

# Nonstationarity in Time Series of State Densities<sup>1</sup>

Yoosoon Chang

*Department of Economics  
Indiana University*

Chang Sik Kim

*Department of Economics  
Sungkyunkwan University*

Joon Y. Park

*Department of Economics  
Indiana University and Sungkyunkwan University*

## Abstract

This paper proposes a new framework to analyze the nonstationarity in the time series of state densities, representing either cross-sectional or intra-period distributions of some underlying economic variables. We regard each state density as a realization of Hilbertian random variable, and use a functional time series model to fit a given time series of state densities. This allows us to explore various sources of the nonstationarity of such time series. The potential unit roots are identified through functional principal component analysis, and subsequently tested by the generalized eigenvalues of leading components of normalized estimated variance operator. The asymptotic null distribution of the test statistic is obtained and tabulated. We use the methodology developed in the paper to investigate the state densities given by the cross-sectional distributions of individual earnings and the intra-month distributions of stock returns. We find some clear evidence for the presence of strong persistency in their time series.

*This version:* February 5, 2015

*JEL Classification:* C13, C32

*Key words and phrases:* time series of cross-sectional and intra-period distributions, state density, nonstationarity, unit root, functional principal component analysis.

---

<sup>1</sup>We thank a Co-editor and anonymous referees for helpful comments. We are also grateful to Junhui Qian for many useful discussions, and to Richard Blundell, John Chao, Wolfgang Härdle, Hide Ichimura, Oliver Linton, Enno Mammen, James Stock, and other participants at the fifth CIREQ time series conference at Montreal, Conference in Honor of Joel Horowitz at University College London, WIAS seminar at Humboldt University, Alexander von Humboldt Seminar at University of Mannheim, 2012 ASSA Meetings for helpful comments. Chang gratefully acknowledges the financial support from the NSF under Grant SES-0969146.

## 1. Introduction

In the paper, we investigate the nonstationarity of the time series of state densities, which are defined to represent either cross-sectional or intra-period distributions of some underlying economic variables. Examples of such distributions include, among many others, cross-sectional distributions of individual earnings and intra-period distributions of asset returns. Though not directly observable, the state densities may be easily estimated using cross-sectional or intra-period observations. In economic time series analysis, we routinely use the time series observations aggregated across cross-sections or averaged over some time periods. Clearly, we may extract more information on the dynamics of the underlying economic time series, if we study the entire cross-sectional or intra-period distributions rather than their representative values. Note that the observations aggregated cross-sectionally or averaged over time periods may simply be regarded as the means of the corresponding state distributions.

For our purpose, we introduce a novel framework. We regard each state density as a realization of Hilbertian random variable, and use a functional time series model to fit a given time series of state densities. Modeling state density as a functional random variable allows us to explore various sources of nonstationarity in state distribution including their higher moments such as variance, skewness and kurtosis. In contrast, the conventional unit root tests are applied to the cross-sectional aggregates and intra-period averages, and therefore, they only examine the nonstationarity existing in the means of state distributions. Our statistical theory of nonstationary functional time series in the paper builds upon the existing literature on the general statistical theory of stationary functional time series. See, e.g., Bosq (2000) for a detailed introduction to the subject. The reader is also referred to Park and Qian (2012) for the statistical theory of stationary regression with state densities defined similarly as in our paper.

Our testing procedure is simple and straightforward. The potential unit roots are identified through functional principal component analysis, and subsequently tested by the generalized eigenvalues of leading components of normalized estimated variance operator. The asymptotic distribution of our test statistic is free of nuisance parameters and dependent only upon the number of unit roots existing in the underlying time series of state densities. In parallel with the asymptotic distributions of the conventional unit root tests, it is represented by the eigenvalues of a matrix of functional of demeaned standard Brownian motion. The asymptotic critical values of the test are obtained and tabulated in the paper for the number of unit roots up to five. The finite sample performance of our test is evaluated by simulation. The test performs very well in terms of both finite sample size and power even for moderately large samples. The finite sample size becomes close to its nominal value and the finite sample power approaches to unity rather quickly as the sample size increases.

For the purpose of illustration, we use our model and methodology to analyze the nonstationarity in the time series of two different types of state distributions. In the first empirical application, we study the state distributions defined as the cross-sectional distributions of individual earnings, which are obtained from the cross-sectional observations of individual weekly earnings provided at monthly frequency by the Current Population Survey (CPS) data set. In the second empirical application, we investigate the state distributions defined

as the intra-month distributions of stock returns, which are obtained from the S&P 500 index returns at one-minute frequency. The sample sizes of their time series are 247 and 222 spanning the periods from January 1994 to July 2014 and from January 1992 to June 2010, respectively. Our test suggests that there are two unit roots in the cross-sectional distributions of individual earnings, whereas we have only one unit root in the intra-month distributions of stock returns.

The presence of unit roots in the time series of state distributions, of course, implies that the time series of their moments are generally nonstationary. In the paper, we propose a measure to represent the proportion of unit root component in each moment of state distributions. The measure, called the unit root proportion of a moment, takes value 0 if the moment is stationary, and 1 if the moment consists entirely of the unit root component of state distributions. For the cross-sectional distributions of individual earnings, the first four moments all have nonnegligible unit root proportions. In particular, the unit root proportions for the first two moments are substantial. It appears that the volatilities, as well as the means, of earnings distributions are quite persistent. On the other hand, for the intra-month distributions of stock returns, the moments have uneven unit root proportions. The unit root proportions of the odd moments are almost negligible, while those of the even moments are relatively much larger. They are, however, much smaller than those for the cross-sectional distributions of individual earnings.

The rest of the paper is organized as follows. Section 2 introduces the model and preliminary results necessary for the subsequent development of our theory and methodology. Our statistical procedure and asymptotic theory are presented in Section 3. To convey the main idea more effectively, here we assume that the state densities are directly observable. The effect of using estimated state densities on our statistical analysis is investigated and summarized in Section 4. There we show that the error incurred by the estimation of state densities are negligible and have no effect on our asymptotic theory under mild conditions. Section 5 includes two illustrative empirical applications of our model and methodology on the time series of cross-sectional distributions of individual earnings and intra-month distributions of stock returns. The fitted models obtained in Section 5 are used to perform our simulation study on the finite sample performance of our unit root test on the time series of state densities, which is reported in Section 6. Conclusion follows in Section 7. Finally, proofs are collected in Mathematical Appendix.

## 2. Model and Preliminaries

In the paper, we consider a sequence of density functions, which we denote by  $(f_t)$ . For each time  $t = 1, 2, \dots$ , we suppose that there is a distribution represented by probability density  $f_t$ , whose value at ordinate  $s \in \mathbb{R}$  is denoted by  $f_t(s)$ . Throughout the paper, we let

$$w_t = f_t - \mathbb{E}f_t$$

denote a centered density function and treat  $(w_t)$  as functional data taking values in Hilbert space  $H$ , where we define  $H$  to be the set of functions on a compact subset  $K$  of  $\mathbb{R}$  that

have vanishing integrals and are square integrable, i.e.,

$$H = \left\{ w \left| \int_K w(s) ds = 0, \int_K w^2(s) ds < \infty \right. \right\},$$

with inner product  $\langle v, w \rangle = \int v(s)w(s)ds$  for  $v, w \in H$ .<sup>2</sup> We define  $H = H_N \oplus H_S$ , where  $H_N$  and  $H_S$  denote respectively the unit root and stationarity subspaces of  $H$  and assume that, for any nonzero  $v \in H$ , the *coordinate process*

$$\langle v, w_t \rangle$$

has a unit root for all  $v \in H_N$ , while it is stationary for all  $v \in H_S$ . Moreover, we assume that  $H_N$  is  $M$ -dimensional with  $0 \leq M < \infty$ , and we set  $H_N = \emptyset$  and  $H_S = H$  by convention in case  $M = 0$ . Finally, we let  $\Pi_N$  and  $\Pi_S$  be the projections on  $H_N$  and  $H_S$ , respectively, and define

$$w_t^N = \Pi_N w_t \quad \text{and} \quad w_t^S = \Pi_S w_t.$$

Note that  $\Pi_N + \Pi_S = 1$ , so in particular we have  $w_t = w_t^N + w_t^S$ . Here and elsewhere in the paper, the identity operator on  $H$  is denoted simply by 1.

In what follows, we use  $\|\cdot\|$  to denote the usual Hilbert space norm, i.e.,  $\|v\|^2 = \langle v, v \rangle$  for a vector  $v \in H$  and  $\|A\| = \sup_v \|Av\|/\|v\|$  for any linear operator  $A$  on  $H$ . We follow the usual convention and denote by  $A'$  the adjoint of a linear operator  $A$  on  $H$ . If  $A = A'$ , the operator is called self-adjoint. Moreover, we write  $A > 0$  and say that  $A$  is positive definite if and only if  $\langle v, Av \rangle > 0$  for all nonzero  $v \in H$ . For any  $H$ -valued random element  $w$ , the expectation  $\mathbb{E}w$  of  $w$  is more formally defined as a vector in  $H$  satisfying

$$\langle v, \mathbb{E}w \rangle = \mathbb{E}\langle v, w \rangle$$

for all  $v \in H$ , and the variance  $\Sigma$  of  $w$  is given by an operator for which

$$\mathbb{E}\langle v_i, w - \mathbb{E}w \rangle \langle v_j, w - \mathbb{E}w \rangle = \langle v_i, \Sigma v_j \rangle$$

for all  $v_i, v_j \in H$ . The reader is referred to, e.g., Bosq (2000) for a brief introduction to Hilbertian random variables and the definitions of their moments.

**Assumption 2.1** For  $u_t = \Delta w_t$ , we let

$$u_t = \Phi(L)\varepsilon_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i},$$

where we assume that (a)  $\sum_{i=1}^{\infty} i\|\Phi_i\| < \infty$ , (b)  $\Pi_N \Phi(1)$  is of rank  $M$  and  $\Pi_S \Phi(1) = 0$ , and (c)  $(\varepsilon_t)$  is an iid sequence with mean zero and variance  $\Sigma > 0$ , for which we have  $\mathbb{E}\|\varepsilon_t\|^p < \infty$  with some  $p \geq 4$ .

---

<sup>2</sup>Since  $(f_t)$  is a sequence of random densities, we have  $\int f_t(s)ds = 1$  a.s. and  $\int_K \mathbb{E}f_t(s)ds = \mathbb{E} \int_K f_t(s)ds = 1$  for all  $t = 1, 2, \dots$ . Therefore, we require in our subsequent analysis that  $\int_K w(s)ds = 0$ . No other restrictions are necessary for  $(w_t)$ .

Aside from our rank conditions on  $\Phi(1)$ , the process  $(u_t)$  introduced in Assumption 2.1 was studied extensively in Bosq (2000) and we may simply view it as a generalization of the finite-dimensional linear process. The coefficients  $(\Phi_i)$  used to define the process are linear operators in  $H$  and the innovation  $(\varepsilon_t)$  is a sequence of random elements in  $H$ . The main results in the paper hold under milder conditions than those we assume here. For instance, we may allow  $(\varepsilon_t)$  to be martingale differences in place of iid sequences. The stronger conditions are invoked here to derive the explicit rates for some of our theoretical results.

For the process  $(u_t)$  introduced in Assumption 2.1, we may write

$$u_t = \Phi(1)\varepsilon_t + (\bar{u}_{t-1} - \bar{u}_t),$$

where

$$\bar{u}_t = \bar{\Phi}(L)\varepsilon_t = \sum_{i=0}^{\infty} \bar{\Phi}_i \varepsilon_{t-i}$$

with  $\bar{\Phi}_i = \sum_{j=i+1}^{\infty} \Phi_j$ . This representation is widely known as the Beveridge-Nelson decomposition for finite-dimensional linear processes, and studied thoroughly by Phillips and Solo (1992). Due to condition (a), we have  $\sum_{i=0}^{\infty} \|\bar{\Phi}_i\| < \infty$ , and therefore the process  $(\bar{u}_t)$  is well defined.

