^{-1}X_i' \mathbf{M}_{\bar{F}CCE} \nu_i - T^{-1}X_i' (\mathbf{M}_{\bar{F}CCE} - \mathbf{M}_{\bar{F}CCE}) \nu_i, \tag{A109}
\]
where, via (A75),

\[
T^{-1}X_i^\prime(M_{FC} - M_{FCCE})v_i = T^{-1}X_i^\prime D^{CCE}(\hat{F}^{CCE}D^{CCE} - 1)D^{CCE}v_i + T^{-1}X_i^\prime D^{CCE}((\hat{F}^{CCE}D^{CCE} - 1)^{-T}(\hat{F}^{CCE}D^{CCE} - 1)D^{CCE})v_i + T^{-1}X_i^\prime F(C^{CCE}D^{CCE} - 1)D^{CCE}v_i + T^{-1}X_i^\prime F[(D^{CCE}D^{CCE} - 1)^{-1}(C^{CCE}F^{CCE})^{-1}]C^{F}\nu_i
\]

\[= I_1 + \cdots + I_4.\]  

(A115)

The order of \(I_1, \ldots, I_4\) can be obtained by using the same steps as when analyzing \(k_2\). Lemmas CCE6–CCE8 imply

\[||I_1|| = ||T^{-1}X_i^\prime D^{CCE}(\hat{F}^{CCE}D^{CCE} - 1)D^{CCE}v_i|| \leq N^{-2}||NT^{-1}X_i^\prime D^{CCE}|| |(T^{-1}(\hat{F}^{CCE}D^{CCE} - 1)||NT^{-1}D^{CCE}v_i|| = N^{-2}[O_p(1) + O_p(\sqrt{NT^{-1/2}})]^2 O_p(N^{2\alpha}) = O_p(N^{2\alpha - 2}) + O_p(N^{2\alpha - 2/3}T^{-1/2}) + O_p(N^{2\alpha - 2}T^{-1}) \]

(A111)

\[||I_2|| = ||T^{-1}X_i^\prime D^{CCE}(\hat{F}^{CCE}D^{CCE} - 1)(\hat{C}^{F}\nu_i)|| \leq N^{-(\alpha + 1)}T^{-1/2}||NT^{-1}X_i^\prime D^{CCE}|| |(T^{-1}(\hat{F}^{CCE}D^{CCE} - 1)||N^{-1/2}\hat{C}^{F}\nu_i|| = O_p(N^{\alpha - 1}T^{-1/2}) + O_p(N^{\alpha - 1/2}T^{-1}) \]

(A112)

and, additionally using (A76),

\[||I_4|| = ||T^{-1}X_i^\prime F(\hat{C}^{CCE}D^{CCE} - 1)D^{CCE}v_i|| \leq N^{-2\alpha}T^{-1/2}||T^{-1}X_i^\prime F|| |(\hat{C}^{CCE}D^{CCE} - 1)||N^{-1/2}C^{F}v_i|| = O_p(N^{\alpha - 1}T^{-1/2}) + O_p(N^{\alpha - 1/2}T^{-1}) + O_p(N^{\alpha - 1/2}T^{-3/2}) \]

(A113)

\[I_3\] can be decomposed as follows:

\[I_3 = T^{-1}X_i^\prime F(C^{CCE}D^{CCE} - 1)(\hat{C}^{F}D^{CCE} - 1)D^{CCE}v_i + T^{-1}X_i^\prime F(C^{CCE}D^{CCE} - 1)(\hat{C}^{CCE}D^{CCE} - 1)D^{CCE}v_i.\]

(A114)

Via Lemma CCE7, we obtain

\[||T^{-1}X_i^\prime F(C^{CCE}D^{CCE} - 1)(\hat{C}^{F}D^{CCE} - 1)D^{CCE}v_i|| \leq N^{-(\alpha + 1)}||T^{-1}X_i^\prime F|| |(\hat{C}^{CCE}D^{CCE} - 1)||N^{-1}T^{-1}D^{CCE}v_i|| = O_p(N^{\alpha - 1}T^{-1/2}) + O_p(N^{3\alpha - 1/2}T^{-1/2}) + O_p(N^{3\alpha - 1/2}T^{-3/2}) + O_p(N^{3\alpha - 1/2}T^{-3/2}).\]

Note also that

\[||T^{-1}E_f^\prime FC(\hat{C}^{CCE}D^{CCE} - 1)D^{CCE}v_i|| = ||T^{-1}E_f^\prime FC(\hat{C}^{CCE}D^{CCE} - 1)D^{CCE}v_i||\]
\[ \leq N^{\alpha-1}T^{-1/2} \left\| (T^{-1/2}E_iF) (T^{-1/2}E_iF)^\dagger \right\| \left\| (T^{-1/2}F'E) (T^{-1/2}F'E)^\dagger \right\| \]  
[\|NT^{-1}D^{CCE}\nu_i\|]

which, together with Lemma CCE6, implies that the first term in (A114) can be written as

\[ T^{-1}X^\prime_i (M_{FC} - M_{FCCE})^\prime \nu_i \]

\[ = T^{-1}A_i F F (C' F' F C) \nu_i + T^{-1}E_i F F (C' F' F C) \nu_i + O_p(N^{\alpha-1}T^{-1/2}) + O_p(N^{\alpha-1/2}T^{-1}) \]

\[ = N^{-1}A_i^0 ((C_i)^0)^\prime NT^{-1}D^{CCE}\nu_i + O_p(N^{\alpha-1}T^{-1/2}) + O_p(N^{\alpha-1/2}T^{-1}) \]

\[ = N^{-1} \sigma_{ij}^2 A_i^0 ((C_i)^0)^\prime (1, 0)' + O_p((NT)^{-1/2}) + O_p(N^{\alpha-1}T^{-1/2}) \]

\[ + O_p(N^{\alpha-1/2}T^{-1}) \]

where the third equality follows from \((C_i)^0)^\prime F F (C_i)^0)^\prime C_i \nu_i \]

\[ = \frac{N^{-1} \sigma_{ij}^2 A_i^0 ((C_i)^0)^\prime (1, 0)'}{O_p((NT)^{-1/2})} \]

The results in (A111)–(A113) and (A117) imply

\[ T^{-1}X_i^\prime (M_{FC} - M_{FCCE})^\prime \nu_i = I_i + \cdots + I_4 = N^{-1}b_{3CCE,i} + O_p((NT)^{-1/2}) + O_p(R_0) \]

where

\[ R_0 = N^{3\alpha-2} + N^{\alpha-1}T^{-1/2} + N^{3\alpha-2}T^{-1/2} + N^{\alpha-1}T^{-1/2} + N^{3\alpha-2} - ^{-1/2} + N^{\alpha-1/2}T^{-1} \]

\[ = N^{3\alpha-2} + N^{\alpha-1}T^{-1/2} + N^{3\alpha-2} + N^{\alpha-1} - 1/k + N^{3\alpha-2} + N^{\alpha-1} - 1/k \]

\[ + N^{3\alpha-2} + N^{\alpha-1} - 1/k + N^{3\alpha-2} + N^{\alpha-1} - 1/k \]

\[ \leq N^{3\alpha-2} + N^{\alpha-1}T^{-1} + N^{3\alpha-2} + N^{\alpha-1}T^{-1} \]

Equation (A109) can be rewritten as

\[ T^{-1}X_i^\prime M_{FC}^\prime \nu_i = T^{-1}X_i^\prime (M_{FC} - M_{FCCE})^\prime \nu_i \]

\[ = T^{-1}X_i^\prime M_{FC}^\prime \nu_i - N^{-1}b_{3CCE,i} + O_p(R_0) \]

\[ = T^{-1}E_i^\prime \nu_i - T^{-1}E_i^\prime P_{FC}^\prime \nu_i - N^{-1}b_{3CCE,i} + O_p(R_0) \]

where the last step arises from \(M_{FC}X_i = M_{FC}E_i \). Consider \(T^{-1}X_i^\prime P_{FC}^\prime \nu_i \). From \(E(\nu_i^\prime \nu_i) = \sigma_{ij}^2 I_T \) and \(P_{FC}^\prime P_{FC} = P_{FC} \)

\[ E[(T^{-1}E_i^\prime P_{FC}^\prime \nu_i)^\prime (T^{-1}E_i^\prime P_{FC}^\prime \nu_i)'] \]

\[ = T^{-2}E[E_i^\prime P_{FC}E(\nu_i^\prime \nu_i)E_i, F_{FC}]P_{FC}E_i] = T^{-2} \sigma_{ij}^2 E(E_i^\prime P_{FC}E_i) \]

\[ = T^{-2} \sigma_{ij}^2 E[E_i^\prime F (C' F' F C) - 1 \nu_i E_i] \]

where \(E_i^\prime F (C' F' F C) - 1 \nu_i E_i \)
$$= T^{-2} \sigma_{\epsilon, i}^2 \sum_{t=1}^{T} \sum_{i=1}^{T} E[E(\epsilon_{i,t}^t|FC) f_t^i \overline{C(C'F'FC)^{-1}C'}_t]$$

$$= T^{-2} \sigma_{\epsilon, i}^2 \sum_{t=1}^{T} \sum_{i=1}^{T} E[f_t^i \overline{C(C'F'FC)^{-1}C'}_t] = (m + 1)T^{-2} \sigma_{\epsilon, i}^2 = O_p(T^{-2}),$$

where last equality holds because

$$\sum_{t=1}^{T} f_t^i \overline{C(C'F'FC)^{-1}C'}_t = \sum_{t=1}^{T} \text{tr}(f_t^i \overline{C(C'F'FC)^{-1}C'}_t)$$

$$= \text{tr}\left(\sum_{t=1}^{T} (C_0^t)^\prime f_t^i C_0^t ((C_0^t)^\prime F'FC_0^t)^{-1}\right) = \text{tr}(I_{m+1}) = m + 1.$$ 

Thus, since the variance is $O_p(T^{-2})$, we have

$$||T^{-1}E[P_{FC}v_i|| = O_p(T^{-1}). \quad (A121)$$

Taking together the results in (A120) and (A121), we obtain

$$T^{-1}X'_i(M_{FCCE} - M_{FCCE})v_i = T^{-1}E_i^t v_i - N^{-1}b_{3CCE,i} + O_p(R_9) + O_p(T^{-1}). \quad (A122)$$

As for the second result of the lemma, note first that the equivalent to (A116) is

$$(NT)^{-1/2} \sum_{i=1}^{N} X'_i M_{FC} (C'F'FC)^{-1}D_{CCE} v_i$$

$$= \sqrt{TN^{-1/2}} \sum_{i=1}^{N} \Lambda_0^0 (\overline{C_0})^{-1} NT^{-1}D_{CCE} v_i + (NT)^{-1/2} \sum_{i=1}^{N} E[f_t^i \overline{C(C'F'FC)^{-1}D_{CCE} v_i}$$

$$= \sqrt{TN^{-1/2}} \sum_{i=1}^{N} \sigma_{\epsilon, i}^2 (\overline{C_0}) + O_p(N^{-1/2}) + \sqrt{NT}(O_p(N^{1/2}T^{-1/2}) + O_p(N^{1/2}T^{-1/2})), \quad (A123)$$

implying that

$$(NT)^{-1/2} \sum_{i=1}^{N} X'_i (M_{FC} - M_{FCCE})v_i = \sqrt{TN^{-1/2}}b_{3CCE} + O_p(N^{-1/2}) + \sqrt{NT}O_p(R_9). \quad (A124)$$

Furthermore,

$$E\left[\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^t P_{FC} v_i\right)\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^t P_{FC} v_i\right)^\prime\right]$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \sigma_{\epsilon, i}^2 \sum_{t=1}^{T} E[f_t^i \overline{C(C'F'FC)^{-1}C'}_t] = O_p(T^{-1})$$

Hence, $(NT)^{-1/2} \sum_{i=1}^{N} E_i^t P_{FC} v_i = O_p(T^{-1/2})$, and we can therefore conclude that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{FCCE} v_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i^t v_i - \sqrt{TN^{-1/2}}b_{3CCE} + \sqrt{NT}O_p(R_9)$$

$$+ O_p(N^{-1/2}), \quad (A125)$$

where $b_{3CCE} = N^{-1} \sum_{i=1}^{N} b_{3CCE,i}$. 

128
Lemma PC11. Under Assumptions ERR, LAM, RK–PC, and KAP,
\[
\frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime M_{\hat{F}^{PC}} v_i = \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} E_i^\prime v_i - \sqrt{T N^{-1/2}} \tilde{b}_{3PC} + O_p(R_{13}) + O_p(T^{-1/2}),
\]
where
\[
\tilde{b}_{3PC} = \frac{1}{N} \sum_{i=1}^{N} \alpha_{v_i}^2 A_i^0 (Q^0)^{-1} C_i^0 (1, 0)',
\]
\[
R_{13} = \left( N^{-(\alpha+3)/2} R_6 + N^{-(3\alpha+3)/2} \right) \sqrt{T} + N^{\alpha-1/2} + N^{-1} R_{10} + \left( N^{2\alpha-1/2} + N^{\alpha-1} R_{10} \right) R_6
\]
\[+ \left( N^{-2\alpha} + N^{-3\alpha} R_1 \right) \left( N^{-\alpha} + N^{\alpha-1} R_{10} + N^{2\alpha-1} R_6 R_{10} \right) T^{-1/2}
\]
\[+ \left( N^{\alpha} + N^{-1/2} R_6 + R_{10} \right) T^{-1} + N^{-2\alpha+1/2} T^{-3/2}.
\]

Proof of Lemma PC11. Analogous to Proof of Lemma CCE11, we write
\[
\frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime (M_{F\|H} - M_{\hat{F}^{PC}}) v_i = \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime (M_{F\|H}^\prime v_i - \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime (M_{F\|H} - M_{\hat{F}^{PC}}) v_i), (A126)
\]
where
\[
T^{-1} X_i^\prime (M_{F\|H} - M_{\hat{F}^{PC}}) v_i
\]
\[= T^{-1} X_i^\prime D_{PC} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} D_{PC}^\prime v_i + T^{-1} X_i^\prime D_{PC} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} H^\prime F^\prime v_i
\]
\[+ T^{-1} X_i^\prime F_{H\|H} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} D_{PC}^\prime v_i + T^{-1} X_i^\prime F_{H\|H} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} (H^\prime F H^\prime)^{-1} H^\prime F^\prime v_i
\]
\[= l_1 + \cdots + l_4. \quad (A127)
\]
Using Lemmas PC6, PC7, and PC8, we can work out the following orders for \(l_1, l_2, \) and \(l_4:\)
\[
||l_1||
\]
\[= \left\| \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime D_{PC} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} D_{PC}^\prime v_i \right\|
\]
\[\leq \sqrt{T} \frac{1}{N^{3/2}} \sum_{i=1}^{N} ||NT^{-1} X_i^\prime D_{PC}|| ||(T^{-1} (\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1}|| ||NT^{-1} D_{PC}^\prime v_i||
\]
\[= \sqrt{TN^{-3/2}} [O_p(N^{1/2-\alpha} T^{-1/2} R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2})]
\]
\[\times [O_p(N^{-\alpha}) + O_p(R_6)]
\]
\[= O_p(N^{-(2\alpha+1)} R_1) + O_p(N^{-(\alpha+1)} R_1 R_6) + O_p(\sqrt{TN^{-(3\alpha+3)/2}}) + O_p(\sqrt{TN^{-(\alpha+3)/2}} R_6)
\]
\[+ O_p(N^{-\alpha+1}) + O_p(N^{-1} R_6) + O_p(N^{-(\alpha+1)/2} T^{-1}) + O_p(N^{-1/2} T^{-1} R_6), \quad (A128)
\]
\[
||l_2|| = \left\| \frac{1}{\sqrt{N T}} \sum_{i=1}^{N} X_i^\prime D_{PC} ((\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1} H^\prime F^\prime v_i \right\|
\]
\[\leq N^{-(2\alpha+1/2)} \frac{1}{N} \sum_{i=1}^{N} ||NT^{-1} X_i^\prime D_{PC}|| ||(T^{-1} (\hat{F}^{PC})^\prime \hat{F}^{PC})^{-1}|| ||H^0|| ||T^{-1/2} F^\prime v_i||
\]
\[= N^{-(2\alpha+1/2)} [O_p(N^{1/2-\alpha} T^{-1/2} R_1) + O_p(N^{-\alpha/2}) + O_p(\sqrt{NT^{-1/2}}) + O_p(NT^{-3/2})]
\]
\[ = O_p(N^{-3\alpha} T^{-1/2} R_1) + O_p(N^{-(5\alpha+1)/2}) + O_p(N^{-2\alpha} T^{-1/2}) + O_p(N^{-2\alpha+1/2} T^{-3/2}), \quad (A129) \]

and, recalling (A76),

\[
||I_4|| = \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' F \mathbb{H} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\mathbb{H}' F' \mathbb{H})^{-1} \mathbb{H}' F' v_i \right| \\
\leq N^{1/2 - 4\alpha} \frac{1}{N} \sum_{i=1}^{N} ||T^{-1} X_i' F|| ||\mathbb{H}^0|| T ||(\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\mathbb{H}' F' \mathbb{H})^{-1}|| ||T^{-1/2} F' v_i|| \\
= N^{1/2 - 4\alpha} [O_p(N^{-\alpha}) + O_p(T^{-1/2})] O_p(N^{4\alpha-1/2} T^{-1/2} R_{10}) \\
= O_p(N^{-\alpha} T^{-1/2} R_{10}) + O_p(T^{-1} R_{10}). \quad (A130) \]

For \(I_3\),

\[
I_3 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' F \mathbb{H} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} D^{PC} v_i \\
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' F \mathbb{H} (\mathbb{H}' F' \mathbb{H})^{-1} D^{PC} v_i \\
+ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' F \mathbb{H} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\mathbb{H}' F' \mathbb{H})^{-1} D^{PC} v_i, \]

where, by Lemmas PC6 and PC7,

\[
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' F \mathbb{H} ((\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\mathbb{H}' F' \mathbb{H})^{-1} D^{PC} v_i \right| \\
\leq \sqrt{T} N^{-(2\alpha+1/2)} \frac{1}{N} \sum_{i=1}^{N} ||T^{-1} X_i' F|| ||\mathbb{H}^0|| T ||(\hat{F}^{PC})' \hat{F}^{PC})^{-1} - (\mathbb{H}' F' \mathbb{H})^{-1}|| ||NT^{-1/2} D^{PC} v_i|| \\
= \sqrt{T} N^{-(2\alpha+1/2)} [O_p(N^{-\alpha}) + O_p(T^{-1/2})] O_p(N^{4\alpha-1/2} T^{-1/2} R_{10}) [O_p(N^{-\alpha}) + O_p(R_6)] \\
= O_p(N^{-1} R_{10}) + O_p(N^{\alpha-1} R_6 R_{10}) + O_p(N^{\alpha-1} T^{-1/2} R_{10}) + O_p(N^{2\alpha-1} T^{-1/2} R_6 R_{10}). \]

Also,

\[
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i' F \mathbb{H} (\mathbb{H}' F' \mathbb{H})^{-1} D^{PC} v_i \right| \\
= N^{2\alpha - 1/2} \left| \frac{1}{N} \sum_{i=1}^{N} T^{-1/2} E_i' F \mathbb{H} \left[ T^{-1} (\mathbb{H}^0)' F' \mathbb{H}^0 \right]^{-1} NT^{-1} D^{PC} v_i \right| \\
= N^{2\alpha - 1/2} [O_p(N^{-\alpha}) + O_p(R_6)] = O_p(N^{\alpha-1/2}) + O_p(N^{2\alpha-1/2} R_6), \]

130
which, via \((H^{-1})'H'FH(H'F'H)^{-1} = (H^{-1})'(H')^{-1}\) and Lemma PC6, yields

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_iFH(H'F'H)^{-1}D^{PC}v_i
\]

\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} A_iF'H(H'F'H)^{-1}D^{PC}v_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_iFH(H'F'H)^{-1}D^{PC}v_i
\]

\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} A_i(H^{-1})'H'FH(H'F'H)^{-1}D^{PC}v_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_iFH(H'F'H)^{-1}D^{PC}v_i
\]

\[
= \sqrt{TN}^{\alpha-1/2} \frac{1}{N} \sum_{i=1}^{N} A_i[(H^{-1})']NT^{-1}D^{PC}v_i + O_p(N^{\alpha-1/2}) + O_p(N^{2\alpha-1/2}R_6)
\]

\[
= \sqrt{TN}^{-1/2} \bar{b}_{3PC} + O_p(N^\alpha T^{-1}) + O_p(N^{\alpha-1/2}) + O_p(N^{2\alpha-1/2}R_6),
\]

where \(\bar{b}_{3PC} = N^{-1} \sum_{i=1}^{N} \sigma_i^2 A_i(q^0)^{-1}C_i^0(1,0)\). Hence,

\[
I_3 = \sqrt{TN}^{-1/2} \bar{b}_{3PC} + O_p(N^\alpha T^{-1}) + O_p(N^{\alpha-1/2}) + O_p(N^{2\alpha-1/2}R_6) + O_p(N^{-1}R_10)
\]

\[
+ O_p(N^{\alpha-1}R_6R_{10}) + O_p(N^{\alpha-1}T^{-1/2}R_{10}) + O_p(N^{2\alpha-1}T^{-1/2}R_{6R_{10}}).
\]

Taking together equations (A128)–(A131), we obtain

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i(M^F_H - M^F_{\hat{C}C})v_i = I_1 + \cdots + I_4 = \sqrt{TN}^{-1/2} \bar{b}_{3PC} + O_p(R_{13}).
\]

where

\[
R_{13} = (N^{-(\alpha+3)/2}R_6 + N^{-(3\alpha+3)/2}) \sqrt{T} + N^{\alpha-1/2} + N^{-1}R_{10} + (N^{2\alpha-1/2} + N^{\alpha-1}R_{10})R_6
\]

\[
+ (N^{-2\alpha} + N^{-3\alpha}R_1 + (N^{-\alpha} + N^{\alpha-1})R_{10} + N^{2\alpha-1}R_6R_{10}) T^{-1/2}
\]

\[
+ (N^{\alpha} + N^{-1/2}R_6 + R_{10}) T^{-1} + N^{-2\alpha+1/2}T^{-3/2}.
\]

Therefore, since in the case of CCE, \((NT)^{-1/2} \sum_{i=1}^{N} X'_i(M^F_H)^{-1}v_i = (NT)^{-1/2} \sum_{i=1}^{N} E'_i v_i + O_p(T^{-1/2})\), we can show that

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i(M^F_H)^{-1}v_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i(M^F_H)^{-1}v_i - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i(M^F_H - M^F_{\hat{C}C})v_i
\]

\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_i v_i - \sqrt{TN}^{-1/2} \bar{b}_{3PC} + O_p(R_{13}) + O_p(T^{-1/2}).
\]

**Appendix C: Proofs of main results**

**Proof of Theorem 1.** We begin by considering the CCE estimator. The equation for \(y_i\) can now be written as

\[
y_i = X_i \beta + \hat{F}^{CCE}C^{-1} \lambda_i - D^{CCE}C^{-1} \lambda_i + v_i,
\]

where \(D^{CCE} = \hat{F}^{CCE} - C\) is as before. The CCE estimator of \(\beta\) is given by

\[
\hat{\beta}^CCE = \left( \sum_{i=1}^{N} X'_iM^F_{CCE}X_i \right)^{-1} \sum_{i=1}^{N} X'_iM^F_{CCE}y_i,
\]
implying that
\[
\sqrt{NT}(\hat{\beta}_{CCE} - \beta) = \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{CCE}(v_i - D_{CCE}^{-1} \lambda_i). \tag{A135}
\]

Now let \( \hat{b}_{CCE} = b_{1CCE} - b_{2CCE} - b_{3CCE} \). From Lemmas CCE10 and CCE11, we know that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{CCE}(v_i - D_{CCE}^{-1} \lambda_i)
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_i v_i + N^{(\kappa-1)/2} \hat{b}_{CCE}
+ \sqrt{NT}O_p(R_8) + \sqrt{NT}O_p(R_9) + O_p(N^{-1/2}). \tag{A136}
\]

The asymptotic covariance matrix of \((NT)^{-1/2} \sum_{i=1}^{N} E'_i v_i\) is given by \(W = N^{-1} \sum_{i=1}^{N} \sigma_{i,d} \Sigma_{\epsilon,i}\). Moreover, since the fourth-order moments of \(\epsilon_{i,t}\) and \(v_{i,t}\) are bounded by assumption, by a central limit law,
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_i v_i \rightarrow_d N \left( \mathbf{0}, \lim_{N \rightarrow \infty} W \right) \tag{A137}
\]
as \(N, T \rightarrow \infty\). For \(\sqrt{NT}O_p(R_8)\) and \(\sqrt{NT}O_p(R_9)\) to be negligible, according to the definitions of \(R_8\) and \(R_9\), the following conditions have to be satisfied:

(i) \(\kappa < 3 - 4\alpha\),
(ii) \(\alpha < 1\),
(iii) \(\kappa > 0\),
(iv) \(\alpha < 1/2\),
(v) \(\kappa > 2\alpha\),
(vi) \(\kappa > 4\alpha - 1\),
(vii) \(\kappa > 6\alpha - 2\).

If (iv) is satisfied, only (i) and (v) are binding. These restrictions are tighter than those implied by Lemma CCE9 for the denominator of (A135). Hence, under the condition that \(\alpha < 1/2\) and \(\kappa \in K_{CCE} = (2\alpha, 3 - 4\alpha)\), (A135) reduces to
\[
\sqrt{NT}(\hat{\beta}_{CCE} - \beta) = \sum_{\epsilon}^{-1} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_i v_i + N^{(\kappa-1)/2} \hat{b}_{CCE} \right) + o_p(1)
\rightarrow_d N \left( \mathbf{0}, \lim_{N \rightarrow \infty} \sum_{\epsilon}^{-1} W \sum_{\epsilon}^{-1} \right) + \lim_{N \rightarrow \infty} N^{(\kappa-1)/2} \sum_{\epsilon}^{-1} \hat{b}_{CCE},
\]
which requires that \(N, T \rightarrow \infty\).

It remains to consider the PC estimator. Using the same steps as above, we obtain
\[
\sqrt{NT}(\hat{\beta}_{PC} - \beta) = \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{PC} X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{PC}(v_i - D_{PC}^{-1} \lambda_i). \tag{A138}
\]

Define \(\hat{b}_{PC} = b_{1PC} - b_{2PC} - b_{3PC}\). Using Lemmas PC10 and PC11, we obtain
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{PC}(v_i - D_{PC}^{-1} \lambda_i)
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E'_i v_i + \sqrt{TN^{-1/2}} \hat{b}_{PC}
+ O_p(R_{12}) + O_p(R_{13}) + O_p(T^{-1/2}). \tag{A139}
\]
Let us now consider the order of the remainder term in the above expression, which can be expanded in the following fashion:

\[ R_{12} + R_{13} + T^{-1/2} \]
\[ = (N^{-(3\alpha+3)/2} + N^{2\alpha-1/2}R_4 + N^{\alpha-1/2}R_5 + N^{-(\alpha+3)/2}R_6 + N^{2\alpha+1/2}R_{11})\sqrt{T} \]
\[ + N^{\alpha-1/2} + N^{-(3\alpha+2)/2}R_1 + N^{2\alpha-1/2}R_6 + N^{-1}R_{10} + N^{\alpha-1/2}R_5R_{10} + N^{\alpha-1}R_6R_10 + (1 + N^{-2\alpha} + (N^{-(\alpha+1)/2} + N^{-3\alpha})R_1 + N^{2\alpha-1/2}R_2^2 \]
\[ + N^{2\alpha-1}R_6R_{10} + (N^{\alpha-1} + N^{-\alpha})R_{10})T^{-1/2} \]
\[ + (N^{\alpha} + N^{-1/2}R_6 + R_{10})T^{-1} + (N^{-2\alpha+1/2} + N^{-\alpha}R_1)T^{-3/2}. \] (A140)

Consider the coefficient of \( \sqrt{T} \), the terms of which are given by

\[ N^{2\alpha+1/2}R_{11} = N^{2\alpha-7/2} + (N^{3/2} + N^{-(3\alpha+1)})T^{-1/2} + (N^{-1} + N^{-(2\alpha+1/2)} + N^{2\alpha-5/2})T^{-1} \]
\[ + (N^{3\alpha} + N^{1/2})T^{-3/2} + (1 + N^{-2\alpha+1/2} + N^{2\alpha-5/2})T^{-2} + \sqrt{NT^{-5/2}} \]
\[ + N^{2\alpha-1/2}T^{-3} + N^{2\alpha+1/2}T^{-4}, \]

\[ N^{2\alpha-1/2}R_4 = N^{-(\alpha+3/2)} + N^{2\alpha-1/2} + (N^{-1} + N^{-2\alpha-1/2} + N^{2\alpha-3/2})T^{-1/2} \]
\[ + (N^{2\alpha-1} + N^{-1/2})T^{-1} + (N^{2\alpha-1/2} + N^{-\alpha})T^{-3/2} + (N^{1/2-2\alpha} + 1)T^{-2} \]
\[ \sqrt{NT^{-5/2}} , \]

\[ N^{\alpha-1/2}R_5 = N^{\alpha/2-1} + N^{\alpha-1/2}T^{-1/2} + N^{\alpha}T^{-3/2} , \]

and

\[ N^{-(\alpha+3)/2}R_6 = N^{-(\alpha/2+1/2)} + (N^{-(\alpha+3)/2} + N^{-(3\alpha+2)/2})T^{-1/2} + N^{-(\alpha/2+1)}T^{-1} + N^{-(\alpha+1)/2}. \]

Insertion and simplification yield

\[ N^{-(3\alpha+3)/2} + N^{2\alpha+1/2}R_{11} + N^{2\alpha-1/2}R_4 + N^{\alpha-1/2}R_5 + N^{-(\alpha+3)/2}R_6 \]
\[ = N^{-1} + N^{2\alpha-7/2} + (N^{\alpha-1/2} + N^{2\alpha-3/2})T^{-1/2} + N^{2\alpha-1}T^{-1} + \]
\[ + (N^{\alpha} + N^{2\alpha-1/2})T^{-3/2} + N^{1/2-2\alpha}T^{-2} + \sqrt{NT^{-5/2}} + N^{2\alpha+1/2}T^{-4} . \]

Consider next the part that is constant in \( T \). Here

\[ N^{-(3\alpha/2+1)}R_1 = N^{-(9\alpha/2+1)} + N^{-(3\alpha/2+1)} + (N^{-(3\alpha/2+1)} + N^{-7\alpha+1/2})T^{-1/2} \]
\[ + N^{-3(\alpha+1)/2}T^{-1} , \]

\[ N^{2\alpha-1/2}R_6 = N^{2\alpha-1} + (N^{2\alpha-1/2} + N^{\alpha})T^{-1/2} + N^{2\alpha}T^{-1} + N^{2\alpha+1/2}T^{-3/2} , \]

\[ N^{-1}R_{10} = \sqrt{TN^{-3/2} + N^{-(5\alpha+1)} + N^{-1/2}T^{-1/2} , \]

\[ N^{\alpha-1/2}R_5R_{10} = N^{-3/2}\sqrt{T} + N^{\alpha-1} + N^{-1/2}T^{-1/2} + N^{\alpha}T^{-1} + N^{\alpha+1/2}T^{-2} , \]

\[ N^{\alpha-1}R_6R_{10} = \sqrt{TN^{\alpha-2} + N^{-1} + N^{-(\alpha+3/2)} + (N^{\alpha-1} + N^{-5(\alpha+1)/2})T^{-1/2} \]
\[ + (1 + N^{\alpha-1/2})T^{-1} + N^{\alpha-1/2}T^{-3/2} + N^{\alpha+1/2}T^{-2} , \]

giving

\[ N^{\alpha-1/2} + N^{-(3\alpha/2+1)}R_1 + N^{2\alpha-1/2}R_6 + N^{-1}R_{10} + N^{\alpha-1/2}R_5R_{10} + N^{\alpha-1}R_6R_{10} \]
\[ = (N^{-3/2} + N^{\alpha-2})\sqrt{T} + N^{\alpha-1/2} + N^{2\alpha-1} + (N^{2\alpha-1/2} + N^{\alpha})T^{-1/2} \]
\[ + N^{2\alpha}T^{-1} + N^{2\alpha+1/2}T^{-3/2} . \]
The coefficient of $T^{-1/2}$ is

$$1 + N^{-2\alpha} + (N^{-\alpha+1/2} + N^{-3\alpha})R_1 + N^{2\alpha-1/2}R_1^2 + N^{2\alpha-1}R_6R_{10} + (N^{\alpha-1} + N^{-\alpha})R_{10},$$

where

$$\begin{align*}
(N^{-\alpha+1/2} + N^{-3\alpha})R_1 &= N^{-6\alpha} + N^{-1/2 + \alpha} + N^{-\alpha} + (N^{-3\alpha} + N^{-\alpha+1/2})T^{-1/2} + (N^{-3\alpha+1/2} + N^{-\alpha})T^{-1}, \\
N^{2\alpha-1/2}R_1^2 &\leq N^{-4\alpha+1/2} + N^{2\alpha-3/2} + (N^{2\alpha-1/2} + N^{1/2-2\alpha})T^{-1} + N^{2\alpha+1/2}T^{-2},
\end{align*}$$

and

$$\begin{align*}
N^{2\alpha-1}R_6R_{10} &= N^{2\alpha-2}\sqrt{T} + N^{2\alpha-3/2} + N^{\alpha-1} + (N^{-4\alpha+1/2} + N^{2\alpha-1})T^{-1/2} \\
&+ (N^{2\alpha-1/2} + N^{\alpha})T^{-1} + N^{2\alpha}T^{-3/2} + N^{2\alpha+1/2}T^{-2},
\end{align*}$$

and

$$(N^{\alpha-1} + N^{-\alpha})R_{10} = (N^{-\alpha+1/2} + N^{-3\alpha})\sqrt{T} + N^{-(4\alpha+1)} + N^{-6\alpha} + (N^{\alpha-1/2} + N^{1/2-\alpha})T^{-1/2}.\tag{viii}$$

Insertion and simplification yield

$$\begin{align*}
1 + N^{-2\alpha} + (N^{-\alpha+1/2} + N^{-3\alpha})R_1 + N^{2\alpha-1/2}R_1^2 + N^{2\alpha-1}R_6R_{10} + (N^{\alpha-1} + N^{-\alpha})R_{10}
&= (N^{2\alpha-2} + N^{\alpha-3/2} + N^{-\alpha+1/2})\sqrt{T} + N^{2\alpha-3/2} + N^{\alpha-1} + 1 + N^{-2\alpha} \\
&+ (N^{2\alpha-1} + N^{\alpha-1/2} + N^{1/2-\alpha})T^{-1/2} + (N^{2\alpha-1/2} + N^{\alpha})T^{-1} + N^{2\alpha}T^{-3/2} + N^{2\alpha+1/2}T^{-2}.
\end{align*}$$

The coefficient of $T^{-1}$ is simple;

$$N^{\alpha} + N^{-1/2}R_6 + R_{10} = N^{-1/2}\sqrt{T} + N^{\alpha} + \sqrt{NT^{-1/2}},$$

and so is that of $T^{-3/2}$;

$$N^{2\alpha+1/2} + N^{-\alpha}R_1 = N^{2\alpha+1/2} + N^{-\alpha+1/2} + N^{-2\alpha}T^{-1/2} + N^{1/2-\alpha}T^{-1}.$$  \tag{vii}

Putting everything together, the remainder term in $(\text{A139})$ becomes

$$R_{12} + R_{13} + T^{-1/2} = (N^{2\alpha-7/2} + N^{\alpha-2} + N^{\alpha/2-1})\sqrt{T} + N^{2\alpha-1} + N^{\alpha-1/2} \\
+ (N^{2\alpha-1/2} + N^{\alpha} + 1)T^{-1/2} + (N^{2\alpha} + N^{1/2-\alpha})T^{-1} \\
+ N^{2\alpha+1/2}T^{-3/2} \\
= (N^{2\alpha-7/2} + N^{\alpha-2} + N^{\alpha/2-1})N^{\kappa/2} + N^{2\alpha-1} + N^{\alpha-1/2} \\
+ (N^{2\alpha-1/2} + N^{\alpha} + 1)N^{-\kappa/2} + (N^{2\alpha} + N^{1/2-\alpha})N^{-\kappa} \\
+ N^{2\alpha+1/2-3\kappa/2}.\tag{A141}$$

For this term to converge to go to zero, the following restrictions have to be satisfied:

(i) $\kappa < 7 - 4\alpha$,
(ii) $\kappa < 4 - 2\alpha$,
(iii) $\kappa < 2 - \alpha$,
(iv) $\alpha < 1/2$,
(v) $\kappa > 4\alpha - 1$,
(vi) $\kappa > 2\alpha$,
(vii) $\kappa > 0$,
(viii) $\kappa > 1/2 - \alpha$,
(ix) $\kappa > (4\alpha + 1)/3$.

