Won-Ki Seo
94 University Avenue
Kingston, Ontario, K7L 3N6, Canada

November 5, 2019

Recruitment Committee
Department of Agricultural Economics
Texas A&M University

Dear Recruitment Committee,

I am interested in the position of Assistant Professor at Texas A&M University. I have enclosed my current curriculum vita, transcript, job market paper. My area of specialization is econometrics focusing on high-dimensional/functional time series analysis. Letters of recommendation have been sent under separate cover.

I will be available for interviews at the ASSA meetings in January 2020. If you have any questions, please contact my advisors or me using the addresses and numbers listed on my curriculum vitae.

Thank you for your consideration.

Sincerely,

Won-Ki Seo

encl: Curriculum vitae, job market paper, transcripts

WON-KI SEO

DEPARTMENT OF ECONOMICS

QUEEN'S UNIVERSITY

Placement Director	Julie Cullen	(858) 822-2056	jbcullen@ucsd.edu
Placement Coordinator	Cathy Pugh	(858) 534-1867	cpugh@ucsd.edu
Placement Coordinator	Jackie Tam	(858) 822-3502	jytam@ucsd.edu

PERSONAL INFORMATION

Mailing Address : 94 University Avenue, Kingston, Ontario, K7L 3N6, Canada
Email : wkseo@econ.queensu.ca
Website : https://sites.google.com/site/wkseo86/

Marital Status : married to Dakyung Seong (Ph.D candidate at UC Davis), one child (born in 2017)
Citizenship : Republic of Korea

CURRENT POSITION

Sir Edward Peacock Post-doctoral fellow, Department of Economics, Queen's University, 2018-2020

FIELD OF INTEREST

Econometric theory, Time series econometrics, High-dimensional/functional data analysis

EDUCATION

University of California, San Diego, 2014 - 2018
- Ph.D. in Economics, 2018
- Ph.D. Candidate in Economics, 2017

Sungkyunkwan University, 2005-2012 (Military service leave Sep 2007- Sep 2009)
- M.A., Economics, 2012
- B.Ec., Economics, 2011

HONORS & AWARDS

Sir Edward Peacock Postdoctoral Fellowship, Queen's University, 2018-2020
Dissertation Fellowship, UCSD, 2018
C.Phil Fellowship, UCSD, 2017-2018
Summer Research Fellowship, UCSD, 2014-2015
Tuition Scholarship, UCSD, 2014-2018
Simsan Merit Fellowship, Sungkyunkwan University, 2011
Academic Excellence Scholarship, Sungkyunkwan University, 2006-2007, 2009-2010

WORKING PAPERS

- Inference on common stochastic trends in functional time series (Job Market Paper)
- Inference on the dimension of the nonstationary subspace in functional time series (with Morten Nielsen and Dakyung Seong), submitted to *Journal of the American Statistical Association* (Sep 2019).
- Extremal behavior of light-tailed Markov-modulated Lévy processes stopped at a state-dependent Poisson rate (with Brendan Beare and Alexis Akira Toda), submitted to *Econometric Theory* (Apr 2019).
- Cointegration and representation of cointegrated processes in Banach spaces, submitted to *Econometric Theory* (Oct 2019).
- Fredholm inversion around a singularity: application to cointegration in Banach space, work in progress.

PUBLICATIONS

- Representation of I(1) and I(2) Autoregressive Hilbertian processes (with Brendan Beare), *Econometric Theory*, forthcoming.
- Cointegrated linear processes in Bayes Hilbert space (with Brendan Beare), *Statistics and Probability Letters*, 147, 2019, pp. 90-95.
- Cointegrated linear processes in Hilbert space (with Brendan Beare and Juwon Seo), *Journal of Time Series Analysis*, 38 (6), 2017, pp. 1010-1027.

PROFESSIONAL ACTIVITIES

Conference Presentations

2019 Annual conference of the CEA, Banff.

Referee Service

Econometric Theory

TEACHING EXPERIENCE

- Queen's University
  Econ 853 : "Applied Econometrics (time series analysis) " (winter 2019)
- University of California, San Diego
  Teaching assistant, 2015-2018

REFERENCES

Brendan K. Beare, University of Sydney, brendan.beare@sydney.edu.au, +61 02 86279414
James G. MacKinnon, Queen's University, jgm@econ.queensu.ca, +1 613 5332293
Morten Ørregaard Nielsen, Queen's University, mon@econ.queensu.ca, +61 04 78831932
Alexis Akira Toda, University of California San Diego, atoda@ucsd.edu, +1 858 5343383

PARCHMENT ID: 25347415
STUDENT NAME: Seo , Won-Ki
SOCIAL SECURITY NUMBER: ***-**-3338

IDENTIFICATION NUMBER: A53-08-0423
DATE AND TIME PRINTED: 10/19/2019 20:37:58

--DEGREES AWARDED BY OTHER INSTITUTIONS--

BA	02/11	College in Korea, South
MA	08/12	College in Korea, South

STUDENT LEVEL	: Graduate
COLLEGE	: Graduate Division
DEPARTMENT(S)	: Economics
MAJOR(S)	: Economics

------------UCSD DEGREES AWARDED----------

AWARD: Candidate in Philosophy CONFERRED: 12/16/17
TERM: Fall Qtr 2017
COLLEGE: Graduate Division
DEPT: Economics
MAJOR: Economics

AWARD: Doctor of Philosophy CONFERRED: 06/15/18
TERM: Spring Qtr 2018
COLLEGE: Graduate Division
DEPT: Economics
MAJOR: Economics

--------------ACADEMIC EVENTS-------------

PHD QUALIFYING EXAM PASSED	09/15/17
PHD ADVANCED TO CANDIDACY	09/28/17
PHD FINAL EXAMINATION PASSED	04/25/18
DOCTORAL DISSERTATION ACCEPTED	05/17/18

Representation Theory for Cointegrated Functional Time Series

-------------COURSE INFORMATION-----------

Fall Qtr 2014 Graduate

ECON	200A	Microeconomics A	4.00	A-	14.80
ECON	205	Mathematics for Economists	4.00	A	16.00
ECON	210A	Macroeconomics A	4.00	A-	14.80
ECON	220A	Econometrics A	4.00	A+	16.00

TERM CREDITS PASSED : 16.00 TERM GPA CREDITS : 16.00
TERM GRADE POINTS : 61.60 TERM GPA : 3.85

Winter Qtr 2015 Graduate

ECON	200B	Microeconomics B	4.00	A	16.00
ECON	210B	Macroeconomics B	4.00	A	16.00
ECON	220B	Econometrics B	4.00	A	16.00
ECON	280	Computation	2.00	S	0.00

TERM CREDITS PASSED : 14.00 TERM GPA CREDITS : 12.00
TERM GRADE POINTS : 48.00 TERM GPA : 4.00

Spring Qtr 2015 Graduate

ECON	200C	Microeconomics C	4.00	A-	14.80
ECON	210C	Macroeconomics C	4.00	A	16.00
ECON	220C	Econometrics C	4.00	A+	16.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 12.00
TERM GRADE POINTS : 46.80 TERM GPA : 3.90

Fall Qtr 2015 Graduate

ECON	220D	Econometrics D	4.00	A	16.00
ECON	272	Finance:T&T Intrtmprl Asset Pr	4.00	A	16.00
MATH	241A	Functional Analysis	4.00	A	16.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 12.00
TERM GRADE POINTS : 48.00 TERM GPA : 4.00

Winter Qtr 2016 Graduate

ECON	220F	Econometrics F	4.00	A	16.00
ECON	227	Nonprametrc/SemiPrametrc Methd	4.00	A	16.00
ECON	270	Finance:Core Asset Pricing	4.00	A	16.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 12.00
TERM GRADE POINTS : 48.00 TERM GPA : 4.00

Spring Qtr 2016 Graduate

ECON	220E	Econometrics E	4.00	B	12.00
ECON	222C	Workshop in Econometrics	4.00	S	0.00
ECON	500C	Teaching Methods in Economics	4.00	S	0.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 4.00
TERM GRADE POINTS : 12.00 TERM GPA : 3.00

Fall Qtr 2016 Graduate

ECON	222A	Workshop in Econometrics	4.00	S	0.00
ECON	250	Labor Economics	4.00	B-	10.80
ECON	285	Pre-Candidacy Presentation	2.00	S	0.00
ECON	296	Original Research Paper	2.00	A	8.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 6.00
TERM GRADE POINTS : 18.80 TERM GPA : 3.13

Winter Qtr 2017 Graduate

ECON	222B	Workshop in Econometrics	4.00	S	0.00
ECON	267	Topics/Environ & Resource Econ	4.00	A	16.00
ECON	500B	Teaching Methods in Economics	4.00	S	0.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 4.00
TERM GRADE POINTS : 16.00 TERM GPA : 4.00

Spring Qtr 2017 Graduate

ECON	281	Special Topics in Economics	4.00	A	16.00
ECON	285	Pre-Candidacy Presentation	2.00	S	0.00
ECON	286	Grad Resrch Presntation Worksh	3.00	S	0.00
ECON	296	Original Research Paper	3.00	A	12.00

TERM CREDITS PASSED : 12.00 TERM GPA CREDITS : 7.00
TERM GRADE POINTS : 28.00 TERM GPA : 4.00

UC San Diego

TRANSCRIPT OF ACADEMIC RECORD

PARCHMENT ID: 25347415
STUDENT NAME: Seo , Won-Ki
SOCIAL SECURITY NUMBER: ***-**-3338

IDENTIFICATION NUMBER: A53-08-0423
DATE AND TIME PRINTED: 10/19/2019 20:37:58

Fall Qtr 2017 Graduate

ECON	212A	Workshop in Macroeconomics	4.00	S	0.00
ECON	222A	Workshop in Econometrics	4.00	S	0.00
ECON	286	Grad Resrch Presntation Worksh	3.00	S	0.00
ECON	500A	Teaching Methods in Economics	4.00	S	0.00

TERM CREDITS PASSED : 15.00 TERM GPA CREDITS : 0.00
TERM GRADE POINTS : 0.00 TERM GPA : 0.00

Winter Qtr 2018 Graduate

ECON	299	Research in Economics	9.00	S	0.00
ECON	500B	Teaching Methods in Economics	4.00	S	0.00

TERM CREDITS PASSED : 13.00 TERM GPA CREDITS : 0.00
TERM GRADE POINTS : 0.00 TERM GPA : 0.00

Spring Qtr 2018 Graduate

ECON	299	Research in Economics	9.00	S	0.00
ECON	500C	Teaching Methods in Economics	4.00	S	0.00

TERM CREDITS PASSED : 13.00 TERM GPA CREDITS : 0.00
TERM GRADE POINTS : 0.00 TERM GPA : 0.00

------------------SUMMARY-----------------

GRADE OPTION	UC-CRDTS ATTM	UC-CRDTS COMPL	CRDTS PSSD	UC-GPA CRDTS	UC-GRADE POINTS	UC-GPA
Letter	85.00	85.00	85.00	85.00	327.20	3.849
S/U	70.00	70.00	70.00	0.00	0.00	0.000
TOTAL	155.00	155.00	155.00	85.00	327.20	3.849

----------------------End of Student Level----------------------

----------------------End of Transcript----------------------

This official university transcript is certified to be a correct transcript of record. Student in good standing unless otherwise indicated.

Transcript void if altered.

Cindy Lyons
University Registrar
Enrollment Management

This document is printed on blue safety paper with UNIVERSITY OF CALIFORNIA SAN DIEGO *printed in white type across the face of the document. It is official if it bears the seal of the University and the Registrar's signature computer printed in black. If photocopied, the word "VOID" will appear prominently across the face of the document. The use of ink eradicator or eraser will be evident and will eliminate the blue background.*

CREDITS: All credits are in quarter units. Cumulative summaries on this record may reflect adjustments for repeated courses and/or other adjustments made in accordance with UC San Diego academic policies.

TRANSFER CREDIT: Only UC San Diego courses and courses taken under official UC San Diego exchange programs with other institutions appear on the transcript. Only grades earned at UC San Diego, at other UC campuses and under the Education Abroad Program are included in the grade point average. All exchange program and transfer credit is included in credits completed.

GRADE INTERPRETATION: Plus (+) and minus (-) grading was approved for use beginning with courses taken in Fall Quarter 1983. The grade of A+, when awarded, represents extraordinary achievement, but does not receive grade point credit beyond that received for the grade of A.

Grade		Grade Points Per Unit
A+, A, A-	Excellent	4.0, 4.0, 3.7
B+, B, B-	Good	3.3, 3.0, 2.7
C+, C, C-	Fair	2.3, 2.0, 1.7
D	Poor (barely passing)	1.0
F	Fail	0.0
E*	Incomplete	**
I	Incomplete	**
IP	In Progress	**
NP	Not Passing (below C-, undergraduates only)	**
NR***	Grade not reported by Instructor	**
P	Passing (C- or better, undergraduates only)	**
S	Satisfactory (B- or better, graduates only)	**
U	Unsatisfactory (below B-, graduates only)	**
W	Withdrew after 4th week of instruction or after second meeting of some laboratory courses	**
Blank	Grade not reported by Instructor	

* Not used after Spring Quarter 1975
** Not included in grade point average
*** Not used after Winter Quarter 1991

School of Medicine and School of Pharmacy & Pharmaceutical Sciences Grades

H/P/F grading system effective Fall, 1986.

H	Honors	NH	Near Honors
P	Pass	S	Satisfactory
F	Fail	U	Unsatisfactory

DEGREE REQUIREMENTS: Undergraduate students must complete a minimum of 180 quarter units with a grade point average of C or better (2.0), satisfy the University of California requirements in American History and Institutions, Diversity, Equity and Inclusion course, and UC Entry Level Writing Requirement (formerly Subject A), satisfy the respective college General Education requirements, and satisfy all requirements for the major. Graduate students must complete their respective degree programs with a grade point average of B (3.0) or better.

COURSE NUMBERS:

Lower Division
1-99 Designed for freshmen and sophomores.

Upper Division
100-199 Designed for juniors and seniors.

Professional
300-399 Designed for teachers or prospective teachers.

Graduate
200-299 Designed for graduate students.
400-499 Rady School of Management.
500-599 For graduate students only.

School of Global Policy and Strategy
(Formerly Graduate School of International Relations & Pacific Studies)
200-295 Courses satisfying Ph.D. requirements.
400-495 Courses satisfying MPIA requirements.

School of Medicine
200-219 Required core courses in years 1 and 2.
220-244 Required core courses in years 1 and 2, effective Fall 2010.
220-295 Departmental pre-clinical electives.
296 Departmental basic science independent study.
299 Independent Study Project.
400-495 Core and elective clerkships in years 3 and 4.
496 Departmental Independent Study.

School of Pharmacy and Pharmaceutical Sciences
200-299 Courses satisfying Pharm.D. requirements.

UNDERGRADUATES:

Honors: Effective Fall Quarter 1978, 14% of graduating seniors who complete at least 80 A-F graded units are eligible for College Honors. Normally, the top 2% are eligible for summa cum laude, the next 4% for magna cum laude, and the remaining 8% for cum laude. Departmental Honors may be awarded to graduating seniors if they complete a special course of study. Provost Honors are awarded quarterly to students who complete 12 or more A-F graded units with a term grade point average of 3.5 or higher.

Physical Education Courses: Through Fall 1994 credit was awarded for all P.E. courses, but only 3 units of activity courses count toward graduation.

Remedial Courses: Remedial courses completed at UC San Diego count as workload credit toward the satisfaction of the minimum progress requirement and eligibility for financial aid, they are included in the cumulative summaries under UC-CRDTS ATTM and UC-CRDTS COMPL. Remedial courses are not applied toward graduation requirements, and the units are excluded from the CRDTS PSSD and UC-GPA CRDTS summaries.

UNIVERSITY OF CALIFORNIA SAN DIEGO
Office of the Registrar, 9500 Gilman Drive
La Jolla, California 92093-0022
(858) 534-3144 FAX (858) 534-5723
http://registrar.ucsd.edu

Repeat Policy: A student may repeat only those courses for which a grade of D, F, NP, U, or W is recorded on the transcript. Repetition of courses for which a grade of C- or higher was awarded is prohibited, unless the course has been specifically approved by the Academic Senate as repeatable for credit.

The first sixteen units of courses that have been repeated by an undergraduate student and for which the student received a D, F, NP, or U are not used in the cumulative grade-point calculations on the student's transcript.

When present, a repeat code indicates that the student's cumulative summary data has been adjusted in accordance with UC San Diego academic policies on repeated courses. Repeat codes appear at the far right of the course following the grade and grade points earned.

Example: MATH 10A Calculus 4.0 F 0.00 **F1**

REPEAT CODE DESCRIPTIONS:

D1	Repeated D - Removed from GPA
D2	Repeat of D - Removed from Units Passed
DA	Additional Repeated D - Removed from GPA & Units Passed
DX	Repeat of D in Excess of 16 units
F1	Repeated F - Removed from GPA
F2	Repeat of F - Grade A - D Received
FA	Additional Repeated F - Removed from GPA
FF	Repeat of F - Grade F Received
FX	Repeat of F in excess of 16 units - Credit Given
FY	Repeat of F in excess of 16 units - No Credit Given
N1	Repeated NP
N2	Repeat of NP - Grade P Received
NA	Additional Repeated NP
NN	Repeat of NP - Grade NP Received
NX	Repeat of NP in Excess of 16 units - Credit Given
NY	Repeat of NP in Excess of 16 units - No Credit Given
OF[1]	Repeat of D/F - Original Course Deleted - F Received
OL[1]	Repeat of D/F - Original Course Deleted - A - D Received
ON[1]	Repeat of NP - Original Course Deleted - NP Received
OP[1]	Repeat of NP - Original Course Deleted - P Received
RF	Repeatable for Credit - F Received
RL	Repeatable for Credit - A - D Received
RN	Repeatable for Credit - NP Received
RP	Repeatable for Credit - P Received
TC	Repeat of Transfer Credit - No Credit Given
UC	UCSD D/F/NP - Repeated at Other UC Campus (Approved)
UF	Repeat of Course from Other UC - F Received
UL	Repeat of Course from Other UC - A - D Received
UN	Repeat of Course from Other UC - NP Received
UP	Repeat of Course from Other UC - P Received
XC	Repeat in Excess of Course Approval
ZC	No Credit - Repeat of C-/Better or P
**	Manually Adjusted Credit

[1] This policy was valid for courses repeated prior to Fall 1975.

This document contains personally identifiable information from a student's education records. It is protected by the Family Educational Rights and Privacy Act (20 U.S.C. § 1232g) and may not be rereleased without the consent of the student.

162499 REVISED 4/2017
SCRIP-SAFE® Security Products, Inc. Cincinnati, OH

The University of California San Diego is accredited by the Western Association of Schools and Colleges

DEPARTMENT OF ECONOMICS
Dunning Hall
Queen's University
Kingston, Ontario, Canada K7L 3N6
Tel 613 533-2250 FAX 613 533-6668
http://www.econ.queensu.ca

November 4, 2019

Dear Sirs,

Won-Ki Seong is currently a post-doctoral fellow at Queen's. He obtained his doctorate from the University of California, San Diego, in 2018 after just four years in their program. He was a student of Brendan Beare. Won-Ki is an extremely talented theoretical econometrician. He already has a very strong portfolio of both published work and research in progress. Up to this point, all but one of his papers have been concerned with the analysis of non-stationary functional time series.

I am sure that Won-Ki could have obtained a tenure-track (or equivalent) position in 2018, but he chose instead to come to Queen's for two years so that his wife, Dakyung Seong, could finish her dissertation and the two of them could go on the market at the same time. She is indeed on the market this year, and I am also writing a (very positive) letter for her.

During the fifteen months or so that he has been at Queen's, Won-Ki has been very productive. His solely-authored job market paper was conceived and completed during this period. So was his paper with Morten Nielsen and Dakyung Seong, which has recently been submitted to a top statistics journal. I believe that he also did quite a bit of work on some papers that were started earlier, notably the one with Beare and Toda (currently the only paper that does concern functional time series), which was submitted about six months ago.

Won-Ki currently has three publications, all of them joint with his supervisor Brendan Beare, plus two papers under submission and three that have not yet been submitted, including his job-market paper. This is very impressive indeed for someone who obtained his doctorate not much more than a year ago and only began his doctoral studies in 2014.

With one exception, all of Won-Ki's work to date involves functional time series. This is, quite frankly, a field that I knew nothing about until I met him. It involves relatively advanced mathematics, and everything that I know about it now has been learned from Won-Ki and, to a lesser extent, from my colleague Morten Nielsen, who is Won-Ki's coauthor on one paper. I therefore cannot provide a really informed assessment of Won-Ki's papers. However, I am certain that he has made important contributions, as evidenced by the fact that two of his earlier papers have already been published in the *Journal of Time Series Analysis* and in *Econometric Theory.*

Functional time series analysis deals with functions that evolve over time. For example, the term structure of interest rates is a function of term to maturity that can, at least in principle, be observed at any point in time for an infinite number of maturities. As time passes, the entire function evolves stochastically. Although one could simply pick one or a few maturities and use conventional univariate or multivariate time-series methods to model how each of them evolves over time, it is probably more attractive to consider the evolution of the entire term structure. Functional time-series methods make this possible.

Another interesting area where functional time-series methods can be used is modeling the evolution of the climate, with (for example) temperature being being thought of as a function of location on the earth's surface. Functional time-series methods can also be used with high-frequency data where it is natural to divide a series into days or weeks. The data for every day in a sample, for example, can be treated as a single observation on a function of the time of day.

Statisticians and econometricians have been working on models for stationary functional time series for some years now, but Won-Ki's work involves integrated (that is, non-stationary) functional time series. The literature on models for this type of data is pretty recent. There has been a great deal of research on unit roots and cointegration for ordinary time series, and it turns out that similar results can in many cases be obtained for functional time series. Won-Ki's papers have made large strides in this direction. Most of them assume that the functions of interest live in a Hilbert space (that is, a space with an inner product, which can be thought of as the infinite-dimensional analog of Euclidean space), but two of the unpublished ones make the weaker assumption that they live in a Banach space.

I will not attempt to discuss any of Won-Ki's papers in this letter, because they are very far from my own areas of interest and expertise. I am sure that his other letter-writers will discuss many of them. The job-market paper appears to be a very impressive piece of work. Based on seeing him present this paper and several other papers during the past year or so, I am confident that he has a deep understanding of the field in which he works and that he will continue to make valuable contributions to it over the next few years. I expect him to be a very productive and successful researcher. Indeed, he is already a very productive one.

Won-Ki could teach advanced courses in several areas of econometrics and statistics, especially time series. With a bit of preparation time, I think that he could also teach a course in machine learning (last year, he audited a course that I taught, and he got up to speed remarkably quickly). He is, of course, also qualified to teach lower-level courses in many areas. He taught a course on time-series methods in our M.A. program last year, and he will be teaching it again this winter.

I am also confident that Won-Ki will be a good colleague. He has been a very active participant in the Quantitative Workshop since he arrived at Queen's. This year he is organizing the Econometrics Working Group, which is an internal workshop series for faculty and students in the field. During his time here, he has presented a large number of talks in this series.

In summary, I believe that Won-Ki has a great deal of potential as a researcher, as evidenced by his three published papers, the two co-authored papers that are under submission, and his three solely-authored papers, including his very impressive job-market paper. I am also confident that he will be a good teacher and a valued colleague. I recommend him very highly.

Sincerely,

James MacKinnon

Sir Edward Peacock Professor of Econometrics
jgm@econ.queensu.ca
613 533-2293

Inference on Common Stochastic Trends in Functional Time Series

(Job Market Paper)

Won-Ki Seo

Department of Economics, Queen's University

October 29, 2019

Abstract

We provide statistical procedures to identify the number of common stochastic trends embedded in functional time series, which may be especially important in economic applications that are often accompanied by nonstationarity. Each of those is given by sequential applications of a proposed test based on a generalized eigenvalue problem associated with sample covariance operators. In particular, we provide a bottom-up procedure that determines the estimate by sequentially testing the null hypothesis that there are a specified number of common stochastic trends against the alternative that there are more. This is distinct from the existing top-down testing procedures and has theoretical advantages especially in an infinite-dimensional setting. Interestingly, the bottom-up procedure entails a top-down procedure as its reverse, and those may be viewed as two different consequences of a single asymptotic phenomenon. We also find some connections between the existing tests and ours: specifically, we provide tests that are asymptotically or exactly equivalent to the existing tests with slight modifications from the eigenvalue problem that our testing procedures are based on. Our Monte Carlo experiments suggest that the finite-sample performances of the proposed tests are satisfactory. We apply our methodology to two empirical examples: U.S. age-specific shares of full-time employment and hourly real wage densities.

1 Introduction

The subject of time series analysis has conventionally dealt with time series that take values in finite-dimensional Euclidean space. A recent literature on so-called functional time series analysis deals with time series that are assumed to take values in a possibly infinite-dimensional Hilbert or Banach space. Each observation of such a time series is, for instance, a continuous function, a square-integrable function, or a probability density function. An important contribution to the literature is the monograph of Bosq (2000), which provides a rigorous treatment of stationary linear processes taking values in Hilbert and Banach spaces. Moreover, Horváth and Kokoszka (2012) discusses statistical analysis of functional time series and provides various empirical applications.

Developments on the functional time series analysis so far have tended to rely on the assumption of stationarity. To the best of the author's knowledge, only a few recent papers have considered nonstationary functional time series. Given that many economic time series exhibit nonstationarity, such a small size of the literature may be an indication that economists have less paid attention to time series analysis of functional observations. A pioneering paper by Chang et al. (2016) appears to be the first effort to fill this gap. The authors consider density-valued cointegrated time series with finitely many (common) stochastic trends in a Hilbert space of square-integrable functions and provide a statistical testing procedure to determine the number of stochastic trends. In addition, Beare et al. (2017) generalizes the notion of cointegration, introduced by Granger (1981), to an arbitrary complex Hilbert space setting and provides a theoretical treatment of cointegrated linear processes.

As in Chang et al. (2016), this paper considers cointegrated linear processes, driven by finitely many stochastic trends and an infinite-dimensional stationary transitory component, in a real separable Hilbert space $\mathcal{H}$ (the CKP model). When this type of cointegrated linear processes is given, it is important to identify the stochastic trends that dominate the long-run behavior of the process, which in fact reduces to determine the number of stochastic trends (Remark 3.15 in Section 3.4). Formal testing procedures commonly used for this purpose in $\mathbb{R}^n$ are the so-called cointegration rank tests. If such a testing procedure supports that there are r cointegrating relationships, then the number of stochastic trends is simply given by $n - r$ as a natural byproduct. In the CKP model, however, we do not have such a nice correspondence: it should be noted that the cointegration 'rank' is always ∞ while the cointegration 'corank', which indicates the number of stochastic trends, is finite. For this reason, every testing procedure to determine the number of stochastic trend is called a cointegration corank test in the present paper. In spite of their empirical relevance, only a few cointegration corank tests are available in practice; as far as this author knows, those provided in Chang et al. (2016) and Nielsen et al. (2019) are all we can use. A common feature of the existing procedures is that they take a top-down approach: for some $\varphi_{\max}$ that is pre-specified by the researcher, they sequentially test the hypotheses

$$H_0 : \text{cointegration corank} = \varphi_0 \qquad \text{against} \qquad H_1 : \text{cointegration corank} < \varphi_0 \tag{1.1}$$

for $\varphi_0 = \varphi_{\max}, \varphi_{\max} - 1, \ldots, 1$. However, a theoretical concern may arise from this top-down approach. Unlike the 'top' hypothesis is naturally constructed by setting $\varphi_{\max} = n$ in $\mathbb{R}^n$, there is no natural starting point of the testing procedure in an infinite-dimensional setting. We thus need prior information on an upper bound for the cointegration corank to construct the top hypothesis. If $\varphi_{\max}$ is smaller than the true cointegration corank by any chance, then the testing procedure always gives an incorrect estimate. It is also worth to mention that the existing testing procedures tend not to reject the top hypothesis if $\varphi_{\max}$ is less than or equal to the true cointegration corank. This implies that (i) non-rejection of the top hypothesis does not mean that $\varphi_{\max}$ is equal to the true value in general and (ii) a potentially important hypothesis that $\varphi_{\max}$ is greater than or equal to the true cointegration corank is not statistically assessed in the existing procedures. Moreover, it is also possible to have different estimates depending on the researcher's conjecture on $\varphi_{\max}$ even at a fixed significance level with the same choice of tuning parameters (if any). These potential issues will be discussed in Section 5 with empirical examples. For these reasons, we note the usefulness of a new testing procedure based on the opposite selection of hypotheses given as follows.

$$H_0 : \text{cointegration corank} = \varphi_0 \qquad \text{against} \qquad H_1 : \text{cointegration corank} > \varphi_0 \tag{1.2}$$

for $\varphi_0 = 0, 1, 2, \ldots$ successively. That is, the null hypothesis of stationarity is examined first,[1] and then the null hypothesis that there are a specified number of stochastic trends is sequentially tested against the alternative that there are more. A clear theoretical advantage of this testing procedure over top-down procedures is that no prior information is required: note that we always have the 'bottom' hypothesis regardless of dimensionality of $\mathcal{H}$. Therefore, the suggested bottom-up selection of hypotheses (1.2) appears to be more natural in our functional setting that is commonly accompanied by an infinite-dimensional $\mathcal{H}$. Even if we leave this theoretical advantage aside, having an alternative testing procedure would be almost always helpful in empirical applications. A new testing procedure can be used as a complement to the existing top-down procedures in a similar way that the KPSS test of Kwiatkowski et al. (1992) is used together with the unit root tests in examining nonstationarity of scalar-valued time series. Moreover, it is quite clear that the estimated cointegration corank from a bottom-up procedure may be used as a guideline for choosing $\varphi_{\max}$ for top-down procedures. If we restrict our discussion to the usual finite-dimensional

[1] In the CKP model, $\varphi_0 = 0$ in (1.2) corresponds to testing stationarity of functional time series against the I(1) alternative.

Euclidean space, there are already available bottom-up procedures such as those in Harris (1997), Snell (1999), and Nyblom and Harvey (2000), which may be useful as alternatives to widely known top-down procedures such as those in Johansen (1995), Breitung (2002), Shintani (2001), Boswijk et al. (2015), etc. It, however, seems that those bottom-up procedures have been less paid attention from applied researchers so far. This might be partly due to that Johansen's VECM model and related statistical procedures, including a well known top-down procedure to determine the number of stochastic trends, have been successfully and popularly applied to economic time series. It is also worth to note that bottom-up procedures are no more or less natural than top-down procedures if the top hypothesis can be properly constructed without any theoretical concern.

This paper provides a bottom-up testing procedure based on a generalized eigenvalue problem associated with a pair of sample covariance operators (hereafter called a variance ratio-type eigenvalue problem). Our approach is inspired by the work of Nielsen et al. (2019) that provides variance ratio-type tests for the number of stochastic trends in an infinite-dimensional Hilbert space setting: the authors suitably generalize Breitung (2002)'s nonparametric cointegration rank test developed in $\mathbb{R}^n$. The proposed bottom-up test in this paper may be viewed as a generalization of the test of Nyblom and Harvey (2000) since it reduces to their test if $\mathcal{H} = \mathbb{R}^n$. More interestingly, our bottom-up test naturally entails a top-down test as its reverse. To be more specific, the bottom-up test (resp. the entailed top-down test) exploits the limiting behavior of the computed eigenvalues (resp. the reciprocals of the computed eigenvalues) in the same variance ratio-type eigenvalue problem. Our bottom-up test has several attractive features, which are shared by the entailed top-down test. First, it is easy to implement in practice. The proposed variance ratio-type eigenvalue problem is associated with sample covariance operators of functional time series projected onto an ℓ-dimensional subspace for some finite integer ℓ, and the test statistic is simply given by the sum of eigenvalues. Due to the finite dimensionality of the projected time series, the test statistic can be easily computed by standard methods. Moreover, our analysis shows that the asymptotic null distribution of the test statistic does not depend on any nuisance parameters and is given by a functional of Brownian motion taking values in a finite-dimensional subspace. Thus, quantiles of the limiting distribution, hence critical values for the statistic, can be easily simulated. Second, our test does not depend on specific parametric assumptions, so it is widely applicable to general cointegrated linear processes. Due to these attractive features of our test, we expect that it would be an appealing alternative option for applied researchers. Our Monte Carlo experiments suggest that the finite-sample performance of the proposed test is satisfactory. In addition, the entailed top-down test also has its own merits over the existing tests suggested by Chang et al. (2016) and Nielsen et al. (2019). We provide a Monte Carlo evidence suggesting that the test can have better finite-sample properties than the existing ones depending on which DGP is employed.

We then find some theoretical connections between our tests and the existing tests. An important feature of our bottom-up test is that it can be viewed as a generalization of the KPSS stationarity test for functional time series (Horváth et al., 2014; Kokoszka and Young, 2016). Our test for $\varphi_0 = 0$ and the functional KPSS test are asymptotically equivalent, and they can be exactly equal under some specific choice of finite-dimensional projection that is employed for dimensionality reduction of functional time series. This implies that our test for $\varphi_0 = 0$ generally has power against the alternative of various types of nonstationarity such as structural breaks and/or neglected deterministic trends as the functional KPSS test does. Moreover, the top-down test entailed with the proposed bottom-up test is closely related to the variance ratio tests of Nielsen et al. (2019). In fact, the top-down test may be viewed as a generalization of their tests in some aspects even if it does not fully generalize those due to differences in approach and employed assumptions. When it comes to discussion on cointegration corank tests in $\mathcal{H}$, the test of Chang et al. (2016) (hereafter, the CKP test) is not to be missed given that it appears to be the first testing procedure. This paper also shows that the CKP test and its extensions can be obtained from a variance ratio-type eigenvalue problem associated with a different pair of sample covariance operators under a suitable set of assumptions.