It follows from condition (b) that

$$w_t^N = \Pi_N w_t = \Pi_N \Phi(1) \sum_{i=1}^t \varepsilon_i - \Pi_N \bar{u}_t$$

and

$$w_t^S = \Pi_S w_t = -\Pi_S \bar{u}_t$$

ignoring the initial values  $w_0$  and  $\bar{u}_0$  that are unimportant in the development of our asymptotic theory. Therefore, it is clear that  $(w_t^N)$  is an integrated process, while  $(w_t^S)$  is a stationary process. Thus far, we have assumed that the dimension  $M$  of unit roots in  $(w_t)$  is known. Of course, the unit root dimension  $M$  is unknown in practical applications. In the next section, we will explain how to statistically determine  $M$ , as well as how to estimate the unit root and stationarity subspaces  $H_N$  and  $H_S$ .

### 3. Statistical Procedure and Asymptotic Theory

In this section, we introduce our statistical procedures and develop their asymptotic theories.

#### 3.1 Functional Principal Component Analysis

Our testing procedure is based on the functional principal component analysis for the unnormalized sample variance operator  $Q^T$  of  $(w_t)$  defined by

$$Q^T = \sum_{t=1}^T w_t \otimes w_t,$$

where  $T$  is the sample size. Moreover, we write

$$Q^T = T^2 Q_{NN}^T + T Q_{NS}^T + T Q_{SN}^T + T Q_{SS}^T, \quad (1)$$

where

$$\begin{aligned} Q_{NN}^T &= \frac{1}{T^2} \Pi_N \left( \sum_{t=1}^T w_t \otimes w_t \right) \Pi_N = \frac{1}{T^2} \sum_{t=1}^T w_t^N \otimes w_t^N \\ Q_{NS}^T &= \frac{1}{T} \Pi_N \left( \sum_{t=1}^T w_t \otimes w_t \right) \Pi_S = \frac{1}{T} \sum_{t=1}^T w_t^N \otimes w_t^S \\ Q_{SS}^T &= \frac{1}{T} \Pi_S \left( \sum_{t=1}^T w_t \otimes w_t \right) \Pi_S = \frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t^S, \end{aligned}$$

and  $Q_{SN}^T$  is the adjoint of  $Q_{NS}^T$ , i.e.,  $Q_{SN}^T = Q_{NS}^{T'}$ .

To establish our asymptotic results on  $Q^T$ , it is necessary to introduce some new concepts and notations. For a sequence  $(A_T)$  of operators on  $H$ , we let  $A_T \rightarrow_p A$  if  $\|A_T - A\| \rightarrow_p 0$ . Moreover, we define  $B$  to be Brownian motion on the unit root subspace  $H_N$  with variance operator  $\Omega$ , if  $B$  takes values on  $H_N$  and if for any  $v \in H_N$ ,  $\langle v, B \rangle$  is Brownian motion with variance  $\langle v, \Omega v \rangle$ . Naturally, for a random sequence  $(B_T)$  taking values on  $H_N$ , we let  $B_T \rightarrow_d B$ , if for any  $v \in H_N$ ,  $\langle v, B_T \rangle \rightarrow_d \langle v, B \rangle$ . It is straightforward to show that if  $B$  is Brownian motion on  $H_N$  then for any  $(v_i), i = 1, \dots, M$ , in  $H_N$

$$(\langle v_1, B \rangle, \dots, \langle v_M, B \rangle)'$$

becomes an  $M$ -dimensional vector Brownian motion with covariance matrix having the  $(i, j)$ -th entry given by  $\langle v_i, \Omega v_j \rangle$  for  $i, j = 1, \dots, M$ . Furthermore, if  $B_T \rightarrow_d B$  then for any  $(v_i), i = 1, \dots, M$ , in  $H_N$  we have

$$(\langle v_1, B_T \rangle, \dots, \langle v_M, B_T \rangle)' \rightarrow_d (\langle v_1, B \rangle, \dots, \langle v_M, B \rangle)',$$

as can be readily shown using the Cramer-Wold device. Finally, for a sequence of operators  $(C_T)$  on  $H_N$ , we let  $C_T \rightarrow_d C$  for an operator  $C$  on  $H_N$  if  $\langle v_1, C_T v_2 \rangle \rightarrow_d \langle v_1, C v_2 \rangle$  for any  $(v_1, v_2) \in H_N \times H_N$ .

**Lemma 3.1** Let Assumption 2.1 hold. We have

$$Q_{NN}^T \rightarrow_d Q_{NN} = \int_0^1 (W \otimes W)(r) dr, \quad (2)$$

where  $W$  is Brownian motion on  $H_N$  with variance operator  $\Pi_N \Phi(1) \Sigma \Phi(1)' \Pi_N$ . Also, it follows that

$$Q_{SS}^T \rightarrow_p Q_{SS} = \Pi_S \left( \sum_{i=0}^{\infty} \bar{\Phi}_i \Sigma \bar{\Phi}_i' \right) \Pi_S. \quad (3)$$

Moreover, we have

$$Q_{NS}^T, Q_{SN}^T = O_p(1) \quad (4)$$

for all large  $T$ .

Lemma 3.1 establishes the limits and stochastic orders for each of the components appearing in our decomposition (1) of the unnormalized sample variance operator  $Q^T$  of  $(w_t)$ . The normalized sample variance operators  $Q_{NN}^T$  and  $Q_{SS}^T$  have well defined limits, and converge to their limits in distribution and probability respectively on the unit root and stationarity subspaces  $H_N$  and  $H_S$ . Note in particular that  $Q_{NN}^T$  has as its distributional limit a random operator represented by a functional of Brownian motion  $W$  on  $H_N$ , whereas the probability limit of  $Q_{SS}^T$  is given by the operator

$$Q_{SS} = \Pi_S \left( \mathbb{E}(w_t \otimes w_t) \right) \Pi_S$$

on  $H_S$ . The sample covariance operators  $Q_{NS}^T$  and  $Q_{SN}^T$  become negligible asymptotically and do not appear in our subsequent asymptotic results.

Now we define

$$(\lambda_i(Q^T), v_i(Q^T)), \quad i = 1, \dots, T,$$

to be the pairs of the eigenvalues and eigenvectors of  $Q^T$ , where we order  $(\lambda_i(Q^T))$  so that  $\lambda_1(Q^T) \geq \dots \geq \lambda_T(Q^T)$ . Moreover, assuming  $T > M$ , we let

$$H_N^T = \bigvee_{i=1}^M v_i(Q^T),$$

where and elsewhere in the paper  $\bigvee$  signifies span, and denote by  $\Pi_N^T$  the projection on  $H_N^T$ . The subspace  $H_N^T$  spanned by the eigenvectors corresponding to  $M$  largest eigenvalues of  $Q^T$  will be referred to as the sample unit root subspace. Finally, we define the projection on the sample stationarity subspace by  $\Pi_S^T = 1 - \Pi_N^T$ , so that we have  $\Pi_N^T + \Pi_S^T = 1$  analogously as the relationship  $\Pi_N + \Pi_S = 1$ . The following proposition is an immediate consequence of Lemma 3.1.

**Proposition 3.2** Under Assumption 2.1, we have

$$\Pi_N^T = \Pi_N + O_p(T^{-1}) \quad \text{and} \quad \Pi_S^T = \Pi_S + O_p(T^{-1})$$

for all large  $T$ .

Proposition 3.2 implies that the projections  $\Pi_N$  and  $\Pi_S$  on the unit root and stationarity subspaces  $H_N$  and  $H_S$  can be estimated consistently at rate  $T$  respectively by the projections  $\Pi_N^T$  and  $\Pi_S^T$  on the sample unit root and stationarity subspaces  $H_N^T$  and  $H_S^T$ .

In what follows, we let

$$(\lambda_i(Q_{NN}), v_i(Q_{NN})), \quad i = 1, \dots, M,$$

be the nonzero eigenvalues and their associated eigenvectors of  $Q_{NN}$ , which we order

$$\lambda_1(Q_{NN}) \geq \cdots \geq \lambda_M(Q_{NN}).$$

Note that  $Q_{NN}$  is stochastic, and therefore, so are  $((\lambda_i(Q_{NN}), v_i(Q_{NN}))$  for  $i = 1, \dots, M$ . Clearly,  $(v_i(Q_{NN}))$ ,  $i = 1, \dots, M$ , span  $H_N$ . Though the set of vectors  $(v_i(Q_{NN}))$  are given randomly as functions of limit Brownian motion  $W$ , the space spanned by them is nonrandom and uniquely determined. On the other hand, we denote by

$$(\lambda_i(Q_{SS}), v_i(Q_{SS})), \quad i = 1, 2, \dots,$$

the nonzero eigenvalues and their associated eigenvectors of  $Q_{SS}$ , for which we assume

$$\lambda_1(Q_{SS}) \geq \lambda_2(Q_{SS}) \geq \cdots.$$

Since  $Q_{SS}$  is the variance operator of  $(\Pi_S w_t)$ , it is positive semi-definite and nuclear, i.e.,  $\lambda_i(Q_{SS}) \geq 0$  for all  $i$ , and

$$\sum_{i=1}^{\infty} \lambda_i(Q_{SS}) < \infty.$$

In particular,  $\lambda_i(Q_{SS}) \rightarrow 0$  as  $i \rightarrow \infty$ , and the origin is the limit point of the spectrum of  $Q_{SS}$ . The reader is referred to Bosq (2000, Theorem 1.7) for more details.

It can be deduced from Lemma 3.1 and Proposition 3.2 that

**Theorem 3.3** Under Assumption 2.1, we have

$$(T^{-2}\lambda_i(Q^T), v_i(Q^T)) \rightarrow_d (\lambda_i(Q_{NN}), v_i(Q_{NN}))$$

jointly for  $i = 1, \dots, M$ , and

$$(T^{-1}\lambda_{M+i}(Q^T), v_{M+i}(Q^T)) \rightarrow_p (\lambda_i(Q_{SS}), v_i(Q_{SS}))$$

for  $i = 1, 2, \dots$

In the stationarity subspace  $H_S$ , the eigenvectors and eigenvalues of the sample variance operator  $Q^T$  of  $(w_t)$ , if appropriately normalized, converge in probability to the corresponding population eigenvectors and eigenvalues. Following the convention made in Bosq (2000), the eigenvectors  $(v_i(Q_{SS}))$  are identified only up to the spaces spanned by them. For instance, we let  $v_i(Q_{SS})$  and  $-v_i(Q_{SS})$  be identical for  $i = 1, 2, \dots$ . Likewise, if  $\lambda_i(Q_{SS}) = \lambda_{i+j}(Q_{SS})$  for some  $i \geq 1$  and  $1 \leq j \leq J$ , then  $v_i(Q_{SS})$  and  $v_{i+j}(Q_{SS})$ ,  $j = 1, \dots, J$ , are not individually identified and they denote any vectors spanning the eigen-subspace of  $\lambda_i(Q_{SS})$ . In the unit root subspace  $H_N$ , on the other hand, the eigenvectors and normalized eigenvalues of the sample variance operator converge in distribution, and their distributional limits are given by the distributions of eigenvalues and eigenvectors of some functionals of Brownian motion.



### 3.2 Asymptotic Behavior of Coordinate Process

One of the most important implications of Proposition 3.2 is that we may regard  $(v_i(Q^T))$ ,  $i = 1, \dots, M$ , asymptotically as vectors in the unit root subspace  $H_N$ . Indeed, it is straightforward to deduce from Proposition 3.2 that

**Corollary 3.4** Under Assumption 2.1, we have

$$\begin{aligned} \max_{1 \leq t \leq T} |\langle \Pi_N^T v, w_t \rangle - \langle \Pi_N v, w_t \rangle| &= O_p(T^{-1/2}) \\ \max_{1 \leq t \leq T} |\langle \Pi_N^T v, \Delta w_t \rangle - \langle \Pi_N v, \Delta w_t \rangle| &= O_p(T^{-1+1/p}) \end{aligned}$$

for all large  $T$ , uniformly in all  $v \in H$  such that  $\|v\| = 1$ .