Obviously, (iv) provides an upper bound on $\alpha$. Given the admissible values of $\alpha$, only (iii), (viii), and (ix) are binding. Consequently, we have that if $\alpha < 1/2$ and $\kappa \in K_{PC} = (\max(1/2 - \alpha, (4\alpha + 1)/3), 2 - \alpha)$,
then
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' M_{pc}(\nu_i - D_{pc}^{-1} H^{-1} \lambda_i) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i' \nu_i + \sqrt{T N^{-1/2} b_{PC} + o_p(1)}
\]
\[
\to_d N \left( 0, \lim_{N \to \infty} \bar{W} \right)
\]  
(A142)
as \( N, T \to \infty \). The limit of the denominator of the estimator is given by Lemma PC9, a result that holds irrespectively of the value of \( \alpha \) and \( \kappa \). Hence,
\[
\sqrt{NT}(\hat{\beta}_{PC}^p - \beta) \to_d N \left( 0, \lim_{N \to \infty} \sum_{i}^{-1} \bar{W} \bar{W}^{-1} \right) + \lim_{N \to \infty} N^{(\kappa - 1)/2} \sum_{i}^{-1} \bar{b}_{PC},
\]  
(A143)
which again requires \( \alpha < 1/2 \) and \( \kappa \in K_{PC} \). This establishes the required result for the PC estimator, and hence the proof of Theorem 1 is complete.

**Proof of Corollary 1.** Consider the CCE estimator. By using (A135) and Lemmas CCE10 and CCE11, it is not difficult to show that
\[
\hat{\beta}_{CCE}^p - \beta = O_P((NT)^{-1/2}) + O_p(N^{-1}) + O_p(R_8) + O_p(R_9).
\]  
(A144)
For the right-hand side of this expression to be \( o_p(1) \), \( R_8 \) and \( R_9 \) have to be negligible, which in turn imposes the following restrictions:

(i) \( \alpha < 1 \),
(ii) \( \kappa > 6\alpha - 4 \),
(iii) \( \kappa > 2\alpha - 2 \),
(iv) \( \kappa > 4\alpha - 3 \),
(v) \( \kappa > 3\alpha - 3/2 \),
(vi) \( \kappa > \alpha - 1/2 \),
(vii) \( \kappa > 2\alpha - 1 \).

Given that (i) is satisfied, only (ii), (v), and (vi) are binding. Hence, for (A144) to converge to zero, we need \( \alpha < 1 \) and \( \kappa > \max\{6\alpha - 4, \alpha - 1/2, 3\alpha - 3/2\} \), which for the relevant range of values for \( \kappa > 0 \) and \( \alpha \in [0, 1] \) reduces to \( \kappa > \max\{6\alpha - 4, 3\alpha - 3/2\} \).

As for the PC estimator, if we assume that \( \kappa > \max\{2\alpha, 4\alpha - 1\} \), such that Lemma PC1 holds, then by (A139),
\[
\hat{\beta}_{PC}^p - \beta = N^{-1/2} O_p(R_{12} + R_{13}),
\]  
(A145)
where, using the definition of \( R_{12} \) and \( R_{13} \), the remainder term is
\[
N^{-1/2}(R_{12} + R_{13})
\]
\[
= (N^{2\alpha - 7/2} + N^{\alpha - 2} + N^{-1})N^{-1/2} + N^{2\alpha - (3+\alpha)/2} + N^{\alpha - (1+\kappa)/2}
\]
\[
+ (N^{2\alpha - 1/2} + N^{\alpha} + 1)N^{-(1/2+\kappa)} + (N^{2\alpha} + N^{1/2-\alpha})N^{-(1+3\alpha)/2} + N^{2(\alpha-\kappa)}
\]
\[
= N^{2\alpha - (3+\kappa)/2} + N^{2\alpha - (1+\kappa)} + N^{\alpha - (1/2+\kappa)} + N^{2\alpha - (1+3\alpha)/2} + N^{2(\alpha-\kappa)},
\]
which converges to zero in probability if the following statements hold:

(i) \( \kappa > 4\alpha - 3 \),
(ii) \( \kappa > 2\alpha - 1 \),
(iii) \( \kappa > \alpha - 1/2 \),
(iv) \( \kappa > (2\alpha - 1/2)/3 \),
(v) \( \kappa > \alpha \).

However, none of these conditions are stricter than those implied by Lemma PC1. Hence, the values for \( \kappa \) that imply \( \hat{\beta}_{PC}^p - \beta = o_p(1) \) are bounded by \( \kappa > \max\{2\alpha, 4\alpha - 1\} \).
Proof of Theorem 2. Consider the CCE estimator, which is given by

\[ \hat{\beta}_{\text{CCE}}^p = \left( \sum_{i=1}^{N} X_i' M_{\text{CCE}} X_i \right)^{-1} \sum_{i=1}^{N} X_i' M_{\text{CCE}} y_i, \]

Under Assumption HET, this implies

\[ \sqrt{N}(\hat{\beta}_{\text{CCE}}^p - \beta) = \left( \frac{1}{NT} \sum_{i=1}^{N} X_i' \tilde{M}_{\text{CCE}} X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' \tilde{M}_{\text{CCE}} (X_i \xi_i + v_i - D_{\text{CCE}} C^{-1} \lambda_i). \quad (A146) \]

The first term in the numerator is

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' M_{\text{FC}} X_i \xi_i \]

Analogously to \((A78)\) in Proof of Lemma CCE9, we obtain

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' M_{\text{FC}} X_i \xi_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i' M_{\text{FC}} E_i \xi_i \]

\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i' \xi_i - T^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (T^{-1} E_i' F \tilde{C} \tilde{C}^{-1} (T^{-1} E_i' F' F^{-1} C) \xi_i \]

\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\epsilon, i} \xi_i + O_p(T^{-1}). \quad (A148) \]

It is easily seen that the first term on the right-hand side of the above expression is zero in expectation and also that the variance is given by

\[ E \left( \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon, i} \xi_i \right)^{\prime} \left( \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon, i} \xi_i \right) \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon, i} E(\xi_i \xi_i') \Sigma_{\epsilon, i} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon, i} \Sigma_{\epsilon, i}. \quad (A149) \]

The order of the second term on the right-hand side of \((A147)\) can be obtained by using \((A77)\);

\[ \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_i' (M_{\text{CCE}} - M_{\text{FC}}) X_i \xi_i \right| \]

\[ \leq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} T^{-1} ||X_i' (M_{\text{CCE}} - M_{\text{FC}}) X_i|| ||\xi_i|| \]

\[ = O_p(N^{2/3 - 3/2}) + O_p(N^{-1/2}) + O_p(N^{2\alpha - (2 + \kappa)/2}) + O_p(N^{\alpha - (1 + \kappa)/2}) \]

\[ + O_p(N^{-\kappa/2}) + O_p(N^{\alpha - \kappa}) + O_p(N^{2\alpha - \kappa - 1/2}). \quad (A150) \]
Hence, from (A148) and (A150), we have that (A147) reduces to
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{CCE} \Sigma_{\epsilon,i} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \xi_i + O_p(R_{14})
\]  
(A151)

where
\[
R_{14} = N^{2\alpha-3/2} + N^{-1/2} + N^{2\alpha-(2+\kappa)/2} + N^{\alpha-(1+\kappa)/2} + N^{-\kappa/2} + N^{\alpha-\kappa} + N^{2\alpha-\kappa-1/2} + N^{-\kappa}.
\]

The remaining terms in (A146) are obtained by following the same steps as in the proof of Theorem 1. Hence, applying (A136) and the result above, we obtain
\[
\sqrt{N}(\hat{\beta}_{CCE}^p - \beta) = \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{CCE} X_i \xi_i \\
+ \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X'_i M_{CCE} (\nu_i - D^{CCE} \nu^\top \lambda_i) \\
= \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \xi_i \\
+ \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \left( T^{-1/2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \nu'_i \nu_i + N^{-1/2} \Sigma_{CCE} \right) \\
+ O_p(R_{14}) + \sqrt{NO_p(R_8)} + \sqrt{NO_p(R_9)} + O_p(T^{-1/2}) \\
= \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_{CCE} X_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \xi_i + O_p(R_{14}) \\
+ \sqrt{NO_p(R_8)} + \sqrt{NO_p(R_9)} + O_p(N^{-1/2}) + O_p(T^{-1/2}).
\]  
(A152)

Consider \(\sqrt{NR_8}\) and \(\sqrt{NR_9}\). To ensure that these are negligible, we again put restrictions \(\alpha\) and \(\kappa\). In particular, the following restrictions have to be met:

(i) \(\alpha < 3/4\),
(ii) \(\kappa > 6\alpha - 3\),
(iii) \(\kappa > 0\),
(iv) \(\kappa > 2\alpha - 1\),
(v) \(\kappa > 4\alpha - 2\),
(vi) \(\kappa > 3\alpha - 1\),
(vii) \(\kappa > \alpha\),
(viii) \(\kappa > 2\alpha - 1/2\).

Clearly, given that (i) is satisfied, only (ii), (vi), and (vii) are binding. These conditions are, however, more stringent than those implied by Lemma CCE9. We can hence conclude that if \(\alpha < 3/4\) and \(\kappa > \max\{\alpha, 6\alpha - 3, 3\alpha - 1\}\), then \(\sqrt{NR_8}\) and \(\sqrt{NR_9}\) are negligible. Additional conditions are needed for
\( R_{14} \) to go to zero. These can, however, be shown either to be less stringent or to coincide with the above conditions. We can therefore conclude that if \( \alpha < 3/4 \) and \( \kappa > \max\{\alpha, 2\alpha - 1/2\} \), as \( N, T \to \infty \),

\[
\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta) = \sum_{\epsilon}^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{i} \right) + o_p(1)
\]

\[
\to_d N \left( 0, \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{i} \right) \Sigma_{\epsilon, i}^{-1} \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_{i} \right) \Sigma_{\epsilon, i}^{-1} \right).
\]

The proof for the PC estimator is very similar to that provided above for the CCE estimator. The main difference is that instead of (A77) we use (A81):

\[
\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_{i}(M_{\text{FF}} - M_{\text{PC}}) X_{i} \right\|
\]

\[
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} T^{-1}||X_{i}(M_{\text{FF}} - M_{\text{PC}}) X_{i}|| ||\xi_{i}|| = O_p(R_{15}),
\]

where

\[
R_{15} = N^{-2\alpha - 1/2} + N^{-\alpha - 3/2} + N^{-(\alpha + \kappa + 2)/2} + N^{-2(\alpha + 1 + \kappa)/2} + N^{-1/2 - \kappa}
\]

\[
+ N^{-2\alpha - \kappa + 1/2} + N^{-\alpha + 1/2 - 3\kappa/2} + N^{-2\kappa + 1/2}.
\]

By using this, (A139), and the same steps as when considering the CCE estimator, we can show that

\[
\sqrt{N}(\hat{\beta}_{\text{PC}}^p - \beta)
\]

\[
= \left( \frac{1}{NT} \sum_{i=1}^{N} X_{i}{M_{\text{FF}} X_{i}} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_{i}{M_{\text{PC}} X_{i} \xi_{i}}
\]

\[
+ \left( \frac{1}{NT} \sum_{i=1}^{N} X_{i}{M_{\text{PC}} X_{i}} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_{i}{M_{\text{PC}} (\nu_{i} - D_{\text{PC}C^{-1} \lambda_{i}})}
\]

\[
= \left( \frac{1}{NT} \sum_{i=1}^{N} X_{i}{M_{\text{PC}} X_{i}} \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Sigma_{\epsilon, i} \xi_{i} + O_p(R_{15})
\]

\[+ T^{-1/2} [O_p(R_{12}) + O_p(R_{13}) + O_p(T^{-1/2})] + O_p(N^{-1/2}) + O_p(T^{-1/2}). \]  \( \text{(A154)} \)

By using the results provided in the PC part of Proof of Theorem 1,

\[
T^{-1/2} (R_{12} + R_{13} + T^{-1/2})
\]

\[
= N^{2\alpha - 7/2} + N^{\alpha - 2} + N^{\alpha - 1/2} + N^{2\alpha - 1 - \kappa/2} + N^{\alpha - 1/2 - \kappa/2} + (N^{2\alpha - 1/2} + N^{\alpha} + 1) N^{-\kappa}
\]

\[
+ (N^{2\alpha} + N^{1/2 - \alpha}) N^{-3\kappa/2} + N^{2\alpha + 1/2 - 2\kappa}. \]  \( \text{(A155)} \)

The following conditions have to be met for this term to go to zero:

(i) \( \alpha < 7/4 \),
(ii) \( \kappa > 4\alpha - 2 \),
(iii) \( \kappa > 2\alpha - 1/2 \),
(iv) \( \kappa > 4\alpha/3 \),
(v) \( \kappa > (1 - 2\alpha)/3 \),
(vi) \( \kappa > \alpha + 1/4 \).

While condition (i) binds \( \alpha \), (ii), (vi), and (v) bind for \( \kappa \), depending on the value of \( \alpha \). The relevant restrictions to put on \( \kappa \) and \( \alpha \) are therefore given by \( \alpha < 1 \) and \( \kappa > \max\{4\alpha - 2, \alpha + 1/4, (1 - 2\alpha)/3\} \), respectively. Similarly, in order to ensure that \( R_{15} = o(1) \), we need as follows:

(i) \( \kappa > 1/2 - 2\alpha \),
(ii) \( \kappa > (1 - 2\alpha)/3 \),
(iii) \( \kappa > 1/4 \).

Among these conditions, only (i) matters since it is stricter than \( (1 - 2\alpha)/3 \) for \( \alpha \) close to zero. Hence, provided that \( \alpha < 7/4 \) and \( \kappa > \max\{4\alpha - 2, \alpha + 1/4, 1/2 - 2\alpha\} \),

\[
\sqrt{N}(\hat{\beta}_{PC} - \beta) \rightarrow_d N\left(0, \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \Sigma_{\zeta,i} \left(\Sigma_{\epsilon,i}^{-1}\right)\right)
\]
as \( N, T \rightarrow \infty \).

**Proof of Corollary 2.** It follows from Proof of Theorem 2 that

\[
\hat{\beta}_{CCE} - \beta = O_p(N^{-1/2}) + N^{-1/2}O_p(R_{14}) + O_p(R_8) + O_p(R_9) + O_p(N^{-1}) + O_p((NT)^{-1/2}).
\]

As noted in Proof of Theorem 2, the requirements for \( R_{14} \) to go to zero are never more stringent than those required for \( \sqrt{N}R_8 \) and \( \sqrt{N}R_9 \) to go to zero. Thus, it is enough to consider \( R_8 \) and \( R_9 \), for which the conditions are the same as in Proof of Corollary 1. Hence, provided that \( \alpha < 1 \) and \( \kappa > 2\alpha - 1 \),

\[
\hat{\beta}_{CCE} - \beta = o_p(1).
\]

The corresponding requirement for the PC estimator is that

\[
\hat{\beta}_{PC} - \beta = O_p(N^{-1/2}) + N^{-1/2}O_p(R_{15}) + (NT)^{-1/2}[O_p(R_{12}) + O_p(R_{13}) + O_p(T^{-1/2})] + O_p(N^{-1}) + O_p((NT)^{-1/2}).
\]

should be \( o_p(1) \). Now, by using the results provided in Proof of Theorem 2, we have that \( N^{-1/2}R_{15} = o(1) \) for all \( \alpha \in [0, 1] \) and \( \kappa > 0 \). As shown in Proof of Corollary 1, \( \kappa > \alpha \) is required for the remaining terms to converge to zero. However, this condition is not as strict as the one required for Lemma PC1 to hold. Hence, as long as \( \kappa > \max\{2\alpha, 4\alpha - 1\} \), we have

\[
\hat{\beta}_{PC} - \beta = o_p(1).
\]

**Appendix D: Some results for the individual and mean group CCE estimators**

As mentioned in the main text, Pesaran (2006) does not only consider the pooled CCE estimator, but also an individual estimator, henceforth denoted \( \hat{\beta}_{CCE,i} \), and a mean group-type CCE estimator, here denoted \( \hat{\beta}_{CCE}^{mg} \). In this section we report some results for the latter two estimators under Assumption HET.

The individual CCE estimator is given simply by

\[
\hat{\beta}_{CCE,i} = (X_i'M_{CCE}X_i)^{-1}X_i'M_{CCE}y_i,
\]

the asymptotic distribution of which is given in Theorem D1.

**Theorem D1.** Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold, and that \( \alpha < 0.7 \) and \( \kappa \in K_{CCE} = (\max\{0, 6\alpha - 3\}, 4 - 4\alpha) \). Then, as \( N, T \rightarrow \infty \),

\[
\sqrt{T}(\hat{\beta}_{CCE,i} - \beta_i) \rightarrow_d N(0, \sigma^2_{\epsilon,i} \Sigma_{\epsilon,i}) + \lim_{N \rightarrow \infty} N^{(\kappa-2)/2} \Sigma_{\epsilon,i}^{-1} b_{CCE,i}.
\]
Proof of Theorem D1. Analogous to Proof of Theorem 1, we have
\[ \sqrt{T}(\hat{\beta}_{CCE,i} - \beta_i) = (T^{-1}X_i'\Sigma_{CCE}X_i)^{-1}T^{-1/2}X_i'\Sigma_{CCE}(v_i - D^{CCE}_{CCE}^{-1}l_i). \]
From Lemmas CCE10 and CCE11, we know that
\[ T^{-1/2}X_i'\Sigma_{CCE}(v_i - D^{CCE}_{CCE}^{-1}l_i) = T^{-1/2}E_i'v_i + N^{-1}\sqrt{T}b_{CCE,i} + \sqrt{T}O_p(R_8) + \sqrt{T}O_p(R_9) + O_p(T^{-3/2}). \]
where \( b_{CCE,i} = b_{1CCE,i} - b_{2CCE,i} - b_{3CCE,i} \). The asymptotic variance of \( T^{-1/2}E_i'v_i \) is \( \sigma_{\epsilon,i}^2 \Sigma_{\epsilon,i} \), which in conjunction with the assumptions placed on \( E_i \) and \( v_i \) implies
\[ T^{-1/2}E_i'v_i \to_d N(0, \sigma_{\epsilon,i}^2 \Sigma_{\epsilon,i}) \]
as \( N, T \to \infty \). As for the remainder terms in (A157), the restrictions for \( \sqrt{T}R_8 \) and \( \sqrt{T}R_9 \) to converge to zero are as follows:
(i) \( \kappa < 4 - 4\alpha \),
(ii) \( \alpha < 3/2 \),
(iii) \( \kappa > -1 \),
(iv) \( \alpha < 1 \),
(v) \( \alpha < 3/4 \),
(vi) \( \kappa > 2\alpha - 1 \),
(vii) \( \kappa > 4\alpha - 2 \),
(viii) \( \kappa > 6\alpha - 3 \).
The binding conditions for \( \kappa \) are given by (i) and (viii) and these impose tighter restrictions on \( \alpha \) than (v) since equality of (i) and (viii) is given at \( \alpha = 0.7 \). Note that condition (vi) is the same as in Lemma CCE9. Hence, provided that \( \alpha < 0.7 \) and \( \kappa \in (\max(0, 6\alpha - 3), 4 - 4\alpha) \), as \( N, T \to \infty \),
\[ \sqrt{T}(\hat{\beta}_{CCE,i} - \beta_i) = \Sigma_{\epsilon,i}^{-1}T^{-1/2}E_i'v_i + \Sigma_{\epsilon,i}^{-1}N^{-1}\sqrt{T}b_{CCE,i} + o_p(1) \]
\[ \to_d N(0, \sigma_{\epsilon,i}^2 \Sigma_{\epsilon,i}) + \lim_{N,T \to \infty} \Sigma_{\epsilon,i}^{-1}N^{-1}\sqrt{T}b_{CCE,i}, \]
as required.

(Pesaran, 2006, Theorem 1) derives the asymptotic distribution of \( \sqrt{T}(\hat{\beta}_{CCE,i} - \beta_i) \) in the strong factor case. He assumes that \( \sqrt{T}/N = N^{(\kappa-2)/2} \to 0 \) as \( N, T \to \infty \), which is more restrictive than the condition required by Theorem D1. However, we see that if \( \kappa < 2 \) such that \( N^{(\kappa-2)/2} \to 0 \), then according to Theorem D1, \( \sqrt{T}(\hat{\beta}_{CCE,i} - \beta_i) \) is asymptotically unbiased, which is in line with the results of Pesaran (2006). Hence, even in the strong factor case Theorem D1 represents an extension of the existing work in the field; it extends the work of Pesaran (2006) to the case when \( \kappa \in (0, 4) \) such that \( \sqrt{T}/N \) is not necessarily zero. In the non-strong factor case, there is no previous research to which we can refer. However, we see that the impact of \( \alpha \) on the set of allowable values of \( \kappa \), \( K_{CCE} \), is very similar to that found in Theorem 1 for the pooled CCE estimator under Assumption HOM. In particular, the smaller \( \alpha \) (the weaker the factors) is the narrower \( K_{CCE} \) is. In the extreme case when \( \alpha \to 0.7 \), \( K_{CCE} \) collapses to a single value equal to 1.2.

Corollary D1. Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold. Suppose also that \( \kappa > \max\{6\alpha - 4, 3\alpha - 3/2\} \) with \( \alpha < 1 \). Then, as \( N, T \to \infty \),
\[ ||\hat{\beta}_{CCE,i} - \beta_i|| = o_p(1). \]

Proof of Corollary D1. It follows from (A157) and (A159) that
\[ \hat{\beta}_{CCE,i} - \beta_i = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p(R_8) + O_p(R_9) + O_p(T^{-3/2}). \]
It remains to establish conditions under which $R_8$ and $R_9$ are $o(1)$. These are given in Proof of Corollary 1, and hence we can conclude that, if $\alpha < 1$ and $\kappa > 2\alpha - 1$, then

$$\hat{\beta}_{\text{CCE},i} - \beta_i = o_p(1).$$

According to Corollaries 1 and D1, the required conditions on $\alpha$ and $\kappa$ to ensure consistency of the individual CCE estimator is the same as for the pooled CCE estimator.

The mean group CCE estimator is simply the average $\hat{\beta}_{\text{CCE},i}$:

$$\hat{\beta}_{\text{CCE}}^{mg} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{\text{CCE},i}.$$  

**Theorem D2.** Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold. Suppose also that $\alpha < 3/4$ and $\kappa > \max\{\alpha, 6\alpha - 3, 3\alpha - 1\}$. Then, as $N, T \to \infty$,

$$\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta) \to_d N(0, \Sigma_\zeta).$$

**Proof of Theorem D2.** Under Assumption HET, $\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta)$ can be written as

$$\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta) = N^{-1/2} \sum_{i=1}^{N} (\hat{\beta}_{\text{CCE},i} - \beta)$$

$$= N^{-1/2} \sum_{i=1}^{N} \xi_i + \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} (T^{-1}X_i'\hat{M}_{\text{CCE}}X_i)^{-1}T^{-1}X_i'\hat{M}_{\text{CCE}}\nu_i$$

$$- \sqrt{N} \frac{1}{N} \sum_{i=1}^{N} (T^{-1}X_i'\hat{M}_{\text{CCE}}X_i)^{-1}T^{-1}X_i'\hat{M}_{\text{CCE}}D_{\text{CCE}}C^{-1}\lambda_i.$$

Since $T^{-1}X_i'\hat{M}_{\text{CCE}}X_i = O_p(1)$, we can make use of Lemmas CCE10 and CCE11 to obtain

$$\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta) = N^{-1/2} \sum_{i=1}^{N} \xi_i + T^{-1/2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (T^{-1}X_i'\hat{M}_{\text{CCE}}X_i)^{-1}E_i'\nu_i + \sqrt{N}O_p(R_9)$$

$$+ O_p(T^{-1}) + N^{1/2} \sum_{i=1}^{N} N^{-1/2} (T^{-1}X_i'\hat{M}_{\text{CCE}}X_i)^{-1}b_{\text{CCE},i}$$

$$+ \sqrt{N}O_p(R_8). \quad (A160)$$

Now let $\Delta = N^{-1/2} \sum_{i=1}^{N} \xi_i + T^{-1/2}(NT)^{-1} \sum_{i=1}^{N} 1 \Sigma_{\epsilon_i}^{-1}E_i'\nu_i$. Given the assumptions on $E_i$, $\nu_i$, and $\xi_i$, the asymptotic variance of this term is given by

$$E(\Delta\Delta') = \frac{1}{N} \sum_{i=1}^{N} E(\xi_i\xi_i') + T^{-1} \frac{1}{NT} \sum_{i=1}^{N} E \left( \left( \frac{1}{T}X_i'\hat{M}_{\text{CCE}}X_i \right)^{-1}E_i'\nu_i\nu_iE_i' \left( \frac{1}{T}X_i'\hat{M}_{\text{CCE}}X_i \right)^{-1} \right)$$

$$= \Sigma_\zeta + O(T^{-1}).$$

Also, in order to avoid that the remainder terms in (A160) dominate the asymptotic behavior of $\sqrt{N}(\hat{\beta}_{\text{CCE}} - \beta)$, conditions have to be placed on $\alpha$ and $\kappa$. Specifically, to ensure that $\sqrt{NR_8}$ and $\sqrt{NR_9}$ are $o(1)$, we need as follows:

(i) $\alpha < 3/4$,

(ii) $\kappa > 6\alpha - 3$,

(iii) $\kappa > 0$,
(iv) $\kappa > 2\alpha - 1$,
(v) $\kappa > 4\alpha - 2$,
(vi) $\kappa > 3\alpha - 1$,
(vii) $\kappa > \alpha$,
(viii) $\kappa > 2\alpha - 1\kappa$.

It is easy to see that given condition (i) for $\alpha$, only (ii), (vi), and (vii) are binding for $\kappa$. These conditions are, however, more stringent than those implied by Lemma CCE9. We can therefore conclude that, if $\alpha < 3/4$ and $\kappa > \max\{\alpha, 6\alpha - 3, 3\alpha - 1\}$, then

$$\sqrt{N}(\hat{\beta}_{CCE}^{mg} - \beta) = N^{-1/2} \sum_{i=1}^{N} \zeta_i + T^{-1/2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Sigma_{\epsilon,i}^{1} E_i v_i + N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \Sigma_{\epsilon,i}^{1} b_{CCE,i} + o_p(1)$$

as $N, T \to \infty$.

According to Theorems 2 and D2, under Assumption HET, the asymptotic distributions of $\hat{\beta}_{CCE}^{p}$ and $\hat{\beta}_{CCE}^{mg}$ differ only in terms of the asymptotic variance. Note in particular that by the Cauchy–Schwarz inequality, $N^{-1} \sum_{i=1}^{N} \Sigma_{\epsilon,i} \Sigma_{\epsilon,i} - \Sigma_{\epsilon} \Sigma_{\epsilon} \Sigma_{\epsilon}$ is positive semidefinite, implying that $\hat{\beta}_{CCE}^{p}$ is more efficient than $\hat{\beta}_{CCE}^{mg}$.

The conditions on $\alpha$ and $\kappa$ that ensure consistency under Assumption HET are the same for $\hat{\beta}_{CCE}^{p}$ and $\hat{\beta}_{CCE}^{mg}$. These conditions are in turn identical to those of $\hat{\beta}_{CCE,i}^{p}$ and $\hat{\beta}_{CCE}^{p}$ under Assumption HOM. Corollary D2 formalizes this.

**Corollary D2.** Suppose that Assumptions HET, ERR, LAM, RK–CCE, and KAP hold. Suppose also that $\kappa > \max\{6\alpha - 4, 3\alpha - 3/2\}$ with $\alpha < 1$. Then, as $N, T \to \infty$,

$$||\hat{\beta}_{CCE}^{mg} - \beta|| = o_p(1).$$

**Proof of Corollary D2.** According to Proof of Theorem D2,

$$\hat{\beta}_{CCE}^{mg} - \beta = O_p(N^{-1/2} + O_p((NT)^{-1/2}) + O_p(N^{-1}) + O_p(R_8) + O_p(R_9)). \quad (A161)$$

Given that $R_8$ and $R_9$ enter this expression without being multiplied by either $N$ or $T$, the conditions for convergence to zero are the same as in Corollary D1.