We provide empirical illustrations of our methodology. Particularly, we consider a monthly sequence of U.S. age-specific shares of full-time employment curves running from January 1980 to May 2019 and a monthly sequence of U.S. hourly real wage densities running from January 1990 to June 2019. We apply the proposed tests to estimate the cointegration corank and compare the results. In both applications, we found some evidence suggesting that our bottom-up procedure can complement the top-down procedures in determining the cointegration corank in practice.

The remainder of the present paper is organized as follows. We review some essential mathematics in Section 2. In Section 3, we provide our test statistics and explore their asymptotic properties, and then compare those with the existing tests. Section 4 reports Monte Carlo results on the finite-sample performances of the proposed tests. We then apply our methodology to two datasets in Section 5. All proofs and some additional simulation results are given in the appendices.

2 Mathematical preliminaries

2.1 Hilbert space and bounded linear operators

Let $\mathcal{H}$ denote a real separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$. Given a subset $U \subset \mathcal{H}$, we denote the orthogonal complement to U by

$$U^{\perp} = \{x \in \mathcal{H} : \langle x, y\rangle = 0 \text{ for all } y \in U\}$$

and the closure of U by $\operatorname{cl} U$. Given two subspaces $U_1, U_2 \subset \mathcal{H}$, $\mathcal{H}$ is said to be a direct sum of U_1 and U_2, denoted by $\mathcal{H} = U_1 \oplus U_2$, if $U_1 \cap U_2 = \{0\}$ and any element of $x \in \mathcal{H}$ can be written as $x = x_{U_1} + x_{U_2}$ for some $x_{U_1} \in U_1$ and $x_{U_2} \in U_2$.

We let $\mathcal{L}_{\mathcal{H}}$ denote the space of bounded linear operators acting on $\mathcal{H}$, equipped with the operator norm $\|A\|_{\mathcal{L}_{\mathcal{H}}} = \sup_{\|x\|\leq 1} \|Ax\|$. For an operator $A \in \mathcal{L}_{\mathcal{H}}$, we let $\operatorname{ran} A$ (resp. $\ker A$) denote the range (resp. kernel) of A. The dimension of $\operatorname{ran} A$ is called the rank of A and denoted by $\operatorname{rank} A$. Given a subspace $U \subset \mathcal{H}$, we let P^U denote the orthogonal projection onto U.

The adjoint of an operator $A \in \mathcal{L}_{\mathcal{H}}$ is uniquely given and denoted by A^*. If $A = A^*$, then A is said to be self-adjoint. A linear operator $A \in \mathcal{L}_{\mathcal{H}}$ is said to be positive semidefinite if $\langle Ax, x\rangle \geq 0$ for any $x \in \mathcal{H}$, and positive definite if also $\langle Ax, x\rangle \neq 0$ for any nonzero $x \in \mathcal{H}$. Throughout this paper, $x \otimes y(\cdot)$ for $x, y \in \mathcal{H}$ denotes rank one operator $\langle x, \cdot\rangle y$.

An operator $A \in \mathcal{L}_{\mathcal{H}}$ is said to be compact if there exists two orthonormal bases $\{u_j\}_{j\in\mathbb{N}}$ and $\{v_j\}_{j\in\mathbb{N}}$, and a real-valued sequence $\{\gamma_j\}_{j\in\mathbb{N}}$ tending to zero, such that

$$Ax = \sum_{j=1}^{\infty} \gamma_j u_j \otimes v_j(x), \tag{2.1}$$

If A is self-adjoint, then the expansion (2.1) can be simplified as follows:

$$Ax = \sum_{j=1}^{\infty} \gamma_j u_j \otimes u_j(x). \tag{2.2}$$

If A is also positive semidefinite, we may assume that $\gamma_1 \geq \gamma_2 \geq \ldots \geq 0$ in (2.2); see (Bosq, 2000, p. 35).

Sometimes we need to restrict the domain and the codomain of a bounded linear operator. Whenever it is required, we let $A|_{U_1\to U_2}$ denote the operator $A \in \mathcal{L}_{\mathcal{H}}$ whose domain (resp. codomain) is restricted to $U_1 \subset \mathcal{H}$ (resp. $U_2 \subset \mathcal{H}$).

2.2 Generalized inverse operator

If an operator $A \in \mathcal{L}_{\mathcal{H}}$ has closed range, then there exists a unique operator $A^\dagger \in \mathcal{L}_{\mathcal{H}}$ satisfying the so-called Moore-Penrose equations:

$$AA^\dagger A = A, \quad A^\dagger AA^\dagger = A^\dagger, \quad (AA^\dagger)^* = AA^\dagger, \quad (A^\dagger A)^* = A^\dagger A.$$

$A^\dagger$ is called the Moore-Penrose inverse of A, which also satisfies the following:

$$AA^\dagger = P^{\operatorname{ran} A}, \quad A^\dagger A = P^{[\ker A]^\perp}.$$

See Engl and Nashed (1981) for more on Moore-Penrose inverses. It is well known that compact operators (and finite rank operators as a special case) cannot be invertible in an infinite-dimensional setting. However, the Moore-Penrose inverse is well defined for any finite rank operator since its range is always closed. Moreover if $A \in \mathcal{L}_{\mathcal{H}}$ is a finite rank self-adjoint operator, then we may obtain the explicit form of the Moore-Penrose inverse based on the spectral representation of A: for $A = \sum_{j=1}^m \gamma_j u_j \otimes u_j$, $A^\dagger$ is given by

$$A^\dagger = \sum_{j=1}^m \gamma_j^{-1} u_j \otimes u_j$$

Moreover, for any arbitrary self-adjoint, positive semidefinite, and compact operator $A \in \mathcal{L}_{\mathcal{H}}$ whose spectral representation is given by $A = \sum_{j=1}^\infty \gamma_j u_j \otimes u_j$ for some $\gamma_1 \geq \gamma_2 \geq \ldots \geq 0$, we let $A_\ell^\dagger$ be given by

$$A_\ell^\dagger = \sum_{j=1}^\ell \gamma_j^{-1} u_j \otimes u_j$$

and be called an ℓ-regularized inverse of A. Clearly, $A_\ell^\dagger$ may be understood as the partial inverse of A on the restricted domain $U_\ell = \operatorname{span}\{u_1, \ldots, u_\ell\}$.

2.3 $\mathcal{H}$-valued random variables

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the underlying probability triple. Then, an $\mathcal{H}$-valued random variable Z is a measurable map from Ω to $\mathcal{H}$, where $\mathcal{H}$ is understood to be equipped with the Borel σ-field. An $\mathcal{H}$-valued random variable Z is said to be integrable (resp. square-integrable) if $\mathbb{E}\|Z\| < \infty$ (resp. $\mathbb{E}\|Z\|^2 < \infty$). If Z is integrable, there exists a unique element $\mathbb{E}Z \in \mathcal{H}$ such that $\mathbb{E}\langle Z, x\rangle = \langle \mathbb{E}Z, x\rangle$ for any $x \in \mathcal{H}$. The element $\mathbb{E}Z$ is called the expectation of Z.

We let $L^2_{\mathcal{H}}$ denote the space of $\mathcal{H}$-valued random variables Z (identifying random elements that are almost surely equal) that satisfy $\mathbb{E}Z = 0$ and $\mathbb{E}\|Z\|^2 < \infty$. For a random variable $Z \in L^2_{\mathcal{H}}$, we may define its covariance operator as $C_Z = \mathbb{E}[Z \otimes Z]$, which is guaranteed to be self-adjoint, positive semidefinite and compact.

2.4 I(0) sequences in $\mathcal{H}$

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be an independent and identically distributed (iid) sequence in $L^2_{\mathcal{H}}$, and let $\{A_j\}_{j\geq 0} \subset \mathcal{L}_{\mathcal{H}}$ be a sequence satisfying $\sum_{k=0}^{\infty} \|A_k\|^2_{\mathcal{L}_{\mathcal{H}}} < \infty$. Then it can be shown (Bosq, 2000, Lemma 7.1) that for each $t \in \mathbb{Z}$ the series

$$Z_t = \sum_{k=0}^{\infty} A_k \varepsilon_{t-k}$$

is convergent in $L^2_{\mathcal{H}}$. Such a sequence $\{Z_t\}_{t\in\mathbb{Z}}$ is called a linear process in $\mathcal{H}$. More generally, given any $t_0 \in \mathbb{Z} \cup \{-\infty\}$, we call the sequence $\{Z_t\}_{t\geq t_0}$ a linear process in $\mathcal{H}$. Linear processes with square summable coefficients are necessarily stationary.

If $\{A_j\}_{j\geq 0} \subset \mathcal{L}_{\mathcal{H}}$ satisfies $\sum_{j=0}^{\infty} \|A_j\|_{\mathcal{L}_{\mathcal{H}}} < \infty$, then $\{Z_t\}_{t\geq t_0}$ is called a standard linear process in $\mathcal{H}$. In this case the series $A(1) := \sum_{j=0}^{\infty} A_j$ is convergent in $\mathcal{L}_{\mathcal{H}}$, and it can be easily shown that the long-run covariance operator $\Lambda_Z \in \mathcal{L}_{\mathcal{H}}$ is given by

$$\Lambda_Z = \mathbb{E}(Z_t \otimes Z_t) + \sum_{s=1}^{\infty} \left(\mathbb{E}(Z_t \otimes Z_{t-s}) + \mathbb{E}(Z_{t-s} \otimes Z_t)\right) = A(1) C_{\varepsilon_0} A(1)^* \tag{2.3}$$

Note that Λ_Z is also self-adjoint, positive semidefinite, and compact. If $\{Z_t\}_{t\geq t_0}$ is standard linear and its long-run covariance is nonzero, then it is called an I(0) sequence (Beare et al., 2017, Section 3.4).

In this paper, more generally, $\{Z_t\}_{t\geq t_0}$ is said to be I(0) if $\{Z_t - \mu_Z\}_{t\geq t_0}$ is standard linear for some $\mu_Z \in \mathcal{H}$ and its long-run covariance, simply obtained by replacing Z_{t-s} with $Z_{t-s} - \mu_Z$ for all $s \geq 0$ in (2.3), is nonzero.

2.5 Cointegration in $\mathcal{H}$

We briefly introduce cointegrated $\mathcal{H}$-valued I(1) processes and their mathematical properties. Our primary reference is Beare et al. (2017). A sequence $\{X_t\}_{t\geq 0}$ is I(1) if its first differences $\Delta X_t := X_t - X_{t-1}$ satisfy

$$\Delta X_t = \sum_{j=0}^{\infty} \Phi_j \varepsilon_{t-j}, \quad t \geq 1, \tag{2.4}$$

$$\{\varepsilon_t\}_{t\in\mathbb{Z}} \text{ is an iid sequence in } L^2_{\mathcal{H}}, \text{ with positive definite } C_{\varepsilon_0}, \tag{2.5}$$

$$\{\Phi_j\}_{j\geq 0} \text{ satisfies } \sum_{j=0}^{\infty} j\|\Phi_j\|_{\mathcal{L}_{\mathcal{H}}} < \infty \text{ and } \Phi(1) := \sum_{j=0}^{\infty} \Phi_j \text{ is convergent.} \tag{2.6}$$

Then from (2.3), we may deduce that the long-run covariance operator of $\{\Delta X_t\}_{t\geq 1}$ is given by $\Lambda_{\Delta X} = \Phi(1)\Sigma\Phi(1)^*$.

For a sequence satisfying (2.4)-(2.6), it can be shown (Beare et al., 2017, equations (3.4) and (3.5)) that the so-called Beveridge-Nelson decomposition is obtained as follows.

$$\Delta X_t = \Phi(1)\varepsilon_t + \nu_t - \nu_{t-1}, \quad t \geq 1, \tag{2.7}$$

where $\nu_t = \sum_{j=0}^{\infty} \tilde{\Phi}_j \varepsilon_{t-j}$ and $\tilde{\Phi}_j = -\sum_{k=j+1}^{\infty} \Phi_k$. Solutions to the difference equation (2.7) are given by

$$X_t = \mu + \Phi(1)\sum_{s=1}^{t} \varepsilon_s + \nu_t, \quad t \geq 1, \tag{2.8}$$

for some time invariant component $\mu \in \mathcal{H}$.

The cointegrating space (stationary subspace) of $\{X_t\}_{t\geq 0}$ is the collection of elements $x \in \mathcal{H}$ such that the scalar-valued sequence $\{\langle X_t, x\rangle\}_{t\geq 0}$ is stationary for a suitable choice of the initial condition X_0. Moreover, the attractor space (nonstationary subspace) of $\{X_t\}_{t\geq 0}$ is defined by the orthogonal complement of the cointegrating space. For any I(1) sequence satisfying (2.4)-(2.6), it may be easily deduced from Proposition 3.3 of Beare et al. (2017) that the cointegrating space is given by $[\operatorname{ran}\Phi(1)]^{\perp}$ and the attractor space is given by $\operatorname{cl}\operatorname{ran}\Phi(1)$. Those are closed subspaces of $\mathcal{H}$, and we hereafter let $\mathfrak{A}$ (resp. $\mathfrak{C}$) denote the attractor space (resp. the cointegrating space). Clearly, we have the direct sum decomposition $\mathcal{H} = \mathfrak{A} \oplus \mathfrak{C}$.

3 Testing Procedures

3.1 Model and Hypotheses of Interest

We consider a cointegrated sequence satisfying the following conditions.

Assumption M.

(i) *The sequence* $\{X_t\}_{t\geq 1}$ *satisfies* (2.4)-(2.8), *and* $\mathbb{E}\|\epsilon_t\|^4 < \infty$.

(ii) $\dim(\mathfrak{A})$ *is given by some integer* φ *in* $[0,\infty)$.

(iii) $\{\nu_t\}_{t\geq 1}$ *is an I(0) sequence.*

(iv) *The coefficients* $\{\tilde{\Phi}_j\}_{j\geq 0}$ *of* ν_t *satisfies* $\sum_{j=1}^{\infty} j\|\tilde{\Phi}_j\|_{\mathcal{L}_{\mathcal{H}}} < \infty$.

Under Assumption M(i) and (ii), the attractor space $\mathfrak{A}$ is finite-dimensional which is similar to the setting considered in Chang et al. (2016) and Nielsen et al. (2019). A sufficient condition for $\mathfrak{A}$ to be a finite-dimensional subspace is given in Seo (2017), Beare and Seo

(2019), and Franchi and Paruolo (2019) when $\{X_t\}_{t\geq 1}$ is an autoregressive process. Since any finite-dimensional subspace is closed, we have the direct sum decomposition $\mathcal{H} = \operatorname{ran}\Phi(1) \oplus [\operatorname{ran}\Phi(1)]^{\perp}$, where $\operatorname{ran}\Phi(1) = \mathfrak{A}$ and $[\operatorname{ran}\Phi(1)]^{\perp} = \mathfrak{C}$. If $\dim(\mathfrak{A}) = \varphi = 0$, then our time series $\{X_t\}_{t\geq 1}$ becomes I(0) due to (iii). (iv) is employed for our asymptotic analysis for convenience, and it is not restrictive in practice.

Given the direct sum $\mathcal{H} = \mathfrak{A} \oplus \mathfrak{C}$, we may decompose $\{X_t\}_{t\geq 1}$ into a unit root process $\{P^{\mathfrak{A}} X_t\}_{t\geq 1}$ taking values in $\mathfrak{A}$ and an I(0) sequence $\{P^{\mathfrak{C}} X_t\}_{t\geq 1}$ taking values in $\mathfrak{C}$. Then the long-run covariance of the I(0) component $\{P^{\mathfrak{C}} \nu_t\}_{t\geq 1}$ is given by

$$
\begin{aligned}
P^{\mathfrak{C}} \Lambda_{\nu} P^{\mathfrak{C}} &= \mathbb{E}(\tilde{\nu}_t \otimes \tilde{\nu}_t) + \sum_{s=1}^{\infty} \left(\mathbb{E}(\tilde{\nu}_t \otimes \tilde{\nu}_{t-s}) + E(\tilde{\nu}_{t-s} \otimes \tilde{\nu}_t) \right) \\
\tilde{\nu}_t &= P^{\mathfrak{C}} X_t - \mathbb{E}\left(P^{\mathfrak{C}} X_t\right)
\end{aligned}
$$

Let $\{u_j\}_{j\in\mathbb{N}}$ be any arbitrary orthonormal basis of $\mathcal{H}$ satisfying $\operatorname{span}\{u_1, \ldots, u_{\varphi}\} = \mathfrak{A}$. We may understand an $\mathcal{H}$-valued random variable X_t as the following basis expansion

$$
X_t = \sum_{j=1}^{\infty} \langle X_t, u_j \rangle u_j, \quad \sum_{j=1}^{\infty} \langle X_t, u_j \rangle^2 < \infty, \text{ almost surely.}
$$

Due to a well known isomorphism between a Hilbert space and the space of square-summable sequences, $\ell^2(\mathbb{N})$, $\{X_t\}_{t\geq 1}$ may be viewed as the following random square-summable sequences

$$
\Big(\langle X_t, u_j \rangle, \quad \cdots \quad \langle X_t, u_{\varphi} \rangle, \quad \langle X_t, u_{\varphi+1} \rangle, \quad \cdots \Big), \quad t \geq 1
$$

Clearly the first φ components are scalar-valued I(1) processes since

$$
\Delta \langle X_t, u_j \rangle = \langle \Phi(1)\varepsilon_t, u_j \rangle + \langle \nu_t - \nu_{t-1}, u_j \rangle, \quad j = 1, \ldots, \varphi
$$

is stationary, and its long-run covariance, given by $\langle \Lambda_{\Delta X} u_j, u_j \rangle$, is nonzero under our assumptions. On the other hand, the remaining components are stationary since

$$
\langle X_t, u_j \rangle = \langle \mu, u_j \rangle + \langle \nu_t, u_j \rangle, \quad j = \varphi + 1, \ldots \tag{3.1}
$$

and $\{\langle \nu_t, u_j \rangle\}_{t\geq 1}$ is a measurable transformation of a stationary sequence $\{\nu_t\}_{t\geq 1}$. It should be noted that $\{\langle X_t, u \rangle\}_{t\geq 1}$ for $u \in \mathfrak{C}$ may not be I(0) since its long-run covariance, given by $\langle P^{\mathfrak{C}} \Lambda_{\nu} P^{\mathfrak{C}} u, u \rangle$, can be zero if $\ker P^{\mathfrak{C}} \Lambda_{\nu} P^{\mathfrak{C}} \neq \{0\}$. That is, (3.1) may not be I(0) in $\mathbb{R}$ even if $\{\nu_t\}_{t\geq 1}$ is I(0) in $\mathcal{H}$.

From the above discussion, it is deduced that the dimension of $\mathfrak{A}$, denoted by φ, is interpreted as the number of common stochastic trends, generalizing the same notion from the literature on cointegration rank tests in Euclidean space. Note that it may not be proper to call any statistical testing procedure to estimate φ as a cointegration rank test since the cointegration rank is not defined in our model if $\dim(\mathcal{H}) = \infty$ as usually assumed. On the other hand, it is quite proper to call it a cointegration corank test since what we are interested in is the corank of $P^{\mathfrak{C}}$ that is necessarily equal to φ.

In the following subsections, we provide statistical testing procedures to estimate the cointegration corank. As in Chang et al. (2016) and Nielsen et al. (2019), we may first consider the following top-down selection of hypotheses:

$$H_0 : \varphi = \varphi_0 \quad \text{against} \quad H_1 : \varphi < \varphi_0, \tag{3.2}$$

for $\varphi_0 = \varphi_{\max}, \dots, 1$ successively for some pre-specified value of $\varphi_{\max}$. If $\mathcal{H}$ is finite-dimensional, then $\varphi_{\max} = \dim(\mathcal{H})$ is an obvious starting point of the above procedure. However, no such natural choice exists if $\mathcal{H}$ is infinite-dimensional and we do not have prior information on an upper bound of φ. There are a few practical ways to choose $\varphi_{\max}$, suggested by Chang et al. (2016), but they tend to depend on the researcher's subjectivity. On the other hand, consider the following selection of the hypotheses:

$$H_0 : \varphi = \varphi_0 \quad \text{against} \quad H_1 : \varphi > \varphi_0, \tag{3.3}$$

for $\varphi_0 = 0, 1, 2, \dots$ successively. That is, we initially test stationarity of time series, and then sequentially test whether there are more stochastic trends. The estimate is determined by the first non-rejected null hypothesis. Especially in our model that typically considers an infinite-dimensional $\mathcal{H}$, this bottom-up approach has theoretical advantages over the aforementioned top-down procedures; it is stand-alone without the help of any prior information on an upper bound of φ.

3.2 Test Statistics and Asymptotic Analysis

3.2.1 Notations

In this section, we provide our bottom-up and top-down test statistics and establish their asymptotic properties. We first fix notation for the subsequent discussion. Throughout the present paper, let $\int$ mean $\int_0^1$ for notational simplicity. We let $\{W(r)\}_{r\in[0,1]}$ denote the standard Hilbertian Brownian motion (see e.g. Chen and White (1998) for more detailed discussion), and define the following: for $r \in [0, 1]$,

$$B(r) = W(r) - rW(1), \quad \overline{W}(r) = W(r) - \int W(\tau)d\tau, \quad V(r) = \int_0^r \overline{W}(\tau)d\tau. \tag{3.4}$$

Moreover, we let $\{W_\eta(r)\}_{r\in[0,1]}$ denote the standard Brownian motion taking values in $\mathbb{R}^\eta$ for $\eta \in \mathbb{N}$, and define $B_\eta(r)$, $\overline{W}_\eta(r)$, and $V_\eta(r)$ as in (3.4). For continuous Hilbert-valued functions $A(r)$ and $B(r)$ defined for $r \in [0,1]$, we let $\mathcal{S}_{A,B}$ denote the operator

$$\mathcal{S}_{A,B} = \int B(r) \otimes A(r) dr.$$

For a possibly random bounded linear operator A, we let $\{\lambda_j(A)\}_{j\in\mathbb{N}}$ denote the eigenvalues in descending order of A if they exist.

Given functional observations $\{X_t\}_{t=1}^T$, let $Y_t = \sum_{s=1}^t (X_s - \bar{X}_T)$ with $\bar{X}_T = T^{-1}\sum_{t=1}^T X_t$ for $t = 1, \ldots, T$ and $\mathcal{C}_T(s)$ be the sample autocovariance operator at lag s, defined by

$$\mathcal{C}_T(s) = T^{-1} \sum_{t=s+1}^{T} (X_t - \bar{X}_T) \otimes (X_{t-s} - \bar{X}_T) \tag{3.5}$$

We also define

$$\Lambda_T = \mathcal{C}_T(0) + \sum_{s=1}^{T-1} \mathrm{k}\left(\frac{s}{h}\right) (\mathcal{C}_T(s) + \mathcal{C}_T(s)^*) \tag{3.6}$$

$$\mathcal{K}_T = T^{-2} \sum_{t=1}^{T} Y_t \otimes Y_t \tag{3.7}$$

where $\mathrm{k}(\cdot)$ (resp. h) is the kernel (resp. the bandwidth) satisfying the requirements given in Assumption K below:

Assumption K.

(i) $\mathrm{k}(0) = 1$, $\mathrm{k}(u) = 0$ *if* $u > c$ *with some* $c > 0$, k *is continuous on* $[0, c]$.

(ii) $h \to \infty$ *and* $h/T \to 0$ *as* $T \to \infty$.

Many kernel functions used in practice satisfy Assumption K(i). For $\mathrm{k}(\cdot)$ and h satisfying Assumption K, we define

$$m = ch, \qquad c_{\mathrm{k}} = 2\int_0^c \mathrm{k}(\tau) d\tau.$$

3.2.2 Test Statistics

We consider the test of (3.2) or (3.3) for various values of φ_0, and let $\ell(\varphi_0)$ denote a positive integer that is defined for each φ_0. For now, we do not restrict $\ell(\varphi_0)$, but for reasons to become apparent we will require $\ell(\varphi_0) > \varphi_0$ for the proposed bottom-up test and $\ell(\varphi_0) \geq \varphi_0$ or $\ell(\varphi_0) = \varphi_0$ for the top-down tests considered in this paper. Our tests are based on the following variance ratio-type eigenvalue problem:

$$\Pi_T^\ell(\varphi_0)\mathcal{K}_T\Pi_T^\ell(\varphi_0)\phi_{j,T} = \lambda_{j,T}\Pi_T^\ell(\varphi_0)\Lambda_T\Pi_T^\ell(\varphi_0)\phi_{j,T}, \quad \phi_{j,T} \in \operatorname{ran}\Pi_T^\ell(\varphi_0), \tag{3.8}$$
$$\lambda_{1,T} \geq \lambda_{2,T} \geq \cdots \geq \lambda_{\ell(\varphi_0),T},$$

where $\{\Pi_T^\ell(\varphi_0)\}_{T\in\mathbb{N}}$ is a sequence of orthogonal projections onto an $\ell(\varphi_0)$-dimensional subspace $\mathcal{H}_T^\ell(\varphi_0)(= \operatorname{ran}\Pi_T^\ell(\varphi_0))$. Given T, (3.8) may be viewed as a generalized eigenvalue problem defined on a finite-dimensional subspace $\mathcal{H}_T^\ell(\varphi_0)$, so we may easily obtain the eigenvalues and the corresponding eigenvectors using standard methods. We first employ the following high-level conditions.

Assumption P. For $\ell(\varphi_0) > 0$, $\{\Pi_T^\ell(\varphi_0)\}_{T\in\mathbb{N}}$ is a sequence satisfies the following:

(a) For some orthogonal projection $\Pi^\ell(\varphi_0)$ onto $\ell(\varphi_0)$-dimensional subspace $\mathcal{H}^\ell(\varphi_0)$,

$$\left\|\Pi_T^\ell(\varphi_0) - \Pi^\ell(\varphi_0)\right\|_{\mathcal{L}_\mathcal{H}} \to_p 0 \quad \text{as} \quad T \to \infty.$$

(b) $\mathcal{H}^\ell(\varphi_0)$ satisfies

$$\begin{aligned}\dim(\mathcal{H}^\ell(\varphi_0) \cap \mathfrak{A}) &= \min\{\varphi, \ell(\varphi_0)\},\\ \operatorname{rank}(\Pi^\ell(\varphi_0)P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}\Pi^\ell(\varphi_0)) &= \ell(\varphi_0) - \min\{\varphi, \ell(\varphi_0)\}.\end{aligned}$$

We will discuss on data-dependent choices of $\Pi_T^\ell(\varphi_0)$ satisfying Assumption P in Section 3.4. When $\{\Pi_T^\ell(\varphi_0)\}_{T\in\mathbb{N}}$ satisfies Assumption P, we let $\Pi^{\mathfrak{A}}(\varphi_0)$ (resp. $\Pi^{\mathfrak{C}}(\varphi_0)$) denote the orthogonal projection onto $\mathcal{H}^\ell(\varphi_0)\cap\mathfrak{A}$ (resp. $\mathcal{H}^\ell(\varphi_0)\cap\mathfrak{C}$), whose rank is $\min\{\varphi, \ell(\varphi_0)\}$ (resp. $\ell(\varphi_0) - \min\{\varphi, \ell(\varphi_0)\}$). The proposed testing procedures are based on the following limiting behavior of the eigenvalues.

Proposition 3.1. *Let Assumptions M, K and P hold. Then the eigenvalues* $(\lambda_{1,T}, \ldots, \lambda_{\ell(\varphi_0),T})$ *in (3.8) satisfy the following.*

(i) *For* $j = 1, \ldots, \min\{\varphi, \ell(\varphi_0)\}$

$$(T/m)\lambda_{j,T}^{-1} \quad \to_d \quad c_{\mathrm{k}} \cdot \lambda_j\left(\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^\dagger\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0)\right), \quad \text{jointly.}$$

(ii) *For* $j = \min\{\varphi, \ell(\varphi_0)\} + 1, \ldots, \ell(\varphi_0)$,

$$\lambda_{j,T} \quad \to_d \quad \lambda_{j-\min\{\varphi,\ell(\varphi_0)\}}\left(\widetilde{\mathcal{S}}_{B,B}(\varphi_0) - \widetilde{\mathcal{S}}_{B,V}(\varphi_0)\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^\dagger\widetilde{\mathcal{S}}_{V,B}(\varphi_0)\right), \quad \text{jointly.}$$

In the above expressions,

$$\widetilde{\mathcal{S}}_{B,B}(\varphi_0) = \Pi^{\mathfrak{C}}(\varphi_0)\mathcal{S}_{B,B}\Pi^{\mathfrak{C}}(\varphi_0), \;\; \widetilde{\mathcal{S}}_{V,V}(\varphi_0) = \Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{V,V}\Pi^{\mathfrak{A}}(\varphi_0), \;\; \widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0) = \Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{\overline{W},\overline{W}}\Pi^{\mathfrak{A}}(\varphi_0)$$
$$\widetilde{\mathcal{S}}_{B,V}(\varphi_0) = \Pi^{\mathfrak{C}}(\varphi_0)\mathcal{S}_{B,V}\Pi^{\mathfrak{A}}(\varphi_0), \;\; \widetilde{\mathcal{S}}_{V,B}(\varphi_0) = \Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{V,B}\Pi^{\mathfrak{C}}(\varphi_0),$$

The limit in Proposition 3.1(ii) is given by an eigenvalue of a specific random self-adjoint operator whose rank is almost surely given by

$$\kappa(\varphi_0) = \ell(\varphi_0) - \varphi_0,$$

which is hereafter called the slackness of $\Pi_T^{\ell}(\varphi_0)$. If $\ell(\varphi_0) \le \varphi$, then Proposition 3.1 implies that every eigenvalue satisfying (3.8) diverges to infinity as $T \to \infty$. On the other hand, we have convergent eigenvalues if $\ell(\varphi_0) > \varphi$. Then it is deduced that the statistic

$$\mathcal{B}(\varphi_0, \kappa(\varphi_0)) = \sum_{j=\varphi_0+1}^{\ell(\varphi_0)} \lambda_{j,T}, \quad \ell(\varphi_0) > \varphi_0, \quad \varphi_0 = 0, 1, 2, \ldots \tag{3.9}$$

has a well defined limiting distribution if $\varphi_0 \ge \varphi$, and it diverges to infinity if $\varphi_0 < \varphi$. It should be noted that the statistic and its limit depend on the choice of $\Pi_T^{\ell}(\varphi_0)$ for each φ_0. Particularly, $\kappa(\varphi_0)$ affects the limiting distribution, which can be deduced from the fact that the test statistic is given by the sum of the smallest $\kappa(\varphi_0)$ eigenvalues that are nondegenerate in the limit. This dependence is a new aspect arising from dimensionality reduction in an infinite-dimensional Hilbert space and generally absent in a finite-dimensional setting. The argument $\kappa(\varphi_0)$ of $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ is introduced to highlight such dependence.

A consistent testing procedure is proposed as follows.

Corollary 3.1. *Suppose that Assumptions M, K and P hold. Then $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$, which is sequentially defined for $\varphi_0 = 0, 1, 2, \ldots$ in (3.9), satisfies the following.*

(i) *If $\varphi_0 = \varphi$, $\mathcal{B}(\varphi_0, \kappa(\varphi_0)) \to_d \operatorname{tr}(\widetilde{\mathcal{S}}_{B,B}(\varphi_0) - \widetilde{\mathcal{S}}_{B,V}(\varphi_0)\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{V,B}(\varphi_0))$.*

(ii) *If $\varphi_0 < \varphi$, $\mathcal{B}(\varphi_0, \kappa(\varphi_0)) \to_p \infty$.*

Let $\hat{\varphi}_{\mathcal{B}}(\alpha)$ denote the value φ_0 such that $\mathcal{B}(\varphi_0, \kappa(\varphi_0)) \le C_{\mathcal{B}}(\alpha; \varphi_0, \kappa(\varphi_0))$ for the first time, where $C_{\mathcal{B}}(\alpha; \varphi_0, \kappa(\varphi_0))$ is the $(1-\alpha)$-quantile of the limiting distribution in (i). Then

(iii) $\mathbb{P}\{\hat{\varphi}_{\mathcal{B}}(\alpha) = \varphi\} \to 1 - \alpha$ *and* $\mathbb{P}\{\hat{\varphi}_{\mathcal{B}}(\alpha) < \varphi\} \to 0$.

In the above testing procedure, there is no restriction other than $\ell(\varphi_0) > \varphi_0$, so the slackness $\kappa(\varphi_0)$ can be any positive integer for each φ_0. Depending on the researcher's choice of $\kappa(\varphi_0)$, test statistic $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ and its limiting distribution are different. We thus need critical values $C_{\mathcal{B}}(\alpha; \varphi_0, \kappa(\varphi_0))$ depending on both φ_0 and $\kappa(\varphi_0)$ in addition to α.

This can be easily done by a large number of simulations for any specific choice of $\kappa(\varphi_0)$; see Remark 3.4. Instead of reporting all possible choices of $\kappa(\varphi_0)$ for each φ_0, which is of course impossible, we may conveniently assume $\ell(\varphi_0) = \varphi_0 + \varrho$ for some constant ϱ so that $\kappa(\varphi_0) = \varrho$ regardless of φ_0. In addition, ϱ does not need to be a big number since our asymptotic theory only requires $\varrho > 0$, so we may conveniently choose small ϱ in practice. Table 1(a)-(c) report critical values for $\alpha = 0.1, 0.05$ and 0.01 when $\varrho = 1, 2$ and 3.