We may well expect from Corollary 3.4 that the coordinate processes of  $(w_t)$  defined by  $(v_i(Q^T))$  and  $(\Pi_N v_i(Q^T))$ ,  $i = 1, \dots, M$ , yield the same asymptotics. Note that  $v_i(Q^T) = \Pi_N^T v_i(Q^T)$ ,  $i = 1, \dots, M$ , since  $(v_i(Q^T))$ ,  $i = 1, \dots, M$ , are in the sample unit root subspace  $H_N^T$ . The difference between  $\langle v_i(Q^T), w_t \rangle$  and  $\langle \Pi_N v_i(Q^T), w_t \rangle$ ,  $i = 1, \dots, M$ , vanishes completely as the sample size increases up to infinity. Needless to say,  $(\Pi_N v_i(Q^T))$ ,  $i = 1, \dots, M$ , are vectors in the unit root subspace  $H_N$ , and we may therefore assume that  $(v_i(Q^T))$ ,  $i = 1, \dots, M$ , are in  $H_N$  in our subsequent development of the asymptotics for nonstationary coordinate processes.

It should, however, be noted that  $\langle v_i(Q^T), w_t \rangle$ ,  $i = 1, \dots, M$ , in general behave quite differently from unit root processes, even though we may regard  $(v_i(Q^T))$ ,  $i = 1, \dots, M$ , asymptotically as vectors in the unit root subspace  $H_N$ . As shown in Theorem 3.3,  $(v_i(Q^T))$ ,  $i = 1, \dots, M$ , do not converge to any fixed set of vectors and remain to be random in the limit. Therefore, their asymptotic behaviors are generally rather distinctive from those of the coordinate processes  $\langle v_i, w_t \rangle$  defined by a fixed set of vectors  $(v_i)$ ,  $i = 1, \dots, M$ , in  $H_N$ . For instance, we may easily deduce from Lemma 3.1 and Theorem 3.3 that

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \langle v_i(Q^T), w_t \rangle^2 &= \langle v_i(Q^T), Q_{NN}^T v_i(Q^T) \rangle + O_p(T^{-1}) \\ &\rightarrow_d \langle v_i(Q_{NN}), Q_{NN} v_i(Q_{NN}) \rangle = \lambda_i(Q_{NN}) \end{aligned}$$

with  $\lambda_i(Q_{NN})$  being the  $i$ -th largest eigenvalue of  $\int_0^1 (W \otimes W)(r) dr$ , whereas

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T \langle v_i, w_t \rangle^2 &= \langle v_i, Q_{NN}^T v_i \rangle + O_p(T^{-1}) \\ &\rightarrow_d \langle v_i, Q_{NN} v_i \rangle = \int_0^1 W_i^2(r) dr \end{aligned}$$

with  $W_i = \langle v_i, W \rangle$ . Unless  $M = 1$ , the two limit distributions are not identical.

We may now clearly see that the asymptotics of  $\langle v_i(Q^T), w_t \rangle$  are different in general from those of  $\langle v_i, w_t \rangle$  defined by a fixed set of vectors  $(v_i)$ ,  $i = 1, \dots, M$ , in the unit root subspace

$H_N$ , though their sample moments have the same order of magnitudes in probability. In particular,  $\langle v_i(Q^T), w_t \rangle$ ,  $i = 1, \dots, M$ , are not well defined unit root processes. To obtain unit root coordinate processes, we need to find a set of vectors in  $H_N^T$  that converge in probability to a fixed set of vectors in  $H_N$ , and use them to define coordinate processes. In fact, if we fix a set of vectors  $(\bar{v}_i)$ ,  $\bar{v}_i \notin H_S$ ,  $i = 1, \dots, M$ ,<sup>3</sup> and project them on  $H_N^T$  using  $\Pi_N^T$  to obtain  $(\Pi_N^T \bar{v}_i)$  and define

$$\langle \Pi_N^T \bar{v}_i, w_t \rangle, \quad i = 1, \dots, M,$$

then it follows immediately from Corollary 3.4 that

$$\max_{1 \leq t \leq T} |\langle \Pi_N^T \bar{v}_i, w_t \rangle - \langle \Pi_N \bar{v}_i, w_t \rangle| = O_p(T^{-1/2}) \quad (5)$$

$$\max_{1 \leq t \leq T} |\Delta \langle \Pi_N^T \bar{v}_i, w_t \rangle - \Delta \langle \Pi_N \bar{v}_i, w_t \rangle| = O_p(T^{-1+1/p}) \quad (6)$$

for all large  $T$ . Consequently,  $\langle \Pi_N^T \bar{v}_i, w_t \rangle$  behave asymptotically as  $\langle \Pi_N \bar{v}_i, w_t \rangle$ ,  $i = 1, \dots, M$ , which are well defined unit root processes.

As is well expected, we may test for the unit root hypothesis  $H_0 : \alpha_i = 1$  in the regression

$$\langle \Pi_N^T \bar{v}_i, w_t \rangle = \alpha_i \langle \Pi_N^T \bar{v}_i, w_{t-1} \rangle + \epsilon_{it}$$

or

$$\langle \Pi_N^T \bar{v}_i, w_t \rangle = \alpha_i \langle \Pi_N^T \bar{v}_i, w_{t-1} \rangle + \sum_{j=1}^q \beta_{ij} \Delta \langle \Pi_N^T \bar{v}_i, w_{t-j} \rangle + \epsilon_{it}$$

for  $i = 1, \dots, M$ , using the conventional unit root tests such as the augmented Dickey-Fuller (ADF) and Phillips' tests, the details of which we refer to Stock (1994). In fact, it follows directly from (5) and (6) that the unit root tests based on  $\langle \Pi_N^T \bar{v}_i, w_t \rangle$  have the same asymptotics as those of  $\langle \Pi_N \bar{v}_i, w_t \rangle$ ,  $i = 1, \dots, M$ , which are well defined unit root processes yielding the usual unit root asymptotics. The conventional unit root tests applied to  $\langle \Pi_N^T \bar{v}_i, w_t \rangle$ ,  $i = 1, \dots, M$ , are therefore valid asymptotically. However, the tests are not applicable, unless  $M$  is known. In particular, we may not use the tests to determine  $M$ . Besides, we expect the test results to be affected by the choice of vectors  $(\bar{v}_i)$ ,  $i = 1, \dots, M$ , which has to be arbitrary, except for the simple case  $M = 1$ .

### 3.3 Test Statistic and Limit Distribution

To determine the dimension  $M$  of the unit root subspace  $H_N$ , we consider the test of the null hypothesis

$$H_0 : \dim(H_N) = M \quad (7)$$

against the alternative hypothesis

$$H_1 : \dim(H_N) \leq M - 1 \quad (8)$$

---

<sup>3</sup>If we choose  $(\bar{v}_i)$  arbitrarily in  $H$ , the probability of any of  $(\bar{v}_i)$  being picked up from  $H_S$  is zero.

successively downward starting from  $M = M_{\max}$  with some fixed number  $M_{\max}$ .<sup>4</sup> By convention,  $M = 0$  implies that there is no unit root, and the unit root subspace  $H_N$  consists only of the origin. Our estimate for  $M$  is given by  $M_{\min} - 1$ , where  $M_{\min}$  is the smallest value of  $M$  for which the null hypothesis (7) is rejected in favor of the alternative hypothesis (8). Obviously, we may find the true value of  $M$  with asymptotic probability one, if we apply any consistent test in the successive manner as suggested here.

It is clear from Theorem 3.3 that we may use  $(\lambda_i(Q^T))$  to determine the unit root dimension in  $(w_t)$ . As can be readily deduced from Theorem 3.3,  $T^{-2}\lambda_i(Q^T)$  has a nondegenerate asymptotic distribution for  $i = 1, \dots, M$ , whereas it converges to zero in probability for all  $i \geq M + 1$ . Therefore, we may consider

$$\sigma_M^T = T^{-2}\lambda_M(Q^T) \quad (9)$$

to test the null hypothesis (7) against the alternative hypothesis (8) for  $M = 1, 2, \dots$ . It is clear that the test would be consistent, if we reject the null hypothesis in favor of the alternative hypothesis, when the test statistic  $\sigma_M^T$  in (9) takes a small value.<sup>5</sup> Unfortunately, however,  $\sigma_M^T$  has limit null distribution that is generally dependent upon various nuisance parameters, and cannot be used directly to test for the unit root dimension  $M$  without employing some resampling schemes to compute the critical values.<sup>6</sup>

To introduce a test statistic whose limit distribution is free of nuisance parameters, we first let  $(v_i)$ ,  $i = 1, \dots, M$ , be an *arbitrary* set of vectors generating  $H_N$ . Then we define

$$z_t = (\langle v_1, w_t \rangle, \dots, \langle v_M, w_t \rangle)' \quad (10)$$

for  $t = 1, \dots, T$ . As we will show later, the choice of  $(v_i)$ ,  $i = 1, \dots, M$ , is unimportant, and does not affect any of our subsequent developments. If we let  $Q_M^T = Z_T' Z_T$  with  $Z_T = (z_1, \dots, z_T)'$ , it follows from Lemma 3.1 that

$$T^{-2}Q_M^T \rightarrow_d Q_M = \int_0^1 W_M(r)W_M(r)'dr, \quad (11)$$

where  $W_M$  is an  $M$ -dimensional vector Brownian motion with variance  $\Omega_M$ , say, which is often referred to as the long run variance of  $(\Delta z_t)$ . The usual estimate for  $\Omega_M$  is defined as

$$\Omega_M^T = \sum_{|i| \leq \ell} \varpi_\ell(i) \Gamma_T(i),$$

where  $\varpi_\ell$  is a bounded weight function and  $\Gamma_T$  is the usual sample autocovariance function of  $(z_t)$ , i.e.,  $\Gamma_T(i) = T^{-1} \sum_t \Delta z_t \Delta z_{t-i}'$  for  $|i| \leq \ell$ , used to estimate the true autocovariance

---

<sup>4</sup>In practice, the choice of  $M_{\max}$  is not a difficult problem. It is often clearly suggested by the estimated spectrum of  $Q^T$ , which normally has several dominant eigenvalues. We may also rely on various methodologies of identifying dominant eigenvalues in the standard principal component analysis.

<sup>5</sup>As will be shown later, our approach yields a test that can be regarded as a generalization of the Sargan-Bhargava test or the class of unit root tests classified as the SB-class in Stock (1994).