According to Corollaries 2, D1, and D2, if consistency is the only concern, in terms of the allowable values of $\alpha$ and $\kappa$, the choice of which CCE estimator to use is just a matter of personal preference. However, if we are also concerned about the rate at which consistency is achieved, if $\kappa < 1$ such that $\sqrt{T} = N^{\kappa/2} < \sqrt{N}$, then $\hat{\beta}_{CCE}^{p}$ is preferred, whereas if $\kappa \in (1, 2)$, then $\hat{\beta}_{CCE,i}^{p}$ is the preferred choice (if $\kappa \geq 2$, then $\sqrt{T}(\hat{\beta}_{CCE,i}^{p} - \beta_i)$ is no longer asymptotically unbiased).

**Acknowledgments**

A previous version of the article was presented at a seminar at Lund University. The authors would like to thank seminar participants, and in particular Esfandiar Maasoumi (Editor) and two anonymous referees for many valuable comments and suggestions.

**Funding**

Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship. Both authors thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.
References


Paper IV
A Hausman test for cross-section dependence in linear panel regression models

Simon Reese*

March 2, 2017

Abstract

Testing for the presence of cross-section dependence in panel data is becoming an ever more relevant model specification issue given the increasing use of interactive fixed effects models. We contribute to this strand of the literature by proposing a Hausman test for cross-section dependence, based on the CCE estimator of Pesaran (2006). By considering the difference between two estimators, our focus is different from that of the dominating approach which is based on correlations in the model residuals. The distribution of the interactive fixed effects estimator under the null hypothesis is derived using a normalisation technique for the factor estimates to circumvent the problem of asymptotic singularity that occurs when the factos are absent. Simulations reveal that the new test has good performance relative to other tests.

JEL Classification: C12; C13; C33.

Keywords: Cross-section dependence; diagnostic tests; CCE estimation; Hausman test.

1 Introduction

The literature on panel regression models that model cross-section dependence via common factors has been expanding rapidly since the introduction of the common correlated effects (CCE) and principal components estimators of Pesaran (2006) and Bai (2009), respectively. In fact, contributions on so-called factor-augmented panel regression are by now so plentiful that they constitute an own branch in the literature (see Pesaran, 2015, Ch.29, for an overview). Together with the increasing awareness of the potential consequences of neglecting cross-section

*Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. E-mail address: simon.reese@nek.lu.se.
correlation, this development has lead to increased interest in how to properly test for the presence of cross-section dependence.

In this article, we are contributing to the literature on testing for cross-section dependence by suggesting a Hausman test based on Pesaran’s (2006) CCE estimator. Nearly all tests for cross-section dependence are variants of the Lagrange Multiplier (LM) test of Breusch and Pagan (1980) which is constructed from the sum of all squared pairwise cross-section correlations of the regression residuals. These tests are simple to apply and have therefore become very popular in empirical research. The most popular test by far is the CD test of Pesaran (2004) whose key difference to the LM test is the use of a plain sum of all pairwise correlations instead of their square. One exception to this tradition of using LM-based tests is Sarafidis (2009) who use Sargan’s (1988) difference test to test for cross-section dependence. The use of a Hausman’s approach of comparing two different estimators has so far not been explored in this context. Bai (2009) discussed the possibility of using a Hausman test to test for the presence of additive versus interactive fixed effects. However, to the best of our knowledge no attempts have been made so far to use a Hausman approach to test the null hypothesis of no cross-section dependence. The purpose of the present paper is to fill this gap in the literature. We find that the difference between the CCE estimator and the OLS estimator, which is the basic building block of our new test, converges at a faster rate than the estimators themselves. This yields a test statistic that has relatively high power when compared to the popular CD test of Pesaran (2004) while showing good size accuracy.

The balance of the paper is structured as follows: Section 2 defines the model and develops the asymptotic properties of the CCE estimator in the absence of cross-section dependence. Section 3 is dedicated to deriving the Hausman test statistic. The finite sample performance of our suggested test is investigated in Section 4. Finally, Section 5 concludes.

1See Demetrescu and Homm (2016), and Mao (2016), for some recent developments and surveys of the existing literature.
2 The CCE estimator and its properties in the absence of cross-section dependence

Consider the model

\[ y_i = X_i \beta + F \lambda_i + \epsilon_i \]

\[ X_i = F \Lambda_i + E_i, \quad i = 1, \ldots, N \]

(1)

where \( y_i \) is a \( T \)-vector and \( X_i \) a \( T \times m \) matrix of explanatory variables. \( F = \begin{bmatrix} f_1 & \cdots & f_T \end{bmatrix} \) is a \( T \times r \) matrix of latent common factors, and the associated factor loadings \( \lambda_i = \left( \lambda_i^{(c)} \right) \) and \( \Lambda_i = \left( \Lambda_i^{(c,d)} \right) \) are of dimension \( r \times 1 \) and \( r \times m \) respectively. \( \epsilon_i = \begin{bmatrix} \epsilon_{i1} & \cdots & \epsilon_{iT} \end{bmatrix}' \) and \( E_i = \begin{bmatrix} e_{i1} & \cdots & e_{iT} \end{bmatrix}' \) are a random \( T \times 1 \) and \( T \times m \) matrices of idiosyncratic errors. This model is a prototypical version of the class of models advocated in the literature following the influential contribution of Pesaran (2006) and can be extended to allow for additional features. The assumptions we make in this paper are in essence the same as those made by Westerlund and Urbain (2015). Here and throughout \( M \) denotes a generic positive finite number and \( \text{vec} \) the vectorisation operator.

**Assumption 1** (Errors).

(i) \( \epsilon_{it} \sim i.i.d. (0, \sigma^2) \) with \( \mathbb{E}[\epsilon_{it}^6] < M. \)

(ii) \( e_{it} \sim i.i.d. (0, \Sigma) \) with \( \Sigma \) being a positive definite matrix and \( \mathbb{E} \left[ ||e_{it}||^8 \right] < M. \)

**Assumption 2** (Factors). \( f_t \) is covariance stationary with positive definite covariance matrix \( \Sigma_F \) and \( \mathbb{E} \left[ ||f_t||^4 \right] < M. \)

**Assumption 3** (Loadings).

(i) \( \lambda_i \) is random and identically distributed over \( i \) with \( \mathbb{E}[\lambda_i] = \lambda_0, \mathbb{E}[||\lambda_i||^4] < M \) and \( N^{-1} \sum_{i=1}^{N} \lambda_i' \lambda_i \xrightarrow{p} \Sigma_\lambda \) where the latter is a positive semi-definite \( r \times r \) matrix.

(ii) \( \Lambda_i \) is random and identically distributed over \( i \) with \( \mathbb{E}[\Lambda_i] = \Lambda_0, \mathbb{E}[||\Lambda_i||^4] < M \) and \( N^{-1} \sum_{i=1}^{N} \text{vec}(\Lambda_i)\text{vec}(\Lambda_i)' \xrightarrow{p} \Sigma_\Lambda, \) where \( \Sigma_\Lambda \) is a positive semi-definite matrix of dimension \( rm \times rm. \)

**Assumption 4** (Independence). \( \epsilon_i, E_j, \{\lambda_n, \Lambda_n\} \) and \( F \) are independent for all \( i, j, n. \)
Remark 1. Assumption 2 is standard in the literature on factor-augmented regressions and could be extended to more general models. Assumption 3 is fairly general in that it allows for correlated loadings, as opposed to the bulk of the existing literature on the CCE estimator which excludes this case by assuming the loadings to be independently distributed. Assumption 1 is more restrictive than in Pesaran (2006) in that it does not allow for serial correlation and heteroskedasticity. That being said, the i.i.d. framework is intentionally chosen to allow for a clear exposition of the Hausman approach to testing for cross-section dependence which is the key contribution of this paper.

It is well-known that a number of restrictions on the factor loadings are required in order to ensure consistency of the OLS estimator of $\beta$ in (1). Imposing $\lambda_i = 0$, or $\Lambda_i = 0$, or assuming independence of the loadings in $y$ and $X$ together with $\lambda_0 = 0$ and $\Lambda_0 = 0$ are sufficient for this purpose. A violation of these conditions renders the OLS estimator inconsistent and hence requires the use of a different estimator. Pesaran’s (2006) CCE estimator is by far the most popular approach to estimating regression models in panels with cross-section dependence. It is given by

$$\hat{\beta}_{CCE} = \left( \sum_{i=1}^{N} X_i'M_iX_i \right)^{-1} \sum_{i=1}^{N} X_i'M_iy_i,$$

where $M_i = I_T - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$ and $\hat{F} = N^{-1}\sum_{i=1}^{N} [y_i; X_i]$. Consistency and asymptotic normality of this estimator have been shown under quite general assumptions. However, the behaviour of the CCE estimators in the absence of latent factors has not yet been investigated. This is a nontrivial issue which can be appreciated by noting that

$$\hat{F} = \begin{bmatrix} \bar{\epsilon} & \bar{E} \\ \beta & 1_m \end{bmatrix},$$

if $\lambda_i = 0$ and $\Lambda_i = 0 \forall i$ where $[\bar{\epsilon}; \bar{E}] = N^{-1}\sum_{i=1}^{N} [\epsilon_i; E_i]$. By Assumption 1, this estimate converges to a zero matrix, so that asymptotically the inverse of $M_{\hat{F}}$ is not defined. To solve this problem, we can use the fact that the space spanned by a set of vectors does not depend on their scaling. Knowing that $[\bar{\epsilon}; \bar{E}]$ converges to zero at rate $\sqrt{N}$, it is possible to regularize the projection matrix by using the rescaled factor estimator $\sqrt{N}\hat{F}$ whose second moment matrix is asymptotically non-singular. This leads to the asymptotic properties of the CCE estimators in the absence of factors, which are summarized in Theorem 1.
Theorem 1. Under Assumptions 1–4 and assuming that $\lambda_i = 0$ and $\Lambda_i = 0$ \( \forall i \),

$$\sqrt{NT}(\hat{\beta}^{CCE} - \beta) = \sqrt{NT}(\hat{\beta}^{OLS} - \beta) + O_p(N^{-1/2}) + O_p(T^{-1/2}).$$

where $\sqrt{NT}(\hat{\beta}^{OLS} - \beta) \overset{D}{\to} N(0, \Sigma^{-1} \sigma^2)$ as $N, T \to \infty$.

Remark 2. In the absence of cross-section dependence, the model above is fairly restrictive since both the dependent variable and the covariates have expected value zero. However, it is straightforward to extend the result to a model of the form

$$y_i = \iota \gamma_i + X_i \beta + \epsilon_i$$
$$X_i = \iota \Gamma_i + E_i,$$

where $\gamma_i$ and $\Gamma_i$ are individual-specific constants. Theorem 1 can be shown to hold after conducting the within transformation on the model above. In this case, the CCE estimator converges to the Fixed Effects estimator.

Convergence of the CCE estimator to the OLS estimator in the absence of cross-section dependence is a surprising result, given that the former projects the data off an \( m + 1 \) dimensional space spanned by the factor estimates. In finite samples this projection has an effect on $\hat{\beta}^{CCE}$ since the idiosyncratic variation in all variables is by construction correlated with the factor estimate which is their cross-section average. However, the contribution of each cross-section to $\hat{F}$ decreases in $N$, so that in the limit $M\hat{F}$ is orthogonal to $E_i$ and $\epsilon_i$. Thus, the additional projection involved in the CCE estimator has a negligible effect. The fact that the CCE estimator converges to the OLS estimator also means that $\beta^{CCE}$ is asymptotically efficient in the absence of cross-section dependence while being inefficient in finite samples.

Furthermore, Theorem 1 does not restrict the relative expansion rate of $N$ and $T$. This result differs from Theorem 4 in Pesaran (2006), which requires $T/N \to 0$ when the data are characterized by a factor structure. As shown in Westerlund and Urbain (2015, Thm. 1), this restriction serves to avoid an asymptotic bias of order $T/N$, a term which is a function of the true factor loadings and which consequently disappears if all loadings are equal to zero.

3 The Hausman test for cross-section dependence

Hausman (1978) proposed a general framework for misspecification testing that avoids the need of precisely specifying the alternative hypothesis by focusing on the difference between
two estimators of which one is efficient, but inconsistent under the alternative hypothesis, whereas the other is always consistent. Taking the difference between the CCE estimator and the OLS estimator fits this framework and can be applied for testing cross-section dependence. The results of Theorem 1 point out a peculiarity of the CCE estimator. The context investigated by Hausman (1978) is one in which the consistent estimator is also inefficient. Under this assumption, the difference between two estimators that are \( \sqrt{NT} \)-consistent is \( O_p((NT)^{-1/2}) \).

However, the CCE estimator is asymptotically efficient since it converges to the OLS estimator so that

\[
\hat{\beta}_{CCE} - \hat{\beta}_{OLS} = \left( \hat{\beta}_{CCE} - \beta \right) - \left( \hat{\beta}_{OLS} - \beta \right) = O_p(N^{-1/2}T) + O_p(NT^{-1/2})
\]

by Theorem 1. This result suggests that tests based on the difference \( \hat{\beta}_{CCE} - \hat{\beta}_{OLS} \) should be more powerful than tests based on \( \hat{\beta}_{OLS} \) alone. Additionally, it emphasizes that the difference needs to be scaled up beyond the factor \( \sqrt{NT} \) to prevent it from converging to zero under the null hypothesis. An appropriate scaling is either \( \sqrt{NT} \) or \( N\sqrt{T} \), but each of these two choices entails a restriction on the relative expansion rate of \( N \) and \( T \). The bulk of the literature on factor-augmented regressions requires that \( T/N \to c < \infty \). In agreement with this condition, we choose the scaling \( \sqrt{NT} \) which yields

\[
\sqrt{NT} \left( \hat{\beta}_{CCE} - \hat{\beta}_{OLS} \right) = O_p(1) + O_p(\sqrt{T/N}).
\]

Theorem 2 below provides the asymptotic distribution of this expression.

**Theorem 2.** Suppose that assumptions 1–3 hold and that \( \lambda_i = 0, \Lambda_i = 0 \) \( \forall i = 1, \ldots, N \). Under the additional condition \( T/N \to c < \infty \) we have,

\[
\sqrt{NT} \left( \hat{\beta}_{CCE} - \hat{\beta}_{OLS} \right) \overset{D}{\to} N(0, (1 + m + c)\sigma^2 \Sigma^{-1}). \tag{3}
\]

The result above is convenient in two regards. First, it shows that the relatively heavy scaling does not introduce an asymptotic bias, so that a Hausman test statistic based on (3) does not need to be adjusted. Second, under the assumptions of this paper \( \text{Var}[\sqrt{NT}(\hat{\beta}_{CCE} - \hat{\beta}_{OLS})] \) is a fairly simple function of the covariance matrices of all the variables at hand. This expression can be substituted into the inverse in the Hausman test statistic instead of following the standard practice of using the difference \( \text{Var}[\hat{\beta}_{CCE}] - \text{Var}[\hat{\beta}_{OLS}] \). The infeasible Hausman
test statistic is hence defined as
\[ NT^2(\hat{\beta}^{\text{CCE}} - \hat{\beta}^{\text{OLS}})'((1 + m + T/N)\sigma^2\Sigma^{-1})^{-1}(\hat{\beta}^{\text{CCE}} - \hat{\beta}^{\text{OLS}}), \quad (4) \]

which, as a consequence of Theorem 2 is asymptotically \( \chi^2(m) \)-distributed. Under the null hypothesis of no cross-section dependence, consistent estimators of \( \sigma^2 \) and \( \Sigma \) respectively are straightforward to construct from the OLS residuals and the explanatory variables or analogously from the remainder after projecting off \( \hat{\beta} \). Using a direct estimator of the specific asymptotic covariance matrix in Theorem 2 has the additional advantage of providing a covariance matrix estimator that is positive (semi-)definite by construction, which is not necessarily the case when using the standard estimator based on \( \text{Var}[\hat{\beta}^{\text{CCE}}] - \text{Var}[\hat{\beta}^{\text{OLS}}] \). The feasible Hausman test statistic \( H \) is therefore given by
\[ H = NT^2(\hat{\beta}^{\text{CCE}} - \hat{\beta}^{\text{OLS}})'((1 + m + T/N)\sigma^2\hat{\Sigma}^{-1})^{-1}(\hat{\beta}^{\text{CCE}} - \hat{\beta}^{\text{OLS}}), \quad (5) \]

where \( \sigma^2 \) and \( \hat{\Sigma} \) are consistent estimators of \( \sigma^2 \) and \( \Sigma \). Theorem 3, expresses its properties under both the null and the alternative hypothesis. Here, and in the following, \( \lambda_i^{(r_1)} \) denotes element \( r_1 \) of the \( r \)-vector \( \lambda_0 \) whereas \( \Lambda_i^{(r_1,m_1)} \) is element \( (r_1,m_1) \) of the \( r \times m \) matrix \( \Lambda_0 \).

**Theorem 3.** Suppose that Assumptions 1–4 are satisfied, that \( T/N \to c < \infty \) and that the test statistic \( H \) is used to test
\[ H_0 : \lambda_i = 0, \Lambda_i = 0 \quad \forall i \]
against \( H_1 : \exists i = 1, \ldots, N \) s.t. \( \lambda_i \neq 0 \) or \( \Lambda_i \neq 0 \).

Then the following hold:
1. Under \( H_0 \), \( H \to \chi^2(m) \).
2. Under \( H_1 \), \( H = O_p(NT) \) if \( N_1 / N \) is bounded away from zero where \( N_1 \) is the number of cross-sections satisfying \( \lambda_i^{(r_1)} \neq 0 \) and \( \Lambda_i^{(r_1,m_1)} \neq 0 \) for some \( r_1 = 1, \ldots, r \) and \( m_1 = 1, \ldots, m \).
3. Under \( H_1 \), \( H = O_p(T) \) if \( N_1 / N \) is bounded away from zero, \( \text{Cov}[\lambda_i^{(r_1)}; \Lambda_i^{(r_1,m_1)}] = 0 \) and either \( \lambda_0^{(r_1)} = 0 \) or \( \Lambda_0^{(r_1,m_1)} = 0 \).

As Theorem 3 makes clear, the Hausman test for cross-section dependence requires that at least one factor has an impact on both \( y \) and \( X \). This is more restrictive than tests for cross-section dependence which are based on taking the sum of the pairwise correlation coefficients.
These latter tests have power even if only $y$ is cross-sectionally correlated. However, it is generally assumed nowadays that if cross-section dependence is suspected, most, if not all, variables are affected\(^2\). For this reason, the additional restrictions of the Hausman test for cross-section dependence can be expected to be of little practical relevance.

The most striking result of Theorem 3 is that $H$ has a higher asymptotic order than the popular CD test of Pesaran (2004) which diverges at the rate $N \sqrt{T}$ (see Pesaran, 2015, Theorem 3). The Hausman testing framework therefore provides a powerful alternative in panels where $T \geq N$. In order to achieve the divergence rate $O_p(NT)$, the Hausman test for cross-section dependence does not require all assumptions that are needed for the CCE estimator to be consistent. This concerns especially the so-called rank condition (see e.g. Pesaran, 2006, eq. 21) which effectively states $rk(\lambda_0; \Lambda_0) = r \leq m + 1$. $H = O_p(NT)$ if the rank of this matrix is at least one and in the case of correlated loadings even if $[\lambda_0; \Lambda_0] = 0$. The only restriction that reduces the asymptotic order of $H$ is zero correlation between the factor loadings together with zero expected values for the loadings on factors that affect both $y$ and $X$, a condition under which the OLS estimator is consistent. However, $H$ is still of order $O_p(T)$ in this case, which is a remarkable result for a Hausman test statistic whose power comes from using the fact that the efficient estimator is inconsistent under the alternative hypothesis. It can be explained by noting that the presence of nonzero factor loadings on the same factor for both $y$ and $X$ introduces a second order bias in $\sqrt{NT} (\hat{\beta}_{OLS} - \beta)$. This second-order bias is amplified to order $\sqrt{T}$ as the test statistic is scaled beyond the factor $\sqrt{NT}$.

### 4 Simulation results

In order to verify our theoretical results and to evaluate the performance of our Hausman test against other tests we conduct a small number of Monte Carlo experiments. As competing tests we choose Pesaran’s (2004) CD test, the most popular test in the literature, and the LM test of Schott (2005), which is a simple test that differs from the CD test in that it is constructed from squared cross-section correlations. Two different DGPs are considered. In DGP 1 the so-called rank condition of Pesaran (2006) is violated. The purpose of this choice is to investigate how the Hausman test performs relative to other tests if not all assumptions made for the CCE estimator are fulfilled. Additionally, we choose t-distributed idiosyncratic components with

\(^2\)see e.g. Pesaran et al. (2013) for an argumentation in the context of macroeconomic indicators.
fat tails. DGP 2 represents a situation in which the factor loadings are correlated. The number of repetitions is set to 5000. A full specification of both DGPs is provided in Table 4.

Table 1: DGP specifications for simulation exercise

<table>
<thead>
<tr>
<th>Component</th>
<th>DGP 1</th>
<th>DGP 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(r)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>(\beta)</td>
<td>(1_{m \times 1})</td>
<td>(1_{m \times 1})</td>
</tr>
<tr>
<td>(\epsilon_{i,t})</td>
<td>i.i.d. (t(9))</td>
<td>i.i.d. (N(0, 1))</td>
</tr>
<tr>
<td>(\epsilon_{i,t})</td>
<td>i.i.d. (t(9))</td>
<td>i.i.d. (N\left(\begin{bmatrix} 1 \ -0.5 \ 2 \end{bmatrix}\right))</td>
</tr>
</tbody>
</table>

Additionally for power simulations

\[
\lambda_0 = \iota_3 U_{1 \times 1}(1,2) \quad \text{and} \quad \Lambda_0 = \Omega_{3 \times 1}(\lambda_0) + U_{3 \times 1}(1.5,2.5) \]

\[
\lambda_i = \lambda_0 + U_{3 \times 1}(\pm 0.5,0.5) \quad \Lambda_i = \lambda_i \Omega_{3 \times 1} + U_{3 \times 2}(\pm 0.5,0.5) \]

\[
f_i = \text{diag}(0.5,0.7) f_{i-1} + \eta_i \quad \text{and} \quad \eta_i = \text{diag}(0.5,0.7) \eta_{i-1} + \iota_3 \Omega_{3 \times 2} \times 2(\pm 0.5,0.5) \]

The simulation results are presented in Tables 2-5. The size of all tests, as depicted in Table 2, is close to the nominal value of 5% in both DGPs. The Hausman test seems to be slightly oversized in small T samples. This observation reflects the fact that the variance used in equation (3) is a large sample approximation of the finite sample variance which has additional terms that go to zero as \(T\) increases but that turn out to have some impact in short panels.

Table 2: Size results

<table>
<thead>
<tr>
<th>(N)</th>
<th>(T)</th>
<th>(N)</th>
<th>(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{Hausman CD Schott})</td>
<td>(\text{Hausman CD Schott})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>25</td>
<td>0.0806</td>
<td>0.0566</td>
<td>0.0536</td>
</tr>
<tr>
<td>50</td>
<td>0.0702</td>
<td>0.0546</td>
<td>0.0532</td>
</tr>
<tr>
<td>100</td>
<td>0.0544</td>
<td>0.0518</td>
<td>0.0564</td>
</tr>
<tr>
<td>200</td>
<td>0.0562</td>
<td>0.056</td>
<td>0.058</td>
</tr>
<tr>
<td>50</td>
<td>0.0766</td>
<td>0.053</td>
<td>0.0584</td>
</tr>
<tr>
<td>100</td>
<td>0.0592</td>
<td>0.05</td>
<td>0.0526</td>
</tr>
<tr>
<td>200</td>
<td>0.0512</td>
<td>0.0488</td>
<td>0.0524</td>
</tr>
<tr>
<td>100</td>
<td>0.082</td>
<td>0.0496</td>
<td>0.055</td>
</tr>
<tr>
<td>200</td>
<td>0.0518</td>
<td>0.0502</td>
<td>0.0538</td>
</tr>
<tr>
<td>200</td>
<td>0.0716</td>
<td>0.0518</td>
<td>0.053</td>
</tr>
<tr>
<td>50</td>
<td>0.059</td>
<td>0.0462</td>
<td>0.053</td>
</tr>
<tr>
<td>100</td>
<td>0.0602</td>
<td>0.0438</td>
<td>0.0512</td>
</tr>
<tr>
<td>200</td>
<td>0.049</td>
<td>0.0474</td>
<td>0.0488</td>
</tr>
</tbody>
</table>
Looking at Table 3, which compares the power of all three tests considered, the merits of the Hausman test for cross-section correlation become apparent. The Hausman test has highest power whenever \( T \) is at least as large as \( N \). This is a reflection of Theorem 3 showing that the power of \( H \) is increasing at a higher rate in \( T \) relative to the CD test. However, the latter test is still superior in cases where \( N \) is larger than \( T \). This suggests that the CD test may be more powerful in the classical micro-data panel setting with large \( N \) and small \( T \) but that its performance is inferior to a Hausman testing approach in macro panels where both \( N \) and \( T \) are large. The LM test of Schott (2005) has lowest power among all three test, its rejection rates being surpassed by a large margin. An interesting observation is that the Hausman test performs well even in DGP 1 where the rank condition of Pesaran (2006) is violated, this condition being a requirement for the entire \( r \)-dimensional space spanned by the latent factors to be estimated.

### Table 3: Power results

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.2014</td>
<td>0.141</td>
<td>0.0612</td>
<td>0.1458</td>
<td>0.0978</td>
<td>0.0562</td>
</tr>
<tr>
<td>50</td>
<td>0.3814</td>
<td>0.1996</td>
<td>0.0628</td>
<td>0.235</td>
<td>0.1118</td>
<td>0.0616</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.6782</td>
<td>0.3022</td>
<td>0.0636</td>
<td>0.4396</td>
<td>0.1642</td>
<td>0.0622</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.927</td>
<td>0.4942</td>
<td>0.0906</td>
<td>0.7502</td>
<td>0.242</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.3106</td>
<td>0.305</td>
<td>0.0654</td>
<td>0.1922</td>
<td>0.1578</td>
<td>0.0488</td>
</tr>
<tr>
<td>50</td>
<td>0.6228</td>
<td>0.4592</td>
<td>0.0664</td>
<td>0.4064</td>
<td>0.2478</td>
<td>0.0576</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.9296</td>
<td>0.6484</td>
<td>0.0854</td>
<td>0.7536</td>
<td>0.4086</td>
<td>0.0592</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.9968</td>
<td>0.8294</td>
<td>0.113</td>
<td>0.972</td>
<td>0.617</td>
<td>0.0712</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.472</td>
<td>0.585</td>
<td>0.065</td>
<td>0.2976</td>
<td>0.3562</td>
<td>0.0574</td>
</tr>
<tr>
<td>50</td>
<td>0.8574</td>
<td>0.7882</td>
<td>0.078</td>
<td>0.6494</td>
<td>0.5446</td>
<td>0.064</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.9954</td>
<td>0.9278</td>
<td>0.123</td>
<td>0.956</td>
<td>0.7832</td>
<td>0.0722</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>0.9922</td>
<td>0.2122</td>
<td>0.9992</td>
<td>0.9196</td>
<td>0.0908</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.6794</td>
<td>0.8538</td>
<td>0.0802</td>
<td>0.4766</td>
<td>0.676</td>
<td>0.0718</td>
</tr>
<tr>
<td>50</td>
<td>0.9686</td>
<td>0.9646</td>
<td>0.1154</td>
<td>0.859</td>
<td>0.8678</td>
<td>0.0676</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.9996</td>
<td>0.997</td>
<td>0.2148</td>
<td>0.9968</td>
<td>0.9704</td>
<td>0.0918</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>1</td>
<td>0.4108</td>
<td>1</td>
<td>0.9944</td>
<td>0.1522</td>
<td></td>
</tr>
</tbody>
</table>

We now consider two restricted cases of DGP 1 and 2. The first restriction sets \( \lambda_0 = 0 \) and \( \Lambda_0 = 0 \) except for \( \lambda_0^{(1)} \) and \( \lambda_0^{(1,1)} \) which are defined as in Table 4. All other parameters correspond to the previous specifications in DGPs 1 and 2. This restriction implies a different violation of the rank condition of Pesaran (2006) than the one given in DGP 1. The results reported in Table 4 show that the power of the Hausman test is negatively affected by factor loadings with zero means. Most strikingly, decent performance requires that \( T \) is sufficiently large. On the contrary, in short panels only modest power gains can be achieved as \( N \) increases. With
regards to the unrestricted DGPs considered before, the Hausman test loses its performance advantage relative to the CD test and exhibits rejection rates that are mostly below the latter. However, the performance of Schott’s LM test is still surpassed. Most importantly, the numbers in Table 4 verify the theoretical results of Theorem 3; nonzero expected values for the same factor in the dependent variable and at least one explanatory variable are enough to yield power increases in both $N$ and $T$.

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.0856</td>
<td>0.0714</td>
<td>0.0568</td>
<td>0.092</td>
<td>0.1226</td>
<td>0.0626</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.0982</td>
<td>0.0846</td>
<td>0.06</td>
<td>0.1152</td>
<td>0.1592</td>
<td>0.0576</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.1282</td>
<td>0.1006</td>
<td>0.0594</td>
<td>0.162</td>
<td>0.225</td>
<td>0.061</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>0.188</td>
<td>0.1276</td>
<td>0.06</td>
<td>0.3054</td>
<td>0.3716</td>
<td>0.0752</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.088</td>
<td>0.1102</td>
<td>0.057</td>
<td>0.1056</td>
<td>0.2332</td>
<td>0.0612</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.114</td>
<td>0.1472</td>
<td>0.0572</td>
<td>0.1542</td>
<td>0.3728</td>
<td>0.065</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.2104</td>
<td>0.2164</td>
<td>0.0606</td>
<td>0.328</td>
<td>0.559</td>
<td>0.0678</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>0.4188</td>
<td>0.3218</td>
<td>0.0628</td>
<td>0.6514</td>
<td>0.7752</td>
<td>0.0876</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.1116</td>
<td>0.229</td>
<td>0.0592</td>
<td>0.135</td>
<td>0.5168</td>
<td>0.0704</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.1664</td>
<td>0.331</td>
<td>0.0548</td>
<td>0.2624</td>
<td>0.724</td>
<td>0.0652</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.393</td>
<td>0.5012</td>
<td>0.052</td>
<td>0.5968</td>
<td>0.894</td>
<td>0.08</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>0.7644</td>
<td>0.6856</td>
<td>0.0704</td>
<td>0.9166</td>
<td>0.9804</td>
<td>0.1302</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.1396</td>
<td>0.4728</td>
<td>0.0616</td>
<td>0.1934</td>
<td>0.8238</td>
<td>0.0658</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.286</td>
<td>0.653</td>
<td>0.0558</td>
<td>0.4468</td>
<td>0.9582</td>
<td>0.0906</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>0.6522</td>
<td>0.8302</td>
<td>0.0682</td>
<td>0.845</td>
<td>0.9966</td>
<td>0.1324</td>
</tr>
<tr>
<td>200</td>
<td></td>
<td>0.9582</td>
<td>0.9422</td>
<td>0.0762</td>
<td>0.9944</td>
<td>0.9998</td>
<td>0.252</td>
</tr>
</tbody>
</table>

In the second restriction to DGP 1 and 2, the expected values of all loadings are set to zero. Furthermore, the variance of $\eta_t$ is increased to 1. The restricted version of DGP 1 illustrates the case where $H = O_p(T)$ by Theorem 3. In fact, given that the loadings in this case are uncorrelated, the power of the Hausman test increases mostly in $T$, and not so much in $N$. The same behaviour is observed for the CD test, albeit at rejection levels that are far lower than those of the Hausman test. The bad performance of the CD test reflects a well-known weakness: Given that the test statistic is a simple sum of the pairwise cross-section correlations, it is expected to perform badly if the cross-section correlations cancel out each other (see e.g. Pesaran, 2004). The LM-test of Schott is robust against cross-section correlations that cancel out, which results in the highest rejection rates among all three tests.

In the case of DGP 2, correlation between the factor loadings in $y$ and $X$ is enough to raise the asymptotic order of $H$ to $O_p(NT)$. This is in line the results of Theorem 3 and gets the rejection rates of $H$ closer to those of the LM test which performs best. The CD test statistic is
only of order \( O_p(T) \) and exhibits only fairly modest power gains as \( T \) increases. An explanation of this result is found in noting that the CD test uses only the unexplained variation in \( y \) to test for cross-section dependence. In contrast to the Hausman test which uses all variables, correlation between the loadings in \( y \) and \( X \) can hence not be employed to improve power.