Remark 3.1. It is not difficult to show that the test based on $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ reduces to the test of Nyblom and Harvey (2000) if $\mathcal{H} = \mathbb{R}^n$ and $\Pi_T^\ell(\varphi_0)$ is the identity map on $\mathbb{R}^n$. It further reduces to the test of Kwiatkowski et al. (1992) if $\varphi_0 = 0$ and $n = 1$ additionally hold.

Remark 3.2. (Choice of h for $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ in practice) In finite samples, our choice of h (or equivalently, m) for $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ needs to depend on persistence of the stationary component $\{\nu_t\}_{t\geq 1}$ to have a better finite-sample performance. If we choose too small h when $\{\nu_t\}_{t\geq 1}$ exhibits strong persistence, then $P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}}$ may be a poor estimate of $P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$, which makes our test over-sized in general. To avoid this problem we may choose a bigger bandwidth, but it brings about a loss of power: when H_1 is true, we may deduce from Proposition 3.1 that $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ is divergent and $\mathrm{O}_p(T/m)$, so the rate of divergence is smaller when m/T is bigger. A Monte Carlo evidence of this trade-off between correct size and power is provided in Section 4. A similar discussion in a Euclidean space setting can be found in Section 6 of Kwiatkowski et al. (1992) or Section 2.3 of Nyblom and Harvey (2000).

Note that Proposition 3.1 also describes the limiting behavior of the inverse eigenvalues, from which we can also derive a top-down testing procedure that may be viewed as the reverse of our bottom-up procedure. Consider a sequence of statistics

$$\mathcal{T}(\varphi_0, \kappa(\varphi_0)) = \frac{T}{m}\sum_{j=1}^{\varphi_0} \lambda_{j,T}^{-1}, \quad \ell(\varphi_0) \geq \varphi_0, \quad \varphi_0 = \varphi_{\max}, \varphi_{\max} - 1, \ldots, 1, \tag{3.10}$$

where $\varphi_{\max}$ is a pre-specified integer satisfying $\varphi_{\max} \geq \varphi$. From Proposition 3.1, we may deduce that $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ has a well defined limiting distribution if $\varphi_0 \leq \varphi$, but diverges to infinity if $\varphi_0 > \varphi$. This limiting behavior is the opposite to that of $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$: except for the case when both are convergent, i.e. $\varphi_0 = \varphi$, $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$ (resp. $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$) diverges when $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ (resp. $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$) converges. It should be noted that the slackness $\kappa(\varphi_0)$ does not have any role in the limit as long as $\ell(\varphi_0) \geq \varphi_0$ because the limiting distribution in Proposition 3.1(i) is given by an eigenvalue of a random self-adjoint operator whose rank is φ_0 regardless of $\kappa(\varphi_0)$. Moreover, it is worth mentioning that $\ell(\varphi_0) = \varphi_0$ is allowed, which is not as in (3.9). A consistent testing procedure based on a sequence of $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ is given as follows.

Table 1: Critical values when $\kappa(\varphi_0) = \varrho$

(a) Mean-adjusted, $\varrho = 1$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.3456	0.1625	0.0933	0.0626	0.0467	0.0367	0.0302	0.0254	0.0219
5%	0.4609	0.2201	0.1201	0.0777	0.0567	0.0439	0.0354	0.0295	0.0252
1%	0.7444	0.4014	0.2027	0.1213	0.0830	0.0624	0.0491	0.0394	0.0332

(b) Mean-adjusted, $\varrho = 2$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.6068	0.2970	0.1687	0.1140	0.0845	0.0667	0.0548	0.0465	0.0403
5%	0.7471	0.3797	0.2068	0.1352	0.0982	0.0762	0.0618	0.0519	0.0446
1%	1.0782	0.6261	0.3182	0.1931	0.1332	0.0998	0.0789	0.0647	0.0549

(c) Mean-adjusted, $\varrho = 3$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.8397	0.4246	0.2405	0.1620	0.1208	0.0954	0.0786	0.0667	0.0578
5%	0.9974	0.5328	0.2888	0.1881	0.1372	0.1069	0.0870	0.0733	0.0630
1%	1.3533	0.8232	0.4327	0.2586	0.1795	0.1351	0.1072	0.0883	0.0752

(d) Trend-adjusted, $\varrho = 1$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.1190	0.0848	0.0610	0.0460	0.0365	0.0301	0.0254	0.0220	0.0193
5%	0.1479	0.1055	0.0751	0.0557	0.0435	0.0353	0.0295	0.0252	0.0220
1%	0.2166	0.1590	0.1133	0.0821	0.0618	0.0485	0.0395	0.0332	0.0285

(e) Trend-adjusted, $\varrho = 2$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.2110	0.1514	0.1104	0.0837	0.0664	0.0547	0.0464	0.0402	0.0354
5%	0.2469	0.1789	0.1295	0.0969	0.0755	0.0616	0.0518	0.0445	0.0390
1%	0.3295	0.2450	0.1781	0.1295	0.0990	0.0785	0.0647	0.0546	0.0472

(f) Trend-adjusted, $\varrho = 3$

	$\varphi_0 = 0$	1	2	3	4	5	6	7	8
10%	0.2954	0.2141	0.1570	0.1193	0.0948	0.0784	0.0666	0.0578	0.0509
5%	0.3363	0.2464	0.1801	0.1350	0.1062	0.0866	0.0729	0.0630	0.0552
1%	0.4267	0.3239	0.2380	0.1755	0.1342	0.1069	0.0883	0.0749	0.0648

Notes: Based on 200,000 Monte Carlo replications.

Corollary 3.2. *Suppose that Assumptions M, K and P hold, and $\varphi_{\max} \geq \varphi$ is satisfied. Then $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$, which is sequentially defined for $\varphi_0 = \varphi_{\max}, \ldots, 1$ in (3.10), satisfies the following.*

(i) *If $\varphi_0 = \varphi$, $c_{\mathrm{k}}^{-1}\mathcal{T}(\varphi_0, \kappa(\varphi_0)) \to_d \cdot \operatorname{tr}(\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0))$.*

(ii) *If* $\varphi_0 > \varphi$, $c_{\mathrm{k}}^{-1}\mathcal{T}(\varphi_0, \kappa(\varphi_0)) \to_p \infty$.

Let $\hat{\varphi}_{\mathcal{T}}(\alpha) = \varphi^*(\alpha) - 1$ *and* $\varphi^*(\alpha)$ *denote the smallest value of* φ_0 *such that* $c_{\mathrm{k}}^{-1}\mathcal{T}(\varphi_0, \kappa(\varphi_0)) > C_{\mathcal{T}}(\alpha; \varphi_0)$, *where* $C_{\mathcal{T}}(\alpha; \varphi_0)$ *is the* $(1-\alpha)$*-quantile of the limiting distribution in* (i). *Then*

(iii) $\mathbb{P}\{\hat{\varphi}_{\mathcal{T}}(\alpha) = \varphi\} \to 1 - \alpha$ *and* $\mathbb{P}\{\hat{\varphi}_{\mathcal{T}}(\alpha) > \varphi\} \to 0$.

This top-down procedure has a convenient property compared to the bottom-up procedure: for a given significance level α, critical values $C_{\mathcal{T}}(\alpha; \varphi_0)$ can be simply provided as a function of φ_0, which is because that the limiting distribution does not depend on our choice of $\kappa(\varphi_0)$ but does only on φ_0. However, $\kappa(\varphi_0)$ set to zero or a big number is not recommended in practice; see Remark 3.13.

Remark 3.3. It is worth noting that Corollary 3.2 requires $\varphi_{\max} \geq \varphi$. Proposition 3.1 says that the top-down test statistic $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ has a well defined limiting distribution for $\varphi_0 \leq \varphi_{\max} < \varphi$, which implies that for each of φ_0 the incorrect null hypothesis $H_0 : \varphi = \varphi_0$ tends not to be rejected with probability $(1 - \alpha)$ when T is large enough. This clearly leads to an incorrect determination of the cointegration corank because the misspecified top hypothesis is not likely to be rejected and we tend to conclude that $\hat{\varphi}_{\mathcal{T}} = \varphi_{\max}$ as a consequence. This is a common drawback shared by all the top-down tests considered in the present paper.

Remark 3.4. Note that $\operatorname{ran}\Pi^{\mathfrak{C}}(\varphi_0)$ (resp. $\operatorname{ran}\Pi^{\mathfrak{A}}(\varphi_0)$) is isomorphic to $\kappa(\varphi_0)$-dimensional (resp. φ_0-dimensional) Euclidean space. Then one can show that the asymptotic null distribution $\operatorname{tr}(\widetilde{\mathcal{S}}_{B,B}(\varphi_0) - \widetilde{\mathcal{S}}_{B,V}(\varphi_0)\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{V,B}(\varphi_0))$ is equal in distribution to

$$\operatorname{tr}\left(\int B_{\kappa(\varphi_0)}(r)B'_{\kappa(\varphi_0)}(r)dr - \int B_{\kappa(\varphi_0)}(r)V'_{\varphi_0}(r)\left(\int V_{\varphi_0}(r)V'_{\varphi_0}(r)dr\right)^{-1}\int V_{\varphi_0}(r)B'_{\kappa(\varphi_0)}(r)dr\right), \tag{3.11}$$

and $\operatorname{tr}(\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0))$ is equal in distribution to

$$\operatorname{tr}\left(\left(\int V_{\varphi_0}(r)V'_{\varphi_0}(r)dr\right)^{-1}\int \overline{W}_{\varphi_0}(r)\overline{W}'_{\varphi_0}(r)dr\right), \tag{3.12}$$

It should be noted that (3.11) depends on φ_0 and $\kappa(\varphi_0)$, but (3.12) depends only on φ_0. From a large number of approximate realizations of (3.11) for each pair of φ_0 and $\kappa(\varphi_0)$, we may obtain critical values for $\mathcal{B}(\varphi_0, \kappa(\varphi_0))$. Moreover, critical values for $c_{\mathrm{k}}^{-1}\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ can be similarly obtained from a large number of approximate realizations of (3.12) for each φ_0, which are available from Table 6 of Breitung (2002).

Remark 3.5. (Choice of bandwidth h for $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ in practice) When H_1 is true, $c_{\mathrm{k}}^{-1}\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ diverges to infinity and its stochastic order is $\mathrm{O}_p(T/m)$. Therefore, a bigger bandwidth again entails a loss of power. With a slight modification of our proof of Proposition 3.1, we may obtain an asymptotically equivalent test statistic $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ by replacing Λ_T with $\mathcal{C}_T(0)$ (that is, Λ_T with $h = 0$) in (3.8), and it diverges to infinity with order $\mathrm{O}_p(T)$.[2] Our simulation results show that the test based on $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ does have a reasonable size control overall. Therefore, we expect that choosing moderate h may be preferred to avoid a substantial loss of power.

3.2.3 Inclusion of a linear trend

We assumed that $\{X_t\}_{t\geq 1}$ has a nonzero time invariant component μ since this may be the most common case in practice. However, our discussion can be easily extended to the model with a linear trend. Consider the following unobserved component model.

$$Z_t = \zeta t + X_t, \quad \zeta \in \mathcal{H}, \tag{3.13}$$

In this case, we define the functional residuals from least square estimation as follows.

$$U_t = X_t - \bar{X}_T - \left(t - \frac{T+1}{2}\right) \frac{\sum_{t=1}^T \left(t - \frac{T+1}{2}\right) X_t}{\sum_{t=1}^T \left(t - \frac{T+1}{2}\right)^2}; \tag{3.14}$$

see Kokoszka and Young (2016) for details. We then define trend-adjusted sample covariance operators $\widetilde{\mathcal{C}}_T(s)$, $\widetilde{\Lambda}_T$ and $\widetilde{\mathcal{K}}_T$ by replacing $(X_t - \bar{X}_T)$ with U_t in (3.5)-(3.7). The generalized eigenvalue problem (3.8) is also changed to

$$\Pi_T^\ell(\varphi_0)\widetilde{\mathcal{K}}_T\Pi_T^\ell(\varphi_0)\phi_{j,T} = \lambda_{j,T}\Pi_T^\ell(\varphi_0)\widetilde{\Lambda}_T\Pi_T^\ell(\varphi_0)\phi_{j,T}, \quad \phi_{j,T} \in \operatorname{ran}\Pi_T^\ell(\varphi_0) \tag{3.15}$$
$$\lambda_{1,T} \geq \lambda_{2,T} \geq \cdots \geq \lambda_{\ell(\varphi_0),T},$$

We then obtain the following asymptotic properties of the eigenvalues.

Proposition 3.2. *Let Assumptions M, K and P hold. Then the eigenvalues* $(\lambda_{1,T}, \dots, \lambda_{\ell(\varphi_0),T})$ *in (3.15) satisfy the following.*

(i) *For* $j = 1, \dots, \min\{\varphi, \ell(\varphi_0)\}$

$$(T/m)\lambda_{j,T}^{-1} \quad \to_d \quad c_{\mathrm{k}}\ \lambda_j\left(\widetilde{\mathcal{S}}_{V_2,V_2}(\varphi_0)^\dagger\widetilde{\mathcal{S}}_{W_2,W_2}(\varphi_0)\right), \quad \text{jointly.}$$

(ii) *For* $j = \min\{\varphi, \ell(\varphi_0)\} + 1, \dots, \ell(\varphi_0)$,

$$\lambda_{j,T} \quad \to_d \quad \lambda_{j-\min\{\varphi,\ell(\varphi_0)\}}\left(\widetilde{\mathcal{S}}_{B_2,B_2}(\varphi_0) - \widetilde{\mathcal{S}}_{B_2,V_2}(\varphi_0)\widetilde{\mathcal{S}}_{V_2,V_2}(\varphi_0)^\dagger\widetilde{\mathcal{S}}_{V_2,B_2}(\varphi_0)\right), \quad \text{jointly.}$$

[2] For reasons to become apparent, a detailed discussion on the test statistic is postponed to Section 3.3.2.

In the above expressions, $(W_2(r), r \in [0,1])$ *(resp.* $(B_2(r), r \in [0,1])$*) denotes the detrended Brownian motion (resp. the second-level Brownian bridge) in* $\mathcal{H}$,[3] $V_2(r) = \int_0^r W_2(\tau)d\tau$, *and*

$$\widetilde{\mathcal{S}}_{B_2,B_2}(\varphi_0)=\Pi^{\mathfrak{C}}(\varphi_0)\mathcal{S}_{B_2,B_2}\Pi^{\mathfrak{C}}(\varphi_0),\ \widetilde{\mathcal{S}}_{V_2,V_2}(\varphi_0)=\Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{V_2,V_2}\Pi^{\mathfrak{A}}(\varphi_0),\ \widetilde{\mathcal{S}}_{W_2,W_2}(\varphi_0)=\Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{W_2,W_2}\Pi^{\mathfrak{A}}(\varphi_0)$$
$$\widetilde{\mathcal{S}}_{B_2,V_2}(\varphi_0)=\Pi^{\mathfrak{C}}(\varphi_0)\mathcal{S}_{B_2,V_2}\Pi^{\mathfrak{A}}(\varphi_0),\ \widetilde{\mathcal{S}}_{V_2,B_2}(\varphi_0)=\Pi^{\mathfrak{A}}(\varphi_0)\mathcal{S}_{V_2,B_2}\Pi^{\mathfrak{C}}(\varphi_0),$$

From the limiting behavior of the eigenvalues given in Proposition 3.2, we may easily obtain the corresponding bottom-up and top-down testing procedures as in Corollary 3.1 and 3.2. Critical values for each statistic can be simulated as in Remark 3.4; see our Table 1(c)-(e) for the bottom-up test with $\kappa(\varphi_0) = \varphi_0 + \varrho$ for $\varrho = 1, 2, 3$, and Table 6 of Breitung (2002) for the top-down test, respectively.

3.3 Comparisons with existing tests

3.3.1 Functional KPSS tests

Let us focus on the following null and alternative hypotheses:

$$H_0 : \dim(\mathfrak{A}) = 0 \quad \text{against} \quad H_1 : \dim(\mathfrak{A}) \geq 1. \tag{3.16}$$

Under the null hypothesis, $\{X_t\}_{t\geq 1}$ is simply a linear process with mean μ. Needless to say, our bottom-up test is then understood to be a test of level stationarity of functional time series belonging to a certain class of stochastic processes. In the context of functional time series, there are a few papers that consider tests of the null hypothesis of stationarity such as Horváth et al. (2014), Kokoszka and Young (2016) and Aue and Van Delft (2017). Particularly, the first (resp. the second) paper provides a suitable generalization of the KPSS test of level stationarity (resp. trend stationarity). In each specification of deterministic terms, it can be easily verified that the limiting distribution of our bottom-up test statistic for some choice of $\ell(0)$ (or equivalently $\kappa(0)$) is equal to that of the generalized KPSS test statistic. Moreover, as shown in Remark 3.6 below, they are even numerically identical under some specific choice of $\Pi_T^\ell(0)$. Since our test can be applied to examine more general hypotheses (3.3) than (3.16), it may be viewed as a generalization of their tests. Moreover, it is a natural consequence drawn from the works of Horváth et al. (2014) and Kokoszka and Young (2016) that our test of the null hypothesis in (3.16) would have a good power against the alternative of various types of nonstationarity such as structural breaks and/or unrecognized deterministic trends.

[3] Specifically, $W_2(r) = W(r) + (6r-4)\int W(\tau)d\tau + (6-12r)\int \tau W(\tau)d\tau$ and $B_2(r) = W(r) + (2r - 3r^2)W(1) + (-6r+6r^2)\int W(\tau)d\tau$.

Remark 3.6. We only consider the case when $\{X_t\}_{t=1}^T$ is level stationary. Based on the results in Section 3.2.3, the subsequent arguments may be easily adapted to the case when a linear trend exists. Let $(\tau_{j,T}, u_{j,T})$ is a pair of eigenvalue and eigenvector of Λ_T and $\tau_{1,T} \geq \tau_{2,T} \geq \ldots$. Horváth et al. (2014)'s test statistic can be written as

$$\sum_{j=1}^{\ell(0)} \tau_{j,T}^{-1} \langle \mathcal{K}_T u_{j,T}, u_{j,T} \rangle, \quad \ell(0) > 0,$$

If we let $\Pi_T^\ell(0) = \sum_{j=1}^{\ell(0)} u_{j,T} \otimes u_{j,T}$, then it satisfies all the requirements of Assumption P, which is supported by Proposition 3.5 that will appears later. For this choice of $\Pi_T^\ell(0)$, we can show that $\mathcal{B}(0, \kappa(0))$ is numerically identical to their statistic. To see this, note that $\langle \mathcal{K}_T u_{j,T}, u_{j,T} \rangle = \langle \Pi_T^\ell(0) \mathcal{K}_T \Pi_T^\ell(0) u_{j,T}, u_{j,T} \rangle$ and $\tau_{j,T} u_{j,T} = \Pi_T^\ell(0) \Lambda_T \Pi_T^\ell(0) u_{j,T}$. Then,

$$\begin{aligned}
\sum_{j=1}^{\ell(0)} \tau_{j,T}^{-1} \langle \mathcal{K}_T u_{j,T}, u_{j,T} \rangle &= \sum_{j=1}^{\ell(0)} \langle \Pi_T^\ell(0) \mathcal{K}_T \Pi_T^\ell(0) \left(\Pi_T^\ell(0) \Lambda_T \Pi_T^\ell(0) \right)_{\ell(0)}^\dagger u_{j,T}, u_{j,T} \rangle \\
&= \operatorname{tr} \left(\Pi_T^\ell(0) \mathcal{K}_T \Pi_T^\ell(0) \left(\Pi_T^\ell(0) \Lambda_T \Pi_T^\ell(0) \right)_{\ell(0)}^\dagger \right) \\
&= \operatorname{tr} \left(\left(\Pi_T^\ell(0) \Lambda_T \Pi_T^\ell(0) \right)_{\ell(0)}^\dagger \Pi_T^\ell(0) \mathcal{K}_T \Pi_T^\ell(0) \right) = \sum_{j=1}^{\ell(0)} \lambda_{j,T},
\end{aligned}$$

which is clearly equivalent to our statistic. Choice of $\ell(0)$ is not carefully discussed in the present paper. One reasonable approach in practice may be to select $\ell(0)$ so that roughly $a\%$ of the sample covariance is explained by the first $\ell(0)$ principal components for fixed a; see Section 4 of Horváth et al. (2014) for more details.

3.3.2 Variance ratio tests of Nielsen et al. (2019).

Consider the following variance ratio-type eigenvalue problem:

$$\begin{aligned}
&\Pi_T^\ell(\varphi_0) \mathcal{K}_T \Pi_T^\ell(\varphi_0) \phi_{j,T} = \lambda_{j,T} \Pi_T^\ell(\varphi_0) \mathcal{C}_T(0) \Pi_T^\ell(\varphi_0) \phi_{j,T}, \quad \phi_{j,T} \in \operatorname{ran} \Pi_T^\ell(\varphi_0) \\
&\lambda_{1,T} \geq \lambda_{2,T} \geq \cdots \geq \lambda_{\ell(\varphi_0),T}
\end{aligned} \tag{3.17}$$

Clearly, (3.17) is a special case of (3.8), obtained by employing $h = 0$ for Λ_T. With only a slight modification of the proof of Proposition 3.1, the following asymptotic properties of the eigenvalues may be deduced.

Proposition 3.3. *Let Assumptions M, K and P hold. Then the eigenvalues* $(\lambda_{1,T}, \ldots, \lambda_{\ell(\varphi_0),T})$ *in (3.8) satisfy the following.*

(i) *For* $j = 1, \dots, \min\{\varphi, \ell(\varphi_0)\}$

$$T\,\lambda_{j,T}^{-1} \quad \to_d \quad \lambda_j\left(\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0)\right), \quad \text{jointly.}$$

(ii) *For* $j = \min\{\varphi, \ell(\varphi_0)\} + 1, \dots, \ell(\varphi_0)$,

$$\lambda_{j,T} \quad \to_d \quad \lambda_{j-\min\{\varphi,\ell(\varphi_0)\}}\left(\widetilde{\mathcal{S}}\right), \quad \text{jointly.}$$

for some finite rank self-adjoint operator $\widetilde{\mathcal{S}}$. *If* $\mathbb{E}[\nu_{t-s} \otimes \nu_t] = \mathbb{E}[\nu_t \otimes \nu_{t-s}] = 0$ *for* $s \neq 0$, *then* $\widetilde{\mathcal{S}} = \widetilde{\mathcal{S}}_{B,B}(\varphi_0) - \widetilde{\mathcal{S}}_{B,V}(\varphi_0)\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{V,B}(\varphi_0)$ *which is given in Proposition 3.1*(ii).

Proposition 3.3 suggests testing procedures to determine the cointegration corank based on the following test statistics.

$$\mathcal{B}_0(\varphi_0, \kappa(\varphi_0)) = \sum_{j=\varphi_0+1}^{\ell(\varphi_0)} \lambda_{j,T}, \quad \ell(\varphi_0) > \varphi_0, \quad \varphi_0 = 0, 1, 2, \dots \tag{3.18}$$

$$\mathcal{T}_0(\varphi_0, \kappa(\varphi_0)) = T\sum_{j=1}^{\varphi_0} \lambda_{j,T}^{-1}, \quad \ell(\varphi_0) \geq \varphi_0, \quad \varphi_0 = \varphi_{\max}, \varphi_{\max} - 1, \dots, 1. \tag{3.19}$$

Corollary 3.3. *Suppose that Assumptions M*(i)-(iii) *and P hold. If* $\mathbb{E}[\nu_{t-s} \otimes \nu_t] = \mathbb{E}[\nu_t \otimes \nu_{t-s}] = 0$ *for* $s \neq 0$, *then* $\mathcal{B}_0(\varphi_0, \kappa(\varphi_0))$, *which is sequentially defined for* $\varphi_0 = 0, 1, 2, \dots$ *in* (3.18), *satisfies the following.*

(i) *If* $\varphi_0 = \varphi$, $\mathcal{B}_0(\varphi_0, \kappa(\varphi_0)) \to_d \mathrm{tr}(\widetilde{\mathcal{S}}_{B,B}(\varphi_0) - \widetilde{\mathcal{S}}_{B,V}(\varphi_0)\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{V,B}(\varphi_0))$.

(ii) *If* $\varphi_0 < \varphi$, $\mathcal{B}_0(\varphi_0, \kappa(\varphi_0)) \to_p \infty$.

Let $\tilde{\varphi}_{\mathcal{B}}(\alpha)$ *denote the value* φ_0 *such that* $\mathcal{B}_0(\varphi_0, \kappa(\varphi_0)) \leq C_{\mathcal{B}}(\alpha; \varphi_0, \kappa(\varphi_0))$ *for the first time, where* $C_{\mathcal{B}}(\alpha; \varphi_0, \kappa(\varphi_0))$ *is the same to that in Corollary 3.1. Then*

(iii) $\mathbb{P}\{\tilde{\varphi}_{\mathcal{B}}(\alpha) = \varphi\} \to 1 - \alpha$ *and* $\mathbb{P}\{\tilde{\varphi}_{\mathcal{B}}(\alpha) < \varphi\} \to 0$.

Corollary 3.4. *Suppose that Assumptions M*(i)-(iii) *and P hold, and* $\varphi_{\max} \geq \varphi$ *is satisfied. Then* $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$, *which is sequentially defined for* $\varphi_0 = \varphi_{\max}, \dots, 1$ *in* (3.19), *satisfies the following.*

(i) *If* $\varphi_0 = \varphi$, $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0)) \to_d \mathrm{tr}(\widetilde{\mathcal{S}}_{V,V}(\varphi_0)^{\dagger}\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0))$.

(ii) *If* $\varphi_0 > \varphi$, $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0)) \to_p \infty$.

Let $\tilde{\varphi}_{\mathcal{T}}(\alpha) = \varphi^*(\alpha) - 1$ *and* $\varphi^*(\alpha)$ *denote the smallest value of* φ_0 *such that* $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0)) > C_{\mathcal{T}}(\alpha; \varphi_0)$, *where* $C_{\mathcal{T}}(\alpha; \varphi_0)$ *is the same to that in Corollary 3.2. Then*

(iii) $\mathbb{P}\{\tilde{\varphi}_{\mathcal{T}}(\alpha) = \varphi\} \to 1 - \alpha$ *and* $\mathbb{P}\{\tilde{\varphi}_{\mathcal{T}}(\alpha) > \varphi\} \to 0$.

The top-down test obtained in Corollary 3.4 is asymptotically equivalent to the test of Nielsen et al. (2019). Depending on how to choose $\Pi_T^\ell(\varphi_0)$ in the variance ratio-type eigenvalue problem (3.17), $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ could be numerically identical to any of the two test statistics suggested by Nielsen et al. (2019) or be something else that has different finite-sample properties; see Remarks 3.9-3.11 in Section 3.4 for details. The suggested bottom-up procedure given in Corollary 3.3 requires $\{\nu_t\}_{t\geq 1}$ to be serially uncorrelated, which may restrict its use to very limited circumstances in practice. It may be deduced from Proposition 3.3 that $\mathcal{B}_0(\varphi_0, \kappa(\varphi_0))$ converges to a well defined limiting distribution even if $\{\nu_t\}_{t\geq 1}$ has serial correlation. However, the limit depends on the covariance operator and the long-run covariance operator of $\{\nu_t\}_{t\geq 1}$, which are unknown; see our proof of Corollary 3.3 for more details.

Remark 3.7. Nielsen et al. (2019) provides trend-adjusted test statistics and their asymptotic null distributions for the case when the time series of interest includes a linear trend. Corollary 3.4 can be extended to the case with only a slight modification as in Section 3.2.3. Then obviously, the top-down test obtained as a result of such an extension becomes asymptotically equivalent to those provided by Nielsen et al. (2019).

Remark 3.8. Assumption M(iv) is not required in Corollaries 3.4 and 3.3 as long as $\Pi_T^\ell(\varphi_0)$ satisfies Assumption P. However, it will be required for our suggested data-dependent construction of $\Pi_T^\ell(\varphi_0)$ satisfying Assumption P.

3.3.3 Test of Chang et al. (2016)

In the functional time series literature, Chang et al. (2016) appears to be the first that provides a cointegration corank test (the CKP test). With a slight modification in our framework, we can derive a test that is similar to the CKP test, and its extension as well. We consider a different variance ratio-type eigenvalue problem motivated by Shintani (2001) and Chang et al. (2016) as follows.

$$\Pi_T^\ell(\varphi_0)\mathcal{C}_T(0)\Pi_T^\ell(\varphi_0)\phi_{j,T} = \lambda_{j,T}\Pi_T^\ell(\varphi_0)\Lambda_{\Delta X,T}\Pi_T^\ell(\varphi_0)\phi_{j,T}, \quad \phi_{j,T} \in \operatorname{ran}\Pi_T^\ell(\varphi_0) \tag{3.20}$$
$$\lambda_{1,T} \geq \lambda_{2,T} \geq \cdots \geq \lambda_{\ell(\varphi_0),T}$$

where,

$$\mathcal{C}_{\Delta X,T}(s) = T^{-1}\sum_{t=s+1}^{T} \Delta X_t \otimes \Delta X_{t-s}$$
$$\Lambda_{\Delta X,T} = \mathcal{C}_{\Delta X,T}(0) + \sum_{s=1}^{T-1} \mathrm{k}\left(\frac{s}{h}\right)\left(\mathcal{C}_{\Delta X,T}(s) + \mathcal{C}_{\Delta X,T}(s)^*\right)$$

To conveniently utilize the existing asymptotic results developed in Phillips (1995) and Shintani (2001), we will replace Assumption K with the following:

Assumption K1.

(i) $\mathrm{k}(\cdot)$ *is twice continuously differentiable with:* (a) $\mathrm{k}(0) = 1$, $\mathrm{k}'(\cdot) = 0$, $\mathrm{k}''(\cdot) \neq 0$; *and either* (b) $\mathrm{k}(x) = 0$, $x \geq 1$ *with* $\lim_{x \to 1} \mathrm{k}(x)/(1-x)^2 = constant$, *or* (c) $\mathrm{k}(x) = O(x^{-2})$, *as* $x \to 1$.

(ii) $h = a_1 T^{a_2}$ *for some* $a_1 > 0$ *and* $a_2 \in (0, 0.5)$.

We then obtain the following asymptotic properties of the eigenvalues.

Proposition 3.4. *Suppose that Assumptions M, K1 and P hold. Then the eigenvalues* $(\lambda_{1,T}, \ldots, \lambda_{\ell(\varphi_0),T})$ *in (3.20) satisfy the following.*

(i) *For* $j = 1, \ldots, \min\{\varphi, \ell(\varphi_0)\}$

$$T^{-1}\lambda_{j,T} \quad \to_d \quad \lambda_j \left(\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0)\right), \quad \text{jointly.}$$

(ii) *For* $j = \min\{\varphi, \ell(\varphi_0)\} + 1, \ldots, \ell(\varphi_0)$,

$$T^{-1}\lambda_{j,T} \quad \to_p \quad 0, \quad \text{jointly.}$$

Based on Proposition 3.4, we define two test statistics

$$\mathcal{T}_{\mathrm{S}}(\varphi_0, \kappa(\varphi_0)) = T \sum_{j=1}^{\varphi_0} \lambda_{j,T}^{-1}, \quad \ell(\varphi_0) \geq \varphi_0, \tag{3.21}$$

$$\mathcal{T}_{\mathrm{CKP}}(\varphi_0) = T^{-1} \min_{1 \leq j \leq \varphi_0} \lambda_{j,T}, \quad \ell(\varphi_0) = \varphi_0. \tag{3.22}$$

Then the top-down procedure based on each of the above test statistics is given as follows.

Corollary 3.5. *Let Assumptions M, K1 and P hold. Then* $\mathcal{T}_{\mathrm{S}}(\varphi_0, \kappa(\varphi_0))$ *and* $\mathcal{T}_{\mathrm{CKP}}(\varphi_0)$, *which are sequentially defined for* $\varphi_0 = \varphi_{\max}, \ldots, 1$ *in (3.21) and (3.22), satisfy the following.*

(i) *If* $\varphi_0 = \varphi$, $\mathcal{T}_{\mathrm{S}}(\varphi_0, \kappa(\varphi_0)) \to_d \mathrm{tr}(\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0)^{\dagger})$ *and* $\mathcal{T}_{\mathrm{CKP}}(\varphi_0) \to_d \min_{1 \leq j \leq \varphi_0} \lambda_j \left(\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}(\varphi_0)\right)$.

(ii) *If* $\varphi_0 > \varphi$, $\mathcal{T}_{\mathrm{S}}(\varphi_0, \kappa(\varphi_0)) \to_p \infty$ *and* $\mathcal{T}_{\mathrm{CKP}}(\varphi_0) \to 0$.

Let $\check{\varphi}_{\mathcal{T}}(\alpha) = \varphi^*(\alpha) - 1$ *and denote* $\varphi^*(\alpha)$ *the smallest value of* φ_0 *such that* $\mathcal{T}_{\mathrm{S}}(\varphi_0, \kappa(\varphi_0))$ *or* $\mathcal{T}_{\mathrm{CKP}}(\varphi_0)$ *is on the rejection region at significance level* α. *Then*

(iii) $\mathbb{P}\{\check{\varphi}_{\mathcal{T}}(\alpha) = \varphi\} \to 1 - \alpha$ *and* $\mathbb{P}\{\check{\varphi}_{\mathcal{T}}(\alpha) > \varphi\} \to 0$.