<sup>6</sup>Of course, it may be possible to use the test statistic  $\sigma_M^T$  in (9) with critical values computed from some resampling schemes such as bootstrap and subsampling. However, this will not be further discussed in the paper.

function  $\Gamma$ . The lag truncation number  $\ell$  is set to increase as  $T$ .<sup>7</sup>

Now we define

$$(\lambda_i(Q_M^T, \Omega_M^T), v_i(Q_M^T, \Omega_M^T)), \quad i = 1, \dots, M,$$

to be the pairs of generalized eigenvalues and eigenvectors of  $Q_M^T$  with respect to the consistent estimate  $\Omega_M^T$  of  $\Omega_M$ . It follows immediately from (11) and continuous mapping theorem that

$$(T^{-2}\lambda_i(Q_M^T, \Omega_M^T), v_i(Q_M^T, \Omega_M^T)) \rightarrow_d (\lambda_i(Q_M, \Omega_M), v_i(Q_M, \Omega_M)) \quad (12)$$

jointly for  $i = 1, \dots, M$ , where

$$(\lambda_i(Q_M, \Omega_M), v_i(Q_M, \Omega_M)), \quad i = 1, \dots, M,$$

are the pairs of generalized eigenvalues and eigenvectors of  $Q_M$  with respect to  $\Omega_M$ . Moreover, if we let

$$(\lambda_i(Q_M^*), v_i(Q_M^*)), \quad i = 1, \dots, M,$$

be the eigenvalue and eigenvector pairs of

$$Q_M^* = \int_0^1 W_M^*(r)W_M^*(r)'dr, \quad (13)$$

where  $W_M^* = \Omega_M^{-1/2}W_M$  is the standard  $M$ -dimensional vector Brownian motion, i.e.,  $M$ -dimensional vector Brownian motion with identity covariance matrix, then it can be easily seen that

$$\lambda_i(Q_M, \Omega_M) = \lambda_i(Q_M^*), \quad \Omega_M^{1/2}v_i(Q_M, \Omega_M) = v_i(Q_M^*), \quad (14)$$

for  $i = 1, \dots, M$ . In particular, we have

$$T^{-2}\lambda_i(Q_M^T, \Omega_M^T) \rightarrow_d \lambda_i(Q_M^*) \quad (15)$$

jointly for  $i = 1, \dots, M$ , due to (12) and (14). The distributions of  $(\lambda_i(Q_M^*)), i = 1, \dots, M$ , are free of any nuisance parameters, and can be tabulated by simulations.

It is important to note that the pairs of generalized eigenvalues and eigenvectors

$$(\lambda_i(Q_M^T, \Omega_M^T), v_i(Q_M^T, \Omega_M^T)), \quad (\lambda_i(Q_M, \Omega_M), v_i(Q_M, \Omega_M))$$

for  $i = 1, \dots, M$  are determined uniquely regardless of the choice of  $(v_i)$  generating  $H_N$ , which we introduce in (10). To see this more clearly, we let  $(\bar{v}_i)$  be another set of vectors generating  $H_N$ . Moreover, we define  $\bar{z}_t = (\langle \bar{v}_1, w_t \rangle, \dots, \langle \bar{v}_M, w_t \rangle)'$  for  $t = 1, \dots, T$  and  $\bar{Q}_M^T = \bar{Z}_T' \bar{Z}_T$  with  $\bar{Z}_T = (\bar{z}_1, \dots, \bar{z}_T)'$ . Then we have a nonsingular  $M$ -dimensional matrix  $U_M$  such that  $\bar{z}_t = U_M z_t$ . However, we have  $\bar{Z}_T = Z_T U_M'$ , and

$$T^{-2}\bar{Q}_M^T \rightarrow_d \bar{Q}_M = \int_0^1 \bar{W}_M(r)\bar{W}_M(r)'dr,$$

---

<sup>7</sup>There exists a large literature on the consistent estimation of  $\Omega_M$ , for the introduction to which the reader is referred to, e.g., Andrews (1991). For all applications in the paper, we use the Parzen window with the Andrews' automatic bandwidth.

where  $\bar{W}_M = U_M W_M$  is  $M$ -dimensional vector Brownian motion with variance  $\bar{\Omega}_M = U_M \Omega_M U_M'$ . We may also easily see that  $\bar{\Omega}_M^T = U_M \Omega_M^T U_M'$  if we use the same estimator for  $\Omega_M$  and  $\bar{\Omega}_M$ . Consequently, it follows readily that the generalized eigenvalues and eigenvectors of  $\bar{Q}_M^T$  and  $\bar{Q}_M$  with respect to  $\bar{\Omega}_M^T$  and  $\bar{\Omega}_M$  are the same respectively as those of  $Q_M^T$  and  $Q_M$  with respect to  $\Omega_M^T$  and  $\Omega_M$ .

The generalized eigenvalues  $(\lambda_i(Q_M^T, \Omega_M^T))$ ,  $i = 1, \dots, M$ , are based on  $(z_t)$  in (10), which are not observable since it consists of coordinate processes given by a set of vectors  $(v_i)$  spanning  $H_N$ . Therefore, we consider

$$\tilde{z}_t = (\langle v_1(Q^T), w_t \rangle, \dots, \langle v_M(Q^T), w_t \rangle)' \quad (16)$$

for  $t = 1, \dots, T$ , in place of  $(z_t)$  in (10). Moreover, we define  $\tilde{Q}_M^T = \tilde{Z}'_T \tilde{Z}_T$  with  $\tilde{Z}_T = (\tilde{z}_1, \dots, \tilde{z}_T)'$ , and

$$\tilde{\Omega}_M^T = \sum_{|i| \leq \ell} \varpi_\ell(i) \tilde{\Gamma}_T(i),$$

accordingly as  $\Omega_M^T$ , where the autocovariance function  $\Gamma_T(i)$  of  $(z_t)$  in  $\Omega_M^T$  is replaced by that of  $(\tilde{z}_t)$ ,  $\tilde{\Gamma}_T(i) = T^{-1} \sum_t \Delta \tilde{z}_t \Delta \tilde{z}'_{t-i}$ .

To test the null hypothesis (7) against the alternative hypothesis (8), we propose to use

$$\tau_M^T = T^{-2} \lambda_M \left( \tilde{Q}_M^T, \tilde{\Omega}_M^T \right). \quad (17)$$

For the test statistic  $\tau_M^T$ , we have

**Theorem 3.5** Let Assumption 2.1 hold. Under the null hypothesis (7), we have

$$\tau_M^T \rightarrow_d \lambda_M(Q_M^*).$$

Moreover, we have

$$\tau_M^T \rightarrow_p 0$$

under the alternative hypothesis (8).

Theorem 3.5 shows that the limit null distribution of our test statistic  $\tau_M^T$  is given by the distribution of the smallest eigenvalue of the integrated product moment of  $M$ -dimensional standard vector Brownian motion on the unit interval, which was introduced earlier in (13). For  $M = 1$ , the limit null distribution of  $\tau_M^T$  reduces to that of the Sargan-Bhargava statistic. Our test may therefore be regarded as a generalization of the Sargan-Bhargava test. The limit distribution is easily derived from (15), once we show that the replacements of  $Q_M^T$  by  $\tilde{Q}_M^T$  and  $\Omega_M^T$  by  $\tilde{\Omega}_M^T$  have no effect on the limit distribution of the generalized eigenvalue  $\lambda_M(Q_M^T, \Omega_M^T)$ . This is well expected from Corollary 3.4, since  $(\tilde{z}_t)$  and  $(z_t)$  are uniformly close for  $t = 1, \dots, T$ , if we choose  $v_i = \Pi_N v_i(Q^T)$ ,  $i = 1, \dots, M$ , in defining  $(z_t)$ .<sup>8</sup> As discussed, the distribution of  $\lambda_M(Q_M^*)$  is free of nuisance parameters, and depends only

---

<sup>8</sup>Recall that our test statistic  $\tau_M^T$  is invariant with respect to the choice of  $(v_i)$ , and we may choose an arbitrary set of  $(v_i)$  as long as they span the unit root subspace  $H_N$ .

upon  $M$ . Therefore, the asymptotic critical values of our test statistic  $\tau_M^T$  can be tabulated for each  $M$ . Theorem 3.5 also establishes that the test based on  $\tau_M^T$  is consistent.<sup>9</sup>

It is possible to more specifically test the null hypothesis  $H_0$  in (7) against the alternative hypothesis  $H_A : \dim(H_N) = N$  for some  $N < M$ . In this case, it would clearly be more powerful to use the test statistic

$$\tau_{M,N}^T = T^{-2} \sum_{i=N+1}^M \lambda_i \left( \tilde{Q}_M^T, \tilde{\Omega}_M^T \right).$$

As in the proof of Theorem 3.5, we may easily show that

$$\tau_{M,N}^T \rightarrow_d \sum_{i=N+1}^M \lambda_i (Q_M^*),$$

under the null hypothesis  $H_0$ , and that  $\tau_{M,N}^T \rightarrow_p 0$  under the alternative hypothesis  $H_A$ . The limit distributions of  $\tau_{M,N}^T$  are also free of nuisance parameters and can be tabulated for various combinations of  $M$  and  $N$ .

### 3.4 Unit Root Moment Decomposition

Define

$$\iota_\kappa(s) = s^\kappa - \frac{1}{|K|} \int_K s^\kappa ds$$

for  $\kappa = 1, 2, \dots$ , where  $|K|$  denotes the Lebesgue measure of  $K$ . Note that

$$\langle \iota_\kappa, w_t \rangle = \int_K s^\kappa f_t(s) ds - \mathbb{E} \int_K s^\kappa f_t(s) ds$$

represents the random fluctuation of the  $\kappa$ -th moment of the distribution given by probability density ( $f_t$ ). We may decompose  $\iota_\kappa$  as  $\iota_\kappa = \Pi_N \iota_\kappa + \Pi_S \iota_\kappa$ , from which it follows that

$$\|\iota_\kappa\|^2 = \|\Pi_N \iota_\kappa\|^2 + \|\Pi_S \iota_\kappa\|^2 = \sum_{i=1}^M \langle \iota_\kappa, v_i \rangle^2 + \sum_{i=M+1}^{\infty} \langle \iota_\kappa, v_i \rangle^2,$$

where  $(v_i)$ ,  $i = 1, 2, \dots$ , is an orthonormal basis of  $H$  such that  $(v_i)_{1 \leq i \leq M}$  and  $(v_i)_{i \geq M+1}$  span  $H_N$  and  $H_S$ , respectively.

Clearly, we may employ the ratio

$$\pi_\kappa = \frac{\|\Pi_N \iota_\kappa\|}{\|\iota_\kappa\|} = \sqrt{\frac{\sum_{i=1}^M \langle \iota_\kappa, v_i \rangle^2}{\sum_{i=1}^{\infty} \langle \iota_\kappa, v_i \rangle^2}} \quad (18)$$

---

<sup>9</sup>The discriminatory power of  $\tau_M^T$  is expected to be lower than that of  $\sigma_M^T$ , since  $\sigma_M^T = O_p(T^{-1})$  under the alternative hypothesis (8). Certainly, this is the price we have to pay to make the statistic  $\tau_M^T$  free of nuisance parameters in limit distribution.

to measure the proportion of the component of  $\iota_\kappa$  lying in  $H_N$ . If, for instance,  $\iota_\kappa$  is entirely in  $H_N$ , we have  $\pi_\kappa = 1$ . On the other hand, we have  $\pi_\kappa = 0$  if  $\iota_\kappa$  is entirely in  $H_S$ . In the paper, we use  $\pi_\kappa$  to represent the proportion of unit root component in the  $\kappa$ -th moment of probability densities  $(f_t)$ . The  $\kappa$ -th moment of  $(f_t)$  has more dominant unit root component as  $\pi_\kappa$  tends to unity, whereas it becomes more stationary as  $\pi_\kappa$  approaches to zero. Of course, it is more difficult to predict the  $\kappa$ -th moment of  $(f_t)$  if  $\pi_\kappa$  is closer to unity. In contrast, the  $\kappa$ -th moment of  $(f_t)$  is easier to predict if  $\pi_\kappa$  is small. In what follows, we will refer to  $\pi_\kappa$  simply as the *unit root proportion* of the  $\kappa$ -th moment.