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
<th>Hausman</th>
<th>CD</th>
<th>Schott</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.153</td>
<td>0.064</td>
<td>0.5656</td>
<td>0.2196</td>
<td>0.067</td>
<td>0.6164</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.1966</td>
<td>0.0864</td>
<td>0.9162</td>
<td>0.331</td>
<td>0.0858</td>
<td>0.9558</td>
</tr>
<tr>
<td>100</td>
<td>289</td>
<td>0.289</td>
<td>0.1192</td>
<td>0.9994</td>
<td>0.5012</td>
<td>0.1316</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>4104</td>
<td>0.4104</td>
<td>0.184</td>
<td>1</td>
<td>0.6536</td>
<td>0.2028</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.1556</td>
<td>0.063</td>
<td>0.8872</td>
<td>0.3004</td>
<td>0.0692</td>
<td>0.921</td>
</tr>
<tr>
<td>50</td>
<td>2072</td>
<td>0.2072</td>
<td>0.086</td>
<td>0.9986</td>
<td>0.4896</td>
<td>0.0894</td>
<td>0.9996</td>
</tr>
<tr>
<td>100</td>
<td>3168</td>
<td>0.3168</td>
<td>0.1238</td>
<td>1</td>
<td>0.6858</td>
<td>0.1328</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>4672</td>
<td>0.4672</td>
<td>0.1728</td>
<td>1</td>
<td>0.845</td>
<td>0.2032</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>0.147</td>
<td>0.0556</td>
<td>0.9946</td>
<td>0.411</td>
<td>0.073</td>
<td>0.9972</td>
</tr>
<tr>
<td>50</td>
<td>2168</td>
<td>0.2168</td>
<td>0.0846</td>
<td>1</td>
<td>0.6314</td>
<td>0.082</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>3502</td>
<td>0.3502</td>
<td>0.1154</td>
<td>1</td>
<td>0.8398</td>
<td>0.1352</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>5294</td>
<td>0.5294</td>
<td>0.1928</td>
<td>1</td>
<td>0.9434</td>
<td>0.2018</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>0.1592</td>
<td>0.0598</td>
<td>1</td>
<td>0.5244</td>
<td>0.0642</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>2282</td>
<td>0.2282</td>
<td>0.0848</td>
<td>1</td>
<td>0.7484</td>
<td>0.0882</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>3794</td>
<td>0.3794</td>
<td>0.1208</td>
<td>1</td>
<td>0.9252</td>
<td>0.1334</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>5722</td>
<td>0.5722</td>
<td>0.1826</td>
<td>1</td>
<td>0.9806</td>
<td>0.1994</td>
<td>1</td>
</tr>
</tbody>
</table>

### 5 Conclusion

This article develops a Hausman test for cross-section dependence based on the CCE estimator of Pesaran (2006). We find that under the null hypothesis of no cross-section dependence the CCE estimator converges to the OLS estimator and that the difference between the two converges to zero at a rate higher than \( \sqrt{NT} \). This finding is used to improve the power of the Hausman test by scaling the test statistic with a higher factor and has two important consequences. First, the test statistic may diverge even in cases where the OLS estimator is not inconsistent because the additional scaling amplifies a second order bias in the OLS estimator. Second, under the alternative hypothesis the the Hausman test for cross-section dependence diverges at a higher rate than the popular CD test of Pesaran (2004). This provides a powerful alternative to the latter test especially in classical macroeconomic panels where both dimensions are large. Simulation results suggest that the test has high power while keeping its size close to the nominal value.
References


A  Notation

In the following lemmas and proofs, individual vectors and single elements of matrices will be used. The notation follows the standard practice of writing matrices as bold capital letters and vectors as bold lowercase letters as far as possible and the indexing is as follows:

- The $T \times m$ matrix $E_i$ has vectors $e_i^{(k)}$ and $e_i^{(l)}$, which are of dimension $T \times 1$ and $m \times 1$, respectively. Individual elements are denoted $e_i^{(k)}$.

- $\Sigma$ has $k$-vectors $\Sigma^{(k)}$ and elements $\Sigma^{(k,l)}$. $(\Sigma^{-1})^{(k)}$ and $(\Sigma^{-1})^{(k,l)}$ are vectors and elements of $\Sigma^{-1}$.

- $\Lambda_i$ has $r$-vectors $\Lambda_i^{(m_1)}$ and elements $\Lambda_i^{(r_1,m_1)}$. The indexing of $\lambda_i$ is analogous.

B  Additional lemmas

**Lemma 1.** Under assumption 1,

$$\frac{(NT)^{-2}}{N} \sum_{i,j,m} E_i^{(k)} e_j^{(l)} e_m^{(l)} E_i = N^{-1} \Sigma^{(k)} \Sigma^{(l)} + T^{-1} \Sigma^{(k,l)} \Sigma + O_p(N^{-1/2} T^{-1})$$ (6)

$$\frac{(NT)^{-2}}{N} \sum_{i,j,m} E_i e_j e_m E_i = T^{-1} \sigma^2 \Sigma + O_p(N^{-1/2} T^{-1})$$ (7)

$$\frac{(NT)^{-2}}{N} \sum_{i,j,m} E_i^{(k)} e_j^{(l)} e_m^{(l)} E_i = O_p(N^{-1} T^{-1}) + O_p(N^{-1/2} T^{-1})$$ (8)

$$\frac{(NT)^{-2}}{N} \sum_{i,j,m} E_i^{(k)} e_j^{(l)} e_m^{(l)} E_i = O_p(N^{-1} T^{-1}) + O_p(N^{-1/2} T^{-1})$$ (9)
Lemma 2. Under assumption 1,
\[
\mathbb{E} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j e_m e_i \right] = 0 \quad (10)
\]
\[
\text{Var} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j e_m e_i \right] = (N^{-1} + T^{-1}) \Sigma \sigma^6 \quad (11)
\]
\[
\mathbb{E} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = 0 \quad (12)
\]
\[
\text{Var} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = N^{-1} \Sigma^{(l,k)} \Sigma^{(l,j)} \sigma^2 + T^{-1} \Sigma \Sigma^{(k,k)} \Sigma^{(l,l)} \sigma^2 \quad (13)
\]
\[
\mathbb{E} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = (NT)^{-1/2} \Sigma^{(l,j)} \sigma^2 \quad (14)
\]
\[
\text{Var} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = O(T^{-1}) \quad (15)
\]
\[
\mathbb{E} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = \sqrt{\frac{T}{N}} \Sigma^{(l,k)} \sigma^2 \quad (16)
\]
\[
\text{Var} \left[ (NT)^{-3/2} \sum_{i,j,m} E'_i e_j^{(k)} e_m^{(l)} e_i \right] = O_p(T^{-1}) \quad (17)
\]

Lemma 3. Let \( V = (v^{(c,d)}) = \frac{1}{T} \tilde{F} \tilde{F} \). Under assumption 1,
\[
v^{(1,1)} = \sigma^2 + O_p(T^{-1/2})
\]
\[
\sqrt{T} v^{(1,k+1)} = \frac{1}{N \sqrt{T}} \sum_{i,j} \sum_{t=1}^T e_t^{(k)} e_{it} = O_p(1)
\]
\[
v^{(k+1,1)} = v^{(1,k+1)}
\]
\[
v^{(k+1,l+1)} = \Sigma^{(k,l)} + O_p(T^{-1/2}),
\]
for \( k, l = 1, \ldots, m \).

Lemma 4. Let \( W = (w^{(c,d)}) = \left( \frac{1}{T} \tilde{F} \tilde{F} \right)^{-1} \). Under assumption 1,
\[
w^{(1,1)} = \sigma^{-2} + o_p(1)
\]
\[
\sqrt{T} w^{(k,1)} = -\sigma^{-2} (\Sigma^{-1})^{(k, \cdot)} \frac{1}{N \sqrt{T}} \sum_{i,j} \sum_{t=1}^T e_{it} e_{jt} = O_p(1)
\]
\[
w^{(1,k+1)} = w^{(k+1,1)}
\]
\[
w^{(k+1,l+1)} = (\Sigma^{-1})^{(k,l)} + o_p(1)
\]
for \( k, l = 1, \ldots, m \).
Lemma 5. Let
\[ h_1 = \Sigma^{-1} \sigma^{-2} \frac{1}{N^{3/2} T} \sum_{i,j,m} \sum_{t,s} e_{i,t} e_{m,s} e_{i,s}, \]
\[ h_{2,kl} = \Sigma^{-1} (\Sigma^{-1})^{(k,l)} \frac{1}{N^{3/2} T} \sum_{i,j,m} \sum_{t,s} e_{i,t} e_{j,t} e_{m,s} e_{i,s}, \]
\[ h_3 = \Sigma^{-1} \frac{1}{N \sqrt{T}} \sum_{i,j} \sum_{t=1}^T e_{i,t} e_{j,t}. \]

Under assumptions 1 and 4,
\[ \text{Cov}[h_1; h_{2,kl}] = \frac{T}{N} \sigma^2 \Sigma^{-1} \sigma^{-1} (\Sigma^{-1})^{(k,l)} \Sigma^{-1} \Sigma, \]
(18)
\[ \text{Cov}[h_1; h_3] = \sqrt{\frac{T}{N}} \sigma^2 \Sigma^{-1}, \]
(19)
\[ \text{Cov}[h_{2,kl}; h_3] = \sqrt{\frac{T}{N}} \sigma^2 \Sigma^{-1} (\Sigma^{-1})^{(k,l)} \Sigma^{-1} \Sigma, \]
(20)

Proof of Lemma 1. Consider the expected value of the first result. We have
\[
\mathbb{E}[(NT)^{-2} \sum_{i,j,m} E[e_j^{(k)} e_m^{(l)}]^t e_i] = (NT)^{-2} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}[e_{i,t} e_{j,t}^{(k)} e_{i,t}^{(l)} e_{j,t}^t] \]
\[ + (NT)^{-2} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}[e_{j,t}^{(k)} e_{j,t}^{(l)}] \mathbb{E}[e_{i,t} e_{j,t}^t], \]
\[ = N^{-1} \Sigma^{(k)} \Sigma^{(l)} + T^{-1} \Sigma^{(k,l)} \Sigma + O((NT)^{-1}), \]
where all remaining terms are equal zero by the law of iterated expectations. Deviations from the expected values above are of order \( O_p(N^{-1/2} T^{-1}) \), provided that \( \mathbb{E} [||e_{i,t}||^8] < M < \infty \), so that the entire term reduces to
\[
(NT)^{-2} \sum_{i,j,m} E[e_j^{(k)} e_m^{(l)}]^t e_i = N^{-1} \Sigma^{(k)} \Sigma^{(l)} + T^{-1} \Sigma^{(k,l)} \Sigma + O_p(N^{-1/2} T^{-1})
\]
We continue with the second result where
\[
\mathbb{E}[(NT)^{-2} \sum_{i,j,m} E[e_j e_m e_i] = (NT)^{-2} \sum_{i=1}^N \sum_{j \neq i} \sum_{t=1}^T \mathbb{E}[e_{i,t} e_{j,t}] \mathbb{E}[e_{i,t} e_{j,t}^t] + 0 \]
\[ = T^{-1} \sigma^2 \Sigma. \]
so that

$$\mathbb{E}[NT^{-2} \sum_{i,j,m} E[e_i e_m' E_i]] = T^{-1} \sigma^2 \Sigma + O_p(N^{-1/2}T^{-1}).$$

Next, note that the third result has an expected value of zero and that

$$\mathbb{V} \text{ar}[(NT)^{-2} \sum_{i,j,m} E[e_i^{(k)} e_m^{(l)} E_i]] = (NT)^{-4} \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i^{(k)} e_j^{(k)} e_m^{(l)} e_{i,t} e_{j,s} e_{m,s} e_{o,q} e_{p,r} e_{n,q}]$$

$$= (NT)^{-4} \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i^{(k)} e_j^{(k)} e_m^{(l)} e_{i,s} e_{j,t} e_{m,s} e_{o,p} e_{r,q} e_{n,q}]$$

$$= (NT)^{-4} \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i^{(k)} e_j^{(k)} e_m^{(l)} e_{i,s} e_{j,t} e_{m,s} e_{o,p} e_{r,q} e_{n,q}]$$

$$+ (NT)^{-4} \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i^{(k)} e_j^{(k)} e_m^{(l)} e_{i,s} e_{j,t} e_{m,s} e_{o,p} e_{r,q} e_{n,q}]$$

$$= O(N^{-1}T^{-2}) + O(N^{-2}T^{-1}),$$

implying that

$$(NT)^{-2} \sum_{i,j,m} E[e_j e_m' E_i] = O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}).$$

In the same fashion, it can be shown that

$$(NT)^{-2} \sum_{i,j,m} E[e_j e_m' E_i] = O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}).$$

\[ \square \]

**Proof of Lemma 2.** Consider the first result. By independence of \( E \) and \( \epsilon \), the expected value of

$$O(N^{-3/2}) \sum_{i,j,m} E[e_i e_m' e_i]$$

is zero. In order to obtain an expression of the variance matrix, note that

$$\mathbb{V} \text{ar}[(NT)^{-3/2} \sum_{i,j,m} E[e_i e_m' e_i]] = \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i e_j e_m e_{i,t} e_{j,s} e_{m,s} e_{o,q} e_{p,r} e_{n,q}]$$

$$= (NT)^{-3} \sum_{i,j} \sum_{t,s} \mathbb{E}[e_i e_j e_{i,t} e_{j,s} e_{m,s} e_{o,q} e_{p,r} e_{n,q}]$$

$$+ (NT)^{-3} \sum_{i,j,m} \sum_{t,s} \mathbb{E}[e_i e_j e_{m,s} e_{o,q} e_{p,r} e_{n,q}]$$

$$= (N^{-1} + T^{-1}) \Sigma \sigma^6,$$
where the above expression involves some simplification for notational convenience since e.g. 
$$\sum_{i,j}^{N} \sum_{t,s,q}^{T} E[e_{it} e'_{it}] E[e_{is}^2] E[e_{iq}^2]$$ involves the term $(NT)^{-3} \sum_{i=1}^{N} \sum_{t=1}^{T} E[e_{it} e'_{it}] E[e_{it}^2]$. However, the asymptotic order of this latter and related terms is dominated by the terms written out above.

Next, note that $E[(NT)^{-3/2} \sum_{i,j,m}^{N} E_{i}^{(k)} e_{m}^{(l)} \epsilon_{i}] = 0$ due to independence of $E$ and $\epsilon$. The variance is given by

$$\text{Var}[(NT)^{-3/2} \sum_{i,j,m}^{N} E_{i}^{(k)} e_{m}^{(l)} \epsilon_{i}]$$

$$= (NT)^{-3} \sum_{i,j,m}^{N} \sum_{t,s,q}^{T} E[e_{it} e'_{it} e_{is} e_{iq} e_{mp} e_{rs}] E[\epsilon_{i} \epsilon_{p}]$$

$$= (NT)^{-3} \sum_{i,j,m}^{N} \sum_{t,s,q}^{T} E[e_{it} e_{is}] E[e_{ix}^2] E[|e_{is}|^2] + (NT)^{-3} \sum_{i,j,m}^{N} \sum_{t,s,q}^{T} E[e_{it} e_{is}] E[|e_{ix}|^2] E[|e_{is}|^2] E[\epsilon_{i}^2]$$

$$= N^{-1} \sum_{x}^{T} \sum_{x}^{M} \sum_{x}^{Q} \sum_{T}^{(x)} \sigma^2 + T^{-1} \sum_{x}^{T} \sum_{x}^{M} \sum_{x}^{Q} \sum_{T}^{(x)} \sigma^2$$

where we take into account some abuse of notation which asymptotically is of no relevance.

For the third result, note that

$$E[(NT)^{-3/2} \sum_{i,j,m}^{N} E_{i}^{(k)} e_{m}^{(l)} \epsilon_{i}] = (NT)^{-3/2} \sum_{i=1}^{N} \sum_{t=1}^{T} E[e_{it} e_{it}^2] E[\epsilon_{i}^2]$$

$$= (NT)^{-1/2} \sum_{x}^{T} \sum_{x}^{M} \sum_{x}^{Q} \sum_{T}^{(x)} \sigma^2$$

Hence,

$$(NT)^{-3/2} \sum_{i,j,m}^{N} E_{i}^{(k)} e_{m}^{(l)} \epsilon_{i} = (NT)^{-1/2} E[e_{i} e_{i}^2] E[\epsilon_{i}^2] + (NT)^{-3/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( e_{it} e_{it}^2 - E[e_{it} e_{it}^2] E[\epsilon_{i}^2] \right)$$

The term in the last line has a variance of order $O(T^{-1})$, which follows from application of a Lindeberg-Levi CLT. Finally, the expected values of the fourth term are

$$E[(NT)^{-3/2} \sum_{i,j,m}^{N} E_{i}^{(k)} e_{m}^{(l)} \epsilon_{i}] = (NT)^{-3/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E[e_{it} e_{is} e_{j} e_{is}]$$

$$= \sqrt{\frac{T}{N}} \sum_{x}^{T} \sum_{x}^{M} \sum_{x}^{Q} \sum_{T}^{(x)} \sigma^2.$$
Proof of Lemma 3. The result follows from
\[
\frac{1}{T} \tilde{F}' \tilde{F} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \left[ e_{i,t} e_{j,t}' - \mathbb{1}_{(i=j)} \sigma^2 \right] e_{i,t} e_{j,t}'
\]
where the asymptotic order of the second matrix in line 2 follows from application of a Lindeberg-Levi CLT. \(\square\)

Proof of Lemma 4. Let \( V \) from Lemma 3 be partitioned
\[
V = \begin{bmatrix} V_1 & V_2' \\ V_2 & V_3 \end{bmatrix},
\]
where \( V_1 = O_p(1) \) is a scalar, \( V_2 = O_p(T^{-1/2}) \) a \( m \)-vector and \( V_3 = O_p(1) \) an invertible \( m \times m \) matrix. Next, define \( W = \left( \frac{1}{T} \tilde{F} \tilde{F} \right)^{-1} \). Using Harville (1997, eq.5.16a), we can write
\[
W = \begin{bmatrix} W_1 & W_2' \\ W_2 & W_3 \end{bmatrix} = \begin{bmatrix} V_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -V_1^{-1}V_2' \\ V_2 \end{bmatrix} (V_3 - V_2V_1^{-1}V_2')^{-1} \begin{bmatrix} -V_2V_1^{-1} \\ V_2 \end{bmatrix}
\]
Now, \( V_2V_1^{-1}V_2' = O_p(T^{-1}) \) and \( V_3 = \Sigma + O_p(T^{-1/2}) \). Together with the fact that \( \Sigma \) is positive definite by assumption, this implies \( W_3 = (V_3 - V_2V_1^{-1}V_2')^{-1} = \Sigma^{-1} + O_p(T^{-1/2}) \) by Kari-biyik et al. (2016, Th. 1). Further noticing that \( V_2 = O_p(T^{-1/2}) \) yields \( W_1 = \sigma^{-2} + O(T^{-1/2}) \). Using these two latter results, we obtain \( \sqrt{T}W_2 = -\Sigma^{-1} \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} e_{i,t} e_{j,t} \sigma^{-2} + O_p(T^{-1/2}) \). Considering the individual elements of \( W_1, W_2 \) and \( W_3 \) yields the results of this lemma. \(\square\)

Proof of Lemma 5. First,
\[
\text{Cov} [ h_{1l}, h_{2kl} ] = \Sigma^{-1} \frac{1}{N^3T^2} \sum_{i,j,m} \sum_{t,s} \sum_{r,s} \sigma^{-2} \mathbb{E} \left[ e_{i,t} e_{j,t} e_{m,s} e_{r,s} e_{n,t} e_{n,s} e_{p,r} e_{p,s} e_{p,t} e_{p,s} \right] \Sigma_{(k,l)}^{-1} \Sigma^{-1}
\]
\[
= \Sigma^{-1} \sigma^{-2} \left( \frac{1}{N^3T^2} \sum_{i,j,m} \sum_{t,s} \sum_{r,s} \mathbb{E} \left[ e_{i,t} e_{j,t} e_{m,s} e_{r,s} e_{n,t} e_{n,s} e_{p,r} e_{p,s} \right] \right) \Sigma_{(k,l)}^{-1} \Sigma^{-1}
\]
\[
= \frac{T}{N} \sigma^2 \Sigma^{-1} \mathbb{E} (\Sigma^{-1}(\Sigma^{-1}(\Sigma^{-1})) (\Sigma^{-1}(\Sigma^{-1})) \Sigma^{-1}
\]
Second,
\[
\text{Cov}[h_1; h_3] = \sigma^{-2} \Sigma^{-1} \frac{1}{N^{5/2}T^{3/2}} \sum_{i,j}^{N} \sum_{t,s}^{T} \mathbb{E} \left[ e_{i,t} e_{j,t} e_{m,s} e_{s,t} e_{p,q} e'_{n,q} \right] \Sigma^{-1}
\]
\[
= \sigma^{-2} \Sigma^{-1} \frac{1}{N^{5/2}T^{3/2}} \sum_{i,j}^{N} \sum_{t,s}^{T} \mathbb{E} \left[ e_{i,t} e'_{n,q} \right] \mathbb{E} \left[ e_{j,t} \right] \mathbb{E} \left[ e_{m,s} \right] \Sigma^{-1}
\]
\[
= \sqrt{\frac{T}{N}} \sigma^{-2} \Sigma^{-1}.
\]

Third,
\[
\text{Cov}[h_{2,kl}; h_3] = \Sigma^{-1} \Sigma_{(k,l)}^{-1} \frac{1}{N^{5/2}T^{3/2}} \sum_{i,j}^{N} \sum_{t,s}^{T} \mathbb{E} \left[ e_{i,t} e'_{j,t} e_{m,s} e_{s,t} e_{p,q} e'_{n,q} \right] \Sigma^{-1}
\]
\[
= \Sigma^{-1} \Sigma_{(k,l)}^{-1} \frac{1}{N^{5/2}T^{3/2}} \sum_{i,j}^{N} \sum_{t,s}^{T} \mathbb{E} \left[ e_{i,t} e'_{j,t} \right] \mathbb{E} \left[ e_{m,s} e'_{n,q} \right] \mathbb{E} \left[ e_{s,t} \right] \Sigma^{-1}
\]
\[
= \sqrt{\frac{T}{N}} \sigma^{-2} \Sigma^{-1} \Sigma_{(k,l)}^{-1} \Sigma_{(l,k)}^{-1} \Sigma^{-1}
\]

\[\square\]

C Proofs on main results

Proof of Proposition 1. Under the assumptions of Theorem 1, the model reduces to
\[
y_i = X_i \beta + \epsilon_i
\]
\[
X_i = E_{i,\epsilon}
\]
so that the CCE estimator is effectively given by
\[
\hat{\beta}^{CCE} = \left( \sum_{i=1}^{N} E_i^t \hat{M}_E E_i \right)^{-1} \left( \sum_{i=1}^{N} E_i^t \hat{M}_E (E_i \beta + \epsilon_i) \right),
\]
where
\[
\hat{M}_E = \begin{bmatrix} \mathbb{E} & \mathbb{E} \\ \mathbb{0} & \mathbb{0} \end{bmatrix}
\]
\[
= \begin{bmatrix} \mathbb{0} & \mathbb{1} \end{bmatrix} \begin{bmatrix} \beta & \mathbb{I}_m \end{bmatrix}, \tag{22}
\]

In addition to regularizing the projection matrix by scaling \( \hat{M}_E \) with the factor \( \sqrt{N} \), we use the fact that orthogonal projection matrices are invariant to invertible transformations of the vectors from which they are constructed see e.g. (Davidson and Mackinnon, 2009, Ch. 2.3). The
last matrix on the right hand side of (22) is by construction invertible. Hence, (24) will be
analysed as a function of the alternative estimator

$$\tilde{F} = \sqrt{N} \begin{bmatrix} \varphi & \mathbf{E} \end{bmatrix}. \quad (23)$$

 Appropriately centered and scaled, the CCE estimator can be expressed as

$$\sqrt{NT} (\hat{\beta}^{CCE} - \beta) \quad (24)$$

$$= \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{M}_{\bar{F}} \mathbf{E}_i \right)^{-1} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \bar{F} \mathbf{e}_i \right)$$

$$= \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i - \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{M}_{\bar{F}} \mathbf{e}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \mathbf{P}_{\bar{F}} \mathbf{e}_i$$

$$+ \left[ \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{M}_{\bar{F}} \mathbf{e}_i \right)^{-1} - \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i \right)^{-1} \right] \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i$$

$$= \sqrt{NT} (\hat{\beta}^{OLS} - \beta) - g_1 + g_2. \quad (25)$$

The asymptotic properties of $\sqrt{NT} (\hat{\beta}^{OLS} - \beta)$ are straightforward to derive. First,

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{e}_i \mathbf{e}_t' = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Sigma + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{e}_i \mathbf{e}_t' - \Sigma)$$

$$= \Sigma + O_p((NT)^{-1/2}), \quad (26)$$

by application of a Lindeberg-Levi CLT. Second,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i \overset{D}{\rightarrow} N(0, \Sigma \sigma^2), \quad (27)$$

by the same reasoning, which yields $\sqrt{NT} (\hat{\beta}^{OLS} - \beta) \overset{D}{\rightarrow} N(0, \sigma^2 \Sigma^{-1})$. Next, consider $g_2$, which can be written

$$g_2 = \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{M}_{\bar{F}} \mathbf{E}_i \right)^{-1} \left[ \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{P}_{\bar{F}} \mathbf{E}_i \right] \left( \frac{1}{NT} \sum_{i=1}^{N} \mathbf{E}' \mathbf{E}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{E}' \mathbf{e}_i. \quad (28)$$
Here, let $P_\tilde{T} = T^{-1}\tilde{F}W\tilde{F}'$ where $W = (T^{-1}\tilde{F}'\tilde{F})^{-1}$ is a square matrix with elements $w^{(c,d)}, c, d = 1, \ldots, m + 1$. Using this notation, we can write

$$\frac{1}{NT}\sum_{i=1}^{N} E_i'P_\tilde{T}E_i = \frac{1}{NT^2}\sum_{i=1}^{N} E_i'FWF'E_i$$

$$= w^{(1,1)} \frac{1}{T^2} \sum_{i=1}^{N} E_i'e'e'E_i + \sum_{k,l} w^{(k+1,j+1)} \frac{1}{T^2} \sum_{i=1}^{N} E_i'e^{(k)}e^{(l)'}E_i$$

$$+ \sum_{k=1}^{m} w^{(k+1,1)} \frac{1}{T^2} \sum_{i=1}^{N} E_i'e^{(k)}e'E_i + \sum_{l=1}^{m} w^{(1,j+1)} \frac{1}{T^2} \sum_{i=1}^{N} E_i'e^{(l)}e'E_i$$

$$= \sigma^{-2} \frac{1}{(NT)^2} \sum_{i,j,m} E_ie_je_m' + \sum_{k,l} (\Sigma^{-1})^{(k,l)} \frac{1}{(NT)^2} \sum_{i,j,m} E_i'e^{(k)}e^{(l)'}E_i$$

$$- \frac{1}{\sqrt{T}} \sum_{k=1}^{m} \left( \sigma^{-2}(\Sigma^{-1})^{(k,c)} \frac{1}{N\sqrt{T}} \sum_{i,j} E_{ij}e_{ij} \right) \frac{1}{(NT)^2} \sum_{i,j,m} E_i'e^{(k)}e_m'E_i$$

$$- \frac{1}{\sqrt{T}} \sum_{l=1}^{m} \left( \sigma^{-2}(\Sigma^{-1})^{(l,c)} \frac{1}{N\sqrt{T}} \sum_{i,j} E_{ij}e_{ij} \right) \frac{1}{(NT)^2} \sum_{i,j,m} E_i'e^{(l)}e_m'E_i + o_p(T^{-1}).$$

(29)

where the last step follows from Lemma 4. Using Lemma 1 and the facts $\sum_{k,l} \Sigma^{(k)}(\Sigma^{-1})^{(k,l)}\Sigma^{(l,c)} = \Sigma\Sigma^{-1}\Sigma$ and $\sum_{k,l} (\Sigma^{-1})^{(k,l)}\Sigma^{(k,c)} = tr(\Sigma^{-1}\Sigma)$, we can simplify (29) to

$$\frac{1}{NT}\sum_{i=1}^{N} E_i'P_\tilde{T}E_i = T^{-1}(1 + m)\Sigma + N^{-1}\Sigma + O_p(N^{-1/2}T^{-1}).$$

(30)

This result implies

$$\left\| \frac{1}{NT}\sum_{i=1}^{N} E_i'P_\tilde{T}E_i \right\| = O_p(T^{-1}) + O_p(N^{-1}),$$

(31)

from which we get

$$\frac{1}{NT}\sum_{i=1}^{N} E_i'P_\tilde{T}E_i = \frac{1}{NT}\sum_{i=1}^{N} E_i'E_i + O_p(T^{-1}) + O_p(N^{-1}).$$

(32)

Taking together (26), (27), (31) and (32) yields

$$\|g_2\| = O_p(T^{-1}) + O_p(N^{-1}).$$

(33)
Next, consider $g_1$. Analogously to (29), we can write

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i P_{\hat{f}} \epsilon_i = \sigma^{-2} \frac{1}{(NT)^{3/2}} \sum_{i,j,m} E_i e_j e_m e_i + \sum_{k,l} (\Sigma^{-1})^{(k,l)} \frac{1}{(NT)^{3/2}} \sum_{i,j,m} E_i e_j e_m e_i
$$

$$
- \frac{1}{\sqrt{T}} \sum_{k=1}^{m} \left( \sigma^{-2} (\Sigma^{-1})^{(k,k)} \frac{1}{N \sqrt{T}} \sum_{i,j,l=1}^{T} e_{ij} e_{jl} \right) \frac{1}{\sqrt{N}} \sum_{i,j,m} E_i e_j e_m e_i
$$

$$
- \frac{1}{\sqrt{T}} \sum_{k=1}^{m} \left( \sigma^{-2} (\Sigma^{-1})^{(l,l)} \frac{1}{N \sqrt{T}} \sum_{i,j,l=1}^{T} e_{ij} e_{jl} \right) \frac{1}{\sqrt{N}} \sum_{i,j,m} E_i e_j e_m e_i + o_p(1).
$$

(34)

Here, by (16), (17) and the fact that $\sum_{k=1}^{m} (\Sigma^{-1})^{(k,k)} = \Sigma^{-1}$,

$$
\sum_{k=1}^{m} \left( \sigma^{-2} (\Sigma^{-1})^{(k,k)} \frac{1}{N \sqrt{T}} \sum_{i,j,l=1}^{T} e_{ij} e_{jl} \right) \frac{1}{\sqrt{N}} \sum_{i,j,m} E_i e_j e_m e_i = \sqrt{\frac{T}{N}} \frac{1}{N \sqrt{T}} \sum_{i,j,l=1}^{T} e_{ij} e_{jl} + o_p(T^{-1/2})
$$

with an analogous result for the last line in (34). Hence, the entire expression can be simplified to

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} E_i P_{\hat{f}} \epsilon_i = \sigma^{-2} \frac{1}{NT} \sum_{i,j,m} E_i e_j e_m e_i + \sum_{k,l} (\Sigma^{-1})^{(k,l)} \frac{1}{NT} \sum_{i,j,m} E_i e_j e_m e_i
$$

$$
- N^{-1/2} \frac{1}{N \sqrt{T}} \sum_{i,j,l=1}^{T} e_{ij} e_{jl} + o_p(1)
$$

$$
= O_p(N^{-1/2}) + O_p(T^{-1/2})
$$

(35)

using Lemma 2. Together with (32), this result implies $||g_1|| = O_p(N^{-1/2}) + O_p(T^{-1/2})$ which together with (33) this yields

$$
\sqrt{NT} (\hat{\beta}^{CCE} - \beta) = \sqrt{NT} (\hat{\beta}^{OLS} - \beta) + O_p(N^{-1/2}) + O_p(T^{-1/2}),
$$

where the asymptotic distribution of the latter has been determined from (26) and (27). □

**Proof of Theorem 2.** Using (24) we can write

$$
\sqrt{NT} (\hat{\beta}^{CCE} - \hat{\beta}^{OLS}) = -\sqrt{T} g_1 + \sqrt{T} g_2
$$

(36)