The test based on $\mathcal{T}_{\text{CKP}}(\varphi_0)$ is asymptotically equivalent to the CKP test. They can become exactly identical by choosing some specific choice of $\Pi_T^{\ell}(\varphi_0)$; see Remark 3.12 in Section 3.4. In addition, the test based on $\mathcal{T}_{\text{S}}(\varphi_0, \kappa(\varphi_0))$ may be viewed as a generalization of the cointegration rank test proposed by Shintani (2001).

It should be noted that the test based on $\mathcal{T}_{\text{CKP}}(\varphi_0)$ does not allow a positive slackness, which is the reason why the argument $\kappa(\varphi_0)$ is dropped. If $\kappa(\varphi_0) > 0$, the test statistic converges to zero even when the null hypothesis is true. On the other hand, the test based on $\mathcal{T}_{\text{S}}(\varphi_0, \kappa(\varphi_0))$ allows $\kappa(\varphi_0) > 0$, which may be a significant advantage over $\mathcal{T}_{\text{CKP}}(\varphi_0)$; see Remark 3.13 in Section 3.4.

3.4 Practical choices of $\Pi_T^{\ell}(\varphi_0)$

All the tests considered in the previous sections are derived under the high-level conditions given in Assumption P. In this section, we provide practical choices of $\Pi_T^{\ell}(\varphi_0)$ that satisfies Assumption P under some additional low-level conditions. We first show that $\Pi_T^{\ell}(\varphi_0)$ can be chosen to be the projection onto the span of the eigenvectors of Λ_T corresponding to the first $\ell(\varphi_0)$ leading eigenvalues, which may be the most reasonable choice for our test statistics. We need the following condition.

Assumption P1. *For each φ_0 and $\ell(\varphi_0)$, $P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$ has at least $\ell(\varphi_0)$ positive eigenvalues.*

Assumption P1 is very mild in our functional setting since φ_0 and $\ell(\varphi_0)$ are finite and small in general, while the number of positive eigenvalues of $P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$ is expected to be infinite or large finite. The following proposition proposes a practical data-dependent choice of $\Pi_T^{\ell}(\varphi_0)$.

Proposition 3.5. *Let Assumptions M, K and P1 hold. $(u_{1,T}, \ldots, u_{\ell(\varphi_0),T})$ be the eigenvectors corresponding to the first $\ell(\varphi_0)$ leading eigenvalues of Λ_T. Define*

$$\Pi_T^{\ell}(\varphi_0) = \sum_{j=1}^{\ell(\varphi_0)} u_{j,T} \otimes u_{j,T}. \tag{3.23}$$

Then $\Pi_T^{\ell}(\varphi_0)$ satisfies all the requirements of Assumption P.

In our proof of Proposition 3.5, it is shown that $(u_{1,T}, \ldots, u_{\varphi,T})$ converges to an orthonormal basis of $\mathfrak{A}$. Then we show that the remaining eigenvectors $(u_{\varphi+1,T}, \ldots, u_{\ell(\varphi_0),T})$ converges to the eigenvectors of $P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$ corresponding to the first $\ell(\varphi_0)$ eigenvalues, where Assumption M(iv) has an essential role: it guarantees some weak dependence condition on $\{P^{\mathfrak{C}}\nu_t\}_{t\geq 1}$ that is sufficient to have a stronger convergence result $\|P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}} - P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} \to_p 0$.[4] From

[4] Convergence in norm between two compact operators imply convergence of their eigenelements (Lemma 4.2-4.4 of Bosq (2000))

the convergence of the eigenvectors, we may deduce that $\Pi_T^\ell(\varphi_0)$ constructed as in (3.23) satisfies Assumption P.

We next show that the eigenvectors of $\mathcal{C}_T(0)$ can be used to construct $\Pi_T^\ell(\varphi_0)$ as in (3.23), which is not a new result but a direct consequence from Theorem 4.2 of Chang et al. (2016) and the following convenient assumption.

Assumption P2. *For each φ_0 and $\ell(\varphi_0)$, the eigenvectors $(u_1, \ldots, u_{\ell(\varphi_0)})$ of $P^{\mathfrak{C}} C_{\nu_0} P^{\mathfrak{C}}$ corresponding the first $\ell(\varphi_0)$ leading eigenvalues satisfy that $u_j \notin \ker \Lambda_\nu$ for all $j = 1, \ldots, \ell(\varphi_0)$.*

Proposition 3.6. *Let Assumptions M(i)-(iii) and P2 hold, and let $(u_{1,T}, \ldots, u_{\ell(\varphi_0),T})$ be the eigenvectors corresponding to the first $\ell(\varphi_0)$ leading eigenvalues of $\mathcal{C}_T(0)$. Define*

$$\Pi_T^\ell(\varphi_0) = \sum_{j=1}^{\ell(\varphi_0)} u_{j,T} \otimes u_{j,T}.$$

Then $\Pi_T^\ell(\varphi_0)$ satisfies all the requirements of Assumption P.

In our proof of Proposition 3.6, it is shown that $(u_{1,T}, \ldots, u_{\varphi_0,T})$ converges to an orthonormal basis of $\mathfrak{A}$, and $(u_{\varphi_0+1,T}, \ldots, u_{\ell(\varphi_0),T})$ converges to the eigenvectors of $P^{\mathfrak{C}} C_{\nu_0} P^{\mathfrak{C}}$. The latter does not imply Assumption P(ii) in general since there could exist a nonzero eigenvector x of $P^{\mathfrak{C}} C_{\nu_0} P^{\mathfrak{C}}$ satisfying $x \in \ker \Lambda_\nu$. The role of Assumption P2 is to exclude this possibility.

Before concluding this section, we present several important remarks on the choice of $\Pi_T^\ell(\varphi_0)$.

Remark 3.9. Clearly, Assumption P1 is a weaker requirement than Assumption P2, so the choice given in Proposition 3.5 may be preferred in general. However, it should be noted that Proposition 3.6 does not require Assumption M(iv), which implies that the test based on $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ given in Corollary 3.4 can be implemented in practice regardless of whether the condition is assumed or not. In fact, the choice of $\Pi_T^\ell(\varphi_0)$ given in Proposition 3.6 makes $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ exactly equivalent to one of the proposed top-down tests by Nielsen et al. (2019). It is also worth noting that Nielsen et al. (2019) does not employ Assumption M(iv), but instead assume that Λ_ν is positive definite on $\mathfrak{C}$; clearly, this is a stronger condition than Assumption P2. We thus conclude that Corollary 3.4 with the choice of $\Pi_T^\ell(\varphi_0)$ given in Proposition 3.6 corresponds to a different derivation of a test of Nielsen et al. (2019).

Remark 3.10. In Nielsen et al. (2019), they require $\Pi_T^\ell(\varphi_0)$ to converge in operator norm to $\Pi^\ell(\varphi_0)$ whose range includes $\mathfrak{A}$. This is a weaker requirement than Assumption P employed for all the tests considered in the present paper. An important consequence of the difference is that they can show that $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ with $\Pi_T^\ell(\varphi_0)$ constructed from the eigenvectors of $\mathcal{K}_T$ as in (3.23) satisfies the asymptotic properties given in Corollary 3.4. The proof of this result requires a different asymptotic analysis from ours; see Section 3 and Appendix of Nielsen et al. (2019).

Remark 3.11. An attractive feature of the tests of Nielsen et al. (2019), i.e. those based on $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ with $\Pi_T^\ell(\varphi_0)$ constructed from $\mathcal{C}_T(0)$ and $\mathcal{K}_T$, is that estimation of any long-run covariance operator is not required; see Remarks 3.9 and 3.10. On the other hand, all the other tests considered in this paper require estimation of the long-run covariance operator of $\{\nu_t\}_{t\geq 1}$ or $\{\Delta X_t\}_{t\geq 1}$. Since accurate estimation of the long-run covariance is known to be difficult in practice, the tests of Nielsen et al. (2019) would be still appealing to the applied researchers.

Remark 3.12. The choice $\Pi_T^\ell(\varphi_0)$ given in Proposition 3.6 makes $\mathcal{T}_{\text{CKP}}(\varphi_0)$ become numerically identical to the CKP test statistic provided that both employ the same long-run covariance operator $\Lambda_{\Delta X,T}$.

Remark 3.13. As discussed in Remark 10 of Nielsen et al. (2019), an essential requirement for the existing cointegration corank tests to work well in finite samples is that $\operatorname{ran}\Pi_T^\ell(\varphi_0)$ approximately includes $\mathfrak{A}$. It seems clear that estimation of the attractor space with φ-dimensional projection is much more difficult than estimation of any superspace including the attractor space with $(\varphi + \varrho)$-dimensional projection for some positive integer ϱ. This is the reason why $\kappa(\varphi_0) > 0$ is recommended in finite samples even if the asymptotic null distribution does not depend on $\kappa(\varphi_0)$. In addition, it is not recommended to set $\kappa(\varphi_0)$ too big either, which is due to inaccuracy of eigenvalues of a high-dimensional covariance matrix, see Section 3.5 of Nielsen et al. (2019).

Remark 3.14. (Inclusion of a linear trend) Suppose that our time series satisfies (3.13). Then we may simply replace Λ_T (resp. $\mathcal{C}_T(0)$) with $\widetilde{\Lambda}_T$ (resp. $\widetilde{\mathcal{C}}_T(0)$) in Proposition 3.5 (resp. 3.6) and use the associated eigenvectors to construct $\Pi_T^\ell(\varphi_0)$. Our proofs of the propositions include short discussions on the case that the time seres of interest includes a linear trend.

Remark 3.15. (Estimation of $\mathfrak{A}$) Oftentimes, we are interested in estimation of $\mathfrak{A}$ instead of its dimension. Proposition 3.5 (resp. 3.6) shows that the first φ eigenvectors of Λ_T (resp. $\mathcal{C}_T(0)$) converge in probability to an orthonormal basis of $\mathfrak{A}$. Therefore, estimation of $\mathfrak{A}$ reduces to that of the cointegration corank. Nielsen et al. (2019) also shows that the first φ eigenvectors of $\mathcal{K}_T$ converge to an orthonormal basis of $\mathfrak{A}$, see Theorem 3 and Remark 11 of their paper.

4 Monte Carlo simulations

We established asymptotic properties of our test statistics in the previous sections. From a practical point of view, it is also of interest to check how those are revealed with sample sizes that are typical for economic or statistical time series. For this purpose, we report the result of a Monte Carlo experiment that is similarly designed to that of Nielsen et al. (2019).

We consider the functional AR(1) model with unit root. Let $\{\zeta_j\}_{j=1}^{20}$ denote the functions defined as follows: for $x \in [0,1]$,

$$\zeta_j(x) = \begin{cases} \sqrt{2}\sin(2\pi j x), & \text{for } j \text{ is odd}, \\ \sqrt{2}\cos(2\pi j x), & \text{for } j \text{ is even}. \end{cases}$$

Adding $\zeta_{21}(x) = \mathbb{1}\{x \in [0,1]\}$, the collection $\{\zeta_j\}_{j=1}^{21}$ becomes the first 21 Fourier basis functions of $\mathcal{H} = L^2[0,1]$. We let $\mathcal{I}$ denote the collection of integers given by $\mathcal{I} = \{1, \ldots, 21\}$, and define a subset $\mathcal{I}_1$ of $\mathcal{I}$ as follows.

$$\mathcal{I}_1 = \{1, 2, \ldots, 10\}.$$

We hereafter let $\{[1], \ldots, [\varphi]\}$ denote randomly selected $\varphi(\leq 5)$ elements from $\mathcal{I}_1$ and let $([\varphi+1], \ldots, [21])$ denote randomly ordered elements of $\mathcal{I} \setminus \{[1], \ldots, [\varphi]\}$. We generate the functional time series as

$$X_t = \mu + \sum_{j=1}^{21} \theta_j \langle \zeta_{[j]}, X_{t-1} - \mu \rangle \zeta_{[j]} + \varepsilon_t \tag{4.1}$$

where $\{\varepsilon_t\}_{t=1}^T$ is a sequence of standard Brownian Bridges independent across t, and

$$\theta_j = \begin{cases} 1 & \text{for } j \leq \varphi, \\ \theta^{(j-\varphi)} & \text{for } j > \varphi, \end{cases} \tag{4.2}$$

for $\theta = 0.1$ and 0.7. The mean function μ in (4.1) is set to a linear combination of $\{\zeta_j\}_{j=1}^{21}$: specifically $\mu(x) = \sum_{j=1}^{21} g_j \zeta_j(x)$ for $x \in [0,1]$, where $\{g_j\}_{j=1}^{21}$ are independent standard normal random variables. Given (4.1) and (4.2), it may be deduced that the cointegration corank is φ and the attractor space $\mathfrak{A}$ is given by the span of $\{\zeta_{[1]}, \ldots, \zeta_{[\varphi]}\}$. Note that, μ and $\mathfrak{A}$ are different across different realizations of the DGP (4.1), which may be helpful to mitigate the effect caused by specifying those. It should also be noted that this setting allows the attractor space to include frequently oscillating functions on $[0,1]$, e.g. $\zeta_{10}(x) = \sqrt{2}\cos(10\pi x), x \in [0,1]$. As observed by Nielsen et al. (2019), this tends to make $\Pi_T^\ell(\varphi_0)$ obtained from the eigenvectors of Λ_T or $\mathcal{C}_T(0)$ a less accurate estimate of $\Pi^\ell(\varphi_0)$, which in turn makes finite-sample properties of a cointegration corank test poor. The functional observations are constructed by smoothing $\{X_t\}_{t=1}^T$ observed at 200 regularly spaced points of $[0,1]$ using 30 cubic B-spline basis functions. Throughout our simulation study, the total number of replications for each statistic is 4,000.

In Appendix C, we consider a few different modified DGPs and report the simulation results. First, we let the attractor space be fixed to the span of $\{\zeta_1, \ldots, \zeta_\varphi\}$. In this case, the attractor includes more smooth (less oscillating) functions, so $\Pi^\ell_T(\varphi_0)$ tends to be more accurate estimate of $\Pi^\ell(\varphi_0)$. In addition, we modify the DGP (4.1) by setting $\{\zeta_j\}_{j=1}^{21}$ to the first 21 Legendre polynomials as in Nielsen et al. (2019) and examine finite-sample properties of the tests. We then also fix the attractor space to the span of the first φ Legendre polynomials to see how this change affects finite-sample properties.

4.1 Finite-sample performance of the bottom-up test

We investigate the finite-sample performance of the proposed bottom-up test (the $\mathcal{B}$ test) for $T = 150, 350$ and 750. The baseline bandwidth, denoted h_b, is set to $0.6\,T^{4/9}$. The sample long-run covariance operator Λ_T is computed with the Parzen kernel $\mathrm{k}(\cdot)$ and the selected bandwidth.[5] To see the effect of bandwidth choice on finite-sample properties, we also consider $h = c\,h_b$ for $c = 1.5$ and $1/1.5$, respectively. We assume $\ell(\varphi_0) = \varphi_0 + 2$ as suggested in Nielsen et al. (2019) for their top-down tests, so the slackness $\kappa(\varphi_0)$ is 2 regardless of φ_0. Based on the selected slackness and Proposition 3.5, the projection $\Pi^\ell_T(\varphi_0)$ for each φ_0 is constructed from the eigenvectors corresponding to the first $\varphi_0 + 2$ eigenvalues of Λ_T. To calculate finite-sample power of the test, we generate (4.1) with $\varphi = \varphi_0 + q$ for $q = 1, 2$.

Table 2 summarizes our simulation results for different values of θ and T. The $\mathcal{B}$ test with the baseline bandwidth has excellent size control for all considered values of φ_0; a slight over-rejection is reported only when $\theta = 0.7$, $\varphi = 0$ and $T = 150$. Finite-sample power tends to be decreasing in φ_0, and increasing in any of T and q, which are easily expected from the asymptotic properties of the test.[6] Especially when $T = 150$ and $q = 1$ or 2, the test loses power more dramatically as φ_0 gets larger, which suggests that the cointegration corank tends to be underestimated when it is bigger and T is smaller. However, the lack of power rapidly disappears as T gets larger, so the test appears to have reasonable finite-sample size and power when $T \geq 350$ in our simulation.

The effect of bandwidth choice on finite-sample size and power of the test coincides with our insight from asymptotic theory (Remark 3.2). When $\theta = 0.7$, our bottom-up test with $h = h_b/1.5$ exhibits a slight over-rejection, but has higher power; see Table 2(d). If we considered θ higher than 0.7, over-rejection would be more severe. When $\theta = 0.1$, on the other hand, a smaller bandwidth choice seems to increase finite-sample power without causing over-rejection, which is also as expected in Remark 3.2; see also Table 2(c). Provided that

[5] Then $\mathrm{k}(\cdot)$ and $h = h_b$ satisfy both of Assumptions K and K1.

[6] Even if we only report the case when $q = 1$ and 2 in Table 2, finite-sample power was investigated for all positive integer $q \leq 4$ and found they tend to monotonically increase in q.

our test with the baseline bandwidth is not severely over-sized, it is expected that choosing $h = h_b \times 1.5$ would make our test too conservative. This can be checked in Table 2(e) and (f); the reported finite-sample size is below the nominal level when $\varphi_0 \geq 1$ and the finite-sample power for each φ_0 is lower compared to those obtained with the baseline bandwidth.

Our simulation results, including those reported in Appendix C for a few different DGPs, suggest that the proposed $\mathcal{B}$ test overall works well.

4.2 Finite-sample performances of the top-down tests

We next check the finite-sample performances of our top-down tests based on $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$ and $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ (the $\mathcal{T}$ test and the $\mathcal{T}_0$ test) that are given by the reverses of the bottom-up tests in Corollaries 3.1 and 3.3, respectively. Based on Remark 3.5, we employ a smaller bandwidth $h = 0.6\,T^{2/9}$ with the Parzen kernel[7] to compute Λ_T, which seems to be enough for the proposed top-down tests not to be severely over-sized. We also assume $\ell(\varphi_0) = \varphi_0 + 2$, and the projection $\Pi_T^{\ell}(\varphi_0)$ is constructed from the first $\varphi_0 + 2$ leading eigenvectors of Λ_T based on Proposition 3.5. For comparison purposes, the tests based on two another versions of $\mathcal{T}_0(\varphi_0, \kappa(\varphi_0))$ associated with $\Pi_T^{\ell}(\varphi_0)$ constructed from the eigenvectors of $\mathcal{K}_T$ and $\mathcal{C}_T(0)$ respectively, are also considered; they are the variance ratio tests suggested by Nielsen et al. (2019), and denoted by $\text{NSS}_{\mathcal{K}}$ and $\text{NSS}_{\mathcal{C}}$, respectively. $\text{NSS}_{\mathcal{K}}$ is preferred to $\text{NSS}_{\mathcal{C}}$ by the authors due to its better finite-sample performance. To calculate finite-sample power of each test, we generate (4.1) with $\varphi = \varphi_0 - 1$.

We also provide the results for the test based on $\mathcal{T}_{\text{S}}(\varphi_0, \kappa(\varphi_0))$ (the $\mathcal{T}_{\text{S}}$ test) which is obtained from a different variance ratio-type eigenvalue problem. The projection $\Pi_T^{\ell}(\varphi_0)$ is constructed in the same way to that of $\mathcal{T}(\varphi_0, \kappa(\varphi_0))$, and the long-run covariance $\Pi_T^{\ell}(\varphi_0)\Lambda_{\Delta X,T}\Pi_T^{\ell}(\varphi_0)$ is calculated with the Parzen kernel and the bandwidth choice proposed by Andrews (1991). In this section, we do not report the results of the CKP test and the test based on $\mathcal{T}_{\text{CKP}}(\varphi_0)$ (the $\mathcal{T}_{\text{CKP}}$ test) since they tend to be severely over-sized for the considered DGP even when $T = 750$. This may be due to inaccuracy of $\Pi_T^{\ell}(\varphi_0)$; see Remark 3.13. However, we need to take into consideration that the CKP test is developed for a density-valued cointegrated time series and the attractor space $\mathfrak{A}$ is in general expected to be given by the span of smooth functions, which is contrary to that our DGP allows $\mathfrak{A}$ to include frequently oscillating functions. If we fix the attractor space to the span of $\{\zeta_1, \ldots, \zeta_\varphi\}$ and $T = 750$, then the CKP test is not severely over-sized unless $\varphi \geq 4$ and the $\mathcal{T}_{\text{CKP}}$ test is only slightly over-sized for all considered φ. If we set $\{\zeta_j\}_{j=1}^{21}$ to the first 21 Legendre polynomials and fix the attractor space to the span of $\{\zeta_1, \ldots, \zeta_\varphi\}$, then both tests work well even when T is moderate; see Appendix C for more details.

Table 3 reports finite-sample sizes and powers for different values of θ and T. $\text{NSS}_{\mathcal{C}}$ may

[7] $c_{\text{k}} = 0.75$ for the Parzen kernel.

Table 2: Simulation results for the bottom-up test with $\kappa(\varphi_0) = 2$

(a) $\theta = 0.1$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	4.7	2.8	2.2	1.5	1.2	1.0
350	4.3	3.6	3.4	3.0	2.5	2.3
750	5.2	4.8	4.6	4.2	3.5	3.7
			power, $q = 1$			
150	88.0	59.3	41.9	27.6	17.8	10.5
350	95.3	83.8	78.8	72.4	65.7	58.7
750	98.2	94.5	93.4	93.0	90.6	89.7
			power, $q = 2$			
150	98.4	82.7	65.5	50.8	32.9	23.8
350	100	99.4	97.8	95.9	90.4	83.0
750	100	100	100	100	100	100

(b) $\theta = 0.7$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	7.7	4.4	3.5	3.0	1.8	2.0
350	6.8	4.8	4.7	4.8	4.6	4.3
750	6.1	5.2	5.4	4.6	4.4	5.3
			power, $q = 1$			
150	87.8	59.2	43.8	29.7	18.3	12.0
350	95.0	84.4	78.7	75.2	67.6	58.2
750	98.1	94.0	94.2	93.2	92.5	90.2
			power, $q = 2$			
150	98.0	83.2	66.4	50.9	35.3	25.1
350	99.9	98.1	96.4	92.9	87.7	79.2
750	100	100	99.9	99.8	99.6	98.9

(c) $\theta = 0.1$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.0	3.8	3.6	2.8	2.2	1.9
350	4.7	4.8	4.0	4.3	3.9	3.6
750	4.9	5.1	4.3	4.4	4.3	4.2
			power, $q = 1$			
150	93.8	78.5	66.1	51.6	39.0	30.7
350	98.4	94.0	93.2	92.5	90.0	86.6
750	99.6	98.6	98.6	99.0	99.0	98.6
			power, $q = 2$			
150	99.5	93.9	88.2	76.0	65.9	54.9
350	100	99.8	99.6	99.1	98.6	96.4
750	100	100	100	100	100	100

(d) $\theta = 0.7$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	10.8	8.0	7.5	8.6	7.5	6.4
350	9.0	7.2	6.9	9.0	8.5	9.8
750	7.0	7.0	6.6	7.0	7.2	7.5
			power, $q = 1$			
150	93.9	77.4	64.4	54.5	41.8	34.4
350	98.4	94.4	93.4	91.9	89.4	85.3
750	99.5	98.8	98.8	99.0	98.9	98.8
			power, $q = 2$			
150	99.2	93.1	86.1	77.6	65.7	57.1
350	100	99.7	99.3	98.8	97.6	96.0
750	100	100	100	100	100	99.9

(e) $\theta = 0.1$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	3.9	2.0	1.2	1.0	0.5	0.2
350	5.4	3.0	1.9	2.0	1.5	1.2
750	4.6	4.4	3.2	3.2	2.6	2.9
			power, $q = 1$			
150	78.6	38.9	18.8	9.3	4.4	2.3
350	89.8	67.0	53.4	42.4	33.2	23.8
750	95.9	84.3	80.0	75.6	69.8	62.8
			power, $q = 2$			
150	95.2	63.9	38.9	20.8	9.7	5.8
350	99.6	95.7	90.3	82.0	67.4	51.3
750	99.9	99.6	99.4	99.4	98.9	97.6

(f) $\theta = 0.7$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.4	2.5	1.0	0.7	0.6	0.4
350	5.4	3.7	2.5	2.5	1.8	2.0
750	5.0	3.7	3.5	3.1	2.8	2.9
			power, $q = 1$			
150	79.0	39.8	19.7	9.6	4.8	2.7
350	91.3	67.0	55.2	44.2	33.2	24.0
750	95.8	83.7	80.5	74.8	67.1	62.6
			power, $q = 2$			
150	94.9	63.0	37.7	20.7	9.3	5.5
350	99.5	93.8	86.3	76.6	60.2	46.0
750	99.9	99.6	99.0	97.9	97.1	94.8

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 + q$.

be the least preferred in practice if we have only a moderate number of observations: it exhibits a severe over-rejection when T is small and φ is high. On the other hand, $\mathrm{NSS}_{\mathcal{K}}$ is relatively conservative. When T is small, it is under-sized at every considered values of φ and displays relatively lower power, which is more prominent when $\theta = 0.7$. However, this is not a general finite-sample property of $\mathrm{NSS}_{\mathcal{K}}$: it turns out in Appendix C that a moderate change of the simulation DGP can make $\mathrm{NSS}_{\mathcal{K}}$ reject the correct null hypothesis more frequently than the $\mathcal{T}$ test or the $\mathcal{T}_0$ test does. Finite-sample sizes and powers of the $\mathcal{T}$ test and the $\mathcal{T}_0$ test seem to be between those of $\mathrm{NSS}_{\mathcal{K}}$ and $\mathrm{NSS}_{\mathcal{C}}$. More specifically, the reported finite-sample size of each of the suggested tests is close to the nominal size even when $T = 150$, and has better power than $\mathrm{NSS}_{\mathcal{K}}$. The $\mathcal{T}_0$ test exhibits a slight over-rejection when $\varphi \geq 4$. Among the considered tests, the $\mathcal{T}_{\mathrm{S}}$ test is the most conservative when T is moderate; it has smaller finite-sample size and power than those of $\mathrm{NSS}_{\mathcal{K}}$. Given the results reported in Table 3, the $\mathcal{T}$ test or the $\mathcal{T}_0$ test would be the most preferred in practice. Even if the considered top-down tests show different finite-sample properties when T is moderate, they become quite similar to each other when $T = 750$, as suggested by asymptotic theory.

Our simulation results reported in Tables 2 and 3 suggest that the proposed top-down procedures based on the $\mathcal{T}$ test and the $\mathcal{T}_0$ test may help the correct determination of the cointegration corank due to their higher power against the closest alternative hypothesis in small samples. For example, when $\theta = 0.7$, $T = 150$ and the true cointegration corank is 4, the rejection frequency of $\mathcal{B}(3, 2)$ with the baseline bandwidth is 29.7%, but those of $\mathcal{T}(5, 2)$ and $\mathcal{T}_0(5, 2)$ are 55.2% and 70.1%, respectively. This implies that there is a better chance to have the true cointegration corank from the proposed top-down procedures starting at $\varphi_{\max} = 5$. Such a gain appears to be prominent when T is small, but become smaller as T increases. This is a clear evidence suggesting that the top-down procedures are still useful in practice as long as a proper choice of $\varphi_{\max}$ is given.

5 Empirical applications

5.1 Age-specific shares of full-time workers

We first apply our methodology to the time series of U.S. age-specific shares of full-time workers, which may be interpreted as a measure of the quality of employment, observed monthly from January 1980 to May 2019. The raw data is available from the CPS at https://www.ipums.org/. The Bureau of Labor Statistic (BLS) defines a full-time worker as who works 35 hours or more a week. However for economic or noneconomic reasons,[8] such as vacation, health issues, weather, etc., a worker who 'usually' works full time can be

[8]See https://www.bls.gov/cps/cpsaat20.htm

Table 3: Simulation results for top-down tests

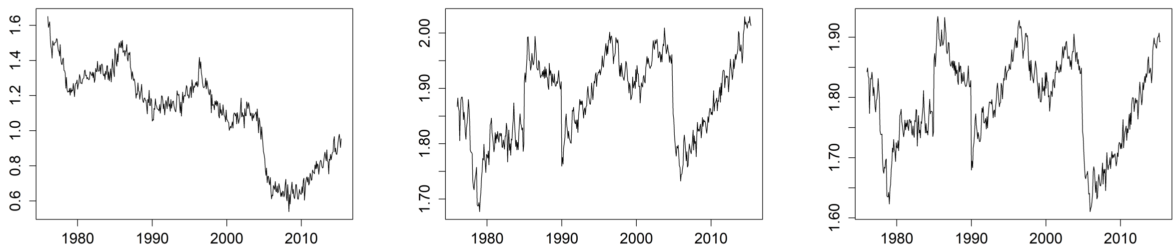
		$\theta = 0.1$, size							$\theta = 0.7$, size				
T	Test	$\varphi_0 = 1$	2	3	4	5	T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	5.0	4.2	4.8	4.7	6.4	150	$\mathcal{T}$	4.3	4.3	4.8	5.1	4.8
	$\mathcal{T}_0$	5.1	6.0	6.8	8.9	9.0		$\mathcal{T}_0$	4.6	5.0	6.6	8.0	7.7
	$\text{NSS}_{\mathcal{K}}$	3.7	2.4	2.0	1.9	1.3		$\text{NSS}_{\mathcal{K}}$	2.2	1.5	0.7	1.0	0.6
	$\text{NSS}_{\mathcal{C}}$	10.5	22.2	34.7	43.8	49.8		$\text{NSS}_{\mathcal{C}}$	10.7	22.0	32.7	38.2	38.1
	$\mathcal{T}_{\text{S}}$	3.4	1.0	0.2	0.0	0.0		$\mathcal{T}_{\text{S}}$	3.8	2.7	0.8	0.3	0.1
350	$\mathcal{T}$	4.8	4.3	3.7	3.8	3.6	350	$\mathcal{T}$	4.4	3.9	3.8	3.8	3.2
	$\mathcal{T}_0$	5.5	4.8	4.8	4.5	4.9		$\mathcal{T}_0$	5.0	4.2	4.2	4.8	4.6
	$\text{NSS}_{\mathcal{K}}$	4.8	4.4	3.7	3.6	3.8		$\text{NSS}_{\mathcal{K}}$	3.8	3.8	3.0	3.2	2.5
	$\text{NSS}_{\mathcal{C}}$	5.5	7.4	10.3	13.8	17.5		$\text{NSS}_{\mathcal{C}}$	5.6	9.0	13.7	16.9	18.8
	$\mathcal{T}_{\text{S}}$	4.6	3.0	2.5	1.4	0.6		$\mathcal{T}_{\text{S}}$	3.8	3.2	2.0	1.6	0.6
750	$\mathcal{T}$	5.0	4.9	4.4	4.6	4.0	750	$\mathcal{T}$	5.2	4.5	4.1	3.7	4.0
	$\mathcal{T}_0$	5.6	5.6	5.0	4.8	5.2		$\mathcal{T}_0$	5.1	4.8	4.8	4.7	4.7
	$\text{NSS}_{\mathcal{K}}$	4.8	4.8	4.9	4.2	4.7		$\text{NSS}_{\mathcal{K}}$	5.8	4.0	4.3	4.8	4.8
	$\text{NSS}_{\mathcal{C}}$	5.2	5.0	5.1	4.9	5.3		$\text{NSS}_{\mathcal{C}}$	5.6	5.3	4.9	6.0	6.8
	$\mathcal{T}_{\text{S}}$	4.5	4.2	3.6	3.1	1.7		$\mathcal{T}_{\text{S}}$	3.6	3.8	2.9	2.5	2.2

		$\theta = 0.1$, power							$\theta = 0.7$, power				
T	Test	$\varphi_0 = 1$	2	3	4	5	T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	99.0	94.4	92.5	91.8	89.8	150	$\mathcal{T}$	87.0	79.4	70.5	64.5	55.2
	$\mathcal{T}_0$	99.8	97.7	97.8	97.3	97.2		$\mathcal{T}_0$	91.0	85.0	79.8	76.3	70.1
	$\text{NSS}_{\mathcal{K}}$	73.4	59.2	53.2	49.8	44.1		$\text{NSS}_{\mathcal{K}}$	47.2	32.4	23.8	19.0	13.4
	$\text{NSS}_{\mathcal{C}}$	99.9	97.4	98.0	98.4	97.9		$\text{NSS}_{\mathcal{C}}$	95.5	88.3	87.3	85.4	82.1
	$\mathcal{T}_{\text{S}}$	100	48.0	9.6	1.4	0.1		$\mathcal{T}_{\text{S}}$	100	69.7	20.7	5.4	1.0
350	$\mathcal{T}$	100	99.2	99.3	99.5	99.8	350	$\mathcal{T}$	98.0	95.8	94.9	93.5	91.9
	$\mathcal{T}_0$	100	100	99.9	99.9	100		$\mathcal{T}_0$	99.0	98.0	97.4	96.6	96.3
	$\text{NSS}_{\mathcal{K}}$	99.8	93.8	95.0	96.1	96.2		$\text{NSS}_{\mathcal{K}}$	94.7	82.8	81.9	79.4	74.8
	$\text{NSS}_{\mathcal{C}}$	100	100	99.9	99.9	100		$\text{NSS}_{\mathcal{C}}$	99.7	98.7	98.4	97.8	97.0
	$\mathcal{T}_{\text{S}}$	100	100	94.7	57.0	24.6		$\mathcal{T}_{\text{S}}$	100	100	98.7	78.9	45.8
750	$\mathcal{T}$	100	100	100	100	100	750	$\mathcal{T}$	99.9	99.4	99.5	99.6	99.4
	$\mathcal{T}_0$	100	100	100	100	100		$\mathcal{T}_0$	100	99.8	99.9	99.9	99.8
	$\text{NSS}_{\mathcal{K}}$	100	100	99.6	99.8	100		$\text{NSS}_{\mathcal{K}}$	100	98.2	98.0	98.5	98.4
	$\text{NSS}_{\mathcal{C}}$	100	100	100	100	100		$\text{NSS}_{\mathcal{C}}$	100	99.9	99.8	99.8	100
	$\mathcal{T}_{\text{S}}$	100	100	100	100	98.5		$\mathcal{T}_{\text{S}}$	100	100	100	100	99.9

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 - 1$.

classified as a part-time worker, and vice versa. Therefore, the computed shares of full-time workers based on the BLS's definition may not be clearly interpreted as a measure of the quality of employment. To avoid this, we redefine a full-time worker as who usually work 35 hours or more a week, which is revealed by the variable 'wkstat' provided in the CPS. We construct functional observations from the raw data as in Nielsen et al. (2019) that considers the time series of U.S. age-specific employment rates. For age a, the age-specific share of

Figure 1: Age group characteristics



(a) younger age group (b) older age group (c) all ages

full-time workers is calculated by

$$X_{a,t} = \frac{\sum_{i=1}^{n_t} w_{i,t} Z_{i,t} \mathbb{1}\{a_{i,t} = a\}}{\sum_{i=1}^{n_t} w_{i,t} \mathbb{1}\{a_{i,t} = a\}}$$

where n_t is the number of employed individuals aged 20 to 64 at time t and $w_{i,t}$, $a_{i,t}$ and $Z_{i,t}$ denote the design weight (WTFINL in CPS), age, and full-time status of individual i at time t, respectively. $Z_{i,t} = 1$ if individual i works usually full time and 0 otherwise. Then the time series of shares of full-time workers specific to each age $\{X_{a,t}, a = 20, \ldots, 64\}$ are seasonally adjusted by the software package provided by the US Census Bureau. The age-specific full time share of full-time worker takes values between 0 and 1 by construction, so we consider the logit transformation $\psi(X_{a,t})$ instead of $X_{a,t}$ as in Nielsen et al. (2019). We then obtain the functional observations $\{X_t(u)\}_{t\geq 1}$ for $u \in [20, 64]$ by smoothing $\psi(X_{a,t})$ over a using 30 cubic B-spline basis functions.