The unit root proportion  $\pi_\kappa$  of the  $\kappa$ -th moment defined in (18) is of course not directly applicable, since  $H_N$  and  $H_S$  are unknown. However, we may use its sample version

$$\pi_\kappa^T = \frac{\sqrt{\sum_{i=1}^M \langle \iota_\kappa, v_i(Q^T) \rangle^2}}{\sqrt{\sum_{i=1}^T \langle \iota_\kappa, v_i(Q^T) \rangle^2}}, \quad (19)$$

and other notations are as defined in (18). The sample version  $\pi_\kappa^T$  in (19) of  $\pi_\kappa$  in (18) will be referred to as the *sample* unit root proportion of the  $\kappa$ -th moment. We may readily show that the sample unit root proportion  $\pi_\kappa^T$  is a consistent estimator for the original unit root proportion  $\pi_\kappa$ . In fact, it follows immediately from Proposition 3.2 that

$$\pi_\kappa^T = \pi_\kappa + O_p(T^{-1})$$

for all  $\kappa = 1, 2, \dots$

## 4. Models with Estimated Densities

Usually, the state densities  $(f_t)$  are not directly observed and should therefore be estimated using the data, either cross-sectional or high frequency observations, that we assume to be available for each time period. Therefore, we denote by  $(\hat{f}_t)$  the estimated density functions and let

$$\hat{w}_t = \hat{f}_t - \frac{1}{T} \sum_{t=1}^T \hat{f}_t \quad (20)$$

be the demeaned density estimate for  $t = 1, \dots, T$ . It is well expected that the replacement of the original centered density  $(w_t)$  with the demeaned density estimate  $(\hat{w}_t)$  does not affect the asymptotic theory as long as the number of cross-sectional or high frequency observations available in each time period to estimate  $(\hat{f}_t)$  is large enough relative to the number  $T$  of time series observations. This was indeed shown by Park and Qian (2012) for a stationary functional regression model. In this section, we show that our asymptotic theories developed in the previous section continue to hold even when we use  $(\hat{w}_t)$  in the place of  $(w_t)$ . To develop subsequent asymptotic theories, we let  $\Delta_t = \hat{f}_t - f_t$  for  $t = 1, \dots, T$  and assume

**Assumption 4.1** Let (a)  $\sup_{t \geq 1} \|\Delta_t\| = O_p(1)$ , and (b)  $T^{-1} \sum_{t=1}^T \|\Delta_t\| \rightarrow_p 0$ .

The conditions introduced in Assumption 4.1 appear to be very mild and satisfied widely in many potential applications. Clearly, condition (a) would hold in general, since both  $(f_t)$  and  $(\hat{f}_t)$  are proper density functions. Moreover, condition (b) is expected to be met whenever the number  $N$  of observations that we use to estimate the state density at each period is sufficiently large relative to the sample size  $T$  for the time series of state density. In fact, if the standard second-order kernel is used with an optimal bandwidth to estimate the state densities, we would normally expect to have  $\sup_{1 \leq t \leq T} \mathbb{E}\|\Delta_t\| = O(a_T/b_N)$  for some  $(a_T)$  and  $(b_N)$  such that  $a_T \rightarrow \infty$  and  $b_N \rightarrow \infty$  respectively as  $T \rightarrow \infty$  and  $N \rightarrow \infty$ . It is well known that under very general regularity conditions we have  $\mathbb{E}\|\Delta_t\| = O(N^{-2/5})$  for each  $t$ , if the state distributions are defined as cross-sectional distributions, and independent and identically distributed observations are available to estimate them for each period. The same result holds if the state distributions are given by intra-period distributions, as long as within each period the underlying economic variables can be regarded as stationary processes satisfying some general mixing conditions.<sup>10</sup> This is shown in, e.g., Bosq (1998) and Hansen (2008).<sup>11</sup>

Now we redefine all sample statistics introduced in Section 3 to construct the feasible version of our unit root test statistic using  $(\hat{w}_t)$  introduced in (20), in place of  $(w_t)$ . As discussed, our testing procedure is based on the functional principal component analysis for the sample variance operator of  $(w_t)$ . Define

$$\hat{Q}_T = \sum_{t=1}^T \hat{w}_t \otimes \hat{w}_t,$$

which we may write similarly as in (1)

$$\hat{Q}_T = T^2 \hat{Q}_{NN}^T + T \hat{Q}_{NS}^T + T \hat{Q}_{SN}^T + T \hat{Q}_{SS}^T,$$

where  $\hat{Q}_{NN}^T$ ,  $\hat{Q}_{NS}^T$ ,  $\hat{Q}_{SN}^T$  and  $\hat{Q}_{SS}^T$  are defined in the same way as their counterpart components of  $Q^T$  given below (1) in Section 3, except that they are all sample product moments based on the demeaned density estimates  $(\hat{w}_t)$ . In what follows, we will refer to  $\hat{Q}^T$  as the unnormalized estimated variance operator of  $(w_t)$ .

In the subsequent development of our theory, we let

$$\left( \lambda_i(\hat{Q}_T), v_i(\hat{Q}_T) \right), \quad i = 1, \dots, T,$$

be the pairs of the eigenvalues and eigenvectors of  $\hat{Q}_T$ , where we order  $\lambda_1(\hat{Q}_T) \geq \dots \geq \lambda_T(\hat{Q}_T)$ . Moreover, for  $T > M$ , we use the eigenvectors  $v_i(\hat{Q}_T)$  corresponding to the  $M$  largest eigenvalues  $\lambda_i(\hat{Q}_T)$  of  $\hat{Q}_T$  to define

$$\hat{H}_N^T = \bigvee_{i=1}^M v_i(\hat{Q}_T).$$

---

<sup>10</sup>Such processes are often called locally stationary.

<sup>11</sup>The result holds under quite general conditions. For instance, Honda (2009) shows that we may even allow for processes with infinite variances.



We call  $\hat{H}_N^T$  the estimated unit root subspace and denote by  $\hat{\Pi}_N^T$  the projection on  $\hat{H}_N^T$ . Finally, we let  $\hat{\Pi}_S = 1 - \hat{\Pi}_N$ , so that we have  $\hat{\Pi}_N^T + \hat{\Pi}_S^T = 1$  analogously with the relationship  $\Pi_N + \Pi_S = 1$ .

**Lemma 4.1** Let Assumptions 2.1 and 4.1 hold. We have

$$\hat{Q}_{NN}^T \rightarrow_d \underline{Q}_{NN} = \int_0^1 (\underline{W} \otimes \underline{W})(r) dr$$

with  $\underline{W}(r) = W(r) - \int_0^1 W(s) ds$  and

$$\hat{Q}_{SS}^T \rightarrow_p Q_{SS},$$

where the limit Brownian motion  $W$  and operator  $Q_{SS}$  are defined in Lemma 3.1. Moreover, we have

$$\hat{Q}_{NS}^T, \hat{Q}_{SN}^T = O_p(1)$$

for all large  $T$ .

Lemma 4.1 presents the basic asymptotics for the estimated variance operator  $\hat{Q}^T$  of  $(w_t)$ . This corresponds to Lemma 3.1, where we establish the basic asymptotics for the sample variance operator  $Q^T$  of  $(w_t)$ . The estimated variance operator  $\hat{Q}^T$  differs from the sample variance operator  $Q^T$  in two aspects. First, the state density  $(f_t)$  used to define  $Q^T$  is replaced by the estimated state density  $(\hat{f}_t)$  for  $\hat{Q}^T$ . Second,  $\hat{Q}^T$  is defined with the sample mean  $T^{-1} \sum_{t=1}^T \hat{f}_t$  instead of the expectation of state density  $\mathbb{E}f_t$  used in the definition of  $Q^T$ . As can be clearly seen from the proof of Lemma 4.1, the replacement of  $(f_t)$  by  $(\hat{f}_t)$  becomes negligible and does not affect any of our asymptotic theory under our assumptions. This is true regardless of the stationarity/nonstationarity of the time series of state density. However, the use of the sample mean of the state density in place of its expectation has no asymptotic effect only for the stationary component of state density. For the nonstationary component, it yields different asymptotics. Note that the limit Brownian motion appeared in Lemma 3.1 is replaced by the demeaned Brownian motion in Lemma 4.1.

**Theorem 4.2** Let Assumptions 2.1 and 4.1 hold. We have

$$\left( T^{-2} \lambda_i(\hat{Q}_T), v_i(\hat{Q}_T) \right) \rightarrow_d \left( \lambda_i(\underline{Q}_{NN}), v_i(\underline{Q}_{NN}) \right)$$

jointly for  $i = 1, \dots, M$ , and

$$\left( T^{-1} \lambda_{M+i}(\hat{Q}_T), v_{M+i}(\hat{Q}_T) \right) \rightarrow_p \left( \lambda_i(Q_{SS}), v_i(Q_{SS}) \right)$$

for  $i = 1, 2, \dots$

The results in Theorem 4.2 are completely analogous to those in Theorem 3.3. The only difference in our asymptotic results in Theorem 4.2 is that we have  $\underline{Q}_{NN}$  defined with the

demeaned Brownian motion  $\underline{W}$  instead of  $Q_{NN}$  defined with the undemeaned Brownian motion  $W$ . As discussed, the appearance of  $\underline{W}$  is due to the use of the sample mean of state density in lieu of its expectation.

Based on the asymptotic results in Lemma 4.1 and Theorem 4.2, we may construct a feasible version of our statistic  $\tau_M^T$  defined in (17) to test for the existence of unit roots in the time series of state density. The feasible test statistic, which we introduce below, is defined using the demeaned state density estimate  $(\hat{w}_t)$  and the estimated variance operator  $\hat{Q}_T$  of  $(w_t)$ . Let

$$\hat{z}_t = \left( \langle v_1(\hat{Q}_T), \hat{w}_t \rangle, \dots, \langle v_M(\hat{Q}_T), \hat{w}_t \rangle \right)' \quad (21)$$

for  $t = 1, \dots, T$ , similarly as  $(\tilde{z}_t)$  in (16), and define  $\hat{Z}_T = (\hat{z}_1, \dots, \hat{z}_T)'$  and

$$\hat{Q}_M^T = \hat{Z}_T' \hat{Z}_T, \quad \hat{\Omega}_M^T = \sum_{|i| \leq \ell} \varpi_\ell(i) \hat{\Gamma}_T(i), \quad (22)$$

where  $\hat{\Gamma}_T(i)$  is the sample autocovariance function of  $(\hat{z}_t)$ , viz.,  $\hat{\Gamma}_T(i) = T^{-1} \sum_t \Delta \hat{z}_t \Delta \hat{z}_{t-i}'$ . Moreover, similarly as before, we define  $\lambda_M(\hat{Q}_M^T, \hat{\Omega}_M^T)$  to be the smallest generalized eigenvalue of  $\hat{Q}_M^T$  with respect to  $\hat{\Omega}_M^T$ . Our feasible version  $\hat{\tau}_M^T$  of the test statistic  $\tau_M^T$  is now given by

$$\hat{\tau}_M^T = T^{-2} \lambda_M(\hat{Q}_M^T, \hat{\Omega}_M^T), \quad (23)$$

which corresponds to  $\tau_M^T$  in (17).

To effectively present the asymptotics of the feasible test statistic  $\hat{\tau}_M^T$ , it is necessary to introduce some additional notations. In parallel with  $Q_M^*$  defined in (13), we define

$$\underline{Q}_M^* = \int_0^1 \underline{W}_M^*(r) \underline{W}_M^*(r)' dr \quad (24)$$

with  $\underline{W}_M^*(r) = W_M^*(r) - \int_0^1 W_M^*(s) ds$ , where  $W_M^*$  is the standard  $M$ -dimensional vector Brownian motion introduced earlier in Section 3. Now we have

**Theorem 4.3** Let Assumptions 2.1 and 4.1 hold. Under the null hypothesis (7), we have

$$\hat{\tau}_M^T \rightarrow_d \lambda_M(\underline{Q}_M^*).$$

Moreover, we have

$$\hat{\tau}_M^T \rightarrow_p 0$$

under the alternative hypothesis (8).

Theorem 4.3 shows that the limit distribution of the feasible test statistic  $\hat{\tau}_M^T$  is given by the distribution of the smallest eigenvalue of the integrated product moment of the demeaned  $M$ -dimensional standard vector Brownian motion on the unit interval, which is defined in (24). Given the limit distribution of  $\tau_M^T$  in Theorem 3.5, this is well predicted from our earlier results in Lemma 4.1 and Theorem 4.2. Theorem 4.3 also establishes that the test based on the feasible statistic  $\hat{\tau}_M^T$  is consistent as was the case for our original test.