With $g_1$ and $g_2$ as defined in the proof of Theorem 1. Consider $\sqrt{T} g_2$. First, note that $E[(NT)^{-1/2} \sum_{i=1}^{N} E_i] = 0$ and that the other expressions in $g_2$ converge to nonrandom matrices. Thus, $g_2$ does not introduce any bias in (36). Furthermore, (33) implies $||\sqrt{T} g_2|| = O_p(\sqrt{TN^{-1}}) + O_p(T^{-1/2})$.
which is negligible under the assumptions on the relative expansion rate on $N$ and $T$ made in Theorem 2. Next, consider $\sqrt{T}g_1$. By (32), (26) and (35),

$$\sqrt{T}g_1 = h_1 + h_2 - \sqrt{\frac{T}{N}}h_3 + o_p(1)$$

where

$$h_1 = \Sigma^{-1} \sigma^{-2} \frac{1}{N^{3/2}T} \sum_{i,j,m} \sum_{t,s} e_{i,t} \epsilon_{j,t} \epsilon_{m,s} \epsilon_{i,s}$$

$$h_{2,kl} = \Sigma^{-1} (\Sigma^{-1})^{(k,l)} \frac{1}{N^{3/2}T} \sum_{i,j,m} \sum_{t,s} e_{i,t} \epsilon_{j,t}^{(k)} \epsilon_{m,s} \epsilon_{i,s}$$

$$h_3 = \Sigma^{-1} \frac{1}{N^{1/2}T} \sum_{i,j} \sum_{l=1}^{T} e_{i,l} \epsilon_{j,l}.$$

By Lemma 3 and the independence of $E_i$ and $\epsilon_i$, the expected values of all three terms above are zero. Hence, a Lindeberg-Levi CLT applies to all three expressions, proving asymptotic normality. Next, consider $\text{Var}[\sqrt{T}g_1]$. The variances of $h_1$ and $h_{2,(k,l)}$ are straightforwardly derived from (11) and (13). $\text{Cov}[h_{2,(k,l)}, h_{2,(k',l')}]$ for $k, k', l, l' = 1, \ldots, m$ are almost identical to $\text{Var}[h_{2,(k,l)}]$ with the only difference being the superscripts on $\epsilon$. Moreover, it is straightforward to show that

$$\text{Var} \left[ \left( \frac{1}{N^{1/2}T} \sum_{i,j} \sum_{l=1}^{T} e_{i,l} \epsilon_{j,l} \right) \right] = \sigma^2 \Sigma.$$

(37)

The remaining required covariances are given by Lemma 5. This gives us

$$\text{Var}[\sqrt{T}g_1] = \text{Var}[h_1] + \frac{T}{N} \text{Var}[h_3] - 2 \sqrt{\frac{T}{N}} \text{Cov}[h_1; h_3] + \sum_{k,j,k',l'} \text{Cov}[h_{2,kl}, h_{2,k'l'}]$$

$$+ 2 \sum_{k,j} \left( \text{Cov}[h_1; h_{2,kl}] - \sqrt{\frac{T}{N}} \text{Cov}[h_{2,kl}; h_3] \right) + o_p(1)$$

Here, from (11), (19) and (37),

$$\text{Var}[h_1] + \frac{T}{N} \text{Var}[h_3] - 2 \sqrt{\frac{T}{N}} \text{Cov}[h_1; h_3]$$

$$= \left( 1 + \frac{T}{N} \right) \sigma^2 \Sigma^{-1} + \frac{T}{N} \sigma^2 \Sigma^{-1} - 2 \frac{T}{N} \sigma^2 \Sigma^{-1}$$

$$= \sigma^2 \Sigma^{-1}.$$
Furthermore, (13) yields
\[ \sum_{k,l,k',l'} m \text{Cov}[h_{2,k,l}, h_{2,k',l'}] = T/N \Sigma^{-1} \sum_{k,j} \Sigma^{(k,j)}(\Sigma^{-1})^{(k,j)}(\Sigma^{-1})(\Sigma^{-1})^{(l',k')} \sigma^2 \Sigma^{-1} \]
\[ + \Sigma^{-1} \sum_{k,l} \Sigma^{(k,l)}(\Sigma^{-1})(\Sigma^{-1})(\Sigma^{-1})^{(l',k')} \sigma^2 \Sigma^{-1} \]
\[ = (m + T/N) \sigma^2 \Sigma^{-1}, \]
where we use the identities \( \sum_{k,l} (\Sigma^{-1})(\Sigma^{-1})(\Sigma^{-1})^{(l',k')} = tr(\Sigma^{-1} \Sigma^{-1} \Sigma) = m \) and \( \sum_{k,l} (\Sigma^{-1})(\Sigma^{-1})(\Sigma^{-1})^{(l',k')} = \Sigma \Sigma^{-1} \Sigma. \) Additionally, from (18) and (20),
\[ \sum_{k,l} \left( \text{Cov} [h_{1}; h_{2,k,l}] - \sqrt{T/N} \text{Cov} [h_{2,k,l}; h_{3}] \right) = 0 \]
Taking these three results together, we obtain
\[ \text{Var}[\sqrt{T}g_1] = (1 + m + T/N) \sigma^2 \Sigma^{-1} + o_p(1). \]
Given that the asymptotic order of \( g_2 \) reflects its variance and recalling the assumption \( T/N \rightarrow c < \infty, \) this yields
\[ \text{Var}[\sqrt{NT}(\hat{\beta}^{CCE} - \hat{\beta}^{OLS})] = (1 + m + T/N) \sigma^2 \Sigma^{-1} + O_p(T^{-1/2}), \]
which provides the last required result for Theorem 2

**Proof of Theorem 3.** The first result follows from noting that, by consistency, the true variance matrices \( \Sigma \) and \( \sigma^2 \) are the probability limits of \( \hat{\Sigma} \) and \( \hat{\sigma}^2. \) Given asymptotic normality of the difference \( \sqrt{NT}(\hat{\beta}^{CCE} - \hat{\beta}^{OLS}), \) is asymptotically the sum of \( m \) uncorrelated, standard normally distributed random variables. This expression is hence by definition \( \chi^2(m) \) distributed.

For the behaviour of the Hausman test statistic under the alternative hypothesis, we investigate the equivalent of (36) in the presence of cross-section dependence. This gives us
\[ \sqrt{NT}(\hat{\beta}^{CCE} - \hat{\beta}^{OLS}) \]
\[ = \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_F X_i \right)^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i P_F X_i \right) \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i X_i \right)^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i y_i \right) \]
\[ - \left( \frac{1}{NT} \sum_{i=1}^{N} X'_i M_F X_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X'_i P_F y_i \]
\[ = \tilde{g}_1 - \tilde{g}_2, \]
where we focus exclusively on \( \tilde{g}_1 \) since it is the leading term. It is straightforward to show that 
\[
\frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i = O_p(1) \quad \text{and} \quad \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{M}_i \mathbf{X}_i = O_p(1)
\]
which we skip at this point for the sake of brevity. Furthermore, we have
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{y}_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (\hat{\mathbf{F}}_i + \mathbf{E}_i)' (\hat{\mathbf{F}}_i + \mathbf{e}_i).
\]
(41)

Here the only relevant term is \( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Lambda_i' \mathbf{F} \mathbf{F} \Lambda_i \). Representatively for all elements in this vector, the first element can be written
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (\Lambda_i' \mathbf{F} \mathbf{F} \Lambda_i) = \sqrt{NT} \text{tr} \left( N^{-1} \sum_{i=1}^{N} \Lambda_i' \Lambda_i \right) (T^{-1} \mathbf{F} \mathbf{F})
\]
Now \( (T^{-1} \mathbf{F} \mathbf{F}) \) converges to a positive definite matrix under assumption 2. Consequently, at least the diagonal of \( \Sigma_F \) must be nonzero. Under this minimal requirement, we have
\[
\text{tr} \left( N^{-1} \sum_{i=1}^{N} \Lambda_i' \Lambda_i \right) (T^{-1} \mathbf{F} \mathbf{F}) = \sum_{r=1}^{r_1} \left( N^{-1} \sum_{i=1}^{N} \lambda_i^{(r_1)} \lambda_i^{(r_1,1)} \right) \Sigma_F^{(r_1,1)},
\]
where additional terms appear if the nondiagonal elements of \( \Sigma_F \) are different from zero. Now let \( N_1 \) be the number of cross-sections that satisfy \( \lambda_i^{(r_1)} \neq 0 \) and \( (\Lambda_i^{(r_1,1)}) \neq 0 \) for some \( r_1 \). Ensuring that the trace above is \( O_p(1) \) requires that \( N_1 / N \) is bounded away from zero. This entails
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} (\Lambda_i' \mathbf{F} \mathbf{F} \Lambda_i) = O_p(\sqrt{NT}).
\]
The asymptotic order reduces to \( O_p(\sqrt{T}) \) if the expected value of \( \lambda_i^{(r_1)} \lambda_i^{(r_1,1)} \) is zero. This case may appear if the loadings are uncorrelated and at least one of \( \lambda_0^{(r_1)} \) or \( \lambda_0^{(r_1,1)} \) is zero.

Next, turn to
\[
\sqrt{T} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{P}_i \mathbf{X}_i = \sqrt{T} \frac{1}{NT} \sum_{i=1}^{N} (\hat{\mathbf{F}}_i + \mathbf{E}_i)' \mathbf{P}_i (\hat{\mathbf{F}}_i + \mathbf{E}_i).
\]

Since
\[
\hat{\mathbf{F}} = \left( \mathbf{F} \left[ \begin{array}{cc} \mathbf{X} & \mathbf{X} \\ \mathbf{P} & \mathbf{E} \end{array} \right] \right) \left[ \begin{array}{c} 1 \\ \beta \\ 0 \end{array} \right] = \hat{\mathbf{F}}_0 \left[ \begin{array}{c} 1 \\ \beta \\ 0 \end{array} \right]
\]
where the last matrix is nonsingular by construction, we can write the above expression in terms of \( \hat{\mathbf{F}}_0 \) instead of \( \hat{\mathbf{F}} \). Doing so, we are again only interested in \( \sqrt{T} \frac{1}{NT} \sum_{i=1}^{N} \Lambda_i' \mathbf{F} \mathbf{P}_i \hat{\mathbf{F}}_0 \mathbf{F} \Lambda_i \).
Representatively for all elements of this expression we can look at its first element, which is given by
\[
\sqrt{T} \frac{1}{NT} \sum_{i=1}^{N} (\Lambda_i^{(-1)})' F_i P_0 F \Lambda_i^{(-1)} = \sqrt{T} tr \left( N^{-1} \sum_{i=1}^{N} (\Lambda_i^{(-1)})' (\Lambda_i^{(-1)})' T^{-1} F_i P_0 F \right).
\]

It turns out that this matrix is $O_p(\sqrt{T})$ under Assumption 3(iii). The proof considers two cases. First, assume that $rk(\overline{\Lambda}) > 0$. In this case, $\hat{F}_0$ converges to some linear combination of $F$. Consequently, $\overline{F}$ is correlated with $F$ by construction and $P_0 \hat{F}$ cannot annihilate $F$. Next, assume that $\overline{\Lambda} = 0$. Here we can use the fact that the projection matrix is invariant to scaling. We can hence define a scaled factor estimator $\tilde{F}$ satisfying $M \tilde{F} = M \hat{F}_0$ by
\[
\tilde{F} = \sqrt{N} \hat{F}_0,
\]
which yields a set of vectors that do not converge to zero in probability. The columns of $\tilde{F}$ consist of noise that is uncorrelated with $F$ and some random linear combination of $F$. The presence of the latter component is guaranteed if $N_1 / N > 0$. For this reason, $M \hat{F}_0$ cannot annihilate $F$. Consequently, $||T^{-1} F \hat{P}_0 F||$ must be bounded away from zero which implies that $\sqrt{T} \frac{1}{NT} \sum_{i=1}^{N} X_i' P_0 X_i = O_p(\sqrt{T})$.

Taking this latter results together with the asymptotic order of (41), we obtain
\[
\text{Var}[\sqrt{NT}(\hat{\beta}^{CCE} - \hat{\beta}^{OLS})] = O_p(\sqrt{NT})
\] (42)

Since the Hausman test statistic is a quadratic form, the asymptotic order above is squared. Keeping in mind the inverse term within the variance which is of order $O_p(T^{-1})$, the power of the Hausman test increases in $NT$. Lastly, in the case where factors are uncorrelated and either $\lambda_0^{(r_1)} = 0$ or $\Lambda_0^{(r_1,1)} = 0$, the divergence rate reduces to $T$.  \[\square\]
A Factor Analytical Approach to Price Discovery*

JOAKIM WESTERLUND,† ‡ SIMON REESE,† PARESH NARAYAN‡

†Department of Economics, Lund University, P. O. Box 7082 S–220 07, Lund, Sweden (e-mail: joakim.westerlund@nek.lu.se)
‡Centre for Financial Econometrics, Deakin University

Abstract

Existing econometric approaches for studying price discovery presume that the number of markets are small, and their properties become suspect when this restriction is not met. They also require making identifying restrictions and are in many cases not suitable for statistical inference. The current paper takes these shortcomings as a starting point to develop a factor analytical approach that makes use of the cross-sectional variation of the data, yet is very user-friendly in that it does not involve any identifying restrictions or obstacles to inference.

I. Introduction

Financial markets incorporate new information into asset prices by matching buyers and sellers. They thereby facilitate the discovery of what the price of an asset should be. This price discovery role of financial markets can take place across separate exchanges and instruments, as many securities and derivatives based on the same underlying asset may trade on multiple venues. In the case of such a multiplicity, there may be variation in the share with which each market’s trades contribute to discovering the one true price of the underlying asset. The present paper is about the modelling and measuring of these so-called ‘information shares’ (ISs), which are important for both investors concerned with priceornativeness and adverse selection risk, as well as policy makers investigating the determinants of price efficiency.

The measurement of price discovery requires isolating informative price movements from noise. Observed price changes constitute the most obvious indicator of price discovery. However, they form an imperfect measure, as observed prices are susceptible to transitory mispricing, caused by noise trading or temporary order imbalances, for example.

JEL Classification numbers: C12; C13; C33

*Earlier versions of this paper were presented at the 1st Conference on Recent Developments in Financial Econometrics and Applications in Geelong and at the 8th International Conference on Computational and Financial Econometrics in Pisa. The authors would like to thank conference participants, and in particular Morten Ørregaard Nielsen, Debopam Bhattacharya (Editor) and two anonymous referees for many useful comments and suggestions. Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship. Thank you also to the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2014–0112:1.

© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd.
In contrast, when security prices absorb new information due to informed trading, these price changes last permanently. To formalize these ideas, let us denote by \( P_{i,t} \) the price of a particular asset on market \( i = 1, \ldots, N \) at time period \( t = 1, \ldots, T \), and let \( P^*_t \) be the fundamental value of the same asset at the same time period. The structural model we have in mind is taken from the market microstructure literature (see Madhavan, 2000, for a survey), as is standard in studies of price discovery (see Hasbrouck, 1993; De Jong, 2002; de Harris, McInish and Wood, 2002; Lehmann, 2002; De Jong and Schotman, 2010, to mention a few). It is given by

\[
P_{i,t} = P^*_t + E_{i,t},
\]

where \( E_{i,t} \) is an idiosyncratic (or market-specific) term capturing microstructure effects. While \( P^*_t \) is assumed to follow a random walk, \( E_{i,t} \) is stationary. Hence, while shocks to the fundamental value have a permanent effect on prices, the effect of idiosyncratic shocks is transitory. The transitory nature of \( E_{i,t} \) implies that \( P_{i,t} \) will adjust to the fundamental value over time. In fact, as Hasbrouck (1995) points out, under said assumptions, \( P_{i,t} \) and \( P_{j,t} \) are cointegrated with cointegrating vector \((1, -1)'\).

A market is relatively efficient in the price discovery process if it incorporates a larger amount of fundamental shocks than other markets. Hasbrouck (1995) defines the IS of a particular market as the variance share of that market that is attributable to the fundamental value. Unfortunately, as equation (1) makes clear, the fundamental value is not directly observed, but is instead contaminated by additive market-specific pricing errors. Therefore, legitimate inference on the price discovery process cannot be made until the confounding effects of these errors have been appropriately accounted for. The most common approach by far is the one of Hasbrouck (1995), in which the effects of the shocks are retrieved from an estimated reduced form vector error correction model (VECM). This VECM route to the IS has in recent years become very popular, and is by now the workhorse of the industry with a huge number of applications (see e.g. Brenner and Kroner, 1995, for a survey).

But while popular, the VECM approach also has its fair share of drawbacks. First, since the VECM is just a reduced form model, the fundamental shocks cannot be retrieved without suitable identifying restrictions. Hasbrouck (1995) uses Cholesky factorization, which makes the IS dependent on the ordering of the series. With \( N \) markets, there are no less than \( N! \) such orderings, suggesting that without prior knowledge about the appropriate order, the IS is likely to be an uninformative measure. As a remedy, Hasbrouck (1995) suggests reporting upper and lower bounds, as obtained by considering all possible orderings. The resulting largest and smallest ISs for each market constitute the upper and lower bounds, respectively. These bounds can, however, be quite far apart (see e.g. Hasbrouck, 2002; Huang, 2002). This is true when \( N \) is relatively small, and the bounds become wider when \( N \) increases. Specifically, the width depends critically on the covariances of the shocks, whose number increases at the rate \( N^2 \). The number of degrees of freedom will therefore decrease very rapidly with \( N \), leading to increased estimation uncertainty, and hence wider bounds. For example, Booth et al. (2002) use trading intervals averaging about 30 minutes for only two markets, namely, the upstairs and downstairs markets on the Helsinki stock exchange. According to their results, the average IS interval for the downstairs market is [13%, 99.2%], which is clearly too wide for any interesting conclusions (see, for example, Martens, 1998; Tse, 1999; Baillie et al., 2002, for similar findings).
Second, the asymptotic standard errors of the ISs are difficult to obtain, complicating inference. Sapp (2002) proposes bootstrapping the standard errors. However, while certainly feasible, the bootstrap is computationally relatively costly, making it unattractive from an applied point of view. Third, well-specified VECMs tend to be heavily parameterized, making them infeasible in large-$N$ samples, which of course restricts their applicability. Fourth, the small-$N$ requirement means that within the VECM approach more information is not necessarily a good thing. In fact, as the above discussion suggests, the best performance is typically obtained by having $N$ as small as possible. This is unfortunate, for one of the most well-documented features of financial data, especially in the aftermath of the global financial crisis, is the strong co-movements that exist across many markets. Being able to accommodate a relatively large $N$ should therefore lead to more decisive conclusions regarding price discovery.

The above drawbacks recently motivated De Jong and Schotman (2010) to develop a generalized method of moments (GMM)-based approach that does not rely on reduced forms, but instead seeks to infer the structural model in equation (1) directly.\footnote{See Lien and Shrestha (2009), and Yan and Zivot (2010) for other attempts to alleviate the identification problem of the VECM approach.} In addition to the advantage of not having to identify the shocks, the main advantage of estimating equation (1) is that it is parsimonious, as opposed to reduced form VECMs. The GMM approach is therefore suitable even when $N$ is relatively large. The implementation is quite complicated, though, which is probably also the main reason for why the approach has not received much interest in the (applied) literature.

In the present paper, we take the complicated implementation of the GMM approach as our starting point. The purpose is to develop a new approach to price discovery that is simple yet does not suffer from the drawbacks of the VECM approach. As a source of inspiration we consider the growing literature on large-dimension common factor models (see Bai and Ng, 2008, for a survey), within which equation (1) can be seen as a restricted common factor model with a single non-stationary common factor, or cross-section common stochastic trend, $P^*_t$, and unit factor loading. This suggests that equation (1) can be estimated using existing methods for such common factor models. The simplest method by far is the cross-section average (CA) approach first considered by Pesaran (2006) in the context of a factor-augmented panel regression model in stationary variables. As the name suggests, the basic idea is to use the cross-section average of the observed data as an estimator of the common factor. In this paper, we extend the CA approach to the price discovery context. Specifically, a statistical toolbox is provided that enables not only estimation but also testing, and this under fairly general conditions. Unlike VECMs, the new approach exploits the information contained in the cross-sectional dimension of the panel, making it particularly well-suited for large-$N$ panels, although $N$ can also be ‘small’.

The usefulness of the new toolbox is illustrated using three data sets. The first data set covers crude oil prices listed on three exchanges, namely, the US, the UK and Oman exchanges. The findings suggest that the price discovery process is dominated by the Oman market, followed by the UK market. The US market contributes $<4\%$ to price discovery. This result is rather interesting in that it represents the first piece of evidence on the oil price discovery process across countries. Our second application is motivated by the equity
prices of firms that are cross-listed. We pick a firm that has listings on a large number of exchanges. We select Arcelor Mittal, a major steel company, which has its primary listing on the Luxembourg stock exchange and has secondary listings on another eight exchanges. Our results reveal that while the primary exchange dominates, the contribution to price discovery is only around 30%, suggesting that the bulk of price discovery takes place on other exchanges. The third and final application considers historical prices of oats covering 20 Swedish counties between 1830 and 1915. As expected, most of the price discovery takes place in counties located in the south with access to important shipping lanes.

The balance of the paper is organized as follows. In section II, we formalize the common factor model setup and explain how it relates to the IS. Section III presents the econometric results. This section is divided into two parts. The first part presents the asymptotic theory, while the second part is concerned with the small-sample accuracy of this theory. Section IV contains the empirical results. Section V concludes. Proofs of important results are given in Appendix.

II. The common factor model

Assumptions

Consider the panel data variable $X_{i,t}$, observable for $i = 1, \ldots, N$ cross-section units and $t = 1, \ldots, T$ time periods. The data generating process (DGP) of this variable is assumed to be given by the following common factor model:

$$X_{i,t} = \lambda_i F_t + U_{i,t},$$  \hspace{1cm} (2)

where $F_t$ is a common factor with $\lambda_i$ being the associated factor loading and $U_{i,t}$ is an idiosyncratic error term. In terms of the model in equation (1), we have $X_{i,t} = P_{i,t}, F_t = P^*_t$, $\lambda_1 = \cdots = \lambda_N = 1$ and $U_{i,t} = E_{i,t}$. Hence, for the DGP in equation (2) to make sense from an economic point of view, while $F_t$ should follow a random walk, $U_{i,t}$ should be stationary. However, we do not want to force $U_{i,t}$ to be stationary, but would instead like to allow for the possibility. For these reasons, it seems natural to assume that the dynamics of $F_t$ and $U_{i,t}$ are governed by the following equations:

$$F_t = F_{t-1} + \eta_t,$$  \hspace{1cm} (3)

$$U_{i,t} = \rho_i U_{i,t-1} + \epsilon_{i,t},$$  \hspace{1cm} (4)

where $\rho_i \in (-1, 1]$, and $\eta_t$ and $\epsilon_{i,t}$ are error terms that are supposed to satisfy Assumption 1.

Assumption 1.

(i) $E(\eta_t | \mathcal{F}_{t-1}) = 0$, $E(\eta_t^2) = \sigma_\eta^2 > 0$ and $E(|\eta_t|^4) < \infty$, where $\mathcal{F}_t$ is the sigma-field generated by $\{\eta_n\}_{n=1}^T$;

(ii) $a_i(L)e_{i,t} = e_{i,t}$, where $e_{i,t}$ is independent and identically distributed (iid) across $i$ with $E(e_{i,t} | \mathcal{F}_{i,t-1}) = 0$, $E(e_{i,t}^2) = \sigma_{e_{i,t}}^2 > 0$ and $E(|e_{i,t}|^4) < \infty$, where $\mathcal{F}_{i,t}$ is the sigma-field generated by $\{e_{i,n}\}_{n=1}^T$ and $a_i(L) = \sum_{n=0}^p a_{i,n}L^n$ is a polynomial in the lag operator $L$ having all roots outside the unit circle and $a_{i,0} = 1$.

(iii) $E(|F_0|) < \infty$ and $E(|U_{i,0}|) < \infty$ for all $i$. 

© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd
(iv) $\lambda_i$ is either deterministic such that $|\lambda_i| < \infty$ or stochastic such that $E(|\lambda_i|^4) < \infty$ and $\bar{\lambda} = N^{-1} \sum_{i=1}^{N} \lambda_i \neq 0$ for all $N$, including $N \to \infty$;

(v) $e_{i,t}$, $\eta_t$ and $\lambda_i$ are mutually independent.

As alluded to in the above, the model in equations (2)–(4) is almost identical to the one in equation (1), in which $X_{i,t}$ and $F_t$ represent the observed price and its fundamental value respectively. The only differences are that; (i) $\lambda_i$ is not restricted to be equal to one and (ii) $U_{i,t}$ need not be stationary. As mentioned in section I, the otherwise so common stationarity requirement on $U_{i,t}$ implies that prices are cointegrated across markets. Indeed, since under $|\rho_i| < 1$ for all $i$, $X_{i,t} - \lambda_i \lambda_j^{-1} X_{j,t} = U_{i,t} - \lambda_i \lambda_j^{-1} U_{j,t}$ is stationary, $X_{i,t}$ and $X_{j,t}$ are cointegrated with cointegrating vector $(1, - \lambda_i \lambda_j^{-1})'$. Hence, in the terminology of the time series literature, equations (2)–(4) constitute a ‘common trend’ representation of $X_{1,t}, \ldots, X_{N,t}$ (see Stock and Watson, 1988). This also highlights the meaning of the unit loading assumption, as a restriction on the cointegrating vector. The fact that $\bar{\lambda} \neq 0$ means that not all loadings can be zero. Zero loadings are not ruled out, however, provided that there is a non-negligible fraction of loadings for which the average is non-zero. This seems relevant in practice, as some markets may not contribute to price discovery.

The assumption that $\epsilon_{i,t}$ and $\eta_t$ are independent is similar to the restriction employed by Watson (1986) (see Hasbrouck, 1993, for a discussion). In section ‘Comparison with existing models’, we discuss the meaning of this assumption in terms of the structural model in equation (1). It is important to note, however, that the common factor approach that we will be considering works under more general conditions, and that in this sense independence is just a simplifying assumption. A minimal requirement in this regard is that $(NT)^{-1/2} \sum_{j=1}^{N} \sum_{t=1}^{T} \eta_t \epsilon_{i,t}$ satisfies a central limit law. Assumptions 1 (i) and (ii) imply that $\eta_t$ and $e_{i,t}$ are serially uncorrelated in their levels. The assumption that $e_{i,t}$ is serial uncorrelated is not restrictive as $e_{i,t}$ may still be serially correlated in a very general way. The assumption that $\eta_t$ is serial uncorrelated is, on the other hand, standard and not very controversial. Indeed, as Hasbrouck (2002, p. 332) states, ‘it is difficult to conjecture why a permanent non-martingale price component warrants general interest’. If, however, one would like to allow for the possibility of such non-martingale effects, then this can be easily accommodated. This is explained in section III. Note also how Assumptions 1 (i) and (ii) do not rule out serial correlation in the conditional second moment. Conditional heteroskedasticity is therefore not ruled out, which is in agreement with the previous literature on the heteroskedasticity of returns (see e.g. Westerlund and Narayan, 2015).

The quantity of interest is the IS, which in the seminal work of Hasbrouck (1995) is defined as the fraction of the variance of the innovation to the fundamental price component that can be attributed to a particular market. As De Jong and Schotman (2010) show, in the current context, the IS for cross-section unit $i$ is given by

$$IS_i = \frac{\lambda_i^2 \sigma_{U,i}^2}{\sigma_{\eta}^2 + \sum_{j=1}^{N} \lambda_j^2 \sigma_{U,j}^2} = \frac{\lambda_i^2 \sigma_{\eta}^2 \sigma_{U,i}^2}{1 + \sum_{j=1}^{N} \lambda_j^2 \sigma_{\eta}^2 \sigma_{U,j}^2}, \tag{5}$$

where $\sigma_{U,i}^2 = E(U_{i,t}^2)$. The derivation of the formula in equation (5) is quite involved, and is provided in Appendix A of this paper. However, we can see that it makes sense. For example, the less the noise in market $i$, as measured by $\sigma_{\eta}^2$ the higher the IS. Similarly, the stronger the covariance between the price of market $i$ and the efficient price, as captured
by \( \lambda_i \), the higher the IS. For a given \( N \), the measure does not sum up to one. However, asymptotically it does:

\[
\sum_{i=1}^{N} IS_i = \frac{\sum_{i=1}^{N} \lambda_i^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon} \sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon}}{\sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon} [1 + (\sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon}})^{-1}] = \frac{\sum_{i=1}^{N} \lambda_i^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon} \sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon} [1 + o(1)]}{\sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon}} \to 1
\]

as \( N \to \infty \). This suggests that in the large-\( N \) scenario considered here the following panel IS (PIS) may be used:

\[
PIS_i = \frac{\lambda_i^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon}}{\sum_{j=1}^{N} \lambda_j^2 \sigma^2_{\epsilon \epsilon} \sigma^2_{U \epsilon}}.
\]  

(6)

As with \( IS_i \), \( PIS_i \) measures the relative variance of the efficient price innovations in market \( i \) as a fraction of the total variation. The purpose of this paper is to infer \( PIS_i \).

Comparison with existing models

As mentioned in section I, most (if not all) of the existing price discovery literature is based on the structural model in equation (1). This is therefore the model of interest. However, since the efficient price is unobservable, suitable identification and estimation strategies must be put in place before any meaningful measure of price discovery can be constructed. The question is: how to best go about this business? The most common approach is to employ the VECM approach of Hasbrouck (1995). Hence, from a popularity point of view, this is the relevant benchmark. However, we start by considering the approach of Gonzalo and Granger (1995), which is almost as popular and is in many regards very similar to the one of Hasbrouck (1995) (see Lehmann, 2002, for a detailed discussion).

The Gonzalo and Granger (1995) approach is based on the so-called ‘permanent–transitory’ (PT) decomposition of the data (see Quah, 1992). It says that if \( X_{i,t} \) is unit root non-stationary, then \( X_t = (X_{1,t}, \ldots, X_{N,t})' \) may be decomposed as

\[
X_t = PE_t + TR_t,
\]

(7)

where \( PE_t (TR_t) \) is unit root non-stationary (stationary). This decomposition is very similar to the common factor model in equations (2)–(4). In fact, setting \( PE_t = \lambda F_t \) and \( TR_t = U_t = (U_{1,t}, \ldots, U_{N,t})' \), where \( \lambda = (\lambda_1, \ldots, \lambda_N) \), we see that the two models coincide. Even the purpose of Gonzalo and Granger (1995), to estimate \( F_i \), is very similar to the purpose of the present study. The authors recognize the potential of the common factor model approach. However, because of the stationarity restriction that was at the time required for its estimation, the approach was not pursued. As we explain in detail in section III, the factor analytical approach considered in this paper is very general in this regard, and can be used even if there is uncertainty over the order of integration of \( U_{i,t} \).

A crucial difference between the PT approach of Gonzalo and Granger (1995), and the one considered here is how the components of the data are identified. According to the structural model in equation (1), \( P_t^* \) is unit root non-stationary and \( E_{i,t} \) is stationary. In the PT approach, it is this difference in the order of integration that enables separation between \( PE_t = P_t^* \) and \( TR_t = E_t \). By contrast, in the approach considered here it is the common versus idiosyncratic variation that is key; while \( F_t \) is assumed to be common,
$U_{i,t}$ is purely idiosyncratic. It is important to note that the separation into common and idiosyncratic components can be achieved regardless of the order of integration of these components. The factor analytical approach considered here can therefore be thought of as two-step procedure, in which the extraction of the common and idiosyncratic components of the data is just a first step. The second step is then to test the extracted components for unit roots. The fact that nothing is assumed regarding the order of integration of the components is a clear advantage when compared with the PT identification scheme.

An advantage of the PT decomposition is that it does not require that $\Delta PE_t$ and $TR_t$ are independent. In particular, $TR_t$ may be correlated with the lags of $\Delta PE_t$. However, because of the way that $PE_t$ ($TR_t$) is identified as unit root non-stationary (stationary), $\Delta PE_t$ cannot be correlated with the lags of $TR_t$, since then $TR_t$ has a permanent effect on $X_t$. For example, if $\Delta P^*_t = \gamma_tE_{t-1} + e_t$, where $\gamma$ is $N \times 1$ and $e_t$ is uncorrelated with $E_{t-1}$, then $P^*_t = P^*_0 + \sum_{i=1}^t (\gamma' E_{t-i} + e_i)$. Thus, since shocks to $E_t$ have a permanent effect on $P^*_t$, and hence also on $X_t$, the PT decomposition will be unable to separate $P^*_t$ from $E_t$. Certain types of correlated innovations can be accommodated also within the present framework. For example, if $E_{i,t} = \lambda_i \Delta P^*_{t-1} + e_{i,t}$, where $e_{i,t}$ is idiosyncratic, then equation (2) still holds with $F_t = P^*_t + \Delta P^*_{t-1}$ and $U_{i,t} = e_{i,t}$. However, because of the common–idiosyncratic identification scheme, allowing for correlated innovations in the common factor framework is in general more difficult than in the PT framework. Both frameworks therefore have their pros and cons.