In Figure 1 we plot several scalar processes $\{\langle X_t, v\rangle\}_{t\geq 1}$ (called the v-characteristic of $\{X_t\}_{t\geq 1}$ in Franchi and Paruolo (2019)) to explore characteristics of the functional time series for $v = v_1, v_2$ and v_3 defined as follows:

$$v_1(x) = \frac{1}{5}\mathbb{1}\{x \leq 25\}, \quad v_2(x) = \frac{1}{39}\mathbb{1}\{x > 25\}, \quad v_3(x) = \frac{1}{44}, \quad x \in [20, 64]$$

Clearly each characteristic is understood as the average of (logit) shares of full-time workers corresponding to each age group. It should be noted that the characteristic of the younger age group is quite distinct from that of the older age group. So if we only analyze the time series of average shares over all ages, such a distinct characteristic would be discarded. This may be a big advantage of this functional approach. We expect that there exists at least one stochastic trend since each characteristic in Figure 1 appears nonstationary.

In this empirical application, we fix $\ell(\varphi_0)$ to $\varphi_0 + 2$ for all φ_0 and $\mathrm{k}(\cdot)$ to the Parzen kernel. For the $\mathcal{B}$ test, we consider two bandwidth choices $h = 9$ and 12. We also provide the results for the top-down tests $\mathcal{T}$, $\mathcal{T}_0$, NSS$_{\mathcal{K}}$ and NSS$_{\mathcal{C}}$ that are considered in Section 4. $\varphi_{\max} = 6$ is employed for all the top-down tests and $h = 3$ is employed for the $\mathcal{T}$ and $\mathcal{T}_0$

tests. Note that the plot for $\{\langle X_t, v_1\rangle\}_{t\geq 1}$ suggests the possible presence of a long-term trend over time. Thus, it would be more plausible to consider the tests adjusted to a linear trend in this example.

Table 4 reports the test results under two different specifications of deterministic components. Figure 2 displays the functional observations and the characteristics of the functional residuals $\{U_t\}_{t\geq 1}$ given in (3.14) with respect to the first five eigenvectors of $\widetilde{\Lambda}_T$ computed with $h = 12$. The $\mathcal{B}$ test adjusted to a linear trend concludes that the cointegration corank is 3 or 4 at 5% significance level depending on the choice of h. This suggests that $\varphi_{\max} = 6$ would not have a significant chance to be smaller than the true cointegration corank. We thus may not have to worry about our initial choice $\varphi_{\max} = 6$ employed for the top-down tests. Among six different tests adjusted to a linear trend, only the $\mathcal{B}$ test with $h = 12$ concludes that the cointegration corank is 3 at 5% significance level while the others do that it is 4. As a consequence of Proposition 3.5 and Remark 3.14, we may expect that the characteristics with respect to the first φ eigenvectors behave as unit root processes. However, the $u_{4,T}$-characteristic, displayed in Figure 2(e), does seem to be a persistent stationary process rather than a unit root process, at least to some extent. For this reason, we would be inclined to conclude that the considered functional time series are driven by three stochastic trends rather than four. Note that all the considered top-down tests give us different results depending on whether $\varphi_{\max} > 3$ and $\varphi_{\max} = 3$. This suggests that there could be disagreement due to the choice of $\varphi_{\max}$; in this example, the researcher's confidence in nonstationarity of the $u_{4,T}$-characteristic may matter. The $\mathcal{B}$ test with $h = 9$ and 12 also give us different estimates, but it is a rather natural consequence resulting from differences in how high persistence of the underlying stationary component should be accommodated in the model.

5.2 Hourly wage densities

We consider a monthly sequence of U.S. hourly real wage densities running from January 1982 to June 2019, which is a revisit of the empirical application provided in Seo and Beare (2019) with an extended time span. To embed probability density functions (wage densities) into a Hilbert space, we follow the transformation approach proposed by Seo and Beare (2019). Let $\{X_t\}_{t\geq 1}$ denote the time series of wage densities whose support $K \subset \mathbb{R}$ for all t. For the density function $f_{\mathbb{M}}$ of any arbitrary measure $\mathbb{M}$ whose support is K, we define

$$\widetilde{X}_t(u) = \log\left(X_t/f_{\mathbb{M}}\right)(u) - \int \log\left(X_t/f_{\mathbb{M}}\right)(u)d\mathbb{M}(u), \quad u \in K \tag{5.1}$$

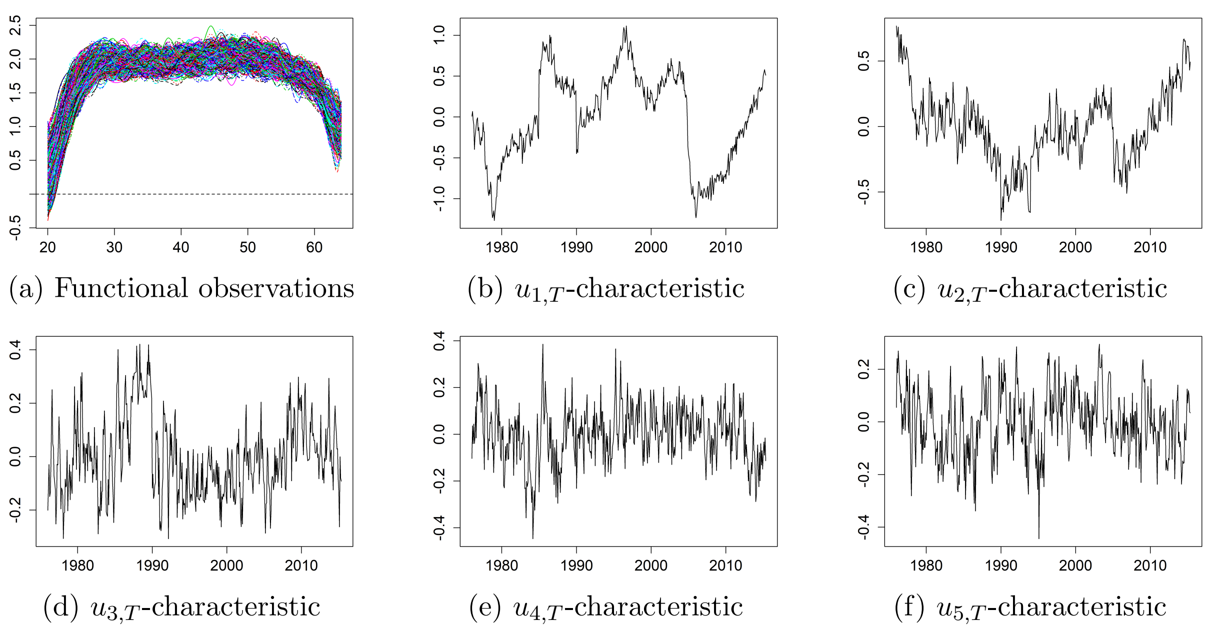
$$X_{\mathbb{M},t}(v) = \widetilde{X}_t \circ Q_{\mathbb{M}}(v), \quad v \in [0,1] \tag{5.2}$$

Table 4: Test results for logit of age-specific shares of full-time employment

Test	m	$\varphi_0 = 0$	1	2	3	4	5	6
				Mean adjusted				
$\mathcal{B}$	9	7.593***	1.787***	0.376***	0.209***	**0.072**	0.045	0.045
	12	5.746***	1.364***	0.356***	0.164**	**0.065**	0.053	0.043
$\mathcal{T}$	3	-	10.53	56.77	382.64	**731.05**	2456.2**	5025.3***
$\mathcal{T}_0$	3	-	10.61	58.27	460.65	**813.08**	3079.2***	6649.1***
$\text{NSS}_{\mathcal{K}}$	-	-	10.65	58.93	389.58	**802.34**	2484.8**	6389.5***
$\text{NSS}_{\mathcal{C}}$	-	-	10.61	57.10	403.88	**689.61**	3058.8***	4879.0***
				Trend adjusted				
$\mathcal{B}$	9	1.811***	0.338***	0.267***	0.097**	**0.053**	0.052	0.037
	12	1.381***	0.413***	0.217***	**0.086***	0.049	0.048	0.044
$\mathcal{T}$	3	-	45.81	256.08	623.99	**1844.7**	3755.8**	7218.0***
$\mathcal{T}_0$	3	-	47.22	292.17	681.65	**2156.5***	4821.9***	9401.7***
$\text{NSS}_{\mathcal{K}}$	-	-	47.75	295.26	666.72	**2084.6***	4405.6***	9644.5***
$\text{NSS}_{\mathcal{C}}$	-	-	47.08	333.44	698.73	**2061.9***	3538.8**	9293.9***

Notes: The number of observations is $T = 473$. We use $*$, $**$, and $***$ to denote rejection at 10%, 5%, and 1% significance level, respectively. The statistic is bold-faced if it is not rejected for the first time at 5% significance level.

Figure 2: Monthly age-specific shares of full-time employment January 1980 to May 2019



(a) Functional observations (b) $u_{1,T}$-characteristic (c) $u_{2,T}$-characteristic

(d) $u_{3,T}$-characteristic (e) $u_{4,T}$-characteristic (f) $u_{5,T}$-characteristic

where, $Q_{\mathbb{M}}$ is the quantile function for $\mathbb{M}$ and $\circ$ denotes composition. Then the transformed series $\{X_{\mathbb{M},t}\}_{t\geq 1}$ are regarded as a time series taking values in $L^2_0[0,1]$, the collection of all

$f \in L^2[0,1]$ satisfying $\int f(x)dx = 0$, which is a Hilbert space. Since the transformation $X_t \mapsto X_{\mathbb{M},t}$ is a bijection,[9] we analyze the time series $\{X_{\mathbb{M},t}\}_{t\geq 1}$ instead of the original density-valued time series $\{X_t\}_{t\geq 1}$.

$X_{\mathbb{M},t}$ for each t is computed in a nearly identical way to that in Seo and Beare (2019). First, individual wages, reported in the CPS database, are adjusted to January 2000 prices using monthly consumer price index obtained from the Federal Reserve Economic Data (FRED). To implement the transformation given in (5.1) and (5.2), the log wage densities are computed by applying local likelihood density estimation (Loader, 1996, 2006), excluding the top and bottom 1.5 % of reported wages due to anomalies associated with the top coding and near-zero wages. After applying the filtering, the number of design-weighted observations used to compute the log wage density in each month 7056-9523. Appendix D briefly summarizes how to obtain the log wage densities with these observations. Measure $\mathbb{M}$, called the reference measure, is set to the log-normal distribution of which parameters are chosen to minimize the mean square distance to the time average wage density $\bar{X}_T$. We then apply the transformations given in (5.1) and (5.2) to obtain $\{X_{\mathbb{M},t}\}_{t=1}^T$. Figure 3(a) and (b) display the original densities and the transformed densities, respectively.

For the representation of functions in $L_0^2[0,1]$, we use 50 quadratic B-spline basis functions projected onto $L_0^2[0,1]$. Even if we focus on the tests that are adjusted to a nonzero mean, we also report the results for those adjusted to a linear trend. It might seem to be unusual to consider a linear trend for a density-valued time series. However, the transformed time series $\{X_{\mathbb{M},t}\}_{t\geq 1}$ may include any arbitrary deterministic component and it is properly interpreted in a Hilbert space of probability density functions; see Seo and Beare (2019). Table 5 reports the test results and Figure 3(c)-(i) display the characteristics of the demeaned transformed series with respect to the first seven eigenvectors of Λ_T computed with $h = 12$.

The $\mathcal{B}$ test adjusted to a nonzero mean with $h = 9$ and 12 conclude that the cointegration corank is 6 at 5% significance level. On the other hand, the top-down tests do not reject the top hypothesis $H_0 : \varphi = 6$, so the estimated cointegration corank is $\varphi_{\max}$ for any $\varphi_{\max} \leq 6$. So the researcher's conjecture on an upper bound of φ may affect the estimate in this case. Non-rejection of the top-hypothesis may be an indication of the possibility that $\varphi_{\max} < \varphi$ (see Remark 3.3). We therefore conclude that $\varphi_{\max}$ may be misspecified and needs to be adjusted to a higher value. It may be deduced from Remark 3.3 that without any prior information on φ, we need to provide $\varphi_{\max}$ high enough so that the first several null hypotheses are rejected to ensure $\varphi_{\max} \geq \varphi$ with high probability. However, as pointed out by Nielsen et al. (2019), it should be noted that it is not a good idea to start with a very large $\varphi_{\max}$ due to potential inaccuracy of eigenvalues of a high-dimensional covariance matrix. Furthermore, note that the sample covariance operators $\Pi_T^\ell(\varphi_0)\mathcal{K}_T\Pi_T^\ell(\varphi_0)$, $\Pi_T^\ell(\varphi_0)\Lambda_T\Pi_T^\ell(\varphi_0)$ and

[9] In fact, it is an isomorphism between $L_0^2[0,1]$ and a Hilbert space of probability density functions whose support is K; see Seo and Beare (2019) for more details.

Table 5: Test results for hourly wages

Test	m	$\varphi_0 = 0$	1	2	3	4	5	6
				Mean adjusted				
$\mathcal{B}$	9	5.612***	1.018***	0.426***	0.335***	0.160***	0.100***	**0.052**
	12	4.255***	0.778***	0.320***	0.269***	0.138***	0.083**	**0.049**
$\mathcal{T}$	3	-	11.09	89.16	341.08	606.0	1083.2	**2143.7**
$\mathcal{T}_0$	3	-	11.33	94.91	359.97	723.1	1505.4	**2934.9**
$\mathrm{NSS}_{\mathcal{K}}$	-	-	11.64	99.65	356.27	829.3	1458.5	**2849.7**
$\mathrm{NSS}_{\mathcal{C}}$	-	-	11.40	95.17	358.50	722.2	1868.8	**2928.6**
				Trend adjusted				
$\mathcal{B}$	9	0.889***	0.593***	0.357***	0.189***	0.136***	0.083***	**0.048***
	12	0.679***	0.456***	0.274***	0.157***	0.115***	0.071**	**0.046**
$\mathcal{T}$	3	-	70.40	271.31	561.37	1057.0	1579.8	**2979.8**
$\widetilde{\mathcal{T}}_0$	3	-	79.64	288.26	614.12	1285.0	2094.9	**4275.6**
$\mathrm{NSS}_{\mathcal{K}}$	-	-	76.32	286.44	613.20	1217.3	2092.3	**4107.4**
$\mathrm{NSS}_{\mathcal{C}}$	-	-	79.78	287.22	615.75	1282.0	2509.2	**4266.9**

Notes: The number of observations is $T = 450$. We use $*$, $**$, and $***$ to denote rejection at 10%, 5%, and 1% significance level, respectively. The statistic is bold-faced if it is not rejected for the first time at 5% significance level.

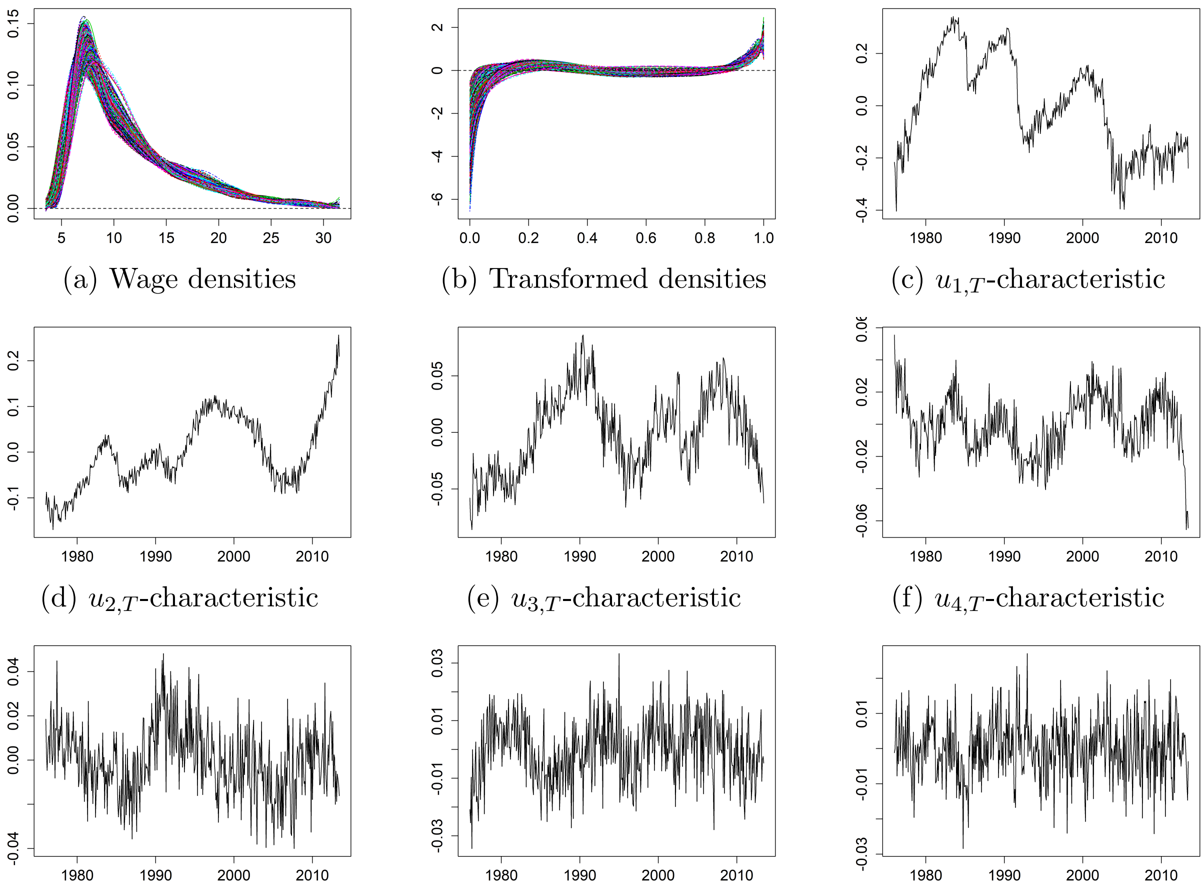
$\Pi_T^\ell(\varphi_0)\mathcal{C}_T(0)\Pi_T^\ell(\varphi_0)$ tend to be nearly singular as φ_0 gets larger.[10] Such a near-singularity can make extreme (small or large) eigenvalues computed with standard methods inaccurate, see e.g. Bunse-Gerstner (1984).[11] Therefore it would be better to choose $\varphi_{\max}$ only slightly bigger than φ if possible. In this situation, we may refer to the estimate from the $\mathcal{B}$ test with different choices of h. Since our simulation study suggests that power of the $\mathcal{B}$ test rapidly increases as H_1 gets farther away from H_0, we expect $\varphi_{\max} \approx 9$ would not have a significant chance to be smaller than the true cointegration corank.

Depending on which testing procedure is used, the cointegration corank is 6 or ≥ 6. However, at least to some extent, the $u_{7,T}$-characteristic with respect to the demeaned transformed series (Figure 3(i)) appears to be a persistent stationary process. We thus would be inclined to conclude that the cointegration corank is 6.

[10]Note $\mathcal{K}_T$, Λ_T and $\mathcal{C}_T(0)$ are all Hilbert-Schmidt operators. This implies the smallest eigenvalues of $\Pi_T^\ell(\varphi_0)\mathcal{K}_T\Pi_T^\ell(\varphi_0)$, $\Pi_T^\ell(\varphi_0)\Lambda_T\Pi_T^\ell(\varphi_0)$ and $\Pi_T^\ell(\varphi_0)\mathcal{C}_T(0)\Pi_T^\ell(\varphi_0)$ tends to zero as φ_0 gets larger.

[11]In this specific empirical example, we found that $\varphi_0 \approx 16$ makes those nearly singular.

Figure 3: Monthly wage densities January 1980 to June 2019



(a) Wage densities (b) Transformed densities (c) $u_{1,T}$-characteristic

(d) $u_{2,T}$-characteristic (e) $u_{3,T}$-characteristic (f) $u_{4,T}$-characteristic

(g) $u_{5,T}$-characteristic (h) $u_{6,T}$-characteristic (i) $u_{7,T}$-characteristic

6 Conclusion

We have proposed testing procedures to determine the cointegration corank of functional time series taking values in a Hilbert space, which can be also used to estimate the dominant components of cointegrated time series in the long-run. Our bottom-up testing procedure generalizing the functional KPSS test and the entailed top-down test given as its reverse may be used together for a better examination of the cointegration corank in practice. We also found some theoretical connections between the existing testing procedures and ours. We applied our methodology to two empirical datasets, U.S. age-specific shares of full-time workers and wage densities, and found evidence of multiple stochastic trends in both applications. Our empirical analysis also illustrates how the proposed bottom-up procedure can complement the top-down procedures in practice.

References

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

AUE, A. AND A. VAN DELFT (2017): "Testing for stationarity of functional time series in the frequency domain," *arXiv preprint arXiv:1701.01741*.

BART, H., I. GOHBERG, M. KAASHOEK, AND A. C. M. RAN (2007): *Factorization of Matrix and Operator Functions: The State Space Method*, Birkhuser Basel.

BEARE, B. K., J. SEO, AND W.-K. SEO (2017): "Cointegrated Linear Processes in Hilbert Space," *Journal of Time Series Analysis*, 38, 1010–1027.

BEARE, B. K. AND W.-K. SEO (2019): "Representation of I(1) and I(2) autoregressive Hilbertian processes," *Econometric Theory*, in press.

BOSQ, D. (2000): *Linear Processes in Function Spaces*, Springer-Verlag New York.

BOSWIJK, H. P., M. JANSSON, AND M. Ø. NIELSEN (2015): "Improved likelihood ratio tests for cointegration rank in the VAR model," *Journal of econometrics*, 184, 97–110.

BREITUNG, J. (2002): "Nonparametric tests for unit roots and cointegration," *Journal of Econometrics*, 108, 343–363.

BUNSE-GERSTNER, A. (1984): "An algorithm for the symmetric generalized eigenvalue problem," *Linear Algebra and its Applications*, 58, 43–68.

CHANG, Y., C. S. KIM, AND J. Y. PARK (2016): "Nonstationarity in time series of state densities," *Journal of Econometrics*, 192, 152 – 167.

CHEN, X. AND H. WHITE (1998): "Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications," *Econometric Theory*, 14, 260–284.

ENGL, H. W. AND M. NASHED (1981): "Generalized inverses of random linear operators in Banach spaces," *Journal of Mathematical Analysis and Applications*, 83, 582 – 610.

FRANCHI, M. AND P. PARUOLO (2019): "Cointegration in functional autoregressive processes," Tech. rep., *Econometric Theory*, in press.

GRANGER, C. W. J. (1981): "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics*, 16, 121 – 130.

Harris, D. (1997): "Principal components analysis of cointegrated time series," *Econometric Theory*, 13, 529–557.

Hörmann, S. and P. Kokoszka (2010): "Weakly dependent functional data," *The Annals of Statistics*, 38, 1845–1884.

Horváth, L. and P. Kokoszka (2012): *Inference for Functional Data with Applications*, Springer-Verlag GmbH.

Horváth, L., P. Kokoszka, and R. Reeder (2013): "Estimation of the mean of functional time series and a two-sample problem," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 103–122.

Horváth, L., P. Kokoszka, and G. Rice (2014): "Testing stationarity of functional time series," *Journal of Econometrics*, 179, 66–82.

Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Kokoszka, P. and G. Young (2016): "KPSS test for functional time series," *Statistics*, 50, 957–973.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics*, 54, 159–178.

Loader, C. R. (1996): "Local likelihood density estimation," *Annals of Statistics*, 24, 1602–1618.

——— (2006): *Local Regression and Likelihood*, Springer New York.

Nielsen, M. Ø., W.-K. Seo, and D. Seong (2019): "Inference on the dimension of the nonstationary subspace in functional time series," QED Working Paper 1420, Queen's University.

Nyblom, J. and A. Harvey (2000): "Tests of common stochastic trends," *Econometric theory*, 16, 176–199.

Phillips, P. (1988): "Spectral Regression for Cointegrated Time Series," Tech. rep., Cowles Foundation for Research in Economics, Yale University.

Phillips, P. C. B. (1995): "Fully modified least squares and vector autoregression," *Econometrica*, 63, 1023–1023.

Seo, W.-K. (2017): "Cointegration and representation of integrated autoregressive processes in function spaces," .

Seo, W.-K. and B. K. Beare (2019): "Cointegrated linear processes in Bayes Hilbert space," *Statistics & Probability Letters*, 147, 90 – 95.

Shintani, M. (2001): "A simple cointegrating rank test without vector autoregression," *Journal of Econometrics*, 105, 337–362.

Skorohod, A. (1983): *Random Linear Operators (Mathematics and its Applications)*, Springer.

Snell, A. (1999): "Testing for r versus r- 1 cointegrating vectors," *Journal of Econometrics*, 88, 151–191.

A Random Linear Operators

In this section, we briefly introduce random linear operators and their convergence for our mathematical proofs given in Section B. Our primary source is Skorohod (1983).

A.1 Random linear operators and convergence

Let A be a mapping satisfying Ax is $\mathcal{H}$-valued random variable for all $x \in \mathcal{H}$ and

(S1) for all $x_1, x_2 \in \mathcal{H}$ and $a, b \in \mathbb{R}$

$$\mathbb{P}\left\{A(ax_1 + bx_2) = aAx_1 + bAx_2\right\} = 1$$

(S2) Ax is continuous in x: for all $\delta > 0$

$$\lim_{x_j \to x} \mathbb{P}\left\{\|Ax_j - Ax\| > \delta\right\} = 0.$$

Then A is said to be a strong random operator. We let $\mathcal{L}^S_{\mathcal{H}}(\Omega)$ denote the set of strong random operators. Two strong random operators A and B are considered to be identical if $\mathbb{P}\{Ax = Bx\} = 1$ for all $x \in \mathcal{H}$.

For $A \in \mathcal{L}^S_{\mathcal{H}}(\Omega)$, $\langle Ax, y\rangle$ defines a real-valued random variable. If A satisfies

(W1) for all $x_1, x_2, y_1, y_2 \in \mathcal{H}$ and $a_1, a_2, b_1, b_2 \in \mathbb{R}$

$$\mathbb{P}\left\{\langle A(a_1x_1 + a_2x_2), b_1y_1 + b_2y_2\rangle = \sum_{j,k=1}^{2} a_j b_k \langle Ax_j, y_k\rangle\right\} = 1$$

(W2) $\langle Ax, y\rangle$ is continuous in (x, y): for all $(x, y) \in \mathcal{H} \otimes \mathcal{H}$ and $\delta > 0$,

$$\lim_{(x_j,y_j)\to(x,y)} \mathbb{P}\left\{|\langle Ax_j, y_j\rangle - \langle Ax, y\rangle| > \delta\right\} = 0,$$

then A is said to be a weak random operator. We let $\mathcal{L}_{\mathcal{H}}^{W}(\Omega)$ denote the set of weak random operators. Two weak random operators A and B are considered to be identical if $\mathbb{P}\{\langle Ax, y\rangle = \langle Bx, y\rangle\} = 1$ for all $x, y \in \mathcal{H}$.

We let $\mathcal{L}_{\mathcal{H}}(\Omega)$ be the set of mappings A from Ω to $\mathcal{L}_{\mathcal{H}}$, such that $\langle Ax, y\rangle$ is Borel measurable for all $x, y \in \mathcal{H}$. Such a map A is called a random bounded linear operator. From the definitions, the following inclusions are easily established.

$$\mathcal{L}_{\mathcal{H}}^{W}(\Omega) \supset \mathcal{L}_{\mathcal{H}}^{S}(\Omega) \supset \mathcal{L}_{\mathcal{H}}(\Omega) \tag{A.1}$$

In this paper, we mainly use four modes of convergence of random linear operators. Let $\{A_j\}_{j\in\mathbb{N}}$ be a sequence of random linear operators. Then we say A_j converges in norm to A_0, and write $A_j \to_{\mathcal{L}_{\mathcal{H}}} A_0$, if

$$\|A_j - A_0\|_{\mathcal{L}_{\mathcal{H}}} \quad \to_p \quad 0.$$

A_j is said to be strongly converge to A_0, and write $A_j \to_s A_0$, if for all $x \in \mathcal{H}$ and $\delta > 0$,

$$\lim_{j\to\infty} \mathbb{P}\left\{\|A_j x - A_0 x\| > \delta\right\} = 0,$$

while it is said to be weakly converge to A_0, and write $A_j \to_w A_0$, if for all $x, y \in \mathcal{H}$ and $\delta > 0$,

$$\lim_{j\to\infty} \mathbb{P}\left\{|\langle A_j x, y\rangle - \langle A_0 x, y\rangle| > \delta\right\} = 0,$$

Moreover, we say A_j weakly converges in distribution to A_0, and write $A_j \to_{wd} A_0$, if, all any k, $x_1, \ldots, x_k$, $y_1, \ldots, y_k \in \mathcal{H}$,

$$\lim_{j\to\infty} \mathbb{E} f\left(\langle A_j x_1, y_1\rangle, \ldots, \langle A_j x_k, y_k\rangle\right) = \mathbb{E} f\left(\langle A_0 x_1, y_1\rangle, \ldots, \langle A_0 x_k, y_k\rangle\right)$$

for any bounded continuous function f on $\mathcal{H}^k$. It is easy to show that weak (resp. strong) convergence implies weak convergence in distribution. Due to the inclusions given in (A.1), the above modes of convergence are all relevant to random bounded linear operators. Moreover, it is deduced that the convergence in norm implies any of strong or weak modes of convergence.

A.2 Characteristic functionals of weak random operators

If $A \in \mathcal{L}_{\mathcal{H}}^{W}(\Omega)$, then $\langle A(x_1 \otimes y_1)x_2, y_2\rangle = \langle x_1, x_2\rangle\langle Ay_1, y_2\rangle$ is defined for all $x_1, x_2, y_1, y_2 \in \mathcal{H}$. Moreover, the trace of $A(x \otimes y)$, denoted by $\mathrm{tr}(A(x \otimes y))$, is given by

$$\mathrm{tr}(A(x \otimes y)) = \sum_{j=1}^{\infty} \langle A(x \otimes y)e_j, e_j\rangle = \langle Ay, x\rangle,$$

where $\{e_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$. Let $\mathcal{L}_{\mathcal{H}}^{0}(\Omega)$ be the set of all operators Υ of the form

$$\Upsilon = \sum_{j=1}^{k} x_j \otimes y_j, \quad x_j, y_j \in \mathcal{H}, \quad j = 1, \ldots, k, \quad k = 1, 2, \ldots$$

For any $\Upsilon \in \mathcal{L}_{\mathcal{H}}^{0}(\Omega)$, it is trivial to show that $\mathrm{tr}(A\Upsilon)$ is well defined. A functional $\Psi_A : \mathcal{L}_{\mathcal{H}}^{0}(\Omega) \to \mathbb{C}$ defined by

$$\Psi_A(\Upsilon) = \mathbb{E}\exp\left(\mathrm{i}\,\mathrm{tr}(A\Upsilon)\right)$$

is called the characteristic functional of A. More detailed mathematical properties of $\Psi_A(\Upsilon)$ can be found in Skorohod (1983).