The consistency is therefore unaffected by using estimated densities and demeaned density estimates of state distributions.

Clearly, the limit null distribution of the test statistic  $\hat{\tau}_M^T$  is free of nuisance parameters and can be obtained through simulation for each  $M$ . In Table 1, we tabulate the simulated critical values for the 1%, 5% and 10% tests based on  $\hat{\tau}_M^T$  for the values of  $M = 1, \dots, 5$ . The reported critical values are obtained by simulating the integrated product moment of the demeaned  $M$ -dimensional standard vector Brownian motion on the unit interval in (24). For the simulation, the Brownian motion is approximated by the standardized partial sum of mean zero i.i.d. normal random variates with sample size 10,000. The actual critical values are computed using 100,000 iterations.

Table 1. Critical Values of the Test Statistic  $\hat{\tau}_M^T$

$M$	1	2	3	4	5
1%	0.0248	0.0163	0.0123	0.0100	0.0084
5%	0.0365	0.0215	0.0156	0.0122	0.0101
10%	0.0459	0.0254	0.0177	0.0136	0.0111

As is well expected, the critical values of the tests based on the statistic  $\hat{\tau}_M^T$  decrease as  $M$  increases. Recall that the limit null distribution of  $\hat{\tau}_M^T$  is given by the smallest eigenvalue of the integrated product moments of the demeaned  $M$ -dimensional vector Brownian motion.

Once we obtain the estimated unit root subspace  $\hat{H}_N^T$ , which is generated by  $\hat{M}_T$ -eigenvectors given by  $v_i(\hat{Q}^T)$ ,  $i = 1, \dots, \hat{M}_T$ , the unit root proportion  $\pi_\kappa$  of the  $\kappa$ -th moment introduced in (18) can be consistently estimated by

$$\hat{\pi}_\kappa^T = \sqrt{\frac{\sum_{i=1}^{\hat{M}_T} \langle l_\kappa, v_i(\hat{Q}^T) \rangle^2}{T} \bigg/ \sum_{j=1}^{\hat{M}_T} \langle l_\kappa, v_j(\hat{Q}^T) \rangle^2}, \quad (25)$$

which will be referred to as the estimated unit root proportion of the  $\kappa$ -th moment. Under our assumptions, it can indeed be readily deduced from our earlier results that  $\hat{\pi}_\kappa^T = \pi_\kappa + o_p(1)$  for all  $\kappa = 1, 2, \dots$

## 5. Empirical Applications

To demonstrate the empirical relevancy of our model, we present two empirical applications. In these applications, we show how to define and estimate the state densities, and test for unit roots in the time series of state densities. As discussed, the state densities represent the intra-period or cross-sectional distributions and we use the intra-period or cross-sectional

observations to estimate them. For the actual estimation of the state densities, we use the Epanechnikov kernel with the optimal bandwidth given in, e.g., Li and Racine (2007).<sup>12</sup> To analyze the nonstationarity of the time series of state densities, we use our statistic developed in the paper to test the null hypothesis (7) against the alternative hypothesis (8). To determine the dimension of the unit root subspace, we apply the test successively downward starting from  $M = M_{\max}$  with  $M_{\max} = 5$ .<sup>13</sup> Once we determine the dimension of the unit root subspace and estimate the unit root subspace, we compute the unit root proportion of the first four moments. As discussed, the unit root proportion of a moment provides the proportion of nonstationary fluctuation in that moment of state distribution.

For the representation of functions in our Hilbert space as numerical vectors, we use a Daubechies wavelet basis. Because wavelets are spatially varying orthonormal bases with two parameters, i.e., location and resolutions, they provide more flexibilities in fitting the state densities in our applications, some of which have severe asymmetry and time-varying support. In fact, the wavelet basis in general yields a much better fit than the trigonometric basis. The Daubechies wavelet is implemented with 1037 basis functions. The data sets we use in our empirical applications have either a censoring problem or extreme outliers. To avoid their undesirable effects on the estimation of state density, we truncate some of the cross-sectional or intra-period observations at the boundaries.<sup>14</sup>

### 5.1 Cross-Sectional Distributions of Individual Earnings

For the first empirical application, we consider the time series of cross-sectional distributions of individual earnings. The cross-sectional observations of individual weekly earnings are obtained at monthly frequency from the Current Population Survey (CPS) data set. The individual weekly earnings are deflated by consumer price index with base year 2005. The data set provides 247 time series observations spanning from January 1994 to July 2014 at monthly frequency, and the number of cross-sectional observations in the data set for each month ranges from 12,180 in April 1996 to 15,826 in October 2001. For confidentiality reasons, individual earnings are topcoded above a certain level. During our sample period, the top code value was revised in 1998 up to \$2,885 from \$1,923.<sup>15</sup> In our empirical analysis, we drop all topcoded individual earnings as well as zero earnings as in Liu (2011) and Shin and Solon (2011).<sup>16</sup> The time series of the estimated densities for cross-sectional

---

<sup>12</sup>The optimal bandwidth for the Epanechnikov kernel, which minimizes the integrated mean squared error, is given by  $h^* = 2.3449 \sigma n^{-1/5}$ , where  $\sigma$  is standard error and  $n$  is sample size.

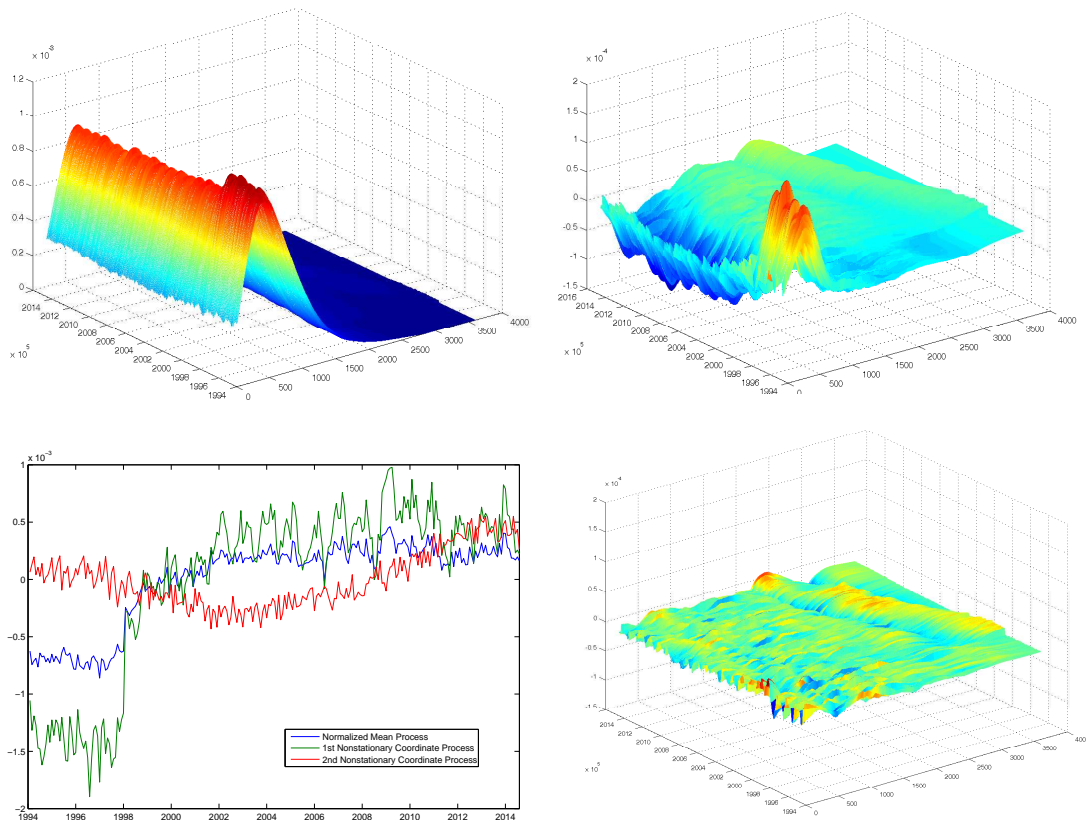
<sup>13</sup>In determining  $M_{\max}$ , we may use various rules that are commonly employed to identify dominant eigenvalues in the principal component analysis. For example, we can retain eigenvectors with eigenvalues greater than the average of all eigenvalues, or choose the eigenvectors needed to explain up to some fixed proportion of the variation in the data.

<sup>14</sup>Indeed, our empirical results are not very sensitive to the presence of censored observations and outliers as long as we use a wavelet basis. In particular, all our empirical results do not change qualitatively if all cross-sectional or intra-period observations are used without truncation. In contrast, they become somewhat sensitive to censored observations and outliers if we use trigonometric basis.

<sup>15</sup>Individual weekly earnings are top coded at \$999 during 1973-88, at \$1,923 during 1989-97, and at \$2,885 beginning in 1998.

<sup>16</sup>It is possible to impute topcoded individual earnings as a fixed multiple above the topcode threshold as in Lemieux (2006) and Autor et al. (2008).

Fig. 1. Distributional Time Series - Individual Earnings



distributions of individual earnings, undemeaned and demeaned, are shown respectively in the top-left and top-right panels of Fig. 1. We may clearly see that the distributions change over time and that there exists some evidence of nonstationarity in the time series of cross-sectional distributions of individual earnings.

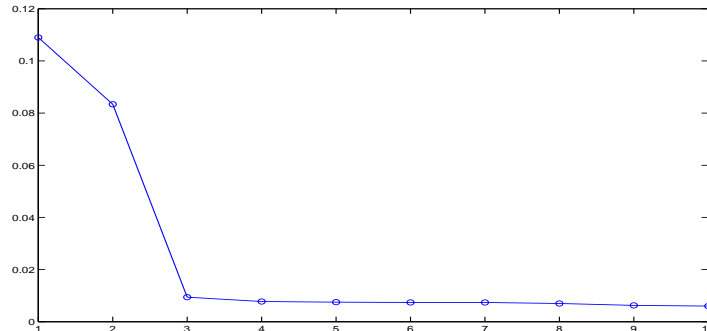
To investigate the unit root dimension in the time series of cross-sectional distributions of individual earnings, we use the feasible statistic  $\hat{\tau}_M^T$  in (23) to test for the null hypothesis (7) against the alternative hypothesis (8) with  $M = 1, \dots, 5$ . The test results are presented below in Table 2.

Table 2. Test Results for Dimension of Unit Root Space - Individual Earnings

$M$	1	2	3	4	5
$\hat{\tau}_M^T$	0.1090	0.0834	0.0094	0.0078	0.0075

Our test, strongly and unambiguously, rejects the null hypothesis (7) against the alter-

Fig. 2. Scree Plot of Eigenvalues - Individual Earnings



native hypothesis (8) successively for  $M = 5, 4, 3$ . Clearly, however, the test cannot reject the null hypothesis (7) in favor of the alternative hypothesis (8) for  $M = 2$ . Therefore, we are led to conclude that there exists two-dimensional unit root subspace, and set  $\hat{M}_T = 2$ , in the time series of cross-sectional distributions of individual earnings. It is also strongly supported by the scree plot of the eigenvalues of the estimated variance operator of the demeaned state densities presented in Fig. 2. The magnitudes of the first two eigenvalues are substantially larger than the others, and we can clearly see that the eigenvalues decrease slowly from the third largest one.

The bottom-left panel of Fig. 1 presents the two nonstationary coordinate processes - the first one in green and the second one in red - along with the normalized aggregated time series of the individual earnings drawn in blue. Both nonstationary coordinate processes, especially the first one, evolve over time like a pure unit root process. The normalized aggregated time series also looks quite persistent, though not as persistent as the first nonstationary coordinate process. This is not surprising since, as we report below, the first moment of income distribution has a substantial unit root proportion, 52%. The time series of the distributions of individual earnings after purging away its nonstationary component is plotted in the bottom-right panel of Fig. 1. It does behave much more stable over time than the time series of the original distributions of individual earnings given in the top two panels of Fig. 1.