The VECM approach of Hasbrouck (1995) is based on the Beveridge–Nelson (BN) decomposition (see Phillips and Solo, 1992, Lemma 2.1), which is in turn similar to the factor framework. The VECM approach of Hasbrouck (1995) is based on the Beveridge–Nelson (BN) decomposition, which can be written alternatively as the vector moving average $\Delta X_t = B(L)v_t$, where $B(L) = \sum_{n=0}^{\infty} B_n L^n$, $B_0 = I_N$ and $v_t$ is iid with mean zero and covariance matrix $\Sigma_v$. By the BN decomposition, $B(L) = B(1) + (1 - L)B'(L)$, where $B'(L)$ has all its roots outside the unit circle. By using this result, backwards substitution and then $X_0 = 0_{N \times 1}$, we obtain the following representation for $X_t$, which is identical to the common trends representation of Stock and Watson (1988):

$$X_t = B(1) \sum_{n=1}^t v_n + B'(L)v_t. \quad (8)$$

Now, since $X_{1,t}, \ldots, X_{N,t}$ are cointegrated, the rank of $B(1)$ is one, that is, there exists an $N \times (N-1)$ matrix $\beta$ such that $\beta' B(1) = 0_{(N-1) \times N}$. Hence, $\beta' X_t = \beta' X_0 + \beta' B'(L)v_t$ is stationary. It is important to note that the above representation has the same form as the PT decomposition with $P_t = B(1) \sum_{n=1}^t v_n$ and $T_t = B'(L)v_t$. Many of the above mentioned drawbacks of the PT approach therefore applies also in case of the VECM approach. The main difference is that unlike in the former approach, in the latter the innovations driving $P_t$ are the same as those driving $T_t$, that is, in the VECM approach $\Delta P_t$ and $T_t$ are perfectly correlated. The drawback of this requirement, which was mentioned also in Section 1, is that the shocks are no longer separable, at least not without imposing further identifying assumptions (see e.g. Lehmann, 2002; De Jong and Schotman, 2010, for discussions).

In view of the above mentioned drawbacks of the PT and VECM approaches, and the results of studies such as Hasbrouck (2002), suggesting that neither approach seems to be
very accurate, it seems natural to seek out other alternatives. The factor analytical approach considered here can be seen as one step in this direction. It should be pointed out, however, that the main advantage is not how the data are decomposed into common and idiosyncratic components, but rather how the new approach enables easy estimation and inference even when \( N \) is relatively large. In section III we elaborate on this.

### III. Econometric results

#### Asymptotic theory

The purpose of this section is to infer \( \hat{PIS}_i \). The idea is the following. Consider the following first-differenced version of equation (2):

\[
x_{i,t} = \tilde{\lambda}_t f_t + u_{i,t},
\]

where \( x_{i,t} = \Delta X_{i,t}, f_t = \Delta F_t \) and \( u_{i,t} = \Delta U_{i,t} \), which are defined for \( t = 2, \ldots, T \). It is important to note that \( f_t = \eta_t \), that is, the common factor in equation (9) is the innovation that drives \( F_t \). Estimation of this model therefore leads naturally to an estimate \( \hat{\lambda}_t \) of \( \tilde{\lambda}_t f_t \), which we can use to infer \( \tilde{\lambda}_t^2 \sigma^2 \eta \) as \( \hat{\lambda}_t^2 \hat{\sigma}^2 \eta \), where \( \hat{\sigma}^2 \eta = T^{-1} \sum_{t=2}^T \hat{f}_t^2 \). The estimation of equation (9) does not lead to an estimator of \( U_{i,t} \); however, it does lead to an estimator \( \hat{u}_{i,t} = x_{i,t} - \hat{\lambda}_t f_t \) of \( u_{i,t} \), which can be accumulated up to levels. The estimator of \( U_{i,t} \) is therefore given by \( \hat{U}_{i,t} = \sum_{n=2}^t \hat{u}_{i,n} \), which can in turn be used to obtain \( \hat{\sigma}^2_{U,i} = T^{-1} \sum_{t=2}^T \hat{U}_{i,t}^2 \). The proposed estimator of \( \hat{PIS}_i \) is given by

\[
\hat{PIS}_i = \frac{\hat{\lambda}_i^2 \hat{\sigma}^2 \hat{\sigma}^{-2}_{U,i}}{\sum_{j=1}^N \hat{\lambda}_j^2 \hat{\sigma}^2 \hat{\sigma}^{-2}_{U,j}}.
\]

Of course, since in the present paper \( U_{i,t} \) is not necessarily stationary, the meaning of \( \hat{PIS}_i \) is not obvious. Therefore, in practice some pre-testing will in general be necessary to ensure valid interpretation. This is discussed in section “Testing for unit roots”. However, we start by discussing the construction of \( \hat{PIS}_i \).

Since equation (9) is nothing but a static common factor model in stationary variables, the estimation can in principle be carried out using the principal components method, which has a long tradition in econometrics and statistics (see Bai, 2003, section 1, for a brief review of this literature). However, preliminary Monte Carlo results suggest that this estimator suffers from poor small-sample performance, especially in the empirically relevant case when \( T > N \). In the present paper we therefore consider an alternative estimator that not only has good small-sample properties, but is also computationally very convenient. In fact, it is difficult to think of a simpler estimator. The idea is to follow Pesaran (2006) and to use \( \hat{f}_t = \hat{x}_t = N^{-1} \sum_{i=1}^N x_{i,t} \) as an estimator of \( f_t \). The estimator of \( \hat{\lambda}_i \) is given by \( \hat{\lambda}_i = (\sum_{t=2}^T \hat{f}_t^2)^{-1} \sum_{t=2}^T x_{i,t} \hat{f}_t = \hat{\sigma}^2 \eta T^{-1} \sum_{t=2}^T x_{i,t} \hat{f}_t \). Given \( \hat{f}_t \) and \( \hat{\lambda}_i \), we can compute \( \hat{U}_{i,t} = \sum_{n=2}^t \hat{u}_{i,n} \), where \( \hat{u}_{i,t} = x_{i,t} - \hat{\lambda}_t \hat{f}_t \) for \( t = 2, \ldots, T \). From \( \hat{f}_t \) and \( \hat{U}_{i,t} \) we compute \( \hat{\sigma}^2_{U,i} \) and \( \hat{\sigma}^2 \eta \), and then \( \hat{PIS}_i \) using the formula in equation (10).
Remark 2. The use of the cross-sectional average as an estimator of the common factor is similar to the approach of Pesaran (2007). The main difference is that while $\bar{x}_t$ is stationary, Pesaran (2007) uses a spurious regression that involves $X_{i,t}$ and $\bar{X}_t = N^{-1} \sum_{i=1}^N X_{i,t}$, leading to a nonstandard asymptotic distribution theory, which in turn makes for relatively complicated implementation. Not only does the differencing–accumulation approach employed here simplify the estimation of the common factor, but it also preserves the properties of the level data. In fact, it is not difficult to show that $T^{-1/2} |\hat{U}_i - U_i| = o_p(1)$, and in this sense observing $\hat{U}_i$ is as good as observing $U_i$. Similarly, defining $\hat{F}_t = \sum_{n=2}^r \hat{f}_n$ for $t = 2, \ldots, T$, we can show that observing $\hat{F}_i$ is as good as observing $F_i$. In section ‘Testing for unit roots’, we elaborate on this.

Remark 3. As mentioned in section ‘Assumptions’, Assumption 1 (i) can be relaxed to accommodate serial correlation in $v_t$. Let us therefore assume that $a(L)v_t = e_t$, where $v_t$ satisfies Assumptions 1 (i) and $a(L) = \sum_{n=0}^\rho a_n L^n$ has all its roots outside the unit circle and $a_0 = 1$. Serial correlation of this type can be ‘filtered away’ prior to the construction of $\widehat{PIS}_i$, which can be accomplished by simply replacing $\hat{f}_i$ by the residual from a regression of $\hat{f}_i$ onto $\hat{f}_{i-1}, \ldots, \hat{f}_{i-p}$.

We are now ready for our first main result.

**Proposition 1.** Under Assumption 1, as $N, T \to \infty$,

$$N|\widehat{PIS}_i - PIS_i| = o_p(1).$$

Proposition 1 can be interpreted as follows. Note that $\sum_{j=1}^N \lambda_j^2 \sigma^2_{v,j} = O_p(N)$, and therefore $PIS_i = o_p(1)$. Hence, as expected, in the current large-$N$ setting, the weight attached to each cross-section unit converges to zero. In fact, as we show in Appendix B (Proof of Proposition 1), $N \cdot PIS_i$ converges to a constant. Proposition 1 says that $N \cdot \widehat{PIS}_i$ converges to the same constant and therefore $N|\widehat{PIS}_i - PIS_i|$ converges to zero. In this sense, $PIS_i$ is consistent for $PIS_i$.

Remark 4. The large-$N$ asymptotic framework used here may appear somewhat strange, in the sense that in applications $N$ is always finite. The reason for working with this framework is the same as why one in the time series literature tends to assume that $T$ is large, that is, it facilitates an analysis of what happens as $N$ increases. In section ‘Testing for unit roots’, we show that while the performance improves as $N$ increases, the new approach works well even when $N$ is relatively small. Hence, as in the empirical VECM-based literature where a large-$T$ method is applied to finite-$T$ data, the approach developed here can and should of course be applied even if $N$ and $T$ are finite. An important difference here is that our approach explores the information contained in both the time series and cross-sectional dimensions of the panel, as compared to VECMs, which only explore the time series information and whose performance actually decreases with increased cross-sectional variation.

Proposition 1 is a pure consistency result that is silent regarding the asymptotic distribution of $\widehat{PIS}_i$. The next proposition can be used as a basis for testing the significance of the individual PISs. As is well known, $f_i$ and $\lambda_i$ are not separately identifiable. This is true in the...
present panel common factor model context, but is known also from the time series cointegration literature (see Escrivano and Peña, 1994, for a detailed discussion). In Appendix B we show that $\hat{\lambda}_i (\hat{f}_i)$ is consistent for $\lambda_i \tilde{\lambda}^{-1} (\tilde{\lambda} \hat{f}_i)$. The fact that $f_i$ and $\lambda_i$ are not separately identifiable is unproblematic for two reasons. First, the product $\hat{\lambda}_i \hat{f}$, still estimates $\lambda_i f_i$, which is what we need. Secondly, similar to, for example, Figuerola-Ferreti and Gonzalo (2010), our interest lies only in testing $\lambda_i = 0$, which is of course unaffected by scaling.

A more serious problem, however, is that for practice. In fact, most (if not all) existing works on price discovery involve samples where $T > N$. Our approach is therefore based on bias-adjustment. Define $\hat{\sigma}_{u,i}^2 = T^{-1} \sum_{t=2}^{T} \hat{u}_{i,t}^2$ and $\hat{\sigma}_{u}^2 = N^{-1} \sum_{i=1}^{N} \hat{\sigma}_{u,i}^2$. The bias-adjusted estimator of $\lambda_i \tilde{\lambda}^{-1}$ is given by

$$\hat{\lambda}_{BA,i} = \hat{\lambda}_i - N^{-1} \hat{\sigma}_{u,i}^2 (\hat{\sigma}_{u,i}^2 - \hat{\lambda}_i \hat{\sigma}_{u}^2).$$

The asymptotic distribution of $\sqrt{T} (\hat{\lambda}_{BA,i} - \lambda_i \tilde{\lambda}^{-1})$ is given in Proposition 2. Before we come to the proposition, however, it is useful to introduce some notation. For a product $a_{i,t} b_{j,t}$ of any two variables $a_{i,t}$ and $b_{j,t}$, we define

$$\omega_{ab,ij}^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(a_{i,t} b_{j,t} a_{i,s} b_{j,s})$$

as the ‘long-run variance’ of $a_{i,t} b_{j,t}$. If $i = j$, then we write $\omega_{ab,ii}^2 = \omega_{ab,ii}^2$. Similarly, if $b_{j,t} = 1$, then we write $\omega_{a_{ii}}^2 = \omega_{a_{ii}}^2$. It is also convenient to introduce $v_{i,t} = (u_{i,t} - \sigma_{u,i}^2)$, $\phi_{u,i}^2 = \lim_{N \to \infty} N^{-1} \sum_{j \neq i}^N \omega_{u,ij}$, $\sigma_{u,i}^2 = E(u_{i,t}^2)$ and $\sigma_{u}^2 = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \sigma_{u,i}^2$.

**Proposition 2.** Under Assumption 1, as $N, T \to \infty$ with $\sqrt{T}/N^{3/2} \to 0$,

$$\lambda_i \sigma_{\\eta}^2 \lambda_i \sigma_{\\eta}^2 \sigma_{u,i}^2 + N^{-1} (\lambda_i \sigma_{\\eta}^2 \sigma_{u,i}^2 + \phi_{u,i}^2) + N^{-2} \omega_{x,i}^2)^{-1/2} \sqrt{T} (\hat{\lambda}_{BA,i} - \lambda_i \tilde{\lambda}^{-1}) \to_d N(0, 1),$$

where $\to_d$ denotes convergence in distribution.

**Remark 5.** As the formula for $\hat{\lambda}_{BA,i}$ suggests, the bias is decreasing in $N$, and, according to Proposition 2, so is the variance. The fact that both the bias and variance of $\hat{\lambda}_{BA,i}$ are decreasing in $N$ is in sharp contrast to the VECM approach, in which the estimation uncertainty is increasing in $N$. In other words, in contrast to the VECM approach, as mentioned in Remark 4, the factor analytical approach considered here exploits the cross-sectional variation of the panel.

**Remark 6.** Unlike Proposition 1, Proposition 2 puts restrictions on the relative expansion rate of $N$ and $T$. The reason for this difference is that while Proposition 2 is an asymptotic distribution result, Proposition 1 is a pure consistency results. In the proof of the latter we show that $N^{3/2} (\tilde{P} S - PIS)$ converges in distribution, but only if $N/T \to 0$ as $N, T \to \infty$. Convergence in distribution is therefore more demanding in terms of the permissible expansion rates of $N$ and $T$ than consistency. The reason being the relatively heavier scaling by $N$. 

© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd
Proposition 2 is useful for constructing test statistics of the null hypothesis of $PI_{Si} = 0$. Indeed, since $\sigma^2_\eta$ and $\sigma^2_\nu_{ij}$ are both positive, for a given $N$, testing $PI_{Si} = 0$ is the same as testing $H_0: \lambda_i = 0$. Proposition 2 implies that under this null,

$$\lambda^2 \sigma^2_\eta (\lambda^2 \sigma^2_\nu_{u,i} + N^{-1} \hat{\phi}^2_{uu,i} + N^{-2} \hat{\omega}^2_{u,i})^{-1/2} \sqrt{T} \hat{\eta}_{BA,i} \to_d N(0, 1) \quad (11)$$

as $N, T \to \infty$ with $\sqrt{T}/N^{3/2} \to 0$. The requirement that $\sqrt{T}/N^{3/2} \to 0$, or $N^3 > T$, represents an improvement over the assumption that $\sqrt{T}/N \to 0$. Interestingly, there is one instance when also the $\sqrt{T}/N^{3/2} \to 0$ requirement can be relaxed. This is when testing $H_0$ and using the following restricted version of $\hat{\lambda}_{BA,i}$:

$$\hat{\lambda}_{BA,i} = \hat{\lambda}_i - N^{-1} \hat{\sigma}_\eta^2 \hat{\sigma}^2_{u,i}.$$

The asymptotic distribution of this estimator under $H_0$ is given in the following corollary to Proposition 2.

**Corollary 1.** Under Assumption 1 and $H_0$, as $N, T \to \infty$,

$$\lambda^2 \sigma^2_\eta (\lambda^2 \sigma^2_\nu_{u,i} + N^{-1} \hat{\phi}^2_{uu,i} + N^{-2} \hat{\omega}^2_{u,i})^{-1/2} \sqrt{T} \hat{\lambda}_{RBA,i} \to_d N(0, 1).$$

A remarkable feature about Corollary 1 is that it holds regardless of the relative expansion rate of $N$ and $T$. Hence, when using $\hat{\lambda}_{RBA,i}$ to infer $H_0$ there are no restrictions on $N$ and $T$.

Inference based on Proposition 2 and/or Corollary 1 requires consistent estimators of $\hat{\phi}^2_{uu,i}$ and $\hat{\omega}^2_{u,i}$. Natural candidates are given by heteroskedasticity and autocorrelation consistent (HAC) estimators similar to those considered by Bai (2003, section 5). Let us therefore denote by $\hat{a}_{i,t} (\hat{b}_{i,t})$ any consistent estimator of $a_{i,t} (b_{i,t})$. The estimator of $\omega^2_{ab,ij}$ is given by

$$\hat{\omega}^2_{ab,ij} = \frac{1}{T} \sum_{t=2}^T \hat{a}_{i,t} \hat{b}_{j,t} + 2 \sum_{k=2}^K \left( 1 - \frac{k}{K + 1} \right) \frac{1}{T} \sum_{t=k+1}^T \hat{a}_{i,t} \hat{b}_{j,t} \hat{a}_{i,t-k} \hat{b}_{j,t-k},$$

where $K$ is a bandwidth parameter satisfying $K/T^{1/4} \to 0$ as $K, T \to \infty$. Hence, $\hat{\omega}^2_{ab,ij}$ is $\hat{\omega}^2_{ab,ij}$ with $\hat{a}_{i,t} \hat{b}_{j,t} = \hat{a}_{i,t} \hat{b}_{j,t}$. Let $\hat{\phi}^2_{uu,i} = N^{-1} \sum_{j \neq i} \hat{\omega}^2_{uu,ij}$. Note that $\hat{f}_t$ estimates $\hat{f}_t$, not $f_t$. Therefore, $\hat{\sigma}_\eta^2$ estimates $\lambda^2 \sigma^2_\eta$. The appropriate statistic to consider for the test of $H_0: \lambda_i = 0$ ($PI_{Si} = 0$) is consequently given by

$$t_{RBA,i} = \hat{\omega}^2_{ab,ij} (\hat{\sigma}_\eta^2 \hat{\sigma}^2_{u,i} + N^{-1} \hat{\phi}^2_{uu,i} + N^{-2} \hat{\omega}^2_{u,i})^{-1/2} \sqrt{T} \hat{\lambda}_{RBA,i},$$

which has a limiting $N(0, 1)$ distribution under $H_0$ as $N, T, K \to \infty$ with $K/T^{1/4} \to 0$. The asymptotic distribution of $t_{RBA,i}$, the $t$-statistic based on $\hat{\lambda}_{RBA,i}$, is the same but only under the additional requirement that $\sqrt{T}/N^{3/2} \to 0$.

The above $t$-statistic is suitable for testing $\lambda_i = 0$ for a single cross-section unit $i$. Testing multiple cross-section units is, however, just as easy. Suppose therefore that we are
interested in testing \( H_0 : \lambda_i = 0 \) for \( i \in S \subset \{1,\ldots,N\} \). Let \( n < N \) denote the cardinality of \( S \). This restriction can be tested using the following Wald-type test statistic:

\[
W_{RBA,S} = \sum_{i \in S} t_{RBA,i}^2
\]

which has a limiting \( \chi^2(n) \) distribution under \( H_0 \).

**Remark 7.** As a measure the fraction of units for which \( \lambda_i \neq 0 \) one may consider the empirical rejection frequency of \( t_{RBA,1}, \ldots, t_{RBA,N} \), as defined by

\[
\hat{R}_{RBA} = \frac{1}{N} \sum_{i=1}^{N} 1(\lvert t_{RBA,i} \rvert > c_2),
\]

where \( 1(A) \) is the indicator function for the event \( A \) and \( c_2 \) is the appropriate right-tail \( \alpha \)-level critical value from \( \mathcal{N}(0, 1) \).

**Remark 8.** The simplicity by which the significance of the PISs can be tested stands in sharp contrast to the standard approach of Hasbrouck (1995), which is not well-suited for testing (see e.g. Figuerola-Ferreti and Gonzalo, 2010, for a discussion).

### Testing for unit roots

The above asymptotic results hold regardless of the value taken by \( \rho_i \in (-1, 1] \). For purpose of interpretation, however, it is important that \( \lvert \rho_i \rvert < 1 \), because only if \( U_{i,t} \) is stationary is it meaningful to refer to \( F_t \) as a ‘fundamental price’. But \( T^{-1/2} \lvert \hat{U}_{i,t} - U_{i,t} \rvert = o_p(1) \) (under \( \rho_i = 1 \)), which means that testing for a unit root in \( \hat{U}_{i,t} \) is asymptotically the same as testing for a unit root in \( U_{i,t} \). The testing of \( \hat{U}_{i,t} \) therefore does not require any specialized techniques, but can be carried out using any existing unit root test, such as the augmented Dickey–Fuller (ADF) test. The fact that \( U_{1,t}, \ldots, U_{N,t} \) are independent means that \( \hat{U}_{1,t}, \ldots, \hat{U}_{N,t} \) are independent too (although strictly speaking the latter independence only holds asymptotically). Hence, when testing \( \hat{U}_{1,t}, \ldots, \hat{U}_{N,t} \) one can also consider so-called ‘first-generation’ panel unit root tests for cross-sectionally independent data, such as the Im–Pesaran–Shin (IPS) test. In section IV we elaborate on this.

That \( F_t \) is unit root non-stationary is necessary for the asymptotic results reported in section ‘Asymptotic theory’ to hold, and is in this sense similar to the assumptions of the previous literature. However, as alluded to in section ‘Comparison with existing models’, in contrast to the PT approach, here the non-stationarity of \( F_t \) is not an identifying assumption, which means that it is testable. Let \( \hat{F}_t = \sum_{n=2}^{T} \hat{f}_n \) be as in Remark 2. By using the results provided in Appendix B of this paper, it can be shown that \( T^{-1/2} \lvert \hat{F}_t \rvert = o_p(1) \). Hence, as expected given that \( \lambda_i \) and \( f_i \) are not separately identifiable, \( F_t \) can only be estimated up to a multiplicative scalar. Of course, if \( F_t \) has a unit root, then so does \( \hat{F}_t \). In other words, for the purpose of testing the unit root restriction, the fact that \( F_t \) is not identified is not a problem. As with \( \hat{U}_{i,t} \), the testing can be carried out using any existing unit root test, such as the ADF test.


Monte Carlo simulations

A small-scale Monte Carlo study was undertaken to investigate the small-sample properties of $\hat{PIS}_1$; however, we start by considering the bias-adjusted (restricted and unrestricted) estimators of $\lambda_i$, which are of course key in inferring $\hat{PIS}_1$. The choice of which cross-section unit $i$ to consider is of course arbitrary. The results reported here are for $i = 1$. The DGP is given by a simplified version of equations (2)–(4), in which both $\eta_i$ and $\epsilon_{it}$ are drawn from $N(0, 1)$.\(^2\) We further set $\rho_1 = \cdots = \rho_N = 0$, which means that prices are martingales. This is realistic if trade is unlimited and costless, but not if trade is costly and/or constrained. We therefore experimented with alternative parameterizations of $\rho_1, \ldots, \rho_N$. The conclusions were, however, unaffected by this, as expected given that the asymptotic properties of $\hat{PIS}_1, \hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ do not depend on $\rho_1, \ldots, \rho_N$. The properties of $\hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ depend on $\lambda_1$ but not on $\lambda_2, \ldots, \lambda_N$, which are also unidentifiable. We therefore follow the bulk of the previous literature on common factor models and draw $\lambda_2, \ldots, \lambda_N$ from the normal distribution (see e.g. Bai, 2003; Pesaran, 2006). In particular, by drawing from $N(1, 1)$, we ensure both that the structural restriction of unit loadings is correct on average, and that Assumption 1 (iv) is fulfilled. Also, depending on the purpose of the experiment, we set $\lambda_1$ to one of 0, 0.25, 0.5, 0.75 or 1. In interest of space we focus on the bias and root mean squared error (RMSE) of the estimators, and the size and power of a nominal 5% level test of the null hypothesis of $H_0 : \hat{\lambda}_1 = 0$.

Table 1 reports size, bias and RMSE for $\hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ when $\lambda_1 = 0$, which is the relevant scenario when wanting to infer if $PIS_1 = 0$. The first thing to notice is that the test statistics of both estimators have a tendency to be oversized; however, this is mainly among the smaller values of $N$ and $T$. Specifically, while the accuracy of the restricted estimator improves very quickly with increases in both $N$ and $T$, for the unrestricted estimator the performance is quite flat in $T$. The reason for the distortions can be found by looking at the bias results; unlike the restricted estimator, for the unrestricted estimator $N = 10$ is apparently too small, and therefore the bias remains even if $T$ increases. Of course, given the $\sqrt{T}/N^{3/2} \to 0$ condition, the fact that the restricted estimator requires $N > 10$ does not come as a surprise.

The power of the tests for different values of $\lambda_1 \neq 0$ are reported in Table 2. Consider the $t$-statistic for testing $H_0$ based on $\hat{\lambda}_{BA,1}$. The numerator of this test statistic is given by $\sqrt{T} \hat{\lambda}_{BA,1} = \sqrt{T} (\hat{\lambda}_{BA,1} - \hat{\lambda}_1) + \sqrt{T} \hat{\lambda}_1$, where the first term is normally distributed by Proposition 2, while the second term is $O_p(\sqrt{T})$ whenever $\hat{\lambda}_1 \neq 0$. The same is true for the $t$-statistic based on $\hat{\lambda}_{RBA,1}$. It follows that under $H_1 : \hat{\lambda}_1 \neq 0$ both $t$-statistics diverge at the rate $\sqrt{T}$. Power should therefore be increasing in $T$ but not in $N$, which is also what we see in the table. We also see that the power of $\hat{\lambda}_{BA,1}$ is generally higher than that of $\hat{\lambda}_{RBA,1}$, which is partly expected given its size distortions under $H_0$ and the fact that the reported powers are not corrected for size.

The results for the PIS are summarized in Figure 1, which reports the absolute bias of $N \cdot \hat{PIS}_1$ as a function of $N$ and $T$ for the above DGP when $\lambda_1 = 1$. In the proof of Proposition 1, we show that $N|\hat{PIS}_1 - PIS_1| = O_p(T^{-1/2}) + O_p(N^{-1/2})$. In accordance with

---

\(^2\)It is common to assume normal innovations (see e.g. de Harris et al., 2002; Hasbrouck, 2002). We have also experimented with $t$-distributed innovations, but without any changes to the conclusions.
**TABLE 1**

*Size, bias and RMSE for $\hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ when $\lambda_1 = 0$*

<table>
<thead>
<tr>
<th>Running</th>
<th>Size</th>
<th>Bias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 10, N$ runs</td>
<td>$N = 10, T$ runs</td>
<td>$N$ and $T$ run</td>
</tr>
<tr>
<td>$\hat{\lambda}_{BA,1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.207</td>
<td>0.176</td>
<td>0.245</td>
</tr>
<tr>
<td>10</td>
<td>0.129</td>
<td>0.095</td>
<td>0.114</td>
</tr>
<tr>
<td>15</td>
<td>0.129</td>
<td>0.129</td>
<td>0.129</td>
</tr>
<tr>
<td>20</td>
<td>0.102</td>
<td>0.081</td>
<td>0.063</td>
</tr>
<tr>
<td>25</td>
<td>0.077</td>
<td>0.081</td>
<td>0.081</td>
</tr>
<tr>
<td>30</td>
<td>0.081</td>
<td>0.102</td>
<td>0.063</td>
</tr>
<tr>
<td>40</td>
<td>0.073</td>
<td>0.105</td>
<td>0.057</td>
</tr>
<tr>
<td>50</td>
<td>0.078</td>
<td>0.118</td>
<td>0.061</td>
</tr>
<tr>
<td>75</td>
<td>0.072</td>
<td>0.125</td>
<td>0.061</td>
</tr>
<tr>
<td>100</td>
<td>0.079</td>
<td>0.114</td>
<td>0.052</td>
</tr>
<tr>
<td>125</td>
<td>0.078</td>
<td>0.122</td>
<td>0.049</td>
</tr>
<tr>
<td>150</td>
<td>0.068</td>
<td>0.118</td>
<td>0.051</td>
</tr>
<tr>
<td>$\hat{\lambda}_{RBA,1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.082</td>
<td>0.107</td>
<td>0.132</td>
</tr>
<tr>
<td>10</td>
<td>0.061</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>15</td>
<td>0.050</td>
<td>0.054</td>
<td>0.045</td>
</tr>
<tr>
<td>20</td>
<td>0.064</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td>25</td>
<td>0.055</td>
<td>0.047</td>
<td>0.054</td>
</tr>
<tr>
<td>30</td>
<td>0.060</td>
<td>0.053</td>
<td>0.048</td>
</tr>
<tr>
<td>40</td>
<td>0.055</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>50</td>
<td>0.063</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>75</td>
<td>0.064</td>
<td>0.047</td>
<td>0.056</td>
</tr>
<tr>
<td>100</td>
<td>0.070</td>
<td>0.042</td>
<td>0.047</td>
</tr>
<tr>
<td>125</td>
<td>0.073</td>
<td>0.037</td>
<td>0.045</td>
</tr>
<tr>
<td>150</td>
<td>0.065</td>
<td>0.034</td>
<td>0.050</td>
</tr>
</tbody>
</table>
A factor analytical approach

TABLE 2

Power for $\hat{\lambda}_{BA,1}$ and $\hat{\lambda}_{RBA,1}$ when $\lambda_1 = 0.25, 0.5, 0.75, 1$

<table>
<thead>
<tr>
<th>Running/$\lambda_1$</th>
<th>$T = 10, N$ runs</th>
<th>$N = 10, T$ runs</th>
<th>$N$ and $T$ runs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25  0.5  0.75  1</td>
<td>0.25  0.5  0.75  1</td>
<td>0.25  0.5  0.75  1</td>
</tr>
<tr>
<td>$\hat{\lambda}_{BA}$</td>
<td>5     0.263  0.368  0.497  0.619</td>
<td>0.201  0.253  0.350  0.446</td>
<td>0.271  0.342  0.423  0.513</td>
</tr>
<tr>
<td></td>
<td>10    0.169  0.276  0.436  0.590</td>
<td>0.169  0.276  0.436  0.590</td>
<td>0.169  0.276  0.436  0.590</td>
</tr>
<tr>
<td></td>
<td>15    0.144  0.258  0.417  0.586</td>
<td>0.179  0.343  0.542  0.729</td>
<td>0.154  0.339  0.555  0.745</td>
</tr>
<tr>
<td></td>
<td>20    0.131  0.247  0.422  0.590</td>
<td>0.205  0.423  0.659  0.837</td>
<td>0.147  0.390  0.638  0.839</td>
</tr>
<tr>
<td></td>
<td>25    0.117  0.225  0.405  0.579</td>
<td>0.220  0.510  0.740  0.900</td>
<td>0.161  0.414  0.724  0.903</td>
</tr>
<tr>
<td></td>
<td>30    0.115  0.225  0.390  0.573</td>
<td>0.243  0.545  0.796  0.928</td>
<td>0.168  0.482  0.787  0.944</td>
</tr>
<tr>
<td></td>
<td>40    0.109  0.212  0.377  0.555</td>
<td>0.297  0.639  0.891  0.970</td>
<td>0.206  0.595  0.894  0.986</td>
</tr>
<tr>
<td></td>
<td>50    0.125  0.228  0.401  0.573</td>
<td>0.336  0.696  0.936  0.989</td>
<td>0.242  0.680  0.943  0.993</td>
</tr>
<tr>
<td></td>
<td>75    0.111  0.213  0.382  0.566</td>
<td>0.460  0.868  0.988  0.997</td>
<td>0.321  0.854  0.994  1</td>
</tr>
<tr>
<td></td>
<td>100   0.111  0.211  0.388  0.581</td>
<td>0.538  0.925  0.998  1</td>
<td>0.422  0.931  1    1</td>
</tr>
<tr>
<td></td>
<td>125   0.109  0.200  0.382  0.573</td>
<td>0.606  0.963  1    1</td>
<td>0.503  0.975  1    1</td>
</tr>
<tr>
<td></td>
<td>150   0.110  0.228  0.401  0.583</td>
<td>0.664  0.989  0.999  0.999</td>
<td>0.562  0.989  1    1</td>
</tr>
<tr>
<td>$\hat{\lambda}_{RBA}$</td>
<td>5     0.125  0.215  0.331  0.453</td>
<td>0.123  0.166  0.250  0.338</td>
<td>0.157  0.205  0.282  0.369</td>
</tr>
<tr>
<td></td>
<td>10    0.096  0.167  0.307  0.462</td>
<td>0.096  0.167  0.307  0.462</td>
<td>0.096  0.167  0.307  0.462</td>
</tr>
<tr>
<td></td>
<td>15    0.095  0.189  0.327  0.488</td>
<td>0.099  0.217  0.413  0.586</td>
<td>0.099  0.242  0.452  0.652</td>
</tr>
<tr>
<td></td>
<td>20    0.097  0.189  0.353  0.524</td>
<td>0.105  0.270  0.505  0.707</td>
<td>0.104  0.311  0.552  0.768</td>
</tr>
<tr>
<td></td>
<td>25    0.091  0.178  0.339  0.513</td>
<td>0.117  0.334  0.577  0.778</td>
<td>0.125  0.349  0.656  0.867</td>
</tr>
<tr>
<td></td>
<td>30    0.086  0.195  0.340  0.515</td>
<td>0.133  0.360  0.664  0.844</td>
<td>0.134  0.427  0.740  0.920</td>
</tr>
<tr>
<td></td>
<td>40    0.092  0.186  0.345  0.520</td>
<td>0.157  0.455  0.749  0.901</td>
<td>0.169  0.553  0.869  0.981</td>
</tr>
<tr>
<td></td>
<td>50    0.109  0.196  0.374  0.543</td>
<td>0.184  0.507  0.833  0.948</td>
<td>0.218  0.645  0.929  0.991</td>
</tr>
<tr>
<td></td>
<td>75    0.099  0.197  0.361  0.545</td>
<td>0.247  0.693  0.929  0.980</td>
<td>0.302  0.837  0.991  1</td>
</tr>
<tr>
<td></td>
<td>100   0.102  0.201  0.378  0.567</td>
<td>0.306  0.799  0.962  0.989</td>
<td>0.404  0.927  1    1</td>
</tr>
<tr>
<td></td>
<td>125   0.103  0.193  0.372  0.558</td>
<td>0.362  0.869  0.984  0.995</td>
<td>0.488  0.974  1    1</td>
</tr>
<tr>
<td></td>
<td>150   0.106  0.220  0.392  0.575</td>
<td>0.414  0.912  0.986  0.993</td>
<td>0.554  0.988  1    1</td>
</tr>
</tbody>
</table>

this, we see that the bias is decreasing in both $N$ and $T$, and also that the rate at which this happens is roughly the same for the two indices.