A.3 Useful Results on Convergence of Random Linear Operators

In this section, we state several lemmas that are useful in our asymptotic analysis. The following Lemmas A.1-A.3 can be found in or deduced from Chapter 3.3 of Skorohod (1983).

Lemma A.1. *For $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}_{\mathcal{H}}^{W}(\Omega)$, $A_j \to_{wd} A_0$ if and only if*

$$\lim_{j\to\infty} \Psi_{A_j}(\Upsilon) = \Psi_{A_0}(\Upsilon)$$

for all $\Upsilon \in \mathcal{L}_{\mathcal{H}}^{0}(\Omega)$.

Lemma A.2 (Cramer-Wold device)**.** *For $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}_{\mathcal{H}}^{W}(\Omega)$, $A_j \to_{wd} A_0$ if and only if*

$$\sum_{j=1}^{k} \langle Ay_j, x_j\rangle \quad \to_d \quad \sum_{j=1}^{k} \langle A_0 y_j, x_j\rangle$$

for any k, $x_1, \ldots, x_k, y_1, \ldots, y_k \in \mathcal{H}$.

Lemma A.3. *Let $\{A_j\}_{j\in\mathbb{N}}$ be a sequence of random operators and $\{\Psi_{A_j}\}_{j\in\mathbb{N}}$ be their characteristic functionals. If the limit*

$$\lim_{j\to\infty} \Psi_{A_j}(\Upsilon) = \Psi_{A_0}(\Upsilon) \quad \textit{for all } \Upsilon \in \mathcal{L}_{\mathcal{H}}^{0}(\Omega)$$

exists and the functional $\Psi_{A_0}(\alpha(x \otimes y))$ is continuous in α, then one can construct a sequence of random operators $\{\widetilde{A}_j\}_{j\in\mathbb{N}}$ and $\widetilde{A}_0$ such that

$$\Psi_{\widetilde{A}_j}(\cdot) = \Psi_{A_j}(\cdot), \quad \Psi_{\widetilde{A}_0}(\cdot) = \Psi_{A_0}(\cdot), \quad A_j \to_w \widetilde{A}_0.$$

Let $\{A_j\}_{j\in\mathbb{N}}$ be a sequence of weak random operators. Then we say that it satisfies stochastic strong boundedness (on the unit sphere) if

$$\sup_j \sup_{\|x\|\le 1} \mathbb{P}\{\|A_j x\| > \alpha\} \to_p 0 \quad \text{and} \quad \sup_j \sup_{\|x\|\le 1} \mathbb{P}\left\{\|A_j^* x\| > \alpha\right\} \to_p 0 \quad \text{as } \alpha \to \infty. \tag{A.2}$$

The following lemma shows that a sequence that weakly converges in distribution to another weak random operator satisfies (A.2).

Lemma A.4. *Suppose that $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}_{\mathcal{H}}^W(\Omega)$ and $A_j \to_{wd} A_0 \in \mathcal{L}_{\mathcal{H}}^W(\Omega)$. Then it satisfies stochastic strong boundedness on the unit sphere.*

Proof. If $A_j \to_{wd} A_0$, then it can be verified (see Ch1, proof of Theorem 3 in Skorohod (1983)) that

$$\sup_j \sup_{\|x\|\le 1, \|y\|\le 1} \mathbb{P}\{|\langle A_j x, y\rangle| > \alpha\} \quad \to_p \quad 0, \quad \text{as } \alpha \to \infty \tag{A.3}$$

For some $x \in \mathcal{H}$ with $\|x\| \le 1$, if we set

$$\begin{aligned} y &= A_j x / \|A_j x\| & &\text{if } A_j x \neq 0, \\ y &= 0 & &\text{if } A_j x = 0, \end{aligned}$$

then the first convergence of (A.2) is deduced.

Moreover, (A.3) may be equivalently written as

$$\sup_j \sup_{\|x\|\le 1, \|y\|\le 1} \mathbb{P}\left\{|\langle x, A_j^* y\rangle| > \alpha\right\} \quad \to \quad 0, \quad \text{as } \alpha \to \infty \tag{A.4}$$

For some $y \in \mathcal{H}$ with $\|y\| \le 1$, if we set

$$\begin{aligned} x &= A_j^* y / \|A_j^* y\| & &\text{if } A_j^* y \neq 0, \\ x &= 0 & &\text{if } A_j^* y = 0, \end{aligned}$$

then the second convergence of (A.2) is deduced. □

We collect several useful results on convergent random operators.

Lemma A.5. *For $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{L}_{\mathcal{H}}^W(\Omega)$, the following hold.*

(i) *If $A_j(\omega) \to_{wd} A$ for some nonrandom operator A, then $A_j(\omega) \to_w A$.*

(ii) *If* $A_j(\omega) \to_w B_j(\omega)$ *and* $A_j(\omega) \to_{wd} A_0(\omega)$*, then* $B_j(\omega) \to_{wd} A_0(\omega)$*.*

(iii) *If* $A_j \to_w A_0$ *and* $B_j \to_{s \text{ or } \mathcal{L}_\mathcal{H}} B_0$*, then* $A_jB_j \to_w A_0B_0$*. If* B_j *is self-adjoint, then* $B_jA_j \to_w B_0A_0$

Proof. (i) and (ii) can be trivially established from the definitions of the modes of convergence. We therefore only prove (iii).

For all $x \in \mathcal{H}$

$$A_jB_jx - A_0B_0x = A_j(B_j - B_0)x + (A_j - A_0)B_0x$$

Then we may easily obtain the following: for all $x, y \in \mathcal{H}$

$$\langle A_j(B_j - B_0)x, y\rangle \leq \|A_j^*y\|\|(B_j - B_0)x\| \tag{A.5}$$

$$\langle (A_j - A_0)B_0x, y\rangle \to_p 0. \tag{A.6}$$

where (A.6) holds since $A_j \to_w A_0$. Moreover, it can be shown that the right hand side of (A.5) converges in probability to zero due to $B_j \to_{s \text{ or } \mathcal{L}_\mathcal{H}} B_0$ and $\|A_j^*y\| = \mathrm{O}_p(1)$, implied by stochastic strong boundedness of weakly convergent sequence of operators (see Lemma A.4).

Now suppose that B_j is self-adjoint for all j. Since $\langle B_jx, y\rangle = \langle x, B_jy\rangle$ for all $x, y \in \mathcal{H}$, self-adjointness of B_0 is easily established. Note that for all $x, y \in \mathcal{H}$,

$$\langle B_jA_jx, y\rangle = \langle A_jx, (B_j - B_0)y\rangle + \langle A_jx, B_0y\rangle.$$

Then it can be shown that

$$\langle A_jx, (B_j - B_0)y\rangle \leq \|A_jx\|\|(B_j - B_0)y\| \to_p 0 \tag{A.7}$$

$$\langle A_jx, B_0y\rangle \to_p \langle A_0x, B_0y\rangle = \langle B_0A_0x, y\rangle \tag{A.8}$$

where (A.7) holds since $B_j \to_{s \text{ or } \mathcal{L}_\mathcal{H}} B_0$ and $\|A_jx\| = \mathrm{O}_p(1)$ (see Lemma A.4) and (A.8) is deduced from $A_j \to_w A_0$. We therefore conclude that $\langle B_jA_jx, y\rangle \to_p \langle B_0A_0x, y\rangle$ for all $x, y \in \mathcal{H}$. □

B Mathematical Proofs

B.1 Preliminary Lemmas

We first establish several lemmas that will be used in the subsequent discussions.

Lemma B.1. *Under Assumptions M and K, the following hold.*

(i) $P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}} \to_{\mathcal{L}_{\mathcal{H}}} P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$.

(ii) $P^{\mathfrak{C}}\widetilde{\Lambda}_T P^{\mathfrak{C}} \to_{\mathcal{L}_{\mathcal{H}}} P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}$.

Proof. To show (i), we note that $P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}}$ is the long-run covariance of the stationary component $\{P^{\mathfrak{C}}\nu_t\}_{t\geq 1}$. Under Assumptions M and K, we apply Theorem 2 of Horváth et al. (2013) to obtain (i).[12] (ii) can be similarly shown by applying Theorem 5.3 of Kokoszka and Young (2016). □

Given the direct sum decomposition $\mathcal{H} = \mathfrak{A} \oplus \mathfrak{C}$ implied by Assumption M, we may view $\mathcal{K}_T$ and Λ_T as the following operator matrices.

$$\mathcal{K}_T = \begin{pmatrix} P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{A}} & P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{C}} \\ P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{A}} & P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{C}} \end{pmatrix}, \qquad \Lambda_T = \begin{pmatrix} P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{A}} & P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{C}} \\ P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{A}} & P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}} \end{pmatrix} \tag{B.1}$$

The following lemma establishes the weak limits in distribution and stochastic orders of the operators in (B.1).

Lemma B.2. *Under Assumptions M and K, we have*

$$\begin{pmatrix} T^{-2}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{A}} & T^{-1}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{C}} \\ T^{-1}P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{A}} & P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{C}} \end{pmatrix} \to_{wd} \begin{pmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} \\ \mathcal{K}_{21} & \mathcal{K}_{22} \end{pmatrix} =: \mathcal{K},$$
$$\begin{pmatrix} (mT)^{-1}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{A}} & (mT)^{-1/2}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{C}} \\ (mT)^{-1/2}P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{A}} & P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}} \end{pmatrix} \to_{wd} \begin{pmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{pmatrix} =: \Lambda,$$

where

$$\begin{aligned} \mathcal{K}_{11} &= P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\mathcal{S}_{V,V}P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}, & \mathcal{K}_{12} &= P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\mathcal{S}_{V,B}P^{\mathfrak{C}}\Lambda_{\nu}^{1/2}P^{\mathfrak{C}}, \\ \mathcal{K}_{21} &= P^{\mathfrak{C}}\Lambda_{\nu}^{1/2}P^{\mathfrak{C}}\mathcal{S}_{B,V}P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}, & \mathcal{K}_{22} &= P^{\mathfrak{C}}\Lambda_{\nu}^{1/2}P^{\mathfrak{C}}\mathcal{S}_{B,B}P^{\mathfrak{C}}\Lambda_{\nu}^{1/2}P^{\mathfrak{C}}, \\ \Lambda_{11} &= c_{\mathrm{k}}P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\mathcal{S}_{\overline{W},\overline{W}}P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}, & \Lambda_{22} &= P^{\mathfrak{C}}\Lambda_{\nu}P^{\mathfrak{C}}. \end{aligned}$$

Proof.

(i) Limit of $T^{-2}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{A}}$: First, we may apply standard asymptotic results for the real-valued sequence $\{\langle Y_t, P^{\mathfrak{A}}x\rangle\}_{t=1}^T$ for any $x \in \mathcal{H}$ as follows.

$$T^{-3/2}\langle Y_t, P^{\mathfrak{A}}x\rangle = \frac{1}{T}\sum_{s=1}^{t}\frac{\langle X_s - \bar{X}_T, P^{\mathfrak{A}}x\rangle}{\sqrt{T}} \to_d s_x\int_0^r \overline{W}_x(\tau)d\tau, \tag{B.2}$$

[12] The summability condition $\sum_{j=1}^\infty j\|\tilde{\Phi}_j\|_{\mathcal{L}_{\mathcal{H}}} < \infty$ establishes that $\{P^{\mathfrak{C}}\nu_t\}_{t\geq 1}$ is L^4-m-approximable (Proposition 2.1 of Hörmann and Kokoszka (2010)) and $M\sum_{j=M+1}^\infty \|\tilde{\Phi}_j\|_{\mathcal{L}_{\mathcal{H}}} \to 0$. With these properties, it is easy to show that all the requirements of Theorem 2 in Horváth et al. (2013) are satisfied for $\{P^{\mathfrak{C}}\nu_t\}_{t\geq 1}$.

where $\overline{W}_x(\tau) = W_x(r) - \int W_x(\tau)d\tau$ and $(W_x(r), r \in [0,1])$ is the standard Brownian motion taking values in $\mathbb{R}$, and

$$s_x^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}\langle \Delta X_t, P^{\mathfrak{A}}x\rangle\langle \Delta X_{t-j}, P^{\mathfrak{A}}x\rangle = \langle x, P^{\mathfrak{A}}\Lambda_{\Delta X}P^{\mathfrak{A}}x\rangle.$$

Since $(P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}W(r), r \in [0,1])$ is Brownian motion taking values in $\mathcal{H}$ and its covariance operator is given by $P^{\mathfrak{A}}\Lambda_{\Delta X}P^{\mathfrak{A}}$, (B.2) may be written as

$$T^{-3/2}\langle P^{\mathfrak{A}}Y_t, x\rangle \to_d \left\langle P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}V(r), x\right\rangle \tag{B.3}$$

Then from a standard continuous mapping theorem, it is deduced that

$$\begin{aligned}\langle T^{-2}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{A}}x, y\rangle &= \frac{1}{T^4}\sum_{t=1}^{T}\langle Y_t, P^{\mathfrak{A}}x\rangle\langle Y_t, P^{\mathfrak{A}}y\rangle \\ &\to_d \left\langle y, P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\left(\int V(r)\otimes V(r)dr\right)P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}x\right\rangle \end{aligned} \tag{B.4}$$

for all $x, y \in \mathcal{H}$. From (B.4) and the Cramer-Wold device (Lemma A.2), we may deduce that

$$T^{-2}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{A}} \quad \to_{wd} \quad P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\left(\int V(r)\otimes V(r)dr\right)P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}$$

(ii) Limit of $P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{C}}$: For any arbitrary $x \in \mathcal{H}$,

$$T^{-1/2}\langle Y_t, P^{\mathfrak{C}}x\rangle = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\langle \nu_t - \bar{\nu}_t, P^{\mathfrak{C}}x\right\rangle \;\to_d\; s_x B_x(r), \tag{B.5}$$

where $(B_x(r), r \in [0,1])$ is the standard Brownian bridge taking values in $\mathbb{R}$ and

$$s_x^2 = \sum_{j=-\infty}^{\infty} \mathbb{E}\langle \nu_t, P^{\mathfrak{C}}x\rangle\langle \nu_{t-j}, P^{\mathfrak{C}}x\rangle = \langle x, P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}x\rangle. \tag{B.6}$$

The above results (B.5) and (B.6) are quite well known for scalar-valued time series, see e.g. Kwiatkowski et al. (1992). As in (B.3), (B.5) may be written as

$$\frac{1}{\sqrt{T}}\langle Y_t, P^{\mathfrak{C}}x\rangle \;\to_d\; \langle P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}B(r), x\rangle.$$

Then from the continuous mapping theorem, it is deduced that

$$\langle P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{C}}x, y\rangle = \frac{1}{T^2}\sum_{t=1}^{T}\langle Y_t, P^{\mathfrak{C}}x\rangle\langle Y_t, P^{\mathfrak{C}}y\rangle$$
$$\to_d \left\langle y, P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}\left(\int B(r)\otimes B(r)dr\right)P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}x\right\rangle. \tag{B.7}$$

Using (B.7) and the Cramer-Wold device (Lemma A.2), we obtain the following.

$$P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{C}} \quad \to_{wd} \quad P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}\left(\int B(r)\otimes B(r)dr\right)P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}$$

(iii) Limit of $T^{-1}P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{A}}$ and $T^{-1}P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{C}}$: From (i) and (ii), the following can be easily established: for any $x, y \in \mathcal{H}$,

$$T^{-1}\langle P^{\mathfrak{C}}\mathcal{K}_T P^{\mathfrak{A}}x, y\rangle = \frac{1}{T}\sum_{t=1}^{T}\frac{\langle Y_t, P^{\mathfrak{A}}x\rangle}{T}\frac{\langle Y_t, P^{\mathfrak{C}}y\rangle}{\sqrt{T}}$$
$$\to_d \quad \left\langle y, P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}\left(\int V(r)\otimes B(r)dr\right)P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}x\right\rangle$$

Similarly,

$$T^{-1}\langle P^{\mathfrak{A}}\mathcal{K}_T P^{\mathfrak{C}}x, y\rangle \to_d \left\langle y, P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\left(\int B(r)\otimes V(r)dr\right)P^{\mathfrak{C}}\Lambda_\nu^{1/2}P^{\mathfrak{C}}x\right\rangle$$

Using the Cramer-Wold device (Lemma A.2), the desired results are easily obtained.

(iv) Limit of $(mT)^{-1}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{A}}$: In our proof of (i), we already established that

$$T^{-1/2}\left\langle X_t - \bar{X}_T, P_{\mathfrak{A}}x\right\rangle \to_d \left\langle P_{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P_{\mathfrak{A}}\overline{W}(r), x\right\rangle$$

From the Skorohod's representation and some algebra, we may deduce the following (see e.g. Phillips (1988)).

$$\left|\frac{1}{T^2}\sum_{t=s+1}^{T}\langle X_t - \bar{X}_T, P_{\mathfrak{A}}x\rangle\langle X_t - \bar{X}_T, P_{\mathfrak{A}}y\rangle - \int\langle\widetilde{W}(r), P_{\mathfrak{A}}x\rangle\langle\widetilde{W}(r), P_{\mathfrak{A}}y\rangle dr\right| = \mathrm{op}(1)$$

where the second term satisfies

$$\int\langle\widetilde{W}(r), P_{\mathfrak{A}}x\rangle\langle\widetilde{W}(r), P_{\mathfrak{A}}y\rangle \quad =_d \quad \int\langle\overline{W}(r), P_{\mathfrak{A}}x\rangle\langle\overline{W}(r), P_{\mathfrak{A}}y\rangle dr.$$

Then the desired result can be deduced as in our previous proofs.

(v) Limit of $(mT)^{-1/2}P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{A}}$ and $(mT)^{-1/2}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{C}}$: For convenience we assume $\mathrm{k}(\cdot) = 1$ on $[0, c]$, which is just to simplify the expressions below. Note that

$$\begin{aligned} &T^{-1}\sum_{s=0}^{m}\sum_{t=s+1}^{T}\langle X_{t-s} - \bar{X}_T, P^{\mathfrak{C}}x\rangle\langle X_t - \bar{X}_T, P^{\mathfrak{A}}y\rangle \\ &\quad \le \max_{1\le t\le T}\left\{\left|\frac{\langle X_t - \bar{X}_T, P^{\mathfrak{A}}y\rangle}{\sqrt{T}}\right|\right\}\left(\frac{1}{\sqrt{T}}\left|\sum_{s=0}^{m}\sum_{t=s+1}^{T}\langle X_{t-s} - \bar{X}_T, P^{\mathfrak{C}}x\rangle\right|\right) \end{aligned} \tag{B.8}$$

Then clearly the right hand side of (B.8) is $\mathrm{O}_{\mathrm{p}}(m)$ (see also proof of Theorem B.4 of Nyblom and Harvey (2000)). From this results, we may easily deduce the desired results.

(vi) Limit of $P^{\mathfrak{C}}\Lambda_T P^{\mathfrak{C}}$: The desired result is implied by Lemma B.1. □

We introduce another useful convergence result, which can be deduced from Theorem 4.2 of Chang et al. (2016).

Lemma B.3. Under Assumption M(i)-(iii), we have

$$\begin{pmatrix} T^{-1}P^{\mathfrak{A}}\mathcal{C}_T(0)P^{\mathfrak{A}} & T^{-1/2}P^{\mathfrak{A}}\mathcal{C}_T(0)P^{\mathfrak{C}} \\ T^{-1/2}P^{\mathfrak{C}}\mathcal{C}_T(0)P^{\mathfrak{A}} & P^{\mathfrak{C}}\mathcal{C}_T(0)P^{\mathfrak{C}} \end{pmatrix} \quad \to_{wd} \quad \begin{pmatrix} \mathcal{C}_{11} & 0 \\ 0 & \mathcal{C}_{22} \end{pmatrix} =: \mathcal{C},$$

where

$$\mathcal{C}_{11} = P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}\mathcal{S}_{\overline{W},\overline{W}}P^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}P^{\mathfrak{A}}, \quad \mathcal{C}_{22} = P^{\mathfrak{C}}C_{\nu_0}P^{\mathfrak{C}}.$$

B.2 Proofs of the main results

We provide proofs of the results provided in Section 3. Throughout this section, we fix φ_0 and write all the mathematical objects of the form $f(\varphi_0)$ simply as f, e.g. $\ell = \ell(\varphi_0)$, $\Pi_T^{\ell} = \Pi_T^{\ell}(\varphi_0)$, $\widetilde{\mathcal{S}}_{B,B} = \widetilde{\mathcal{S}}_{B,B}(\varphi_0)$, etc. It may not cause any confusion and makes us simplify mathematical expressions.

We first note that $\Pi_T^{\ell}\Lambda_T\Pi_T^{\ell}$ and $\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}$ are not invertible in our infinite-dimensional setting, but their $\ell-$regularized inverses are well defined. Therefore, we may obtain the following reformulations of (3.8).

$$\lambda_{j,T}\phi_{j,T} = \left(\Pi_T^{\ell}\Lambda_T\Pi_T^{\ell}\right)_{\ell}^{\dagger}\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}\phi_{j,T}, \quad \phi_{j,T} \in \operatorname{ran}\Pi_T^{\ell} \tag{B.9}$$

$$\gamma_{j,T}\psi_{j,T} = \left(\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}\right)_{\ell}^{\dagger}\Pi_T^{\ell}\Lambda_T\Pi_T^{\ell}\psi_{j,T}, \quad \gamma_{j,T} = \lambda_{j,T}^{-1}, \quad \psi_{j,T} \in \operatorname{ran}\Pi_T^{\ell} \tag{B.10}$$

It is also noteworthy that $\mathcal{H}$ allows the following direct sum decomposition which are repeatedly used in this section.

$$\mathcal{H} = (\operatorname{ran}\Pi_T^{\mathfrak{A}} \cap \mathcal{H}_T^{\ell}) \oplus ((\operatorname{ran}\Pi_T^{\mathfrak{A}})^{\perp} \cap \mathcal{H}_T^{\ell}) \oplus (\mathcal{H}_T^{\ell})^{\perp}, \tag{B.11}$$
$$\mathcal{H} = (\operatorname{ran}\Pi^{\mathfrak{A}} \cap \mathcal{H}^{\ell}) \oplus ((\operatorname{ran}\Pi^{\mathfrak{A}})^{\perp} \cap \mathcal{H}^{\ell}) \oplus (\mathcal{H}^{\ell})^{\perp}, \tag{B.12}$$

Proof of Proposition 3.1

For convenience, we first show (ii).

Proof of (ii) : To show (ii), we assume that $\ell > \varphi$ and reformulate (B.10) as follows.

$$\gamma_{j,T}\psi_{j,T} = \mathcal{R}_T\psi_{j,T}, \quad \psi_{j,T} \in \mathcal{H}_T^{\ell} \tag{B.13}$$
$$\mathcal{R}_T = \left(D_T^2\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_T^2\right)_{\ell}^{\dagger} D_T^2\Pi_T^{\ell}\Lambda_T\Pi_T^{\ell}D_T^2,$$
$$D_T = \begin{pmatrix} T^{-1/2}I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix}$$
$$\gamma_{j,T} = \lambda_{j,T}^{-1}, \quad \gamma_{1,T} \leq \cdots \leq \gamma_{\ell,T}.$$

where I_1, I_2 and I_3 are properly defined identity maps according to the decomposition (B.11).

Given the decomposition as in (B.11), the operator $\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}$ may be viewed as the following operator matrix.

$$\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell} = \begin{pmatrix} \Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & \Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ \Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & \Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.14}$$

Then we have

$$D_T^2\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_T^2 = \begin{pmatrix} T^{-2}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & T^{-1}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ T^{-1}\Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & \Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.15}$$

Since $\Pi_T^{\ell} \to_{\mathcal{L}_{\mathcal{H}}} \Pi^{\ell}$, we may easily deduce that $\Pi_T^{\mathfrak{A}} \to_{\mathcal{L}_{\mathcal{H}}} \Pi^{\mathfrak{A}}$ and $\Pi_T^{\mathfrak{C}} \to_{\mathcal{L}_{\mathcal{H}}} \Pi^{\mathfrak{C}}$. Then from Lemmas A.5 and B.2, we may deduce that the operators in (B.14) satisfy

$$T^{-2}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} \to_{wd} \Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\mathcal{S}_{V,V}\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}} =: \widetilde{\mathcal{K}}_{11} \tag{B.16}$$
$$T^{-1}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} \to_{wd} \Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\mathcal{S}_{B,V}\Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}} =: \widetilde{\mathcal{K}}_{12} \tag{B.17}$$
$$T^{-1}\Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} \to_{wd} \Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}}\mathcal{S}_{V,B}\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}} =: \widetilde{\mathcal{K}}_{21} \tag{B.18}$$
$$\Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} \to_{wd} \Pi_{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi_{\mathfrak{C}}\mathcal{S}_{B,B}\Pi_{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi_{\mathfrak{C}} =: \widetilde{\mathcal{K}}_{22} \tag{B.19}$$

Since Π_T^ℓ satisfies Assumption P for $\ell > \varphi$, we have $\text{rank}(\widetilde{\mathcal{K}}_{11}) = \text{rank}(\widetilde{\mathcal{K}}_{12}) = \varphi$ and $\text{rank}(\widetilde{\mathcal{K}}_{21}) = \text{rank}(\widetilde{\mathcal{K}}_{22}) = \ell - \varphi$ almost surely. We let

$$\widetilde{\mathcal{K}} = \begin{pmatrix} \widetilde{\mathcal{K}}_{11} & \widetilde{\mathcal{K}}_{12} & 0 \\ \widetilde{\mathcal{K}}_{21} & \widetilde{\mathcal{K}}_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.20}$$

denote the limiting operator of (B.15) based on the direct sum decomposition (B.12). It is easy to verify that $\Psi_{\widetilde{\mathcal{K}}}(\alpha x \otimes y)$ is continuous in α from continuity of $\widetilde{\mathcal{K}}$, then Lemma A.3 implies that we can assume $D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \to_w \widetilde{\mathcal{K}}_\diamond$ for some $\widetilde{\mathcal{K}}_\diamond =_{wd} \widetilde{\mathcal{K}}$. Based on Assumption P, let $(e_1, \ldots, e_\ell)$ (resp. $(e_{1,T}, \ldots, e_{\ell,T})$) be an orthonormal basis of Π^ℓ (resp. Π_T^ℓ). Note that $(D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond)x = 0$ if $x \in \text{span}\{e_1, \ldots, e_\ell, e_{1,T}, \ldots, e_{\ell,T}\}^\perp$. From this result, we may deduce that there exist some orthonormal set $(\tilde{e}_{1,T}, \ldots, \tilde{e}_{\ell,T})$, which consists of at most ℓ vectors, satisfying $\tilde{e}_{j,T} \in [\text{ran}\,\Pi_\ell]^\perp$ for all $j = 1, \ldots, \ell$, and

$$\begin{aligned} \|D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond\|_{\mathcal{L}_\mathcal{H}} \leq & \sum_{j=1}^{\ell} |\langle (D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond) e_j, e_j \rangle| \\ & + \sum_{j=1}^{\ell} |\langle (D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond) \tilde{e}_{j,T}, \tilde{e}_{j,T} \rangle|, \end{aligned} \tag{B.21}$$

where the inequality is established from the fact that the operator norm of $D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond$ is bounded above by its trace norm (see equation (1.55) in Bosq (2000)). The first term on the right hand side of (B.21) clearly converges to zero since $\langle (D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond) e_j, e_j \rangle \to_p 0$ for $j = 1, \ldots, \ell$. Since $(I - \Pi^\ell)\tilde{e}_{j,T} = \tilde{e}_{j,T}$ and $\|\tilde{e}_{j,T}\| = 1$ regardless of T, we may deduce the following from the Cauchy-Schwartz inequality and properties of (operator) norm.

$$\begin{aligned} |\langle (D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 - \widetilde{\mathcal{K}}_\diamond) \tilde{e}_{j,T}, \tilde{e}_{j,T} \rangle| &= |\langle (I - \Pi^\ell) D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \tilde{e}_{j,T}, \tilde{e}_{j,T} \rangle| \\ &\leq \|(I - \Pi^\ell)\Pi_T^\ell\|_{\mathcal{L}_\mathcal{H}} \|D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \tilde{e}_{j,T}\| \end{aligned} \tag{B.22}$$

The right hand side of (B.22) converges in probability to zero since $\|D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \tilde{e}_{j,T}\|$ is $\mathrm{O}_p(1)$ by the stochastic strong boundedness on the unit sphere (Lemma A.4) and $\|(I - \Pi^\ell)\Pi_T^\ell\|_{\mathcal{L}_\mathcal{H}} \to_p 0$ by Assumption P. This implies that the second term on the right hand side of (B.21) converges in probability to zero. We therefore conclude that

$$D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \quad \to_{\mathcal{L}_\mathcal{H}} \quad \widetilde{\mathcal{K}}_\diamond \tag{B.23}$$

For some orthonormal basis $(u_{j,T}, j = 1, \ldots, \ell)$ of $\mathcal{H}_T^\ell$ and $(u_j, j = 1, \ldots, \ell)$ of $\mathcal{H}^\ell$, we have the following spectral representations of $D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2$ and $\widetilde{\mathcal{K}}_\diamond$.

$$D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 = \sum_{j=1}^{\ell} \eta_{j,T} u_{j,T} \otimes u_{j,T}, \quad \widetilde{\mathcal{K}}_\diamond = \sum_{j=1}^{\ell} \eta_j u_j \otimes u_j,$$

Then from (B.23) and Lemma 4.2 of Bosq (2000), we may easily deduce that

$$\sup_{1 \le j \le \ell} |\eta_{j,T} - \eta_j| \to 0 \tag{B.24}$$

We note that $\eta_j > 0$ and the associated eigenspace is one-dimensional for each $j = 1, \dots, \ell$ almost surely. Then it is deduced from Lemma 4.3 of Bosq (2000) that the eigenvectors satisfy

$$\|u_{j,T} - \operatorname{sgn}(\langle u_{j,T}, u_j \rangle) u_j\| \to_p 0, \quad \text{for } j = 1, \dots, \ell \tag{B.25}$$

Note that we have

$$\begin{aligned} \|\widetilde{\mathcal{K}}_\diamond^\dagger x - (D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2)_\ell^\dagger x\| \le & \left\| \sum_{j=1}^{\ell} (\eta_j^{-1} - \eta_{j,T}^{-1}) \langle u_j, x \rangle u_j \right\| + \left\| \sum_{j=1}^{\ell} \eta_{j,T}^{-1} \langle u_j, x \rangle (u_j - u_{j,T}) \right\| \\ & + \left\| \sum_{j=1}^{\ell} \eta_{j,T}^{-1} (\langle u_j, x \rangle - \langle u_{j,T}, x \rangle) u_{j,T} \right\| \end{aligned} \tag{B.26}$$

Then it follows straightforwardly from (B.24) and (B.25) that the right hand side of (B.26) converge in probability to zero, meaning that

$$\left(D_T^2 \Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_T^2 \right)_\ell^\dagger \quad \to_s \quad \widetilde{\mathcal{K}}_\diamond^\dagger \tag{B.27}$$

It is worth noting that $\widetilde{\mathcal{K}}_\diamond$ is almost surely invertible on $\mathcal{H}^\ell$, and zero on $(\mathcal{H}^\ell)^\perp$. From the Schur's formula of the inverse of the partitioned operator matrix (e.g. Bart et al. (2007, p. 29)), we may easily show that $\widetilde{\mathcal{K}}_\diamond^\dagger$ is given by

$$\widetilde{\mathcal{K}}_\diamond^\dagger =_{wd} \begin{pmatrix} \Pi^{\mathfrak{A}} A_{11} \Pi^{\mathfrak{A}} & \Pi^{\mathfrak{A}} A_{12} \Pi^{\mathfrak{C}} & 0 \\ \Pi^{\mathfrak{C}} A_{21} \Pi^{\mathfrak{A}} & \Pi^{\mathfrak{C}} A_{22} \Pi^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$\begin{aligned} A_{11} &= \widetilde{\mathcal{K}}_{11}^\dagger + \widetilde{\mathcal{K}}_{11}^\dagger \widetilde{\mathcal{K}}_{12} A_{22} \widetilde{\mathcal{K}}_{21} \widetilde{\mathcal{K}}_{11}^\dagger, & A_{12} &= -\widetilde{\mathcal{K}}_{11}^\dagger \widetilde{\mathcal{K}}_{12} A_{22}, \\ A_{21} &= -A_{22} \widetilde{\mathcal{K}}_{21} \widetilde{\mathcal{K}}_{11}^\dagger, & A_{22} &= \left(\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21} \widetilde{\mathcal{K}}_{11}^\dagger \widetilde{\mathcal{K}}_{12} \right)^\dagger. \end{aligned}$$

Similarly, the operator $\Pi_T^\ell \Lambda_T \Pi_T^\ell$ may be viewed as the following operator matrix.

$$\Pi_T^\ell \Lambda_T \Pi_T^\ell = \begin{pmatrix} \Pi_T^{\mathfrak{A}} \Lambda_T \Pi_T^{\mathfrak{A}} & \Pi_T^{\mathfrak{A}} \Lambda_T \Pi_T^{\mathfrak{C}} & 0 \\ \Pi_T^{\mathfrak{C}} \Lambda_T \Pi_T^{\mathfrak{A}} & \Pi_T^{\mathfrak{C}} \Lambda_T \Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

From Lemma B.2, it is easy to verify that

$$D_T^2 \Pi_T^\ell \Lambda_T \Pi_T^\ell D_T^2 \quad \to_w \quad \widetilde{\Lambda} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Pi^{\mathfrak{C}} \Lambda_\nu \Pi^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{B.28}$$

From (B.27), (B.28) and Lemma A.5(iii), we conclude that

$$\mathcal{R}_T \quad \to_w \quad \widetilde{\mathcal{K}}_\diamond^\dagger \widetilde{\Lambda} = \mathcal{R} =_{wd} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Pi^{\mathfrak{C}} A_{22} \Pi^{\mathfrak{C}} \Lambda_\nu \Pi^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Moreover from similar arguments between equations (B.20) and (B.23), we obtain

$$\mathcal{R}_T \quad \to_{\mathcal{L}_\mathcal{H}} \quad \mathcal{R} \tag{B.29}$$

We also consider the spectral representations of $\mathcal{R}_T$ and $\mathcal{R}$ as follows: for some orthonormal bases $(v_{j,T}, j = 1, \ldots, \ell)$ and $(w_{j,T}, j = 1, \ldots, \ell)$ of $\mathcal{H}_T^\ell$, and $(v_j, j = 1, \ldots, \ell)$ and $(w_j, j = 1, \ldots, \ell)$ of $\mathcal{H}^\ell$

$$\mathcal{R}_T = \sum_{j=1}^{\ell} \tau_{j,T} v_{j,T} \otimes w_{j,T}, \quad \mathcal{R} = \sum_{j=1}^{\ell} \tau_j v_j \otimes w_j,$$

Then from (B.29) and Lemma 4.2 in Bosq (2000), we have

$$\sup_{1 \le j \le \ell} |\tau_{j,T} - \tau_j| \to_p 0$$

The result implies that the eigenvalues of $\mathcal{R}_T$ converges in probability to those of $\mathcal{R}$. Since $\mathcal{R}(\mathcal{H}^\ell \cap \mathfrak{A}) = \{0\}$ and $\dim(\mathcal{H}^\ell \cap \mathfrak{A}) = \varphi$, $(\gamma_{1,T}, \ldots, \gamma_{\varphi,T})$ satisfying (B.13) converge in probability to zeroes. Moreover, $(\gamma_{\varphi+1,T}, \ldots, \gamma_{\ell,T})$ converge in distribution to eigenvalues of $\Pi^{\mathfrak{C}} A_{22} \Pi^{\mathfrak{C}} \Lambda_\nu \Pi^{\mathfrak{C}}$. Therefore we have

$$\gamma_{j,T} \quad \to_d \quad \lambda_{j-\varphi} \left(\Pi^{\mathfrak{C}} \left(\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21} \widetilde{\mathcal{K}}_{11}^\dagger \widetilde{\mathcal{K}}_{12} \right)^\dagger \Pi^{\mathfrak{C}} \Lambda_\nu \Pi^{\mathfrak{C}} \right), \quad j = \varphi + 1, \ldots, \ell \tag{B.30}$$

We note that the following holds.[13]

$$\widetilde{\mathcal{K}}_{11}^{\dagger} = \left(\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\right)^{\dagger} \widetilde{\mathcal{S}}_{VV}^{\dagger} \left(\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\right)^{\dagger} \tag{B.31}$$

This implies that

$$\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21}\widetilde{\mathcal{K}}_{11}^{\dagger}\widetilde{\mathcal{K}}_{12} = \Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}} \left(\widetilde{\mathcal{S}}_{BB} - \widetilde{\mathcal{S}}_{BV}\widetilde{\mathcal{S}}_{VV}^{\dagger}\widetilde{\mathcal{S}}_{VB}\right) \Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}},$$

and therefore

$$\left(\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21}\widetilde{\mathcal{K}}_{11}^{\dagger}\widetilde{\mathcal{K}}_{12}\right)^{\dagger} = \left(\Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}}\right)^{\dagger} \left(\widetilde{\mathcal{S}}_{BB} - \widetilde{\mathcal{S}}_{BV}\widetilde{\mathcal{S}}_{VV}^{\dagger}\widetilde{\mathcal{S}}_{VB}\right)^{\dagger} \left(\Pi^{\mathfrak{C}}\Lambda_{\nu}^{1/2}\Pi^{\mathfrak{C}}\right)^{\dagger}. \tag{B.32}$$

Then (B.30), (B.32) and the properties of an eigenvalue, we have

$$\lambda_j \left(\Pi^{\mathfrak{C}} \left(\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21}\widetilde{\mathcal{K}}_{11}^{\dagger}\widetilde{\mathcal{K}}_{12}\right)^{\dagger} \Pi^{\mathfrak{C}}\Lambda_{\nu}\Pi^{\mathfrak{C}}\right) =_d \lambda_{j-\varphi} \left(\left(\widetilde{\mathcal{S}}_{BB} - \widetilde{\mathcal{S}}_{BV}\widetilde{\mathcal{S}}_{VV}^{\dagger}\widetilde{\mathcal{S}}_{VB}\right)^{\dagger}\right), \tag{B.33}$$

for $j = \varphi + 1, \ldots, \ell$. Then (ii) follows from (B.30) and (B.33).