Now we compute as in (25) with  $\hat{M}_T = 2$  the estimate  $\hat{\pi}_i^T$  of the unit root proportion  $\pi_i^T$  defined in (19) for the first four moments, based on the estimated nonstationarity subspace which we obtain for the time series of cross-sectional distributions of individual earnings. We summarize the results below in Table 3.

Table 3. Estimated Unit Root Proportions in Moments - Individual Earnings

$\hat{\pi}_1^T$	$\hat{\pi}_2^T$	$\hat{\pi}_3^T$	$\hat{\pi}_4^T$
0.5280	0.3388	0.2377	0.1822

It appears that the unit root proportions for the first four moments are all nonnegligibly large. In particular, the unit root proportions for the first two moments are quite substantial. The presence of a substantial unit root proportion in the second moment of the cross-sectional distribution of individual earnings seems to be particularly interesting. Recently, the time series analysis of changes in the volatilities of individual earnings and income have drawn much attention. See, e.g., Dynan et al. (2012), Barth et al. (2010) and Shin and Solon (2011). Nonstationarity in the time series of cross-sectional distributions of individual earnings would certainly make their volatilities more persistent, as evidenced extensively in the existing literature.

### 5.2 Intra-Month Distributions of Stock Returns

For the second empirical application, we consider the time series of intra-month distributions of stock returns. For each month during the period from January 1992 to June 2010, we use S&P 500 index returns at one-minute frequency to estimate 222 densities for the intra-month distributions. The one-minute returns of S&P 500 index are obtained from Tick Data Inc. The number of intra-month observations available for each month varies from 7211 to 9177, except for September 2001, for which we only have 5982 observations.<sup>17</sup> The time series of the estimated densities for intra-month distributions of S&P 500 index returns, undemeaned and demeaned, are shown respectively in the top-left and top-right panels of Fig. 3. The intra-month observations are truncated at 0.50% and 99.5% percentiles before we estimate the state densities. It can be clearly seen that the mean locations and volatility levels, in particular, of intra-month return distributions vary with time. We may also see some strong evidence of nonstationarity in the time series of intra-month return distributions.

To determine the unit root dimension and check the existence of nonstationarity in the time series of intra-month S&P 500 return distributions, we test using the feasible statistic  $\hat{\tau}_M^T$  defined in (23) for the null hypothesis (7) against the alternative hypothesis (8) with  $M = 1, \dots, 5$ . The test results are summarized in Table 4 below.

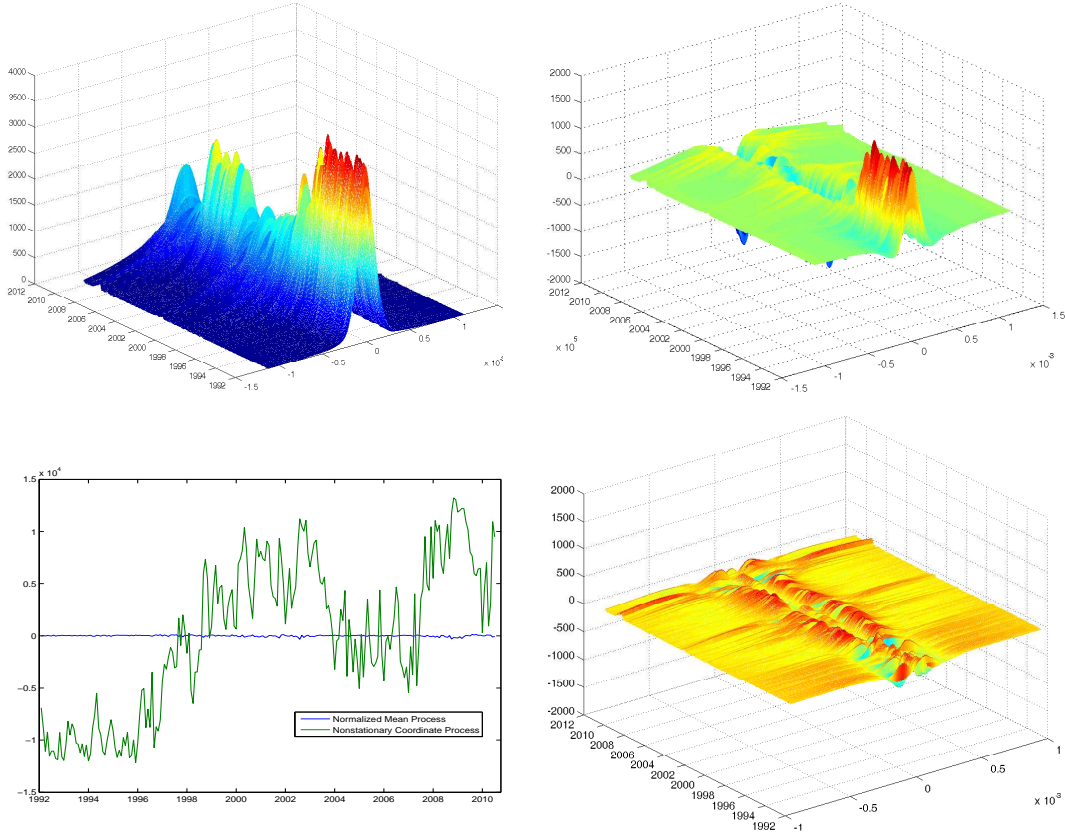
Table 4. Test Results for Dimension of Unit Root Space - S&P 500 Intra-Month Returns

$M$	1	2	3	4	5
$\hat{\tau}_M^T$	0.0612	0.0167	0.0112	0.0107	0.0010

Our test successively rejects the null hypothesis (7) against the alternative hypothesis (8) for  $M = 5, 4, 3, 2$ . However, at 5% level, the test cannot reject the null for  $M = 1$  in favor of the alternative. Our test result implies that there exists one-dimensional unit root, i.e.,  $\hat{M}_T = 1$ , in the time series of intra-month S&P 500 return distributions. The scree plot of the eigenvalues of the estimated variance operator of the state densities presented in Fig. 4 also strongly supports the presence of one dimensional unit root. The magnitude of the first eigenvalue is distinctively larger than all the other eigenvalues.

<sup>17</sup>To avoid the micro-structure noise, we also use the five-minute observations to estimate the intra-month observations. Our empirical results are, however, virtually unchanged.

Fig. 3. Distributional Time Series - S&amp;P 500 Intra-Month Returns

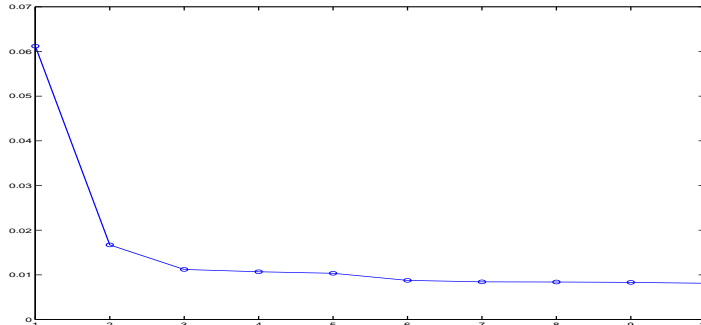


In the bottom-left panel of Fig. 3, we present the time series of the nonstationary coordinate process drawn in green line which behaves like a random walk process, making a stark contrast with the normalized aggregated time series of returns shown in blue line. This is consistent with the negligible value of the unit root proportion, less than 1%, of the first moment we report below. This drastically demonstrates that our distributional analysis gives a new light on the nonstationarity of the given time series, which is not observable from the usual analysis of aggregated time series. Our point is especially well demonstrated here in the time series of the intra-month return distributions. Although the aggregated monthly returns are unambiguously stationary, the nonstationary coordinate process shows some strong evidence of nonstationarity in the distributions of the intra-month return distributions. Therefore, our analysis makes it clear that the intra-month return distributions have a strong persistency that cannot be seen in the time series of their aggregates. As in the individual earnings example, we remove the nonstationary component in the time series of S&P 500 intra-month return distributions, and provide it in the bottom-right panel of Fig. 3, which looks much more stable over time.

Finally, we compute as in (25) with  $\hat{M}_T = 1$  the estimate  $\hat{\pi}_{\kappa}^T$  of the unit root proportion



Fig. 4. Scree Plot of Eigenvalues - S&amp;P 500 Intra-Month Returns



$\pi_{\kappa}^T$  defined in (19) for the first four moments, presuming that we have one-dimensional unit root in the time series of intra-month S&P 500 return distributions. The results are presented below in Table 5.

Table 5. Estimated Unit Root Proportions in Moments - S&amp;P 500 Intra-Month Returns

$\hat{\pi}_1^T$	$\hat{\pi}_2^T$	$\hat{\pi}_3^T$	$\hat{\pi}_4^T$
0.0047	0.2087	0.0039	0.0958

The unit root proportions are in general small for all of the first four moments. This implies that the nonstationarity in the time series of intra-month S&P 500 return distributions is not concentrated in the first four moments. This is in contrast with our first empirical application, where we study the time series of cross-sectional distributions of individual earnings. However, the nonstationarity in the time series of intra-month S&P 500 return distributions is relatively more concentrated in the second and fourth moments, with the unit root proportion of the second moment being the largest. The unit root proportion of the first and third moments are almost negligible, and it appears that nonstationarity does not exist in the first and third moments of the time series of intra-month S&P 500 return distributions. This is well expected, since for many financial time series strong persistency is observed mainly in volatility and kurtosis.

## 6. Monte Carlo Simulation

In this section, we perform a simulation to investigate the finite sample performance of the statistic  $\tau_M^T$  introduced in (17) to test for the null hypothesis (7) against the alternative hypothesis (8). For the simulation, we generate the data using the models that approximate as closely as possible the estimated models we obtained from our empirical applications in the previous section. This is to make our simulation more realistic and practically more relevant. Needless to say, the performance of our test is expected to be varying depending

upon data generating processes. We assume that the state densities are observable, and therefore, our simulation would not provide any information on the effect of estimating unobserved state densities. The numbers of observations  $N$  we use to estimate the state densities in our empirical applications are quite large compared to the sample sizes  $T$  for the time series of state density. The approximation error due to the estimation of state densities should therefore be small and unimportant, if not totally negligible. Several choices of  $T$  between 100 and 500 are considered. Recall that we have  $T = 247$  and  $T = 222$  for each of the two empirical applications reported in the previous section. In all cases we employ our test with 5% significance level, and all the reported results are based on 5,000 iterations.

In our simulation, we directly generate the centered state density ( $w_t$ ), instead of the original state density ( $f_t$ ). This of course causes no loss in generality, since we use the demeaned original state density  $f_t - T^{-1} \sum_{t=1}^T f_t$  that is identical to the demeaned centered state density  $w_t - T^{-1} \sum_{t=1}^T w_t$ . To generate our simulation sample ( $w_t$ ), we let

$$w_t = \sum_{i=1}^I c_{it} v_i, \quad (26)$$

where  $(c_{it})$ ,  $i = 1, \dots, I$  and  $t = 1, \dots, T$ , are scalar stochastic processes and  $(v_i)$ ,  $i = 1, \dots, I$ , nonrandom orthonormal vectors in  $H$ , which are specified more specifically below. Clearly, we have  $c_{it} = \langle v_i, w_t \rangle$ , and  $(c_{it})$  becomes the  $i$ -th coordinate process of  $(w_t)$  for  $i = 1, \dots, I$ . Note that the simulation sample  $(w_t)$  is generated from the  $I$ -dimensional subspace of  $H$  spanned by  $(v_i)$ ,  $i = 1, \dots, I$ . We set  $I = T$ , and let  $(v_i)$  in (26) be the orthonormal eigenvectors associated with nonzero eigenvalues of the estimated variance operator of  $(w_t)$  obtained in our empirical applications. Note that for a sample  $(w_t)$  of size  $T$  we only have  $T$  eigenvectors  $(v_i)$  associated with nonzero eigenvalues of the sample variance operator.