In sum, we find that the proposed estimators of $\lambda_i$ have good small-sample properties. The restricted estimator performs particularly well, and in fact works quite well even when $N$ and $T$ are as small as 5, which makes it widely applicable. This good performance is reflected in the PIS, which is very accurate. In our empirical illustrations, $N < 10$ and $T > 1,000$, suggesting that inference should be based on the restricted estimator. This is confirmed by Table 3, which contains 5% size and power results for some constellations of $N$ and $T$ with $N^3 < T$ (such that the $\sqrt{T/N^3/2} \to 0$ requirement is not met).

IV. Empirical illustrations

The data sets

We demonstrate the usefulness of our new toolbox for price discovery using three data sets, denoted ‘OIL’, ‘EQUITY’ and ‘OATS’. VECMs become impossible when the number of parameters to be estimated is larger than $T$. As Gonzalo and Pitarakis (1999) show,
however, the problems associated with high-dimensionality starts much earlier, causing misleading inference. VECMs are intrinsically high-dimensional due to a quadratically growing parameter space. For instance, fitting a VECM without cointegration and four lags when $N = 10$ requires estimating 400 parameters, excluding estimation of the error covariance matrix. Then there is also the problem of too wide IS bounds when the number of Cholesky orderings is large. With $N = 10$ there are no less than 3,628,800 such orderings to consider. The values of $N$ ($T$) in the OIL, EQUITY and OATS data sets are 3 (1,401), 9 (1,249) and 20 (84) respectively. Hence, except perhaps for OIL, the data sets considered here are not really feasible when using the VECM approach.

The first data set, OIL, contains three crude oil price series listed on the US, the UK and Oman exchanges. The data are sampled on a daily frequency, covering the period 1/06/2007 to 12/10/2012, culminating into 1,401 time series observations per exchange. All oil price series are downloaded from Bloomberg, and are expressed in US dollars. Studies of oil are relatively rare in the price discovery literature, and as far as we know none has yet studied oil price discovery from a cross-country perspective. This first data set is therefore interesting from an empirical perspective, and also because it has the potential to stimulate more research on price discovery in markets other than cross-listed equities.

The EQUITY data set is more conventional in the sense that we have equity prices of a particular firm that is listed in more than one exchange. We choose Arcelor Mittal, a major steel company, which has its primary listing on the Luxembourg stock exchange and has secondary listings on another eight exchanges. The stock price data are daily, covering the period 30/11/2010 to 1/05/2014 for a total of 1,249 observations per series. As with OIL, the EQUITY data are downloaded from Bloomberg, and are expressed in US dollars.
The OATS data set consists of historical oats prices from across Swedish counties, which were at the time considered as suitable ‘market price districts’ (see Jörberg, 1972, for a detailed discussion). In contrast to the first two data sets, the frequency of OATS data sets is annual. However, the sample period is relatively long, from 1830 to 1913, which means that there is a total of 84 observations available for each county. The number of counties in the OATS data set is 20. The prices are expressed in öre (the centesimal subdivision of the Swedish krona) per kilogram. The data are taken from Myrdal (1933) and have been digitized by the authors. The OATS data set is interesting, because, as far as we are aware, none has considered price discovery in such an historical context before. This is somewhat surprising, given that we have a commodity that is traded on multiple markets and where market frictions are likely to be high.

### Results

In Table 4 we report some descriptive statistics. We begin by considering the results for OIL. The mean and median crude oil price is highest for Oman and lowest for the UK. Likewise, the standard deviation of oil price is highest for the UK, followed by Oman, and the US. The lowest (highest) oil price is recorded for the US (UK) at US$33.98 (US$146.08). These differences in the descriptive statistics are suggestive of different price setting mechanisms. Consider next the results for EQUITY. The first thing to note is that the cross-sectional variation of EQUITY is somewhat smaller than that of OIL. For example, while in case

<table>
<thead>
<tr>
<th>$T$</th>
<th>$N$</th>
<th>$\hat{\lambda}_1$</th>
<th>$\hat{\lambda}_{BB,1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>5</td>
<td>0.700 0.999 0.997 0.999 0.998</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>7</td>
<td>0.479 1 0.999 0.999 1</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>9</td>
<td>0.330 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1,250</td>
<td>5</td>
<td>0.725 1 0.999 0.998</td>
<td></td>
</tr>
<tr>
<td>1,250</td>
<td>7</td>
<td>0.523 1 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1,250</td>
<td>9</td>
<td>0.338 0.999 1 1 1</td>
<td></td>
</tr>
<tr>
<td>1,500</td>
<td>5</td>
<td>0.762 1 1 0.999 0.998</td>
<td></td>
</tr>
<tr>
<td>1,500</td>
<td>7</td>
<td>0.562 1 0.999 0.999 0.999</td>
<td></td>
</tr>
<tr>
<td>1,500</td>
<td>9</td>
<td>0.400 0.999 1 0.999 0.999</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 4

Descriptive statistics

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oman</td>
<td>87.919</td>
<td>22.617</td>
<td>86.7</td>
<td>141.2</td>
<td>35</td>
</tr>
<tr>
<td>USA</td>
<td>85.489</td>
<td>20.071</td>
<td>85.745</td>
<td>145.29</td>
<td>33.98</td>
</tr>
<tr>
<td>UK</td>
<td>81.249</td>
<td>23.11</td>
<td>76.99</td>
<td>146.08</td>
<td>36.61</td>
</tr>
<tr>
<td>EQUITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XETRA ARRB GR</td>
<td>14.715</td>
<td>5.331</td>
<td>12.535</td>
<td>27.94</td>
<td>8.523</td>
</tr>
<tr>
<td>NYSE Euronext Amsterdam MT</td>
<td>14.726</td>
<td>5.339</td>
<td>12.555</td>
<td>28.015</td>
<td>8.443</td>
</tr>
<tr>
<td>TOM MTF MT</td>
<td>14.782</td>
<td>5.479</td>
<td>12.535</td>
<td>29.270</td>
<td>8.417</td>
</tr>
<tr>
<td>BATS Europe MTS</td>
<td>14.595</td>
<td>5.094</td>
<td>12.595</td>
<td>28.130</td>
<td>8.487</td>
</tr>
<tr>
<td>OATS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malmöhus county</td>
<td>9.658</td>
<td>2.587</td>
<td>10.400</td>
<td>14.300</td>
<td>4.300</td>
</tr>
<tr>
<td>Kristianstad county</td>
<td>9.957</td>
<td>2.395</td>
<td>10.100</td>
<td>16.300</td>
<td>5.000</td>
</tr>
<tr>
<td>Kalmar county</td>
<td>8.920</td>
<td>1.955</td>
<td>8.850</td>
<td>13.200</td>
<td>5.100</td>
</tr>
<tr>
<td>Gotland county</td>
<td>8.501</td>
<td>2.283</td>
<td>8.400</td>
<td>13.700</td>
<td>4.200</td>
</tr>
<tr>
<td>Gothenburg and Bohus county</td>
<td>9.542</td>
<td>2.304</td>
<td>9.700</td>
<td>14.900</td>
<td>4.600</td>
</tr>
<tr>
<td>Skaraborg county</td>
<td>8.631</td>
<td>1.889</td>
<td>8.800</td>
<td>13.400</td>
<td>4.400</td>
</tr>
<tr>
<td>Södermanland county</td>
<td>9.155</td>
<td>1.705</td>
<td>9.300</td>
<td>12.900</td>
<td>5.800</td>
</tr>
<tr>
<td>Stockholm county</td>
<td>9.626</td>
<td>1.772</td>
<td>9.700</td>
<td>13.500</td>
<td>5.800</td>
</tr>
<tr>
<td>Uppsala county</td>
<td>9.081</td>
<td>1.721</td>
<td>9.350</td>
<td>12.800</td>
<td>5.800</td>
</tr>
<tr>
<td>Örebro county</td>
<td>9.185</td>
<td>1.941</td>
<td>9.350</td>
<td>14.100</td>
<td>5.100</td>
</tr>
<tr>
<td>Västmanland county</td>
<td>9.036</td>
<td>1.822</td>
<td>9.300</td>
<td>14.200</td>
<td>5.300</td>
</tr>
<tr>
<td>Värmland county</td>
<td>9.726</td>
<td>1.891</td>
<td>9.700</td>
<td>14.600</td>
<td>5.800</td>
</tr>
<tr>
<td>Gästleborg county</td>
<td>9.960</td>
<td>1.822</td>
<td>9.800</td>
<td>15.200</td>
<td>6.400</td>
</tr>
<tr>
<td>Västernorrland county</td>
<td>11.142</td>
<td>2.314</td>
<td>11.300</td>
<td>15.500</td>
<td>6.100</td>
</tr>
</tbody>
</table>

Notes: ‘SD’ refers to the estimated standard deviation.

of OIL the mean price ranges between 81.249 and 87.919, in case of EQUITY the range is between 14.595 and 14.782. We also see that the highest price is recorded at the TOM MTF (the Netherlands) exchange and that the lowest price of US$5.094 is recorded at the BATS Europe exchange. The mean of the öre per kilogram prices in the OATS data set falls in the 8.501 (Gotland) to 11.142 (Västernorrland) range. The fact that the mean (and median) price is relatively high in Västernorrland is to be expected given its location in the rural north of Sweden (where oats cannot be grown).

Table 5 reports some ADF and IPS unit root test results for the estimated common and idiosyncratic components respectively. Both tests allow for a constant and a linear
A factor analytical approach

TABLE 5
Unit root test results

<table>
<thead>
<tr>
<th>Country</th>
<th>Test</th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL</td>
<td>Common</td>
<td>ADF</td>
<td>−1.617</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic</td>
<td>IPS</td>
<td>−1.346</td>
</tr>
<tr>
<td>EQUITY</td>
<td>Common</td>
<td>ADF</td>
<td>−1.351</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic</td>
<td>IPS</td>
<td>−7.445</td>
</tr>
<tr>
<td>OATS</td>
<td>Common</td>
<td>ADF</td>
<td>−2.8242</td>
</tr>
<tr>
<td></td>
<td>Idiosyncratic</td>
<td>IPS</td>
<td>−4.8125</td>
</tr>
</tbody>
</table>

Notes: ‘ADF’ and ‘IPS’ refer to the augmented Dickey–Fuller and Im–Pesaran–Shin unit root tests, respectively.

As expected, we see that while for the common component the null hypothesis of a unit root is not rejected, for the idiosyncratic component the evidence against the unit root null is much stronger. The evidence is, however, not overwhelming, and in two cases the conclusion depends on the chosen significance level. One case is for the idiosyncratic component of OIL, in which the evidence against the unit root null is significant at the 10% level but not at the 5% level. Similarly, while common component of OATS is unit root non-stationary at the 10% level, at the 5% level it is stationary. These findings illustrate quite clearly the importance of being able to test rather than to just assume that the idiosyncratic (common) component is in fact stationary (unit root non-stationary).

Consider next the evidence on price discovery. The results from the (best performing) restricted estimator are reported in Table 6. We begin by looking at the results reported for OIL. The Oman exchange has the highest PIS of 51%, followed by the UK exchange with a PIS of about 45%. The US exchange ends up last with a PIS of 3.6% only. The main implication here is that the Oman and UK exchanges dominate the price discovery in the market for crude oil, and that the contribution of the US exchange is only very marginal. These results are supported by a formal test for no price discovery ($\lambda_i = 0$). Indeed, while for Oman and the UK the hypothesis is comfortably rejected, for the US the evidence is less strong, being significant at the 10% level only. The fact that the PIS is highest in Oman is not unexpected. Indeed, the rapid decline in Dubai crude oil output has increased the (potential) importance of Oman in pricing crude oil. Oman has some of the characteristics to enable it to play a dominating role in the price discovery process. For example, the production is not subject to Organization of Petroleum Exporting Countries (OPEC) quotas as Oman is not a member of OPEC, there are no destination restrictions, Oman’s government offers incentives to international oil companies for extraction and development activities, and Oman has a strong market presence in Asia, one of the fastest growing regions in the world.3

The results for EQUITY show how the price discovery process for Arcelor Mittal is dominated by the primary exchange, the Luxembourg stock exchange, which has a PIS of

---

3 In 2012 over 95% of Oman’s oil was exported to Asia with about half of this exported to China.
<table>
<thead>
<tr>
<th>Series</th>
<th>( \hat{P}_{IS} )</th>
<th>( \hat{\epsilon}_{RBA,i} )</th>
<th>SE</th>
<th>t-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OIL Oman</td>
<td>0.5097</td>
<td>0.9698</td>
<td>0.0360</td>
<td>26.9578</td>
<td>0.0000</td>
</tr>
<tr>
<td>USA</td>
<td>0.0360</td>
<td>0.1126</td>
<td>0.0641</td>
<td>1.7555</td>
<td>0.0792</td>
</tr>
<tr>
<td>UK</td>
<td>0.4543</td>
<td>1.0703</td>
<td>0.0395</td>
<td>27.0631</td>
<td>0.0000</td>
</tr>
<tr>
<td>EQUITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XETRA ARRB GR</td>
<td>0.2244</td>
<td>0.9571</td>
<td>0.0198</td>
<td>48.2879</td>
<td>0.0000</td>
</tr>
<tr>
<td>Luxembourg SE MT</td>
<td>0.3039</td>
<td>1.0039</td>
<td>0.0179</td>
<td>56.1553</td>
<td>0.0000</td>
</tr>
<tr>
<td>NYSE Euronext Amsterdam MT</td>
<td>0.0522</td>
<td>1.0645</td>
<td>0.0137</td>
<td>77.5905</td>
<td>0.0000</td>
</tr>
<tr>
<td>TOM MTF MT</td>
<td>0.0073</td>
<td>1.1293</td>
<td>0.0951</td>
<td>11.8703</td>
<td>0.0000</td>
</tr>
<tr>
<td>BATS Europe MTS</td>
<td>0.0005</td>
<td>0.4958</td>
<td>0.0369</td>
<td>13.4304</td>
<td>0.0000</td>
</tr>
<tr>
<td>BATS Europe MT</td>
<td>0.1049</td>
<td>1.0449</td>
<td>0.0134</td>
<td>78.1444</td>
<td>0.0000</td>
</tr>
<tr>
<td>Chi–X Europe MTS</td>
<td>0.0960</td>
<td>0.9455</td>
<td>0.0174</td>
<td>54.4109</td>
<td>0.0000</td>
</tr>
<tr>
<td>Chi–X Europe MTAS</td>
<td>0.1015</td>
<td>1.0457</td>
<td>0.0133</td>
<td>78.7682</td>
<td>0.0000</td>
</tr>
<tr>
<td>Turquoise MT</td>
<td>0.1093</td>
<td>1.0440</td>
<td>0.0134</td>
<td>78.1465</td>
<td>0.0000</td>
</tr>
<tr>
<td>OATS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stockholm city</td>
<td>0.0672</td>
<td>1.2173</td>
<td>0.0655</td>
<td>18.5761</td>
<td>0.0000</td>
</tr>
<tr>
<td>Malmöhus county</td>
<td>0.0131</td>
<td>0.8493</td>
<td>0.0606</td>
<td>14.0124</td>
<td>0.0000</td>
</tr>
<tr>
<td>Kristianstad county</td>
<td>0.0574</td>
<td>1.0976</td>
<td>0.0846</td>
<td>12.9686</td>
<td>0.0000</td>
</tr>
<tr>
<td>Småland county</td>
<td>0.0387</td>
<td>1.2486</td>
<td>0.0909</td>
<td>13.7431</td>
<td>0.0000</td>
</tr>
<tr>
<td>Kalmar county</td>
<td>0.0517</td>
<td>0.9625</td>
<td>0.0842</td>
<td>11.4258</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gotland county</td>
<td>0.0079</td>
<td>0.6840</td>
<td>0.0692</td>
<td>9.8777</td>
<td>0.0000</td>
</tr>
<tr>
<td>Halland county</td>
<td>0.0348</td>
<td>0.9753</td>
<td>0.0667</td>
<td>14.6242</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gothenburg and Bohus county</td>
<td>0.0553</td>
<td>1.1861</td>
<td>0.0735</td>
<td>16.1402</td>
<td>0.0000</td>
</tr>
<tr>
<td>Älvsborg county</td>
<td>0.0397</td>
<td>0.9599</td>
<td>0.0760</td>
<td>12.6378</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skaraborg county</td>
<td>0.0789</td>
<td>1.0965</td>
<td>0.0663</td>
<td>16.5417</td>
<td>0.0000</td>
</tr>
<tr>
<td>Östergötland county</td>
<td>0.0789</td>
<td>1.1387</td>
<td>0.0729</td>
<td>15.6134</td>
<td>0.0000</td>
</tr>
<tr>
<td>Södermanland county</td>
<td>0.0790</td>
<td>0.8493</td>
<td>0.0554</td>
<td>15.3425</td>
<td>0.0000</td>
</tr>
<tr>
<td>Stockholm county</td>
<td>0.0405</td>
<td>0.7757</td>
<td>0.0550</td>
<td>14.0929</td>
<td>0.0000</td>
</tr>
<tr>
<td>Uppsala county</td>
<td>0.0532</td>
<td>1.0308</td>
<td>0.0800</td>
<td>12.8870</td>
<td>0.0000</td>
</tr>
<tr>
<td>Örebro county</td>
<td>0.1334</td>
<td>1.0134</td>
<td>0.0661</td>
<td>15.3284</td>
<td>0.0000</td>
</tr>
<tr>
<td>Västmanland county</td>
<td>0.0572</td>
<td>1.0196</td>
<td>0.0878</td>
<td>11.6172</td>
<td>0.0000</td>
</tr>
<tr>
<td>Värmland county</td>
<td>0.0562</td>
<td>0.9582</td>
<td>0.0737</td>
<td>12.9961</td>
<td>0.0000</td>
</tr>
<tr>
<td>Kopparberg county</td>
<td>0.0223</td>
<td>0.9487</td>
<td>0.1111</td>
<td>8.5404</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gävleborg county</td>
<td>0.0333</td>
<td>1.1738</td>
<td>0.0880</td>
<td>13.3440</td>
<td>0.0000</td>
</tr>
<tr>
<td>Västernorrland county</td>
<td>0.0011</td>
<td>0.3569</td>
<td>0.0635</td>
<td>5.6201</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

That price discovery is dominated by the primary exchange is consistent with the previous literature (see e.g. Solnik, 1996; Bacidore and Sofianos, 2002). However, while the primary exchange dominates, in agreement with the results of, for example, Chen, Choi and Hong (2013), and Eun and Sabherwal (2003), the contribution of the other exchanges is still important. Indeed, the contribution of the XETRA exchange (Germany-based) is almost as high as for the primary exchange with a PIS of 22.4%, and four of five UK exchanges contribute around 10% each. We also notice that the null hypothesis of no price discovery is comfortably rejected at the 1% level for all nine exchanges.

It remains to consider the results for OATS. We see that the price discovery is dominated by Örebro with a PIS of 13.3%, followed by Skaraborg, Östergötland and Södermanland,
each with a PIS of about 8%. These counties are all located in the fertile southern part of Sweden with access to important shipping lanes (such as Kattegatt, Skagerrak, the lake Vänern and the surrounding river system), which were at the time the most important means of transportation of oats. On the other side of the scale, we have counties such as Västernorrland, which we have already mentioned is located in the very north, and Gotland, which is an island.

V. Conclusion

Increasing globalization and financial integration have, together with recent major events such as the global financial crisis, sparked interest in the issue of price discovery. The question is: to what extent do markets contribute to the (fundamental) price of cross-listed assets? The standard econometric approach by which researchers have been trying to answer this question is based on a fitted VECM, from which the ISs of different markets can be obtained. But while very popular, this approach suffers from a number of important shortcomings that are likely to become even more important in the future as more data become available. Specifically, being multivariate time series, the VECM approach is not equipped to handle data where \( N \) is ‘large’, which obviously puts a restriction on applicability. Also, the resulting ISs depend critically on the appropriateness of the required identifying restrictions. The current paper can be seen as a reaction to these shortcomings. The purpose is to develop a new reduced-form approach that makes use of the cross-sectional variation of the data. It should also be user-friendly, and enable both estimation and statistical inference of the ISs. The solution is a factor analytical approach that has its roots in the large and growing literature on large-dimensional common factor models. The asymptotic properties of the approach are derived and evaluated in small samples using Monte Carlo simulation. In the empirical part of the paper we consider three data sets; (i) crude oil prices listed at three national exchanges, (ii) Arcelor Mittal, a major steel company, whose equity is listed at nine exchanges and (iii) historical prices of oats from across 20 Swedish counties.

Appendix A: Derivation of equation (5)

The IS is obtained as the explanatory power of observed price changes in a regression. Let us use \( v_{i,t} = \tilde{\lambda}_i \eta_t + U_{i,t} \) to denote the sum of the noise coming from the efficient price and market microstructure components of the model. Hence, \( v_{i,t} \) can be seen as the shocks to the observed price, \( X_{i,t} \). In vector notation, we have

\[
v_t = \Lambda \eta_t + U_t, \tag{A1}\]

where \( v_t = (v_{1,t}, \ldots, v_{N,t})' \), \( \Lambda = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N)' \) and \( U_t = (U_{1,t}, \ldots, U_{N,t})' \) are all \( N \times 1 \).

The \( N \times N \) covariance matrix of \( v_t \) is given by

\[
\Sigma_v = \sigma_v^2 \Lambda \Lambda' + \Sigma_U, \tag{A2}\]

where, under Assumption 1, \( \Sigma_U \) is diagonal; \( \Sigma_U = \text{diag}(\sigma_{U,1}^2, \ldots, \sigma_{U,N}^2) \). As in Hasbrouck (1995), De Jong and Schotman (2010) consider the following (reverse) population regression relationship between \( \eta_t \) and \( v_t \):

© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd
where $e_t$ is the part of the innovation in the efficient price that is not due to shocks in the observed price, and $\gamma = (\gamma_1, \ldots, \gamma_N)' = \Sigma_v^{-1} \Lambda \sigma^2_n$. The $R^2$ measure from the above regression is

$$R^2 = \frac{\gamma' \Sigma_v^{-1} \gamma}{\sigma^2_n} = \gamma' \Lambda = \sum_{i=1}^{N} \gamma_i \lambda_i.$$  \hfill (A4)

The IS of market $i$ is given simply by $IS_i = \gamma_i \lambda_i$. In what remains, we show that this formula is in fact equal to the one in equation (5). We begin by using the fact that for a nonsingular $N \times N$ matrix $A$ and an $N \times 1$ vector $a$, we have $(A + aa')^{-1} = A^{-1} - (1 + a'A^{-1}a)^{-1} A^{-1} a a' A^{-1}$ (see Abadir and Magnus, 2005, Exercise 4.28). Applying this to $\Sigma_v^{-1}$ in $\gamma$, we obtain

$$\Sigma_v^{-1} = \Sigma_U^{-1} - \frac{1}{1 + \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda} \sigma^2_n \Sigma_U^{-1} \Lambda \Lambda' \Sigma_U^{-1},$$ \hfill (A5)

which in turn implies

$$\gamma' \Sigma_v^{-1} \Lambda \sigma^2_n = \sigma^2_n \Sigma_U^{-1} \Lambda - \frac{1}{1 + \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda} \sigma^2_n \Sigma_U^{-1} \Lambda \Lambda' \Sigma_U^{-1} \Lambda \sigma^2_n$$

$$= \sigma^2_n \Sigma_U^{-1} \Lambda \left(1 - \frac{1}{1 + \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda} \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda \right) = \frac{\sigma^2_n \Sigma_U^{-1} \Lambda}{1 + \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda}. $$ \hfill (A6)

But $\Sigma_U$ is diagonal and therefore $\Sigma_U^{-1} = \text{diag}(\sigma^{-2}_{U,1}, \ldots, \sigma^{-2}_{U,N})$. Hence, denoting by $[a]_n$ the $n$-th element of the vector $a$, we can show that

$$IS_i = \gamma_i \lambda_i = \frac{\sigma^2_n [\Sigma_U^{-1} \Lambda]_i \lambda_i}{1 + \sigma^2_n \Lambda' \Sigma_U^{-1} \Lambda} = \frac{\sigma^2_n \sigma^{-2}_{U,i} \lambda^2_i}{1 + \sum_{n=1}^{N} \sigma^2_n \sigma^{-2}_{U,n} \lambda^2_n},$$ \hfill (A7)

which is the formula given in (5).

**Appendix B: Proofs**

We start with some notation. The model for $x_{i,t}$ can be written in matrix notation as

$$x_t = f \lambda_t + u_t,$$ \hfill (A8)

where $x_t = (x_{i,1}, \ldots, x_{i,T})'$, $f = (f_2, \ldots, f_T)'$ and $u_t = (u_{i,1}, \ldots, u_{i,T})'$ are $(T-1) \times 1$, while $\lambda_t$ is $1 \times 1$. Alternatively, the model can be written as the following $N$-dimensional system:

$$x_t = \hat{f} \lambda_t + u_t,$$ \hfill (A9)

where $x_t = (x_{1,t}, \ldots, x_{N,t})'$ and $u_t = (u_{1,t}, \ldots, u_{N,t})'$ are $N \times 1$, and $\lambda = (\lambda_1, \ldots, \lambda_N)'$ is $N \times 1$. The matrix notation

$$x = f \lambda' + u$$ \hfill (A10)

will also be used, where $x = (x_1, \ldots, x_N)$ and $u = (u_1, \ldots, u_N)$ are $(T-1) \times N$. In what follows the representations in equations (9)–(11) will be used interchangeably.
Denote by \( \bar{a} = N^{-1} \sum_{i=1}^{N} a_i \) the cross-section average of the generic variable \( a_i \). Many of the results can be expressed in terms of \( \hat{\bar{a}} = \frac{1}{N} \sum_{i=1}^{N} a_i \), whose \( t \)-th element is given by \( \hat{f}_i = \hat{\vec{f}}_i = \bar{u}_i \). Before we come to the proofs of Propositions 1 and 2, we state a preliminary lemma.

**Lemma A.1.** Under Assumption 1, as \( N, T \to \infty \),

\[
\sqrt{NT^{-1/2}} f' \bar{u} \to_d N(0, \sigma_a^2 \sigma_u^2),
\]

\[
NT^{-1} u_i^2 \bar{u} \sim \sigma_a^2 + (T^{-1} \omega_v^2 + NT^{-1} \phi_{u_i,i}^2)^{1/2} N(0, 1),
\]

\[
NT^{-1} \bar{u} \bar{u} = \sigma_a^2 + O_p(N^{-1/2}) + O_p(T^{-1/2}),
\]

where \( v_{i,t} = (u_{i,t}^2 - \sigma_{u_i}^2) \) and

\[
\sigma_a^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_{a_i,i}^2,
\]

\[
\omega_v^2 = \lim_{T \to \infty} \frac{1}{T} \sum_{i=2}^{T} \sum_{s=2}^{T} E(v_{i,s} v_{i,t}),
\]

\[
\phi_{u_i,i}^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{j \neq i} \omega_{u_i,j}^2.
\]

**Proof of Lemma A.1.** Consider the first result. Note that

\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} f' u_i \right)^2 \right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} E(f' u_i u'_j) = \frac{1}{NT} \sum_{i=1}^{N} E(f' u_i u'_i)
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s=2}^{T} E(f_i u_{i,t} f_i u_{i,t})
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} E(f_i^2) E(u_{i,t}^2) + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=2}^{T} \sum_{s \neq t} E(f_i) E(u_{i,t} u_{i,s}) E(f_t)
\]

\[
= \sigma_v^2 \frac{1}{N} \sum_{i=1}^{N} \sigma_{u,i}^2 \to \sigma_a^2 \sigma_u^2
\]

as \( N \to \infty \). Application of the Lindeberg–Feller central limit theorem now yields

\[
\sqrt{NT^{-1/2}} f' \bar{u} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} f' u_i \to_d N(0, \sigma_a^2 \sigma_u^2)
\]

(A11)

as \( N, T \to \infty \).
For the second result, from
\[
E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} v_{i,t} \right)^2 \right] = \frac{1}{T} \sum_{t=2}^{T} E(\v_i, \v_i) \\
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{j=2}^{T} \sum_{j \neq i}^{N} u_{i,j} \right)^2 \right] = \frac{1}{NT} \sum_{t=2}^{T} \sum_{j=2}^{T} \sum_{j \neq i}^{N} E(u_{i,j} \v_i, u_{i,j} \v_i),
\]
we obtain
\[
NT^{-1} \tilde{u}^\prime \tilde{u} = \frac{1}{T} \sum_{t=2}^{T} u_{i,t} N \tilde{v}_i = \frac{1}{T} \sum_{t=2}^{T} \sum_{j=1}^{N} u_{i,j} u_{j,t} = \frac{1}{T} \sum_{t=2}^{T} u_{i,t}^2 + \frac{\sqrt{N}}{\sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{t=2}^{T} \sum_{j \neq i}^{N} u_{i,j} u_{j,t}
\]
\[= \sigma_{u,i}^2 + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \sum_{t=2}^{T} u_{i,t}^2 + \frac{\sqrt{N}}{\sqrt{T}} \frac{1}{\sqrt{NT}} \sum_{t=2}^{T} \sum_{j \neq i}^{N} u_{i,j} u_{j,t}
\]
\[\sim \sigma_{u,i}^2 + (T^{-1} \omega_{i,t}^2 + NT^{-1} \phi_{\v_i}^2)^{1/2} N(0, 1),
\]
which requires \(N, T \to \infty\).

The third and final result follows from noting that
\[
NT^{-1} \tilde{u}^\prime \tilde{u} = \frac{N}{T} \sum_{t=2}^{T} u_{i,t} \tilde{v}_i = \frac{1}{NT} \sum_{t=2}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N} u_{j,i} u_{i,t}
\]
\[= \frac{1}{NT} \sum_{t=2}^{T} \sum_{i=1}^{N} u_{i,t}^2 + \frac{1}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} u_{i,j} u_{j,t}
\]
\[= \frac{1}{N} \sum_{i=1}^{N} \sigma_{u,i}^2 + \frac{1}{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=2}^{T} u_{i,t}^2 - \sigma_{u,i}^2 \right) + O_p(T^{-1/2})
\]
\[= \sigma_u^2 + O_p(N^{-1/2}) + O_p(T^{-1/2}). \tag{A13}
\]

This completes the proof of the lemma.