Proof of (i) : First, we only consider the case when $\ell > \varphi$. (B.9) can be reformulated as follows.

$$\begin{aligned}
&\lambda_{j,T}\psi_{j,T} = \mathcal{R}_T\psi_{j,T}, \quad \psi_{j,T} \in \mathcal{H}_T^{\ell} \qquad \text{(B.34)}\\
&\mathcal{R}_T = \left(D_{mT}\Pi_T^{\ell}\Lambda_T\Pi_T^{\ell}D_{mT}\right)_{\ell}^{\dagger} D_{mT}\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_{mT},\\
&D_{mT} = \begin{pmatrix} (mT)^{-1/2}I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix},\\
&\gamma_{j,T} = \lambda_{j,T}^{-1}, \quad \gamma_{1,T} \leq \cdots \leq \gamma_{\ell,T},
\end{aligned}$$

where I_1, I_2 and I_3 are properly defined identity maps according to the decomposition (B.11). Then clearly

$$\frac{m}{T}D_{mT}\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_{mT} = \begin{pmatrix} T^{-2}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & \left(\frac{m}{T}\right)^{1/2} T^{-1}\Pi_T^{\mathfrak{A}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ \left(\frac{m}{T}\right)^{1/2} T^{-1}\Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{A}} & \left(\frac{m}{T}\right) \Pi_T^{\mathfrak{C}}\mathcal{K}_T\Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

Since $\|\Pi_T^{\ell} - \Pi_{\ell}\|_{\mathcal{L}_{\mathcal{H}}} \to_p 0$, clearly $\|\Pi_T^{\mathfrak{A}} - \Pi^{\mathfrak{A}}\|_{\mathcal{L}_{\mathcal{H}}} \to_p 0$ and $\|\Pi_T^{\mathfrak{C}} - \Pi^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} \to_p 0$. Then we may

[13] In general, $(UVU)^{\dagger} \neq U^{\dagger}V^{\dagger}U^{\dagger}$ for $U, V \in \mathcal{L}_{\mathcal{H}}$. If $\operatorname{ran} U = \operatorname{ran} V = (\ker U)^{\perp} = (\ker V)^{\perp}$, then it can be shown that $(UVU)^{\dagger} = U^{\dagger}V^{\dagger}U^{\dagger}$.

deduce from Lemmas A.3, A.5, and B.2 that

$$\frac{m}{T} D_{mT}\Pi_T^\ell \mathcal{K}_T \Pi_T^\ell D_{mT} \quad \to_w \quad \widetilde{\mathcal{K}} =_{wd} \begin{pmatrix} \widetilde{\mathcal{K}}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\widetilde{\mathcal{K}}_{11}$ is given in (B.16). Note that we have $\dim(\mathcal{H}^\ell \cap \mathfrak{A}) = \varphi$, $\operatorname{rank}(\widetilde{\mathcal{K}}_{11}) = \varphi$ almost surely due to that $\Lambda_{\Delta X}^{1/2}$ is bijective as an operator acting on $\mathfrak{A}$. Moreover from Lemmas A.3, A.5, and B.2, we also have

$$D_{mT}\Pi_T^\ell \Lambda_T \Pi_T^\ell D_{mT} \quad \to_w \quad \widetilde{\Lambda} =_{wd} \begin{pmatrix} \widetilde{\Lambda}_{11} & 0 & 0 \\ 0 & \widetilde{\Lambda}_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B.35}$$

where

$$\widetilde{\Lambda}_{11} = c_{\mathrm{k}} \Pi^{\mathfrak{A}} \Lambda_{\Delta X}^{1/2} \Pi^{\mathfrak{A}} \mathcal{S}_{\overline{W},\overline{W}} \Pi^{\mathfrak{A}} \Lambda_{\Delta X}^{1/2} \Pi^{\mathfrak{A}}, \quad \widetilde{\Lambda}_{22} = \Pi^{\mathfrak{C}} \Lambda_\nu \Pi^{\mathfrak{C}} \tag{B.36}$$

From similar arguments between equations (B.20) and (B.27),[14] we may easily obtain

$$\left(D_{mT}\Pi_T^\ell \Lambda_T \Pi_T^\ell D_{mT}\right)_\ell^\dagger \quad \to_s \quad \widetilde{\Lambda}_\diamond^\dagger =_{wd} \begin{pmatrix} \widetilde{\Lambda}_{11}^\dagger & 0 & 0 \\ 0 & \widetilde{\Lambda}_{22}^\dagger & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We therefore conclude from Lemma A.2(iii) that

$$\frac{m}{T}\mathcal{R}_T \quad \to_w \quad \mathcal{R} = \widetilde{\Lambda}_\diamond^\dagger \widetilde{\mathcal{K}} =_{wd} \begin{pmatrix} \widetilde{\Lambda}_{11}^\dagger \widetilde{\mathcal{K}}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Again, from similar arguments to derive (B.23), we have $(m/T)\mathcal{R}_T \to_{\mathcal{L}_\mathcal{H}} \mathcal{R}$, so the eigenvalues of $(m/T)\mathcal{R}_T$ converges to those of $\mathcal{R}$. This implies that the last $\ell - \varphi$ eigenvalues of $\mathcal{R}_T$ converge to zeroes, and

$$(m/T)\lambda_{j,T} \quad \to_d \quad \lambda_j\left(\widetilde{\Lambda}_{11}^\dagger \widetilde{\mathcal{K}}_{11}\right).$$

Then it is easy to show the following from the properties of an eigenvalue.

$$\lambda_j\left(\widetilde{\Lambda}_{11}^\dagger \widetilde{\mathcal{K}}_{11}\right) =_d c_{\mathrm{k}}^{-1} \cdot \lambda_j\left(\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}^\dagger \widetilde{\mathcal{S}}_{VV}\right). \tag{B.37}$$

[14]Note that we need convergence of eigenelements of $D_{mT}\Pi_T^\ell \Lambda_T \Pi_T^\ell D_{mT}$ to those of $\widetilde{\Lambda}_\diamond$ to establish this result. Since an eigenvalue of $\widetilde{\Lambda}_{22}$ could allow multiple eigenvectors, the only difference is that we need to refer to Lemma 4.4 of Bosq (2000) instead of Lemma 4.3.

Then (ii) follows from (B.37).

For the case when $\ell \leq \varphi$, $\widetilde{\Lambda}_{22}$ in the right hand sides of (B.35) is understood as 0, and $\operatorname{rank}(\widetilde{\mathcal{K}}_{11}) = \operatorname{rank}(\widetilde{\Lambda}_{11}) = \ell$ since $\operatorname{rank} \Pi^{\mathfrak{A}} = \ell$. The rest of the proof is trivial, so omitted to save the space. □

Proof of Corollary 3.1

(i) is a direct consequence of Proposition 3.1. Moreover, we may easily deduce the following from Proposition 3.1.

$$\begin{aligned} \lambda_{j,T}^{-1} &= \mathrm{O_p}(m/T) \quad \text{for } j = 1, \ldots, \ell, \quad \text{if } \varphi_0 < \varphi \text{ and } \ell \leq \varphi, \\ \lambda_{j,T}^{-1} &= \mathrm{O_p}(m/T) \quad \text{for } j = \varphi + 1, \ldots, \ell, \quad \text{if } \varphi_0 < \varphi \text{ and } \ell > \varphi. \end{aligned}$$

Thus, if $\varphi_0 < \varphi$, at least one summand of the suggested statistic diverges to infinity. This proves (ii). (iii) follows from (i) and (ii). □

Proof of Corollary 3.2

(i) is a direct consequence of Proposition 3.1. Moreover, the proposition implies that

$$(m/T)\lambda_{j,T} = \mathrm{op}(1) \quad \text{for } j = \varphi + 1, \ldots, \ell, \quad \text{if } \varphi < \varphi_0 \leq \varphi_{\max}.$$

Therefore at least one summand of the suggested statistic diverges to infinity, which proves (ii). (iii) follows from (i) and (ii). □

Proof of Proposition 3.2

The proof only requires a slight and obvious modification from those of Lemma B.2 and Proposition 3.1, hence is omitted. □

Proof of Proposition 3.3

Proof of (i) : We may replace Λ_T with $\mathcal{C}_T(0)$ and m/T with $1/T$ in our proof of Proposition 3.1-(i). Using the result in Lemma B.3, we replace (B.35) and (B.36) with

$$D_T \Pi_T^\ell \mathcal{C}_T(0) \Pi_T^\ell D_T \quad \to_w \quad \widetilde{\mathcal{C}} =_{wd} \begin{pmatrix} \widetilde{\mathcal{C}}_{11} & 0 & 0 \\ 0 & \widetilde{\mathcal{C}}_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$\widetilde{\mathcal{C}}_{11} = \Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\mathcal{S}_{\overline{W},\overline{W}}\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}, \quad \widetilde{\mathcal{C}}_{22} = \Pi^{\mathfrak{C}}C_{\nu_0}\Pi^{\mathfrak{C}}.$$

From nearly identical arguments used in proof of Proposition 3.1(i), we can show that the eigenvalues of $T^{-1}\mathcal{R}_T \coloneqq T^{-1}(D_T\Pi_T^{\ell}\mathcal{C}_T(0)\Pi_T^{\ell}D_T)_{\ell}^{\dagger}D_T\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_T$ converge to those of $\widetilde{\mathcal{C}}^{\dagger}\widetilde{\mathcal{K}}$, which implies that

$$\begin{aligned}
&(1/T)\lambda_{j,T} \to_d \lambda_j\left(\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}^{\dagger}\widetilde{\mathcal{S}}_{VV}\right), \quad j = 1, \dots, \min\{\varphi, \ell\},\\
&(1/T)\lambda_{j,T} \to_p 0, \quad j = \min\{\varphi, \ell\} + 1, \dots, \ell.
\end{aligned}$$

Then the desired results can be deduced from the above properties of the eigenvalues as in Corollary 3.2.

Proof of (ii) : Similarly we replace Λ_T with $\mathcal{C}_T(0)$ in our proof of Proposition 3.1(ii) and we use the convergence result given in Lemma B.3. As a result, (B.28) is replaced with

$$D_T^2\Pi_T^{\ell}\mathcal{C}_T(0)\Pi_T^{\ell}D_T^2 \quad \to_w \quad \widetilde{\mathcal{C}} = \begin{pmatrix} 0 & 0 & 0\\ 0 & \Pi^{\mathfrak{C}}C_{\nu_0}\Pi^{\mathfrak{C}} & 0\\ 0 & 0 & 0\end{pmatrix} \tag{B.38}$$

From nearly identical arguments used in proof of Proposition 3.1(ii), it can be shown that

$$\gamma_{j,T} \quad \to_d \quad \lambda_j\left(\Pi^{\mathfrak{C}}\left(\widetilde{\mathcal{K}}_{22} - \widetilde{\mathcal{K}}_{21}\widetilde{\mathcal{K}}_{11}^{\dagger}\widetilde{\mathcal{K}}_{12}\right)^{\dagger}\Pi^{\mathfrak{C}}C_{\nu}\Pi^{\mathfrak{C}}\right)$$

If $\mathbb{E}[\nu_{t-s} \otimes \nu_t] = \mathbb{E}[\nu_t \otimes \nu_{t-s}] = 0$ for $s \neq 0$, then $C_{\nu_0} = \Lambda_{\nu}$. We then obtain the same limit given in Proposition 3.1(ii). □

Proof of Corollaries 3.4 and 3.3

The proof of Corollary 3.4 (resp. Corollary 3.3) is nearly identical to that of Corollary 3.2 (resp. Corollary 3.1), but using Proposition 3.3 instead of Proposition 3.1.

Proof of Proposition 3.4

Proof of (i) : We first assume that $\ell > \varphi$ and reformulate (3.20) as follows.

$$\begin{aligned}
&\lambda_{j,T}\psi_{j,T} = \mathcal{R}_T\psi_{j,T}, \quad \psi_{j,T} \in \mathcal{H}_T^{\varphi} \\
&\mathcal{R}_T = \left(D_{mT}\Pi_T^{\ell}\Lambda_{\Delta X,T}\Pi_T^{\ell}D_{mT}\right)_{\ell}^{\dagger} D_{mT}\Pi_T^{\ell}\mathcal{C}_T(0)\Pi_T^{\ell}D_{mT}, \\
&D_m = \begin{pmatrix} m^{-1}I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix},
\end{aligned} \tag{B.39}$$

where I_1, I_2 and I_3 are properly defined identity maps according to the decomposition (B.11). We have

$$\frac{m^2}{T}D_m\Pi_T^{\ell}\mathcal{C}_T(0)\Pi_T^{\ell}D_m = \begin{pmatrix} T^{-1}\Pi_T^{\mathfrak{A}}\mathcal{C}_T(0)\Pi_T^{\mathfrak{A}} & \left(\frac{m}{T}\right)\Pi_T^{\mathfrak{A}}\mathcal{C}_T(0)\Pi_T^{\mathfrak{C}} & 0 \\ \left(\frac{m}{T}\right)\Pi_T^{\mathfrak{C}}\mathcal{C}_T(0)\Pi_T^{\mathfrak{A}} & \left(\frac{m^2}{T}\right)\Pi_T^{\mathfrak{C}}\mathcal{C}_T(0)\Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$m^2D_m\Pi_T^{\ell}\Lambda_{\Delta X,T}\Pi_T^{\ell}D_m = \begin{pmatrix} \Pi_T^{\mathfrak{A}}\Lambda_{\Delta X,T}\Pi_T^{\mathfrak{A}} & m\Pi_T^{\mathfrak{A}}\Lambda_{\Delta X,T}\Pi_T^{\mathfrak{C}} & 0 \\ m\Pi_T^{\mathfrak{C}}\Lambda_{\Delta X,T}\Pi_T^{\mathfrak{A}} & m^2\Pi_T^{\mathfrak{C}}\Lambda_{\Delta X,T}\Pi_T^{\mathfrak{C}} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

From similar arguments that used to show $D_T^2\Pi_T^{\ell}\mathcal{K}_T\Pi_T^{\ell}D_T^2 \to_w \widetilde{\mathcal{K}}_{\diamond}$ in our proof of Proposition 3.1, it can be shown that

$$\frac{m^2}{T}D_{mT}\Pi_T^{\ell}\mathcal{C}_T(0)\Pi_T^{\ell}D_{mT} \quad \to_w \quad \widetilde{\mathcal{C}} =_{wd} \quad \begin{pmatrix} \widetilde{\mathcal{C}}_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B.40}$$

where $\widetilde{\mathcal{C}}_{11} = \Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}\mathcal{S}_{\overline{W},\overline{W}}\Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}$. Moreover from Lemma 8.1 of Phillips (1995), Theorem 3.1 of Shintani (2001), and Lemma A.5(i), we may deduce that

$$m^2D_m\Pi_T^{\ell}\Lambda_{\Delta X,T}\Pi_T^{\ell}D_m \quad \to_w \quad \widetilde{\Lambda}_{\Delta X} = \begin{pmatrix} \widetilde{\Lambda}_{\Delta X,11} & 0 & 0 \\ 0 & \widetilde{\Lambda}_{\Delta X,22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B.41}$$

where

$$\widetilde{\Lambda}_{\Delta X,11} = \Pi^{\mathfrak{A}}\Lambda_{\Delta X}^{1/2}\Pi^{\mathfrak{A}}, \quad \widetilde{\Lambda}_{\Delta X,22} = -\mathrm{k}''(0)\Pi^{\mathfrak{C}}\Lambda_{\nu}\Pi^{\mathfrak{C}}$$

Since $\ell > \varphi$, $\tilde{\Lambda}_{\Delta X,22} \neq 0$ and $\text{rank}(\widetilde{\mathcal{C}}_{11}) = \text{rank}(\tilde{\Lambda}_{\Delta X,11}) = \varphi$. From similar arguments between equations (B.20) and (B.27), we also obtain

$$\left(m^2 D_m \Pi_T^\ell \Lambda_{\Delta X,T} \Pi_T^\ell D_m\right)_\ell^\dagger \quad \to_s \quad \widetilde{\Lambda}_{\Delta X}^\dagger.$$

Then we may deduce that the eigenvalues of $T^{-1}\mathcal{R}_T$ converges to those of $\widetilde{\Lambda}_{\Delta X}^\dagger \widetilde{\mathcal{C}}_{11}$; that is, the first φ eigenvalues converge to those of $\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}$.

If $\ell \leq \varphi$, we have $\tilde{\Lambda}_{\Delta X,22} = 0$ and $\text{rank}(\widetilde{\mathcal{C}}_{11}) = \text{rank}(\tilde{\Lambda}_{\Delta X,11}) = \ell$ in (B.40) and (B.41). The rest of the proof is similar, and we conclude that all the eigenvalues of $T^{-1}\mathcal{R}_T$ converges to those of $\widetilde{\mathcal{S}}_{\overline{W},\overline{W}}$.

Proof of (ii) : Given the expressions in (B.40) and (B.41), it can be easily shown that

$$T^{-1}\lambda_{j,T} \quad \to_p \quad 0, \quad j \geq \varphi + 1.$$

□

Proof of Corollary 3.5

(i), (ii) and (iii) may be deduced from Proposition 3.4, so we omit the detailed proof. □

Proof of Proposition 3.5

For convenience we assume $\mathrm{k}(\cdot) = 1$ on $[0, c]$, which just to simplify the expressions below. Our proof may be easily generalized to a general kernel function $\mathrm{k}(\cdot)$ with a trivial modification. From the Cauchy-Schwarz inequality and the properties of norm, we may easily deduce that

$$\begin{aligned} &\|(mT)^{-1}\Lambda_T - (mT)^{-1}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{A}}\|_{\mathcal{L}_\mathcal{H}} \\ &\quad \leq \frac{1}{mT^2}\sum_{s=0}^{m}\sum_{t=1}^{T}\|P^{\mathfrak{A}}(X_{t-s} - \bar{X}_T)\|\|P^{\mathfrak{C}}(X_t - \bar{X}_T)\| \\ &\quad\quad + \frac{1}{mT^2}\sum_{s=0}^{m}\sum_{t=1}^{T}\|P^{\mathfrak{C}}(X_{t-s} - \bar{X}_T)\|\|P^{\mathfrak{A}}(X_t - \bar{X}_T)\| \\ &\quad\quad + \frac{1}{mT^2}\sum_{s=0}^{m}\sum_{t=1}^{T}\|P^{\mathfrak{C}}(X_{t-s} - \bar{X}_T)\|^2 = O_p(T^{-1}), \end{aligned} \tag{B.42}$$

see also (B.8). Then from Lemma 4.3 of Bosq (2000) that the eigenvectors $(u_{1,T}, \ldots, u_{\varphi,T})$ converges to those of $(mT)^{-1}P^{\mathfrak{A}}\Lambda_T P^{\mathfrak{A}}$, whose span is given by $\mathfrak{A}$ (see also Proposition 3.2 and its proof in Chang et al. (2016)). This implies that $P_T^{\mathfrak{A}} = \sum_{j=1}^{\varphi} u_{j,T}$ and $P_T^{\mathfrak{C}} = I - P_T^{\mathfrak{A}}$

satisfy

$$\|P_T^{\mathfrak{A}} - P^{\mathfrak{A}}\|_{\mathcal{L}_{\mathcal{H}}} = \mathrm{O_p}(T^{-1}) \tag{B.43}$$
$$\|P_T^{\mathfrak{C}} - P^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} = \mathrm{O_p}(T^{-1}) \tag{B.44}$$

Then from (B.42), (B.43), (B.44), we may easily deduce that

$$\begin{aligned}
&\|P_T^{\mathfrak{C}} P^{\mathfrak{A}} \Lambda_T P^{\mathfrak{A}} P_T^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} = \mathrm{O_p}\left(\frac{m}{T}\right) \\
&\|P_T^{\mathfrak{C}} P^{\mathfrak{A}} \Lambda_T P^{\mathfrak{C}} P_T^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} = \mathrm{O_p}\left(\frac{m}{T}\right) \\
&\|P_T^{\mathfrak{C}} P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{A}} P_T^{\mathfrak{C}}\|_{\mathcal{L}_{\mathcal{H}}} = \mathrm{O_p}\left(\frac{m}{T}\right) \\
&P_T^{\mathfrak{C}} P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{C}} P_T^{\mathfrak{C}} \to_{\mathcal{L}_{\mathcal{H}}} P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{C}}
\end{aligned}$$

That is, we have

$$P_T^{\mathfrak{C}} \Lambda_T P_T^{\mathfrak{C}} \quad \to_{\mathcal{L}_{\mathcal{H}}} \quad P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{C}}.$$

Note that $P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{C}}$ is the long-run covariance of the stationary component $(\tilde{\nu}_t, t \geq 1)$. Lemma B.1(i) implies that we have $P^{\mathfrak{C}} \Lambda_T P^{\mathfrak{C}} \to_{\mathcal{L}_{\mathcal{H}}} P^{\mathfrak{C}} \Lambda_\nu P^{\mathfrak{C}}$. Then from Lemma 4.3-4.4 of Bosq (2000), we may deduce the $(u_{\varphi+1,T}, \ldots, u_{\ell,T})$ converge to the eigenvectors of $P^{\mathfrak{C}} \Lambda_\nu P^{\mathfrak{C}}$. Then it may be easily shown that Π_T^ℓ satisfy the requirements in Assumption P.

(Inclusion of a linear trend, Remark 3.14) With a slight modification of our Lemma B.2 and the above proof, we can similarly show that the first φ eigenvectors converges to an orthonormal basis of $\mathfrak{A}$. Using Lemma B.1(ii), it can be shown that the eigenvectors of $P^{\mathfrak{C}} \widetilde{\Lambda}_T P^{\mathfrak{C}}$ converge to those of $P^{\mathfrak{C}} \Lambda_\nu P^{\mathfrak{C}}$. The remaining proof is quite apparent as before, so omitted. □

Proof of Proposition 3.6

From Theorem 4.2 of Chang et al. (2016) and Lemma 4.3-4.4 of Bosq (2000), we may deduce that

$$\begin{aligned}
\Pi_T^{\mathfrak{A}} &:= \sum_{j=1}^{\varphi} u_{j,T} \otimes u_{j,T} \quad \to_{\mathcal{L}_{\mathcal{H}}} \quad P^{\mathfrak{A}} \\
\Pi_T^{\mathfrak{C}} &:= \sum_{j=\varphi+1}^{\ell} u_{j,T} \otimes u_{j,T} \quad \to_{\mathcal{L}_{\mathcal{H}}} \quad \sum_{j=\varphi+1}^{\ell} u_j \otimes u_j =: P_{\ell-\varphi}^{\mathfrak{C}},
\end{aligned}$$

where $(u_{\varphi+1}, \ldots, u_{\ell(\varphi_0)})$ are the eigenvectors of $P^{\mathfrak{C}} C_{\nu_0} P^{\mathfrak{C}}$. Clearly Π_T^ℓ converges to Π^ℓ includes $\mathfrak{A}$ as a subspace, so we have

$$\dim(\operatorname{ran}\Pi_\ell \cap \mathfrak{A}) = \dim(\mathfrak{A}).$$

Moreover, $P^{\mathfrak{C}}_{\ell-\varphi}\Lambda_\nu P^{\mathfrak{C}}_{\ell-\varphi}$ is positive definite on $\operatorname{ran} P^{\mathfrak{C}}_{\ell-\varphi}$ by Assumption P2, which implies that $\operatorname{rank}\Pi^\ell P^{\mathfrak{C}}\Lambda_\nu P^{\mathfrak{C}}\Pi^\ell = \ell - \varphi$ hold.

(Inclusion of a linear trend, Remark 3.14) Even if Theorem 4.2 of Chang et al. (2016) does not explicitly consider a linear trend, it only requires a slight and obvious modification to extend their result to the case. Hence, the detailed proof is omitted. □

C Additional Simulation Results

In this section, we modify the DGP in Section 4 by setting $\mathcal{I}_1 = \{1, \ldots, \varphi\}$, i.e. the attractor space now is fixed to the span of $\{\zeta_1, \ldots, \zeta_\varphi\}$. This change reduces complexity of the attractor space in the sense that only less oscillating functions, that can be written as a linear combination of $\{\zeta_j\}_{j=1}^{\varphi}$, can be included as an element. In addition to the tests considered in in Section 4, we also report finite-sample sizes and powers of the CKP test and the $\mathcal{T}_{\text{CKP}}$ test. The long-run covariance $\Pi_T^\ell(\varphi_0)\Lambda_{\Delta X,T}\Pi_T^\ell(\varphi_0)$ for those tests is computed with the Parzen kernel and the bandwidth choice proposed by Andrews (1991). Moreover, for the $\mathcal{T}_{\text{CKP}}(\varphi_0)$ test, the projection $\Pi_T^\ell(\varphi_0)$ is obtained from the first φ_0-leading eigenvectors of Λ_T as in Section 4.2. Tables 6 and 7 report our simulation results. The $\mathcal{B}$ test tends to have better finite-sample power overall, but the reported over-rejection caused by choosing $h = h_b/1.5$ seems to be bigger than that reported for the original DGP. Moreover, all the considered top-down tests in Section 4 work well. Especially it should be noted that $\text{NSS}_{\mathcal{C}}$ for the original DGP was severely over-sized when $T = 150$, but now it has excellent size control and better finite-sample power. The performance of the $\mathcal{T}_0$ test seems to be similar to that of $\text{NSS}_{\mathcal{C}}$, but it displays a lower finite-sample power than $\text{NSS}_{\mathcal{C}}$. In general, $\mathcal{T}$ and $\text{NSS}_{\mathcal{K}}$ may be less preferred to either of the $\mathcal{T}_0$ test and $\text{NSS}_{\mathcal{C}}$. All these significant changes in the finite-sample performances of the top-down tests suggest that complexity of the attractor space is a crucial factor. Meanwhile, the CKP test has a reasonable size control when $\varphi \leq 2$ (resp. $\varphi \leq 3$) when $T \leq 350$ (resp. $T = 750$). The $\mathcal{T}_{\text{CKP}}$ test does not exhibit a severe over-size unless $\theta = 0.7$ and $\varphi = 5$.

We next modify the simulation DGP in Section 4 by setting $\{\zeta_j\}_{j=1}^{21}$ to the first 21 Legendre polynomials. The functional observations are constructed by smoothing $\{X_t\}_{t=1}^{T}$ observed at 200 regularly spaced points of $[0, 1]$ using the first 31 Fourier basis functions. This setting is very similar to that considered in Nielsen et al. (2019), but there is difference in the employed permutation scheme. They permute the first 8 Legendre polynomials and the remainders respectively, and combine those in order. As a result, the first 8 elements $\{\zeta_{[1]}, \zeta_{[2]}, \ldots, \zeta_{[8]}\}$ are always pick from the first 8 Legendre polynomials. Clearly this is a special case of our permutation scheme described in Section 4. Tables 8 and 9 report our simulation results. There is a noticeable difference in finite-sample performance of $\mathrm{NSS}_{\mathcal{K}}$ relative to those of the other tests: the test was relatively conservative in our simulation results from the DGP considered in Section 4, but now it rejects the correct null hypothesis more frequently than the $\mathcal{T}$ test or the $\mathcal{T}_0$ test does.

Lastly, we fix the attractor space to the span of the first φ Legendre polynomials, and investigate the finite-sample performances of the tests, including the CKP test and the $\mathcal{T}_{\text{CKP}}$ test. Tables 10 and 11 summarize our simulation results. For the $\mathcal{B}$ test, the reported over-rejection caused by choosing $h = h_b/1.5$ appears to be bigger compared to the case when $\mathfrak{A}$ is randomly determined, which is similarly observed in Table 6. Note that all the top-down tests have good size control; even if the CKP test seems to be over-sized when $T = 150$ and $\varphi \geq 4$, such a size distortion becomes disappear as T gets larger.

D Log-density estimation

In Section 5.2, each log wage density $\log X_t$ at time t is obtained from the following procedure.

Given survey responses $x_1, \ldots, x_n$ with design weights $w_1, \ldots, w_n$ such that $\sum_{i=1} w_i = n$, we consider the weighted log-likelihood

$$l(X_t) = \sum_{i=1}^{n} w_i \log(X_t(x_i)) - n \left(\int X_t(u) du - 1 \right).$$

Let K be the support of X_t. Under some local smoothness assumptions, we can consider a localized version of the log-likelihood and $\log X_t(u)$ can be locally approximated by a polynomial function, as follows.

$$l_p(X_t)(x) = \sum_{i=1}^{n} w_i \,\mathrm{k}\left(\frac{x_i - x}{h} \right) \mathcal{Q}(x_i - x; \alpha_t) - n \int \mathrm{k}\left(\frac{u - x}{h} \right) \exp(\mathcal{Q}(u - x; \alpha_t)) du \quad \text{(D.1)}$$

where $\mathrm{k}(\cdot)$ is a suitable kernel function, h is a bandwidth which assumed to be fixed, and $\mathcal{Q}(u; \alpha_t)$ is polynomial in u with coefficients α_t.

We set $\mathcal{W}(u) = \frac{70}{81}(1 - |u|^3)^3$ as commonly employed and suggested in Loader (2006), and $\mathcal{Q}(u; \alpha) = \alpha_{0,t} + \alpha_{1,t} u + \alpha_{2,t} u^2$. For fixed $x \in K$, let $(\hat{\alpha}_{0,t}, \hat{\alpha}_{1,t}, \hat{\alpha}_{2,t})$ be the maximizer of (D.1). Then the local likelihood log-density estimate is given by

$$\widehat{\log X}_t(x) = \hat{\alpha}_{0,t}$$

The procedure is repeated for a fine grid of points, and then $\widehat{\log X_t}$ may be obtained from an interpolation method described in (Loader, 2006, Chapter 12). Each log wage density is estimated on $[3.5, 31.5]$ including all the historical observations, and $h = 4.5$ is employed for all t.

Table 6: Simulation results for the bottom-up test with $\kappa(\varphi_0) = 2$ (fixed attractor)

(a) $\theta = 0.1$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	4.5	2.2	3.1	1.9	1.7	1.4
350	4.6	3.6	4.2	3.4	3.4	2.9
750	4.9	4.2	5.0	4.6	4.6	4.6
			power, $q = 1$			
150	87.8	63.5	49.0	39.8	28.7	20.6
350	94.5	83.8	78.0	74.7	68.3	60.3
750	98.4	94.7	94.1	93.7	91.2	89.3
			power, $q = 2$			
150	97.0	95.2	88.8	71.9	50.1	30.7
350	99.9	99.4	99.2	99.3	97.9	92.8
750	100	100	100	100	100	100

(b) $\theta = 0.7$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	8.2	4.4	5.7	4.6	4.0	3.7
350	6.1	4.9	6.3	7.0	6.1	6.4
750	5.7	5.2	6.1	6.3	6.6	6.4
			power, $q = 1$			
150	90.4	63.6	51.4	43.2	31.5	23.2
350	95.8	84.8	78.8	76.2	68.6	62.0
750	98.5	94.1	93.6	93.1	91.6	90.0
			power, $q = 2$			
150	96.5	93.8	87.0	69.3	52.8	34.6
350	99.7	99.4	99.4	97.9	95.7	89.4
750	100	100	100	100	100	99.6

(c) $\theta = 0.1$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	4.9	3.6	3.4	3.4	3.0	2.5
350	4.9	4.4	5.0	4.8	4.2	3.9
750	5.0	4.2	5.0	5.4	4.7	5.2
			power, $q = 1$			
150	94.9	82.8	75.6	70.4	61.3	54.0
350	98.4	94.4	93.3	92.3	90.8	89.2
750	99.4	98.6	98.8	99.0	98.6	98.7
			power, $q = 2$			
150	99.0	99.0	98.4	92.5	82.9	68.6
350	100	100	100	100	99.8	99.2
750	100	100	100	100	100	100

(d) $\theta = 0.7$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	10.5	7.4	9.7	11.7	12.3	12.4
350	8.6	7.5	10.8	11.1	12.1	13.8
750	6.6	6.0	7.8	8.7	10.4	11.4
			power, $q = 1$			
150	95.2	83.5	77.2	71.8	65.8	60.4
350	98.5	94.5	93.2	93.4	91.6	90.6
750	99.3	98.5	98.9	98.9	99.2	99.1
			power, $q = 2$			
150	98.3	98.9	97.1	91.2	83.9	71.8
350	99.9	100	100	99.9	99.6	98.6
750	100	100	100	100	100	100

(e) $\theta = 0.1$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	4.1	1.7	1.2	1.0	0.5	0.2
350	4.4	2.9	2.8	2.3	1.7	2.3
750	3.9	3.6	4.3	4.5	3.6	3.9
			power, $q = 1$			
150	79.0	39.4	22.6	14.4	6.3	3.6
350	90.6	65.7	54.1	45.0	34.6	25.8
750	95.8	84.8	80.3	76.2	69.4	63.3
			power, $q = 2$			
150	93.7	80.4	62.9	36.6	17.5	7.4
350	98.9	96.0	94.0	91.0	82.1	67.7
750	100	99.7	99.5	99.4	98.9	98.6

(f) $\theta = 0.7$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.4	2.0	1.8	1.7	1.1	0.8
350	5.3	3.6	4.3	3.2	2.8	2.3
750	5.5	4.0	4.8	4.8	4.2	3.9
			power, $q = 1$			
150	79.4	42.1	24.2	14.8	8.2	5.2
350	90.7	66.8	53.3	44.2	35.0	26.3
750	95.7	84.3	80.8	75.8	70.4	63.0
			power, $q = 2$			
150	93.1	79.5	63.1	34.6	16.8	7.6
350	98.8	95.8	93.4	87.9	78.0	60.6
750	99.9	99.5	99.4	99.3	98.6	96.8

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 + q$.

Table 7: Simulation results for top-down tests (fixed attractor)

$\theta = 0.1$, size						
T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	5.1	3.2	2.6	2.6	2.2
	$\mathcal{T}_0$	4.8	3.4	3.4	3.7	3.6
	$\text{NSS}_{\mathcal{K}}$	2.3	1.0	0.8	0.5	0.4
	$\text{NSS}_{\mathcal{C}}$	5.2	4.0	3.5	3.7	3.5
	$\mathcal{T}_{\text{S}}$	3.1	1.1	0.4	0.1	0.0
	CKP	5.0	8.3	26.4	2.9	0.1
	$\mathcal{T}_{\text{CKP}}$	4.2	3.6	5.1	2.4	0.6
350	$\mathcal{T}$	4.2	4.4	4.6	3.2	3.2
	$\mathcal{T}_0$	5.2	5.0	4.4	4.5	4.8
	$\text{NSS}_{\mathcal{K}}$	4.2	3.8	3.2	3.2	3.1
	$\text{NSS}_{\mathcal{C}}$	4.5	5.1	4.5	4.4	4.0
	$\mathcal{T}_{\text{S}}$	3.7	2.5	2.0	1.3	0.7
	CKP	4.3	4.9	30.4	66.4	74.9
	$\mathcal{T}_{\text{CKP}}$	4.6	4.4	5.1	6.3	11.8
750	$\mathcal{T}$	5.4	4.1	4.3	3.7	4.0
	$\mathcal{T}_0$	5.0	4.7	4.6	4.9	4.5
	$\text{NSS}_{\mathcal{K}}$	5.0	4.8	4.2	4.3	4.0
	$\text{NSS}_{\mathcal{C}}$	5.2	4.8	4.8	5.0	4.6
	$\mathcal{T}_{\text{S}}$	4.5	4.4	3.2	2.8	2.0
	CKP	4.6	5.0	7.4	19.1	73.9
	$\mathcal{T}_{\text{CKP}}$	5.1	4.7	5.2	5.6	7.7

$\theta = 0.7$, size						
T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	4.3	2.5	2.1	1.8	0.8
	$\mathcal{T}_0$	4.1	3.1	2.8	2.2	2.4
	$\text{NSS}_{\mathcal{K}}$	1.6	0.6	0.4	0.0	0.0
	$\text{NSS}_{\mathcal{C}}$	4.7	4.2	3.1	2.3	3.0
	$\mathcal{T}_{\text{S}}$	3.3	1.2	0.4	0.1	0.0
	CKP	5.3	11.9	42.5	15.5	4.3
	$\mathcal{T}_{\text{CKP}}$	4.4	5.8	11.8	9.5	3.7
350	$\mathcal{T}$	4.1	3.6	3.3	3.0	2.4
	$\mathcal{T}_0$	5.3	4.0	4.6	4.4	3.6
	$\text{NSS}_{\mathcal{K}}$	4.2	2.6	2.5	1.8	1.8
	$\text{NSS}_{\mathcal{C}}$	4.9	4.0	4.5	4.0	3.8
	$\mathcal{T}_{\text{S}}$	4.1	2.6	1.8	1.3	0.7
	CKP	4.2	6.5	35.9	72.7	89.3
	$\mathcal{T}_{\text{CKP}}$	4.6	5.2	9.2	14.2	26.4
750	$\mathcal{T}$	5.0	3.6	4.2	4.3	3.2
	$\mathcal{T}_0$	5.0	4.5	4.7	4.6	4.3
	$\text{NSS}_{\mathcal{K}}$	4.4	4.7	4.0	4.0	3.9
	$\text{NSS}_{\mathcal{C}}$	5.4	5.4	4.2	4.4	4.7
	$\mathcal{T}_{\text{S}}$	5.4	3.0	2.6	2.6	1.9
	CKP	4.5	4.9	10.4	25.7	79.9
	$\mathcal{T}_{\text{CKP}}$	4.4	4.9	6.6	8.5	15.6

$\theta = 0.1$, power						
T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	98.8	93.8	92.2	91.7	87.7
	$\mathcal{T}_0$	99.8	97.3	97.4	97.0	96.7
	$\text{NSS}_{\mathcal{K}}$	73.6	42.2	26.4	19.2	11.0
	$\text{NSS}_{\mathcal{C}}$	99.8	97.5	97.4	98.1	97.2
	$\mathcal{T}_{\text{S}}$	100	52.7	15.4	2.9	0.2
	CKP	100	94.1	42.3	3.2	0.2
	$\mathcal{T}_{\text{CKP}}$	100	92.2	49.7	6.4	0.2
350	$\mathcal{T}$	100	99.4	99.4	99.4	99.7
	$\mathcal{T}_0$	100	100	99.9	99.9	100
	$\text{NSS}_{\mathcal{K}}$	99.9	90.5	91.6	90.4	88.8
	$\text{NSS}_{\mathcal{C}}$	100	100	99.9	100	100
	$\mathcal{T}_{\text{S}}$	100	100	99.2	71.2	32.1
	CKP	100	100	100	98.9	78.1
	$\mathcal{T}_{\text{CKP}}$	100	100	100	99.2	78.7
750	$\mathcal{T}$	100	100	100	100	100
	$\mathcal{T}_0$	100	100	100	100	100
	$\text{NSS}_{\mathcal{K}}$	100	99.6	99.0	99.7	99.8
	$\text{NSS}_{\mathcal{C}}$	100	100	100	100	100
	$\mathcal{T}_{\text{S}}$	100	100	100	100	99.8
	CKP	100	100	100	100	100
	$\mathcal{T}_{\text{CKP}}$	100	100	100	100	100

$\theta = 0.7$, power						
T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	88.8	76.6	64.4	54.1	38.8
	$\mathcal{T}_0$	93.0	83.5	74.0	65.7	53.4
	$\text{NSS}_{\mathcal{K}}$	46.3	17.7	7.7	3.0	1.3
	$\text{NSS}_{\mathcal{C}}$	94.8	87.7	84.8	78.1	73.9
	$\mathcal{T}_{\text{S}}$	100	74.4	27.9	7.6	1.5
	CKP	100	96.8	59.2	16.6	3.6
	$\mathcal{T}_{\text{CKP}}$	100	96.3	67.1	20.0	4.0
350	$\mathcal{T}$	98.7	95.4	93.5	91.4	87.0
	$\mathcal{T}_0$	99.3	97.6	97.0	96.0	94.3
	$\text{NSS}_{\mathcal{K}}$	95.2	76.2	69.5	65.6	56.0
	$\text{NSS}_{\mathcal{C}}$	99.7	99.0	98.2	97.6	97.4
	$\mathcal{T}_{\text{S}}$	100	100	99.9	88.6	59.1
	CKP	100	100	100	99.8	90.7
	$\mathcal{T}_{\text{CKP}}$	100	100	100	99.6	89.6
750	$\mathcal{T}$	100	99.6	99.3	99.2	99.2
	$\mathcal{T}_0$	100	99.7	99.6	99.7	99.5
	$\text{NSS}_{\mathcal{K}}$	99.9	97.0	97.0	97.0	97.4
	$\text{NSS}_{\mathcal{C}}$	100	99.8	99.9	99.8	99.9
	$\mathcal{T}_{\text{S}}$	100	100	100	100	100
	CKP	100	100	100	100	100
	$\mathcal{T}_{\text{CKP}}$	100	100	100	100	100

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 - 1$.

Table 8: Simulation results for $B_{\varphi_0,\kappa(\varphi_0)}$ with $\kappa(\varphi_0) = 2$ (Polynomial)

(a) $\theta = 0.1$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	4.6	2.9	2.6	2.1	1.5	1.3
350	4.0	4.2	3.0	3.7	2.8	2.6
750	4.2	4.8	3.6	4.4	4.1	3.8
			power, $q = 1$			
150	89.2	62.4	43.1	31.4	19.6	11.3
350	95.2	83.8	78.2	74.2	65.4	55.4
750	98.3	94.8	94.3	92.8	91.1	88.2
			power, $q = 2$			
150	98.3	86.7	74.2	57.8	44.1	32.5
350	99.9	99.2	97.9	94.6	90.8	85.2
750	100	100	99.9	100	99.4	98.9

(b) $\theta = 0.7$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	8.1	5.0	4.5	4.0	2.8	2.7
350	6.6	5.3	5.6	5.0	6.0	4.2
750	6.2	5.9	5.5	5.1	5.9	5.5
			power, $q = 1$			
150	88.5	62.2	45.2	33.6	22.8	15.4
350	95.4	84.5	79.1	72.5	65.2	57.2
750	98.1	94.3	94.7	93.4	91.9	89.2
			power, $q = 2$			
150	98.6	85.0	71.4	59.4	44.1	33.8
350	99.9	98.4	96.9	94.5	89.4	84.2
750	100	99.9	99.7	99.8	99.2	98.8

(c) $\theta = 0.1$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.0	4.4	3.5	3.5	2.7	2.4
350	4.8	4.7	4.6	4.5	4.1	4.0
750	5.5	5.2	4.8	4.7	4.9	4.4
			power, $q = 1$			
150	94.6	78.0	67.0	54.2	42.9	32.0
350	98.2	93.5	93.4	91.6	87.7	82.2
750	99.3	98.9	98.1	98.9	98.6	97.9
			power, $q = 2$			
150	99.6	94.0	89.2	82.6	73.1	64.5
350	100	99.9	99.5	99.1	98.1	96.5
750	100	100	100	100	100	99.9

(d) $\theta = 0.7$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	11.4	9.2	9.5	8.2	8.4	9.0
350	8.6	7.9	8.0	9.1	9.9	10.4
750	7.1	7.2	7.2	8.2	8.3	8.8
			power, $q = 1$			
150	94.0	79.1	67.0	59.2	46.2	38.8
350	98.5	94.0	93.0	91.6	87.3	81.4
750	99.6	98.7	98.9	98.8	98.8	97.7
			power, $q = 2$			
150	99.5	94.3	89.4	83.2	74.4	67.4
350	100	99.6	98.8	98.5	97.7	96.7
750	100	99.9	100	100	100	100

(e) $\theta = 0.1$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	3.5	2.0	1.2	0.6	0.5	0.5
350	4.4	3.6	2.7	2.1	2.0	1.6
750	4.6	4.6	3.8	3.0	2.9	3.3
			power, $q = 1$			
150	80.7	39.0	19.5	9.9	5.4	3.3
350	90.1	67.0	53.4	45.2	32.0	23.8
750	95.6	85.1	81.0	76.2	68.1	61.1
			power, $q = 2$			
150	96.5	68.1	46.4	28.2	15.2	8.7
350	99.6	96.1	90.1	81.2	70.2	58.8
750	99.9	99.8	99.6	99.0	97.1	94.6

(f) $\theta = 0.7$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.3	2.6	1.8	1.2	0.8	0.4
350	5.7	4.2	3.1	3.0	2.3	2.0
750	5.8	5.0	3.8	4.1	3.9	3.2
			power, $q = 1$			
150	79.4	41.4	21.0	10.5	5.9	3.2
350	90.1	67.4	54.6	44.4	33.0	24.3
750	95.4	85.4	79.4	75.3	69.8	62.4
			power, $q = 2$			
150	95.8	67.8	43.8	28.0	14.0	8.9
350	99.5	94.4	88.5	78.2	65.2	56.2
750	100	99.6	99.0	98.5	95.9	91.8

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 + q$.

Table 9: Simulation results for top-down tests (Polynomial)

$\theta = 0.1$, size

T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	4.7	4.5	5.3	6.1	7.6
	$\mathcal{T}_0$	4.8	5.4	6.8	8.7	11.4
	$\text{NSS}_{\mathcal{K}}$	4.9	5.0	6.0	10.5	14.9
	$\text{NSS}_{\mathcal{C}}$	7.5	14.2	23.6	31.0	37.1
	$\mathcal{T}_\text{S}$	4.0	1.9	1.5	1.3	1.8
350	$\mathcal{T}$	5.5	4.5	4.2	4.2	4.3
	$\mathcal{T}_0$	5.4	5.2	4.6	5.6	6.2
	$\text{NSS}_{\mathcal{K}}$	4.2	4.5	5.4	7.0	11.4
	$\text{NSS}_{\mathcal{C}}$	5.5	6.1	7.8	12.0	14.4
	$\mathcal{T}_\text{S}$	4.0	3.2	2.4	2.2	2.9
750	$\mathcal{T}$	4.5	4.0	4.3	4.3	3.8
	$\mathcal{T}_0$	5.9	5.4	4.6	4.8	5.2
	$\text{NSS}_{\mathcal{K}}$	5.2	4.2	5.2	6.0	7.9
	$\text{NSS}_{\mathcal{C}}$	5.7	5.0	5.4	6.3	7.2
	$\mathcal{T}_\text{S}$	4.7	4.2	3.7	3.7	3.7

$\theta = 0.7$, size

T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	3.8	4.0	4.4	4.8	5.3
	$\mathcal{T}_0$	4.8	4.4	5.9	8.0	8.4
	$\text{NSS}_{\mathcal{K}}$	4.0	4.3	4.9	6.9	10.2
	$\text{NSS}_{\mathcal{C}}$	7.8	14.0	20.5	25.0	28.1
	$\mathcal{T}_\text{S}$	4.5	3.0	1.8	1.8	1.6
350	$\mathcal{T}$	4.6	4.7	3.8	4.1	4.2
	$\mathcal{T}_0$	5.6	4.9	5.0	5.7	5.7
	$\text{NSS}_{\mathcal{K}}$	4.5	5.2	5.6	6.9	11.4
	$\text{NSS}_{\mathcal{C}}$	6.0	6.0	10.0	13.8	18.2
	$\mathcal{T}_\text{S}$	4.6	4.0	3.4	3.6	3.8
750	$\mathcal{T}$	4.8	4.5	4.0	4.5	4.0
	$\mathcal{T}_0$	5.1	5.1	4.6	4.8	4.9
	$\text{NSS}_{\mathcal{K}}$	5.2	5.0	5.4	6.2	8.5
	$\text{NSS}_{\mathcal{C}}$	5.1	4.7	5.2	6.8	8.4
	$\mathcal{T}_\text{S}$	4.8	3.8	4.3	3.9	4.2

$\theta = 0.1$, power

T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	98.7	93.4	92.2	91.6	88.7
	$\mathcal{T}_0$	99.9	97.6	97.4	97.5	97.0
	$\text{NSS}_{\mathcal{K}}$	99.1	90.6	92.9	93.5	93.4
	$\text{NSS}_{\mathcal{C}}$	99.7	96.8	97.7	98.0	97.4
	$\mathcal{T}_\text{S}$	99.9	50.8	24.2	20.2	21.8
350	$\mathcal{T}$	100	99.3	99.3	99.7	99.5
	$\mathcal{T}_0$	100	100	99.8	99.9	100
	$\text{NSS}_{\mathcal{K}}$	100	99.7	99.7	99.8	99.8
	$\text{NSS}_{\mathcal{C}}$	100	99.9	99.9	100	100
	$\mathcal{T}_\text{S}$	100	100	95.8	69.8	55.7
750	$\mathcal{T}$	100	100	99.9	100	100
	$\mathcal{T}_0$	100	100	100	100	100
	$\text{NSS}_{\mathcal{K}}$	100	100	100	100	100
	$\text{NSS}_{\mathcal{C}}$	100	100	100	100	100
	$\mathcal{T}_\text{S}$	100	100	100	100	99.0

$\theta = 0.7$, power

T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	84.0	77.2	69.8	62.6	54.7
	$\mathcal{T}_0$	88.8	84.0	79.8	75.2	69.0
	$\text{NSS}_{\mathcal{K}}$	80.9	72.6	70.2	68.2	63.2
	$\text{NSS}_{\mathcal{C}}$	94.1	88.0	85.2	83.8	79.8
	$\mathcal{T}_\text{S}$	100	72.7	36.0	24.4	18.0
350	$\mathcal{T}$	98.2	95.6	95.0	93.4	91.6
	$\mathcal{T}_0$	99.5	97.7	97.6	97.0	96.2
	$\text{NSS}_{\mathcal{K}}$	99.0	96.5	95.8	96.0	96.2
	$\text{NSS}_{\mathcal{C}}$	99.0	96.5	95.8	96.0	96.2
	$\mathcal{T}_\text{S}$	100	100	98.8	85.7	69.0
750	$\mathcal{T}$	99.9	99.4	99.5	99.6	99.6
	$\mathcal{T}_0$	100	99.9	99.8	99.9	99.9
	$\text{NSS}_{\mathcal{K}}$	100	99.6	99.6	99.8	99.9
	$\text{NSS}_{\mathcal{C}}$	100	100	99.8	100	99.9
	$\mathcal{T}_\text{S}$	100	100	100	100	99.8

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 - 1$.

Table 10: Simulation results for $B_{\varphi_0,\kappa(\varphi_0)}$ with $\kappa(\varphi_0) = 2$ (Polynomial, fixed attractor)

(a) $\theta = 0.1$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	4.0	4.2	2.9	2.4	1.6	1.2
350	5.6	5.2	4.6	3.8	3.3	2.5
750	4.8	6.0	5.8	4.2	3.6	4.0
			power, $q = 1$			
150	88.2	63.2	49.4	39.0	30.3	18.9
350	95.2	84.6	80.4	75.4	67.0	63.1
750	98.1	94.4	94.0	93.5	91.5	90.3
			power, $q = 2$			
150	98.4	94.7	91.7	86.1	72.1	59.8
350	99.8	99.8	99.4	99.4	99.1	98.0
750	100	100	100	100	100	100

(b) $\theta = 0.7$, baseline bandwidth

T	φ_0=0	1	2	3	4	5
			size			
150	7.5	7.0	6.0	5.5	5.0	3.8
350	6.9	6.9	6.4	6.5	6.8	6.4
750	5.3	6.4	6.2	6.4	6.1	6.0
			power, $q = 1$			
150	89.6	66.5	55.1	47.2	37.2	25.4
350	95.8	85.4	81.8	77.2	74.0	68.2
750	98.6	94.8	93.7	93.9	93.5	91.7
			power, $q = 2$			
150	97.8	95.5	90.9	86.4	69.9	58.1
350	99.6	99.5	99.6	99.4	98.6	97.8
750	100	100	100	100	100	100

(c) $\theta = 0.1$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	4.5	5.2	4.4	3.6	3.3	2.6
350	4.8	5.2	5.0	5.3	4.1	4.5
750	5.2	5.8	6.3	5.4	5.0	4.6
			power, $q = 1$			
150	95.4	80.8	74.6	68.0	61.2	48.6
350	98.2	93.6	92.6	92.3	91.2	89.5
750	99.5	98.3	98.7	98.7	98.5	98.6
			power, $q = 2$			
150	99.3	99.4	99.0	98.3	93.0	88.4
350	99.9	100	100	100	99.9	99.9
750	100	100	100	100	100	100

(d) $\theta = 0.7$, baseline bandwidth/1.5

T	φ_0=0	1	2	3	4	5
			size			
150	12.0	10.4	11.4	13.0	13.2	12.4
350	8.7	8.5	10.4	11.6	12.3	14.0
750	6.7	7.8	8.2	8.9	9.5	10.4
			power, $q = 1$			
150	95.5	85.3	80.5	77.1	70.3	63.5
350	98.3	94.4	94.8	94.9	93.4	92.0
750	99.4	98.3	98.8	99.2	99.3	99.2
			power, $q = 2$			
150	99.0	99.3	99.1	98.1	93.8	89.8
350	100	100	100	100	99.9	100
750	100	100	100	100	100	100

(e) $\theta = 0.1$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	4.2	2.9	1.6	1.0	0.4	0.4
350	4.1	4.0	3.6	2.8	2.0	1.7
750	4.5	5.1	4.1	4.2	3.2	3.0
			power, $q = 1$			
150	79.8	40.8	22.9	14.5	8.5	3.9
350	90.8	67.1	54.4	45.6	35.4	26.6
750	95.4	85.3	79.5	76.2	71.0	65.1
			power, $q = 2$			
150	95.7	81.0	64.7	49.2	31.0	18.0
350	99.3	96.3	94.4	90.8	85.8	78.2
750	99.9	99.7	99.5	99.5	99.0	99.0

(f) $\theta = 0.7$, baseline bandwidth×1.5

T	φ_0=0	1	2	3	4	5
			size			
150	5.8	4.8	2.6	1.8	1.6	1.0
350	5.3	4.7	4.2	3.6	2.3	2.5
750	5.2	5.5	4.8	4.5	4.7	4.0
			power, $q = 1$			
150	80.6	43.8	25.0	16.2	9.6	5.0
350	89.9	66.4	57.2	47.4	38.2	29.4
750	95.4	85.0	79.8	76.8	71.0	64.8
			power, $q = 2$			
150	94.8	81.0	63.7	49.3	30.6	18.3
350	99.3	96.5	93.8	91.9	84.1	76.7
750	100	99.4	99.5	99.4	99.0	98.5

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 + q$.

Table 11: Simulation results for top-down tests (Polynomial, fixed attractor)

		$\theta = 0.1$, size							$\theta = 0.7$, size				
T	Test	$\varphi_0 = 1$	2	3	4	5	T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	4.5	3.1	2.9	2.9	2.4	150	$\mathcal{T}$	3.4	2.7	2.0	1.4	1.0
	$\mathcal{T}_0$	4.8	3.9	4.2	4.7	4.4		$\mathcal{T}_0$	3.8	2.9	2.8	3.0	2.3
	$\mathrm{NSS}_{\mathcal{K}}$	4.2	4.0	3.3	4.0	3.3		$\mathrm{NSS}_{\mathcal{K}}$	3.3	2.8	1.8	2.1	2.2
	$\mathrm{NSS}_{\mathcal{C}}$	4.8	4.1	3.6	3.9	5.0		$\mathrm{NSS}_{\mathcal{C}}$	4.4	3.1	3.4	3.4	4.1
	$\mathcal{T}_{\mathrm{S}}$	2.8	2.1	1.3	1.6	1.2		$\mathcal{T}_{\mathrm{S}}$	2.4	1.2	1.6	1.2	0.7
	CKP	4.0	4.1	4.3	10.0	15.8		CKP	4.4	4.1	5.2	9.8	13.6
	$\mathcal{T}_{\mathrm{CKP}}$	3.8	3.4	3.2	3.2	2.4		$\mathcal{T}_{\mathrm{CKP}}$	4.6	4.2	3.4	3.6	2.4
350	$\mathcal{T}$	4.2	4.2	3.6	3.7	3.4	350	$\mathcal{T}$	4.3	4.0	3.0	3.7	2.8
	$\mathcal{T}_0$	5.1	4.9	4.5	5.0	5.0		$\mathcal{T}_0$	4.7	4.2	4.1	4.0	4.3
	$\mathrm{NSS}_{\mathcal{K}}$	5.0	4.6	4.2	5.8	4.8		$\mathrm{NSS}_{\mathcal{K}}$	4.8	3.8	3.4	4.9	4.2
	$\mathrm{NSS}_{\mathcal{C}}$	5.6	4.2	4.2	4.8	4.8		$\mathrm{NSS}_{\mathcal{C}}$	4.6	4.2	4.0	4.5	4.1
	$\mathcal{T}_{\mathrm{S}}$	3.5	2.8	2.2	3.0	1.8		$\mathcal{T}_{\mathrm{S}}$	2.8	2.2	2.3	2.9	2.4
	CKP	4.9	4.9	4.3	5.6	6.5		CKP	5.2	5.7	5.7	7.0	7.2
	$\mathcal{T}_{\mathrm{CKP}}$	4.6	4.6	4.6	5.2	4.3		$\mathcal{T}_{\mathrm{CKP}}$	4.9	4.7	4.4	6.1	4.4
750	$\mathcal{T}$	4.9	5.0	4.4	4.4	3.4	750	$\mathcal{T}$	5.1	5.0	3.7	4.0	3.6
	$\mathcal{T}_0$	5.4	5.1	4.3	5.2	4.3		$\mathcal{T}_0$	5.2	4.6	4.4	4.4	4.7
	$\mathrm{NSS}_{\mathcal{K}}$	4.6	5.0	5.0	4.3	4.6		$\mathrm{NSS}_{\mathcal{K}}$	4.9	4.7	4.2	5.3	4.2
	$\mathrm{NSS}_{\mathcal{C}}$	4.8	5.3	5.2	4.5	4.4		$\mathrm{NSS}_{\mathcal{C}}$	5.2	4.6	4.4	4.6	4.0
	$\mathcal{T}_{\mathrm{S}}$	3.5	3.5	2.5	4.2	2.4		$\mathcal{T}_{\mathrm{S}}$	2.9	3.0	2.6	3.8	2.7
	CKP	4.4	5.0	4.5	6.3	5.1		CKP	4.4	5.2	5.0	6.0	6.2
	$\mathcal{T}_{\mathrm{CKP}}$	4.6	4.8	4.7	5.4	4.6		$\mathcal{T}_{\mathrm{CKP}}$	5.4	5.0	5.0	6.2	5.6

		$\theta = 0.1$, power							$\theta = 0.7$, power				
T	Test	$\varphi_0 = 1$	2	3	4	5	T	Test	$\varphi_0 = 1$	2	3	4	5
150	$\mathcal{T}$	99.1	92.6	91.3	90.9	87.8	150	$\mathcal{T}$	89.6	75.2	67.0	54.8	41.9
	$\mathcal{T}_0$	99.7	96.6	96.7	96.8	97.2		$\mathcal{T}_0$	92.3	82.4	76.2	70.6	60.0
	$\mathrm{NSS}_{\mathcal{K}}$	99.2	88.6	88.9	88.0	89.0		$\mathrm{NSS}_{\mathcal{K}}$	87.0	67.6	62.8	55.5	49.9
	$\mathrm{NSS}_{\mathcal{C}}$	99.8	96.4	96.8	97.1	97.2		$\mathrm{NSS}_{\mathcal{C}}$	95.1	84.9	83.4	79.6	74.3
	$\mathcal{T}_{\mathrm{S}}$	99.9	96.9	86.4	74.8	59.5		$\mathcal{T}_{\mathrm{S}}$	100	96.9	84.3	65.9	45.0
	CKP	100	99.2	95.7	87.5	73.2		CKP	100	99.3	96.5	86.2	72.8
	$\mathcal{T}_{\mathrm{CKP}}$	100	99.4	96.0	87.0	74.2		$\mathcal{T}_{\mathrm{CKP}}$	100	99.7	96.0	85.0	67.7
350	$\mathcal{T}$	100	98.7	99.2	99.4	99.4	350	$\mathcal{T}$	98.7	94.8	93.5	93.0	90.2
	$\mathcal{T}_0$	100	99.8	99.9	100	100		$\mathcal{T}_0$	99.6	97.2	97.0	96.3	95.8
	$\mathrm{NSS}_{\mathcal{K}}$	100	99.5	99.4	99.7	99.9		$\mathrm{NSS}_{\mathcal{K}}$	99.2	95.8	94.7	95.2	94.1
	$\mathrm{NSS}_{\mathcal{C}}$	100	99.9	99.8	100	99.9		$\mathrm{NSS}_{\mathcal{C}}$	99.7	98.1	97.9	97.8	97.4
	$\mathcal{T}_{\mathrm{S}}$	100	100	100	99.6	98.7		$\mathcal{T}_{\mathrm{S}}$	100	100	100	99.8	99.2
	CKP	100	100	100	100	99.7		CKP	100	100	100	100	99.9
	$\mathcal{T}_{\mathrm{CKP}}$	100	100	100	100	99.6		$\mathcal{T}_{\mathrm{CKP}}$	100	100	100	100	99.9
750	$\mathcal{T}$	100	100	99.9	99.9	100	750	$\mathcal{T}$	100	99.4	99.0	99.5	99.2
	$\mathcal{T}_0$	100	100	100	100	100		$\mathcal{T}_0$	100	99.8	99.6	99.8	99.9
	$\mathrm{NSS}_{\mathcal{K}}$	100	100	100	100	100		$\mathrm{NSS}_{\mathcal{K}}$	100	99.6	99.5	99.9	99.8
	$\mathrm{NSS}_{\mathcal{C}}$	100	100	100	100	100		$\mathrm{NSS}_{\mathcal{C}}$	100	99.8	99.9	99.9	99.9
	$\mathcal{T}_{\mathrm{S}}$	100	100	100	100	100		$\mathcal{T}_{\mathrm{S}}$	100	100	100	100	100
	CKP	100	100	100	100	100		CKP	100	100	100	100	100
	$\mathcal{T}_{\mathrm{CKP}}$	100	100	100	100	100		$\mathcal{T}_{\mathrm{CKP}}$	100	100	100	100	100

Notes: (i) The results are based on 4,000 Monte Carlo replications. (ii) Power is calculated when the cointegration corank is $\varphi_0 - 1$.