**Proof of Proposition 1.** We begin by showing the consistency of \(\hat{\lambda}_i \hat{\sigma}_i^2\). The estimator of \(f_i\) is given by
\[\hat{f}_i = \tilde{x}_i = \tilde{\lambda} f_i + \tilde{u}_i = \tilde{\lambda} f_i + O_p(N^{-1/2}), \tag{A14}\]
which holds point-wise in \(i\). However, it is easy to show that the result also holds uniformly in \(t\). Given \(\hat{f} = (\hat{f}_2, \ldots, \hat{f}_T)'\), we compute \(\hat{\lambda}_i = \hat{\sigma}_i^{-2} T^{-1} \hat{f} ' \tilde{x}_i\). By Lemma A.1,
A factor analytical approach

\[ \hat{\sigma}_n^2 = T^{-1} \hat{f}^2 \]

\[ = T^{-1} (\hat{\lambda} f^2 f + 2 \hat{\lambda} f' \bar{u} + \bar{u}' \bar{u}) \]

\[ = \hat{\lambda} - T^{-1} f' f + O_p((NT)^{-1/2}) + O_p(N^{-1}) \]

\[ = \hat{\lambda} \sigma_n^2 + T^{-1/2} \sqrt{T} (T^{-1} f' f - \sigma_n^2) + O_p((NT)^{-1/2}) + O_p(N^{-1}) \]

\[ = \hat{\lambda} \sigma_n^2 + O_p(T^{-1/2}) + O_p(N^{-1}) \quad (A15) \]

Hence, by Taylor expansion,

\[ \hat{\sigma}_n^{-2} = \hat{\lambda}^{-2} \sigma_n^{-2} + O_p(T^{-1/2}) + O_p(N^{-1}) \quad (A16) \]

It is easy to show that \( T^{-1} f' u_i = O_p(T^{-1/2}) \) and \( T^{-1} u_i' \bar{u} = O_p(N^{-1}) + O_p((NT)^{-1/2}) \), giving

\[ \hat{\lambda}_i = \hat{\sigma}_n^{-2} T^{-1} x'_i \hat{f} \]

\[ = \hat{\sigma}_n^{-2} T^{-1} (\hat{\lambda}_i f + u_i) (\hat{\lambda}_i f + \bar{u}) \]

\[ = \hat{\sigma}_n^{-2} T^{-1} (\hat{\lambda}_i \lambda_i f' f + \hat{\lambda}_i f' \bar{u} + \bar{u}' \bar{u}) \]

\[ = \hat{\lambda}_i \hat{\sigma}_n^{-1} \sigma_n^{-2} T^{-1} f' f + O_p(T^{-1/2}) + O_p(N^{-1}) \]

\[ = \hat{\lambda}_i \hat{\sigma}_n^{-1} + O_p(T^{-1/2}) + O_p(N^{-1}) \quad (A17) \]

This result, together with the fact that \( \hat{f}_i = \hat{\lambda}_i f_i + O_p(N^{-1/2}) \), implies

\[ \hat{\lambda}_i \hat{\sigma}_n^{-1} = [\hat{\lambda}_i \hat{\sigma}_n^{-1} + O_p(T^{-1/2}) + O_p(N^{-1})][\hat{\lambda}_i f_i + O_p(N^{-1/2})] \]

\[ = \hat{\lambda}_i f_i + O_p(N^{-1/2}) + O_p(T^{-1/2}). \quad (A18) \]

By taking squares and then averaging over time, we obtain

\[ \hat{\lambda}_i \sigma_n^2 = \hat{\lambda}_i T^{-1} f' f + O_p(N^{-1/2}) + O_p(T^{-1/2}) \]

\[ = \hat{\lambda}_i \sigma_n^2 + T^{-1/2} \hat{\lambda}_i \sqrt{T} (T^{-1} f' f - \sigma_n^2) + O_p(N^{-1/2}) + O_p(T^{-1/2}) \]

\[ = \hat{\lambda}_i \sigma_n^2 + O_p(N^{-1/2}) + O_p(T^{-1/2}), \quad (A19) \]

which establishes the consistency of \( \hat{\lambda}_i \sigma_n^2 \).

Let \( \hat{r}_i = \hat{\lambda}_i \sigma_n^2 \hat{\sigma}_{U,i}^2 \) and \( \hat{R} = N^{-1} \sum_{i=1}^{N} \hat{r}_i \) with analogous definitions of \( r_i \) and \( R \), but without hats. It can be shown that under \( |\rho| < 1 \), \( \sigma^2_{U,i} = \sigma_{U,i} + O_p(T^{-1/2}) \). This result, together with the one for \( \hat{\lambda}_i \sigma_n^2 \), implies

\[ \hat{r}_i - r_i = O_p(N^{-1/2}) + O_p(T^{-1/2}) \]

with \( \hat{R} - R \) being of the same order. Also, by Taylor expansion,

\[ \hat{R}^{-1} = R^{-1} + O_p(N^{-1/2}) + O_p(T^{-1/2}), \quad (A20) \]
which in turn implies
\[ N(PIS_t - PIS_i) = \hat{r}_i \hat{R}^{-1} - r_i R^{-1} = N^{-1}(\hat{r}_i - r_i)R^{-1} + (\hat{R}^{-1} - R^{-1})N^{-1} \hat{r}_i \]  
(A21)
as was to be shown.

**Proof of Proposition 2.** Consider \( \hat{\lambda}_i = \hat{\sigma}_i^2 T^{-1} \hat{f}' x_i \). The numerator is
\[
T^{-1/2} \hat{f}' x_i = T^{-1/2} \hat{f}' \lambda_i + T^{-1/2} \hat{f}' u_i
\]
\[= T^{-1/2} \hat{f}' \lambda_i \hat{\lambda}^{-1} - T^{-1}(\hat{\lambda}^f + \hat{u}) \hat{\lambda}^{-1} \lambda_i + T^{-1}(\hat{\lambda}^f + \hat{u}) u_i \]  
(A22)
which we can rewrite in the following way:
\[
T^{-1/2} \hat{f}' x_i - \sqrt{T} \hat{\lambda}^{-1} \lambda_i \hat{\sigma}^2_q
\]
\[= \sqrt{T} N^{-1} \left( \sigma^2_{u_i} - \lambda_i \hat{\lambda}^{-1} \sigma^2_q \right) - \lambda_i \hat{\lambda}^{-1} N^{-3/2} \sqrt{NT} (NT^{-1} \hat{u} \hat{u} - \sigma^2_{u_i}) + \hat{\lambda} T^{-1/2} \hat{f}' u_i
\]
\[-\lambda_i N^{-1/2} (\sqrt{NT} T^{-1/2} \hat{u} f) + \sqrt{T} N^{-1} (NT^{-1} \hat{u} u_i - \sigma^2_{u_i}). \]

By using calculations similar to those used in Proof of Lemma A.1, it is possible to show
that \( \sqrt{NT} (NT^{-1} \hat{u} \hat{u} - \sigma^2_{u_i}) = O_p(\sqrt{T}) + O_p(\sqrt{N}) \), implying
\[
T^{-1/2} \hat{f}' x_i - \sqrt{T} \hat{\lambda}^{-1} \lambda_i \hat{\sigma}^2_q
\]
\[= \sqrt{T} N^{-1} \left( \sigma^2_{u_i} - \lambda_i \hat{\lambda}^{-1} \sigma^2_q \right) - \lambda_i \hat{\lambda}^{-1} N^{-3/2} \sqrt{NT} (NT^{-1} \hat{u} \hat{u} - \sigma^2_{u_i}) + \hat{\lambda} T^{-1/2} \hat{f}' u_i
\]
(A23)
\[-\lambda_i N^{-1/2} (\sqrt{NT} T^{-1/2} \hat{u} f) + \sqrt{T} N^{-1} (NT^{-1} \hat{u} u_i - \sigma^2_{u_i}) + O_p(N^{-3/2} \sqrt{T}) + O_p(N^{-1}). \]

By assumption, \( T^{-1/2} \hat{f}' u_i \to_d N(0, \sigma^2_{u_i} \sigma^2_q) \) as \( T \to \infty \), and by Lemma A.1, \( \sqrt{NT} T^{-1/2} \hat{u} f \to_d N(0, \sigma^2_{u_i} \sigma^2_q) \) and \( NT^{-1} \hat{u} u_i - \sigma^2_{u_i} \sim (T^{-1} \omega^2_{\sigma_{u_i}} + NT^{-1} \phi^2_{\sigma_{u_i}})^{1/2} N(0, 1) \). We now show that these three terms are asymptotically uncorrelated. We begin with the covariance between
\( T^{-1/2} \hat{f}' u_i \) and \( N^{-1/2} (\sqrt{NT} T^{-1/2} \hat{u} f) \), which is given by
\[
E[(T^{-1/2} \hat{f}' u_i) N^{-1/2} (\sqrt{NT} T^{-1/2} \hat{u} f)] = E \left( \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} f_t u_{t,i} \hat{u}_s f_s \right)
\]
\[= E \left( \frac{1}{NT} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j=1}^{N} f_t u_{t,i} u_{s,j} f_s \right)
\]
\[= \frac{1}{N} \frac{1}{T} \sum_{t=2}^{T} E(f_t^2 u_{t,i}^2) + \frac{1}{NT} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j=1}^{N} E(f_t u_{t,i} u_{s,j} f_s)
\]
\[= N^{-1} \sigma_{u_i}^2 \sigma_{u_i}^2 = O(N^{-1}), \]
which is obviously negligible. Also, \( NT^{-1}\tilde{u}'_i - \sigma^2_{a,i} \) is uncorrelated with \( T^{-1/2}\tilde{u}'_i \) and \( T^{-1/2}f' u_i \);

\[
E[(T^{-1/2}\tilde{u}'_i - \sigma^2_{a,i})] = E \left( \frac{N}{T^{3/2}} \sum_{t=2}^{T} \sum_{s=2}^{T} \tilde{u}_t f_t \tilde{u}_s u_{t,s} \right) - T^{-1/2}\sigma^2_{a,i} E (\tilde{u}' f) = \frac{1}{NT^{3/2}} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j=1}^{N} \sum_{l=1}^{N} E(u_{t,j} u_{s,j} u_{t,l} u_{s,l}) E(f_t) = 0,
\]

\[
E[(T^{-1/2}f' u_i)(NT^{-1}\tilde{u}'_i - \sigma^2_{a,i})] = E \left( \frac{N}{T^{3/2}} \sum_{t=2}^{T} \sum_{s=2}^{T} f_t u_{t,j} \tilde{u}_s u_{t,s} \right) - T^{-1/2}\sigma^2_{a,i} E (f' u_i) = \frac{1}{T^{3/2}} \sum_{t=2}^{T} \sum_{s=2}^{T} \sum_{j=1}^{N} E(f_t) E(u_{t,j} u_{s,j} u_{t,s}) = 0.
\]

These results imply that

\[
T^{-1/2}f' x_i - \sqrt{T} \tilde{\lambda}^{-1}_a \tilde{\sigma}^2_{\eta} - \sqrt{T} N^{-1}(\sigma^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a}) = \sqrt{T} N^{-1} \left( \tilde{\sigma}^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a} \right) - N^{-1} \sqrt{T} \tilde{\sigma}^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a} + N^{-3/2} \tilde{\sigma}^2_{a} \tilde{\lambda}_a \tilde{\sigma}^2_{a} + N^{-1} \sqrt{T} \left( \tilde{\sigma}^2_{a} - \tilde{\lambda}_a \tilde{\sigma}^2_{a} \right) = \sqrt{T} N^{-1} \left( \tilde{\sigma}^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a} \right) + O_p(N^{-1}) + O_p(\sqrt{T} N^{-3/2}).
\]

Hence, provided that \( \sqrt{T} N^{-3/2} = o(1) \),

\[
T^{-1/2}f' x_i - \sqrt{T} \tilde{\lambda}_a \tilde{\sigma}^2_{\eta} - \sqrt{T} N^{-1}(\sigma^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a}) 
\approx [\tilde{\lambda}_a \tilde{\sigma}^2_{\eta} \tilde{\sigma}^2_{a,i} + N^{-1}(\tilde{\lambda}_a \tilde{\sigma}^2_{\eta} \tilde{\sigma}^2_{a} + \tilde{\lambda}_a \tilde{\sigma}^2_{a,i} + N^{-2}\tilde{\sigma}^2_{a,i})]^{1/2} N(0,1).
\]

Another application of the results provided as a part of Proof of Proposition 1 yields \( \tilde{\sigma}^2_{u,i} = \tilde{\lambda}_a \tilde{\sigma}^2_{\eta} + O_p(N^{-1/2}) + O_p(T^{-1/2}) \). It follows that, with \( \tilde{\lambda}_{B,i} = \tilde{\lambda}_a - \tilde{\sigma}^2_{u,i} \),

\[
\sqrt{T} (\tilde{\lambda}_{B,i} - \tilde{\lambda}_a) = \tilde{\sigma}^2_{u,i} - \tilde{\lambda}_a \tilde{\sigma}^2_{a,i} = \tilde{\sigma}^2_{u,i} T^{-1/2} f' x_i - \tilde{\lambda}_a \tilde{\sigma}^2_{a,i} \tilde{\lambda}_a \tilde{\sigma}^2_{a,i}.
\]
\[
\begin{align*}
&= \frac{\hat{x}_i}{\hat{\sigma}_{u,i}^2} \left[ T^{1/2} \hat{\lambda}_i - \sqrt{T} N^{1/2} \left( \tilde{\lambda}_i - \hat{\lambda}_i \right) \right] - \sqrt{T} \hat{\lambda}_i \left( \frac{\hat{\sigma}_{u,i}^2}{\hat{\lambda}_i} \right) + O_p(T^{-1/2}) + O_p(1)
\end{align*}
\]

which again requires $\sqrt{T} N^{-3/2} = o(1)$. The required result is implies by this.

\textbf{Proof of Corollary 1.} This proof follows almost immediately from that of Proposition 2. The only difference is that the bias-correction in $\hat{\lambda}_{RBA,i}$ does not involve $\hat{\lambda}_i$, which in turn eliminates the need for the requirement that $\sqrt{T} N^{-3/2} = o(1)$.

\textbf{Final Manuscript Received: January 2017}

\section*{References}


© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd

© 2017 The Department of Economics, University of Oxford and John Wiley & Sons Ltd
**Lund Economic Studies**

1. Guy Arvidsson  
   Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg  
   A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund  
   Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling  
   Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström  
   Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh  
   Straffrätt och samhällesekonomi, 1973
7. Ulf Jakobsson och Göran Norman  
   Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö  
   Immigration och samhällesekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson  
   Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart  
    Search and the Organization of Marketplaces, 1975
11. S Enone Metuge  
    An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson  
    Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg  
    Effekter av ATP - en samhällesekonomisk studie, 1977
14. Krister Hjalte  
    Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson  
    Social Justice and Fair Distributions, 1977
16. Curt Wells  
    Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren  
    Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl  
19. Inga Persson-Tanimura  
    Studier kring arbetsmarknad och information, 1980
20. Bengt Turner  
    Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson  
22. Daniel Boda Ndlela  
    Dualism in the Rhodesian Colonial Economy, 1981
23. Tom Alberts  
   Agrarian Reform and Rural Poverty: A Case Study of Peru, 1981
24. Björn Lindgren  
   Costs of Illness in Sweden 1964-75, 1981
25. Göte Hansson  
26. Noman Kanafani  
   Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg  
   Inkomsteffekter av invandring, 1983
28. Stefan Hedlund  
   Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson  
   Hushållen och kreditpolitiken. En studie av kreditrestriktioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson  
31. Bengt Assarsson  
   Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato  
   Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard  
   Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos  
   Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle  
36. Peter Stenkula  
   Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens  
   Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren  
39. Jan Petersson  
   Erik Lindahl och Stockholmskolors dynamiska metod, 1987
40. Yves Bourdet  
41. Krister Andersson and Erik Norrman  
42. Tohmas Karlsson  
43. Rosemary Vargas-Lundius  
   Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
<table>
<thead>
<tr>
<th></th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>44.</td>
<td>Lena Ekelund Axelson</td>
<td>Structural Changes in the Swedish Marketing of Vegetables, 1991</td>
</tr>
<tr>
<td>47.</td>
<td>Johan Torstensson</td>
<td>Factor Endowments, Product Differentiation, and International Trade, 1992</td>
</tr>
<tr>
<td>48.</td>
<td>Tarmo Haavisto</td>
<td>Money and Economic Activity in Finland, 1866-1985, 1992</td>
</tr>
<tr>
<td>49.</td>
<td>Ulf Grönkvist</td>
<td>Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992</td>
</tr>
<tr>
<td>50.</td>
<td>Evelyne Hangali Maje</td>
<td>Monetization, Financial Development and the Demand for Money, 1992</td>
</tr>
<tr>
<td>51.</td>
<td>Michael Bergman</td>
<td>Essays on Economic Fluctuations, 1992</td>
</tr>
<tr>
<td>52.</td>
<td>Flora Mndeme</td>
<td>Development Strategy and Manufactured Exports in Tanzania, 1992</td>
</tr>
<tr>
<td>56.</td>
<td>Lisbeth Hellvin</td>
<td>Trade and Specialization in Asia, 1994</td>
</tr>
<tr>
<td>57.</td>
<td>Sören Höjgård</td>
<td>Long-term Unemployment in a Full Employment Economy, 1994</td>
</tr>
<tr>
<td>58.</td>
<td>Karolina Ekholm</td>
<td>Multinational Production and Trade in Technological Knowledge, 1995</td>
</tr>
<tr>
<td>59.</td>
<td>Fredrik Andersson</td>
<td>Essays in the Economics of Asymmetric Information, 1995</td>
</tr>
<tr>
<td>60.</td>
<td>Rikard Althin</td>
<td>Essays on the Measurement of Producer Performance, 1995</td>
</tr>
<tr>
<td>62.</td>
<td>Kristian Bolin</td>
<td>An Economic Analysis of Marriage and Divorce, 1996</td>
</tr>
<tr>
<td>64.</td>
<td>Hossein Asgharian</td>
<td>Essays on Capital Structure, 1997</td>
</tr>
<tr>
<td>65.</td>
<td>Hans Falck</td>
<td>Aid and Economic Performance - The Case of Tanzania, 1997</td>
</tr>
<tr>
<td>66.</td>
<td>Bengt Liljas</td>
<td>The Demand for Health and the Contingent Valuation</td>
</tr>
</tbody>
</table>
67. Lars Pålsson Syll  Utility Theory and Structural Analysis, 1997
68. Richard Henricsson  Time Varying Parameters in Exchange Rate Models, 1997
70. Lars Nilsson  Essays on North-South Trade, 1997
74. Dan Anderberg  Essays on Pensions and Information, 1998
76. Hans-Peter Bermin  Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
82. Mattias Ganslandt  Games and Markets – Essays on Communication, Coordination and Multi-Market Competition, 1999
83. Carl-Johan Belfrage  Essays on Interest Groups and Trade Policy, 1999
84. Dan-Olof Rooth  Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85. Karin Olofsdotter  Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
87. Peter Martinsson  Stated preference methods and empirical analyses of equity in health, 2000
89. Hanna Norberg  Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90. Åsa Hansson  Limits of Tax Policy, 2000
<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>Henrik Amilon</td>
<td>Essays on Financial Models, 2000</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Mattias Lundbäck</td>
<td>Asymmetric Information and The Production of Health, 2000</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Mattias Persson</td>
<td>Portfolio Selection and the Analysis of Risk and Time Diversification, 2001</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>Pontus Hansson</td>
<td>Economic Growth and Fiscal Policy, 2002</td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>Joakim Gullstrand</td>
<td>Splitting and Measuring Intra-Industry Trade, 2002</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Andreas Graflund</td>
<td>Financial Applications of Markov Chain Monte Carlo Methods, 2002</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>Therése Hindman Persson</td>
<td>Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Göran Hjelm</td>
<td>Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>Klas Rikner</td>
<td>Sickness Insurance: Design and Behavior, 2002</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Thomas Ericson</td>
<td>Essays on the Acquisition of Skills in Teams, 2002</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>Thomas Elger</td>
<td>Empirical Studies on the Demand for Monetary Services in the UK, 2002</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Helena Johansson</td>
<td>International Competition, Productivity and Regional Spillovers, 2003</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>Fredrik Gallo</td>
<td>Explorations in the New Economic Geography, 2003</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Fredrik CA Andersson</td>
<td>Interest Groups and Government Policy, A Political Economy Analysis, 2003</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Petter Lundborg</td>
<td>Risky Health Behaviour among Adolescents, 2003</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>Ingemar Bengtsson</td>
<td>Central bank power: a matter of coordination rather than money supply, 2003</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Andreas Bergh</td>
<td>Distributive Justice and the Welfare State, 2003</td>
<td></td>
</tr>
<tr>
<td>No.</td>
<td>Author</td>
<td>Title</td>
<td>Year</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>116</td>
<td>Staffan Waldo</td>
<td>Efficiency in Education - A Multilevel Analysis</td>
<td>2003</td>
</tr>
<tr>
<td>117</td>
<td>Mikael Stenkula</td>
<td>Essays on Network Effects and Money</td>
<td>2004</td>
</tr>
<tr>
<td>118</td>
<td>Catharina Hjortsberg</td>
<td>Health care utilisation in a developing country - The case of Zambia</td>
<td>2004</td>
</tr>
<tr>
<td>120</td>
<td>Mårten Wallette</td>
<td>Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training</td>
<td>2004</td>
</tr>
<tr>
<td>121</td>
<td>Tommy Andersson</td>
<td>Essays on Nonlinear Pricing and Welfare</td>
<td>2004</td>
</tr>
<tr>
<td>122</td>
<td>Kristian Sundström</td>
<td>Moral Hazard and Insurance: Optimality, Risk and Preferences</td>
<td>2004</td>
</tr>
<tr>
<td>123</td>
<td>Pär Torstensson</td>
<td>Essays on Bargaining and Social Choice</td>
<td>2004</td>
</tr>
<tr>
<td>124</td>
<td>Frederik Lundtofte</td>
<td>Essays on Incomplete Information in Financial Markets</td>
<td>2005</td>
</tr>
<tr>
<td>125</td>
<td>Kristian Jönsson</td>
<td>Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data</td>
<td>2005</td>
</tr>
<tr>
<td>126</td>
<td>Henrik Andersson</td>
<td>Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden</td>
<td>2005</td>
</tr>
<tr>
<td>127</td>
<td>Björn Ekman</td>
<td>Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries</td>
<td>2005</td>
</tr>
<tr>
<td>128</td>
<td>Ulf G Erlandsson</td>
<td>Markov Regime Switching in Economic Time Series</td>
<td>2005</td>
</tr>
<tr>
<td>129</td>
<td>Joakim Westerlund</td>
<td>Essays on Panel Cointegration</td>
<td>2005</td>
</tr>
<tr>
<td>130</td>
<td>Lena Hiselius</td>
<td>External costs of transports imposed on neighbours and fellow road users</td>
<td>2005</td>
</tr>
<tr>
<td>131</td>
<td>Ludvig Söderling</td>
<td>Essays on African Growth, Productivity, and Trade</td>
<td>2005</td>
</tr>
<tr>
<td>132</td>
<td>Åsa Eriksson</td>
<td>Testing and Applying Cointegration Analysis in Macroeconomics</td>
<td>2005</td>
</tr>
<tr>
<td>133</td>
<td>Fredrik Hansen</td>
<td>Explorations in Behavioral Economics: Realism, Ontology and Experiments</td>
<td>2006</td>
</tr>
<tr>
<td>135</td>
<td>Christoffer Bengtsson</td>
<td>Applications of Bayesian Econometrics to Financial Economics</td>
<td>2006</td>
</tr>
<tr>
<td>137</td>
<td>Fredrik Wilhelmsson</td>
<td>Trade, Competition and Productivity</td>
<td>2006</td>
</tr>
<tr>
<td>138</td>
<td>Ola Jönsson</td>
<td>Option Pricing and Bayesian Learning</td>
<td>2007</td>
</tr>
<tr>
<td>Number</td>
<td>Name</td>
<td>Title</td>
<td>Year</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>139</td>
<td>Ola Larsson</td>
<td>Essays on Risk in International Financial Markets</td>
<td>2007</td>
</tr>
<tr>
<td>140</td>
<td>Anna Meyer</td>
<td>Studies on the Swedish Parental Insurance</td>
<td>2007</td>
</tr>
<tr>
<td>141</td>
<td>Martin Nordin</td>
<td>Studies in Human Capital, Ability and Migration</td>
<td>2007</td>
</tr>
<tr>
<td>142</td>
<td>Bolor Naranhuu</td>
<td>Studies on Poverty in Mongolia</td>
<td>2007</td>
</tr>
<tr>
<td>143</td>
<td>Margareta Ekbladh</td>
<td>Essays on Sickness Insurance, Absence Certification and Social Norms</td>
<td>2007</td>
</tr>
<tr>
<td>144</td>
<td>Erik Wengström</td>
<td>Communication in Games and Decision Making under Risk</td>
<td>2007</td>
</tr>
<tr>
<td>145</td>
<td>Robin Rander</td>
<td>Essays on Auctions</td>
<td>2008</td>
</tr>
<tr>
<td>146</td>
<td>Ola Andersson</td>
<td>Bargaining and Communication in Games</td>
<td>2008</td>
</tr>
<tr>
<td>147</td>
<td>Marcus Larson</td>
<td>Essays on Realized Volatility and Jumps</td>
<td>2008</td>
</tr>
<tr>
<td>148</td>
<td>Per Hjertstrand</td>
<td>Testing for Rationality, Separability and Efficiency</td>
<td>2008</td>
</tr>
<tr>
<td>149</td>
<td>Fredrik NG Andersson</td>
<td>Wavelet Analysis of Economic Time Series</td>
<td>2008</td>
</tr>
<tr>
<td>150</td>
<td>Sonnie Karlsson</td>
<td>Empirical studies of financial asset returns</td>
<td>2009</td>
</tr>
<tr>
<td>151</td>
<td>Maria Persson</td>
<td>From Trade Preferences to Trade Facilitation</td>
<td>2009</td>
</tr>
<tr>
<td>152</td>
<td>Eric Rehn</td>
<td>Social Insurance, Organization and Hospital Care</td>
<td>2009</td>
</tr>
<tr>
<td>153</td>
<td>Peter Karpestam</td>
<td>Economics of Migration</td>
<td>2009</td>
</tr>
<tr>
<td>154</td>
<td>Marcus Nossman</td>
<td>Essays on Stochastic Volatility</td>
<td>2009</td>
</tr>
<tr>
<td>155</td>
<td>Erik Jonasson</td>
<td>Labor Markets in Transformation: Case Studies of Latin America</td>
<td>2009</td>
</tr>
<tr>
<td>156</td>
<td>Karl Larsson</td>
<td>Analytical Approximation of Contingent Claims</td>
<td>2009</td>
</tr>
<tr>
<td>157</td>
<td>Therese Nilsson</td>
<td>Inequality, Globalization and Health</td>
<td>2009</td>
</tr>
<tr>
<td>158</td>
<td>Rikard Green</td>
<td>Essays on Financial Risks and Derivatives with Applications to Electricity Markets and Credit Markets</td>
<td>2009</td>
</tr>
<tr>
<td>159</td>
<td>Christian Jörgensen</td>
<td>Deepening Integration in the Food Industry – Prices, Productivity and Export</td>
<td>2010</td>
</tr>
<tr>
<td>160</td>
<td>Wolfgang Hess</td>
<td>The Analysis of Duration and Panel Data in Economics</td>
<td>2010</td>
</tr>
<tr>
<td>161</td>
<td>Pernilla Johansson</td>
<td>From debt crisis to debt relief: A study of debt determinants, aid composition and debt relief effectiveness</td>
<td>2010</td>
</tr>
<tr>
<td>162</td>
<td>Nils Janlöv</td>
<td>Measuring Efficiency in the Swedish Health Care Sector</td>
<td>2010</td>
</tr>
</tbody>
</table>
163. Ai Jun Hou  
Essays on Financial Markets Volatility, 2011

164. Alexander Reffgen  

165. Johan Blomquist  
Testing homogeneity and unit root restrictions in panels, 2012

166. Karin Bergman  
The Organization of R&D - Sourcing Strategy, Financing and Relation to Trade, 2012

167. Lu Liu  
Essays on Financial Market Interdependence, 2012

168. Bujar Huskaj  
Essays on VIX Futures and Options, 2012

169. Åsa Ljungvall  
Economic perspectives on the obesity epidemic, 2012

170. Emma Svensson  
Experimenting with Focal Points and Monetary Policy, 2012

171. Jens Dietrichson  
Designing Public Organizations and Institutions: Essays on Coordination and Incentives, 2013

172. Thomas Eriksson  
Empirical Essays on Health and Human Capital, 2013

173. Lina Maria Ellegård  
Political Conflicts over Public Policy in Local Governments, 2013

174. Andreas Hatzigeorgiou  

175. Gustav Kjellsson  
Inequality, Health, and Smoking, 2014

176. Richard Desjardins  
Rewards to skill supply, skill demand and skill match-mismatch – Studies using the Adult Literacy and Lifeskills survey, 2014

177. Viroj  

Jienwatcharamongkhol

178. Anton Nilsson  
Health, Skills and Labor Market Success, 2014

179. Albin Erlanson  
Essays on Mechanism Design, 2014

180. Daniel Ekeblom  
Essays in Empirical Expectations, 2014

181. Sofie Gustafsson  
Essays on Human Capital Investments: Pharmaceuticals and Education, 2014

182. Katarzyna Burzynska  
Essays on Finance, Networks and Institutions, 2015
<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>183</td>
<td>Mingfa Ding</td>
<td>Corporate Ownership and Liquidity in China’s Stock Markets</td>
<td>2015</td>
</tr>
<tr>
<td>184</td>
<td>Anna Andersson</td>
<td>Vertical Trade</td>
<td>2015</td>
</tr>
<tr>
<td>185</td>
<td>Cecilia Hammarlund</td>
<td>Fish and Trips in the Baltic Sea – Prices, Management and Labor Supply</td>
<td>2015</td>
</tr>
<tr>
<td>186</td>
<td>Hilda Ralsmark</td>
<td>Family, Friend, or Foe?</td>
<td>2015</td>
</tr>
<tr>
<td>187</td>
<td>Jens Gudmundsson</td>
<td>Making Pairs</td>
<td>2015</td>
</tr>
<tr>
<td>189</td>
<td>Ida Lovén</td>
<td>Education, Health, and Earnings – Type 1 Diabetes in Children and Young Adults</td>
<td>2015</td>
</tr>
<tr>
<td>190</td>
<td>Caren Yinxia Nielsen</td>
<td>Essays on Credit Risk</td>
<td>2015</td>
</tr>
<tr>
<td>191</td>
<td>Usman Khalid</td>
<td>Essays on Institutions and Institutional change</td>
<td>2016</td>
</tr>
<tr>
<td>193</td>
<td>Milda Norkute</td>
<td>A Factor Analytical Approach to Dynamic Panel Data Models</td>
<td>2016</td>
</tr>
<tr>
<td>194</td>
<td>Valeriia Dzhamalova</td>
<td>Essays on Firms’ Financing and Investment Decisions</td>
<td>2016</td>
</tr>
<tr>
<td>195</td>
<td>Claes Ek</td>
<td>Behavioral Spillovers across Prosocial Alternatives</td>
<td>2016</td>
</tr>
<tr>
<td>196</td>
<td>Graeme Cokayne</td>
<td>Networks, Information and Economic Volatility</td>
<td>2016</td>
</tr>
<tr>
<td>197</td>
<td>Björn Thor Arnarson</td>
<td>Exports and Externalities</td>
<td>2016</td>
</tr>
<tr>
<td>198</td>
<td>Veronika Lunina</td>
<td>Multivariate Modelling of Energy Markets</td>
<td>2017</td>
</tr>
<tr>
<td>199</td>
<td>Patrik Karlsson</td>
<td>Essays in Quantitative Finance</td>
<td>2017</td>
</tr>
<tr>
<td>200</td>
<td>Hassan Sabzevari</td>
<td>Essays on systemic risk in European banking</td>
<td>2017</td>
</tr>
<tr>
<td>201</td>
<td>Margaret Samahita</td>
<td>Self-Image and Economic Behavior</td>
<td>2017</td>
</tr>
<tr>
<td>202</td>
<td>Aron Berg</td>
<td>Essays on informational asymmetries in mergers and acquisitions</td>
<td>2017</td>
</tr>
<tr>
<td>203. Simon Reese</td>
<td>Estimation and Testing in Panel Data with Cross-Section Dependence, 2017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>204. Karl McShane</td>
<td>Essays on Social Norms and Economic Change, 2017</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>