The coordinate processes  $(c_{it})$  in (26) are also specified similarly as the estimated coordinate processes obtained from our empirical applications. For both of our empirical applications, all information order selection criteria including AIC and BIC applied to the estimated coordinate processes yield simple AR(1) and AR(2) models respectively for the stationary and nonstationary coordinate processes.<sup>18</sup> Accordingly, we let

$$(c_{it} - \alpha_i c_{i,t-1}) = \beta_i (c_{i,t-1} - \alpha_i c_{i,t-2}) + \eta_{it} \quad (27)$$

for  $i = 1, \dots, M$ , and

$$c_{it} = \alpha_i c_{i,t-1} + \eta_{it} \quad (28)$$

for  $i = M + 1, \dots, I$ , where  $(\eta_{it})$  are independent normal random variates with mean zero and variance  $(\sigma_i^2)$  for  $i = 1, \dots, I$ . The parameter values for  $(\beta_i)$ ,  $i = 1, \dots, M$ ,  $(\alpha_i)$ ,  $i = M + 1, \dots, I$ , and  $(\sigma_i^2)$ ,  $i = 1, \dots, I$ , are all set to be our estimates for the estimated coordinate processes of our empirical applications. The estimates for  $(\beta_i)$  are obtained with the restriction  $\alpha_i = 1$  imposed for  $i = 1, \dots, M$ . Respectively for our first and second empirical models, we set  $M = 2$  and  $M = 1$  in (27) and (28).

---

<sup>18</sup>The results are also entirely robust to the choice of maximum orders for the order selection criteria.

Table 6. Rejection Probabilities of  $\tau_M^T$  in Empirical Model of Individual Earnings

		$M$				
$M_0$	$T$	1	2	3	4	5
2	100	0.0204	0.0664	0.7429	0.7146	0.4724
	200	0.0245	0.0490	0.9966	0.9995	0.9907
	300	0.0255	0.0551	1.0000	1.0000	1.0000

			$\alpha$			
$M$	$M_0$	$T$	0.80	0.85	0.90	0.95
2	1	100	0.2724	0.1076	0.0348	0.0114
		300	1.0000	0.9998	0.9078	0.2378
		500	1.0000	1.0000	0.9998	0.7626
	0	100	0.5498	0.2358	0.0838	0.0166
		300	1.0000	1.0000	0.9924	0.4452
		500	1.0000	1.0000	1.0000	0.9540
1	0	100	0.8376	0.5498	0.2294	0.0416
		300	1.0000	1.0000	0.9988	0.6802
		500	1.0000	1.0000	1.0000	0.9930

### 6.1 Empirical Model of Individual Earnings

The simulation results for our first empirical model on the cross-sectional distributions of individual earnings are tabulated in Table 6. We denote by  $M_0$  the number of unit roots in our simulated samples ( $w_t$ ) to distinguish it from the hypothesized number  $M$  of unit roots for the test of the null hypothesis (7) against the alternative hypothesis (8) based on  $\tau_M^T$ . Our simulation samples for the first empirical model are obtained from the coordinate processes generated by (27) and (28) with  $M = 2$ . To obtain simulation samples with  $M_0 = 2$ , we may simply set  $\alpha_1 = \alpha_2 = 1$  in (27). On the other hand, we set  $\alpha_1 = 1$  and let  $\alpha_2 = \alpha$ , where  $\alpha$  takes values from 0.80 to 0.95 with increment 0.05, in (27) and (28) to obtain simulation samples with  $M_0 = 1$ . Finally, for the simulation samples with  $M_0 = 0$ , we set  $\alpha_1 = \alpha_2 = \alpha$  in (28) with  $\alpha$  again taking values from 0.80 to 0.95 with increment 0.05.

In the upper panel, we set the model exactly as we find in our empirical analysis with  $M_0 = 2$ , and present the rejection probabilities of  $\tau_M^T$  for various choices of  $M$ ,  $M = 1, \dots, 5$ . Under our setting, we expect the null hypothesis (7) to be rejected for all  $M \geq 3$ . Overall, the finite sample powers of the test are quite good even for moderately large  $T$ . The powers of the test are almost 100% for all  $M$  as long as  $T$  is greater than 200. If  $T$  is 100, the test loses power rather dramatically as  $M$  increases. Therefore, if  $T$  is small, it does not appear to be practically meaningful to use the test for large values of  $M$ . The lack of power in

the test for small  $T$  however quickly disappears as  $T$  gets large. On the other hand, the rejection probabilities are quite close to the nominal size 5% unless  $T$  is very small, in which case the test tends to under-reject the null hypothesis. In sum, our empirical finding for the existence of nonstationarity in the cross-sectional distributions of individual earnings seem to be well supported by our simulation results here.

In the lower panel, we further investigate the finite sample powers of our test against the models with roots in the vicinity of unity. As alternative models, we consider both cases with  $M_0 = 1$  and  $M_0 = 0$ . Our simulation results are largely as expected. In all cases, the powers of the test decrease as  $\alpha$  approaches to unity and it becomes harder to discriminate the null and alternative hypotheses. Moreover, the test becomes more powerful rather quickly as the sample size increases in every case we consider. This is true uniformly in  $\alpha$ . However, it appears that the finite sample powers of the test are not large enough to effectively distinguish the near unit roots from the exact unit roots. In particular, when  $T$  is very small, the test has virtually no power against the near unit roots and rarely rejects the null hypothesis of unit roots, regardless of the true number  $M_0$  of unit roots and the hypothesized number  $M$  of unit roots.

## 6.2 Empirical Model of Stock Returns

Now we use our second empirical model on the intra-month distributions of stock returns to evaluate the finite sample performance of our test  $\tau_M^T$  for the null hypothesis (7) against the alternative hypothesis (8). The simulation results are tabulated in Table 7. As in the previous section, we use  $M_0$  to denote the number of unit roots in our simulated samples ( $w_t$ ), whereas  $M$  signifies the hypothesized number of unit roots for the test. Our simulation samples for the second empirical model are obtained from the coordinate processes generated by (27) and (28) with  $M = 1$  as in the second empirical model. It is straightforward to obtain simulation samples with  $M_0 = 1$ , since we may simply set  $\alpha_1 = 1$  in (27). To obtain simulation samples with  $M_0 = 0$ , on the other hand, we set  $\alpha_1 = \alpha$ , where  $\alpha$  takes values from 0.80 to 0.95 with increment 0.05, in (28). This is completely analogous to our earlier setup for the first empirical model.

For the simulation results in the upper panel, we use the model we find in our empirical analysis with  $M_0 = 1$  and compute the rejection probabilities of  $\tau_M^T$  for various choices of  $M$ ,  $M = 1, \dots, 5$ . Consequently, we expect the null hypothesis (7) to be rejected for all  $M \geq 2$ . The finite sample powers of the test are quite good and have almost perfect power even when  $T$  is only moderately large, i.e.,  $T = 200$ . As in our simulation results for the first empirical model, the power of the test is poor and not satisfactory when  $T$  is small. However, the problem quickly disappears as  $T$  increases. The performance of the test is rather satisfactory also in terms of finite sample sizes. The sizes of the test are nearly exact even in very small samples. The actual rejection probabilities are indeed almost identical to the nominal size for  $T = 100$  as well as for bigger samples. This is in contrast with our earlier results for the first empirical model, where we observe a clear tendency for the test to under-reject when  $T$  is small.

As in the simulations for the first empirical model, we further examine the finite sample powers of the test against stationary models with roots in the neighborhood of unity.

Table 7. Rejection Probability of  $\tau_M^T$  in Empirical Model of Stock Returns

		$M$				
$M_0$	$T$	1	2	3	4	5
1	100	0.0512	0.9696	0.6021	0.2725	0.0708
	200	0.0584	1.0000	1.0000	1.0000	0.9851

			$\alpha$			
$M$	$M_0$	$T$	0.80	0.85	0.90	0.95
1	0	100	0.8682	0.6198	0.3008	0.0718
		300	0.9990	0.9992	0.9980	0.7582
		500	1.0000	1.0000	1.0000	0.9920

The overall aspects of our simulation results for the second empirical model are essentially identical to those of our previous results for the first empirical model. Again, the power of the test decreases as  $\alpha$  approaches to unity, and the power loss is particularly severe if  $T$  is small. The finite sample powers, however, rapidly increase uniformly in  $\alpha$  as the sample size increases. The performance of the test appears to be slightly better for the second empirical model compared with the first.

## 7. Conclusion

In this paper, we consider testing for nonstationarity for the time series of state distributions, which can be either cross-sectional or intra-period distributions of some underlying economic variable. The state densities are regarded as Hilbertian random variables, and we employ the functional principal component analysis to construct a statistic to test for unit roots in the time series of state densities. Our test is nuisance parameter free and its critical values are tabulated in the paper. Once we estimate the unit root subspace, we may compute the unit root proportions in the moments of state distributions. We apply our methodology to analyze nonstationarity in the time series of two different types of state distributions, i.e., the cross-sectional distributions of individual earnings and intra-month distributions of stock returns. In both cases, we could find some clear evidence for the presence of nonstationarity. The presence of nonstationarity in the time series of state distributions yields some important implications, both economic and statistical, which can certainly be further explored. This will be reported in our subsequent work.

## Mathematical Appendix

**Proof of Lemma 3.1** Define the stochastic process  $W_T$  on  $[0, 1]$  by

$$W_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} H_N \Delta w_t,$$

which takes values on  $H_N$ . We will establish that

$$W_T \rightarrow_d W, \tag{29}$$

where  $W$  is Brownian motion on  $H_N$  with variance operator  $\Pi_N \Phi(1) \Sigma \Phi(1)' \Pi_N$ . To derive (29), it suffices to show that

$$\langle v, W_T \rangle \rightarrow_d \langle v, W \rangle \tag{30}$$

for  $v \in H_N$  chosen arbitrarily.

For any  $v \in H_N$ , we have

$$\langle v, W_T(r) \rangle = T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, H_N \Delta w_t \rangle = T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, \Delta w_t \rangle. \tag{31}$$

Observe that

$$\langle v, \Delta w_t \rangle = \langle v, \Phi(1) \varepsilon_t \rangle + \langle v, \bar{u}_{t-1} - \bar{u}_t \rangle, \tag{32}$$

and therefore, for  $r \in [0, 1]$ , it follows that

$$T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, \Delta w_t \rangle = T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, \Phi(1) \varepsilon_t \rangle - T^{-1/2} \langle v, \bar{u}_{[Tr]} \rangle \tag{33}$$

ignoring  $\bar{u}_0$ . However, we have

$$\begin{aligned} \mathbb{P} \left\{ \sup_{r \in [0,1]} T^{-1/2} |\langle v, \bar{u}_{[Tr]} \rangle| > K \right\} &= \mathbb{P} \left\{ \max_{1 \leq t \leq n} T^{-1/2} |\langle v, \bar{u}_t \rangle| > K \right\} \\ &\leq \sum_{t=1}^T \mathbb{P} \left\{ T^{-1/2} |\langle v, \bar{u}_t \rangle| > K \right\} \\ &= T \mathbb{P} \left\{ T^{-1/2} |\langle v, \bar{u}_t \rangle| > K \right\} \\ &\leq T^{1-p/2} (1/K^p) \mathbb{E} |\langle v, \bar{u}_t \rangle|^p \rightarrow 0 \end{aligned}$$

for any constant  $K > 0$  and for  $p > 2$  such that  $\mathbb{E} \|\varepsilon_t\|^p < \infty$ , since  $\mathbb{E} \|\varepsilon_t\|^p < \infty$  implies  $\mathbb{E} \|\bar{u}_t\|^p < \infty$ , as can be shown similarly as in, e.g., Chang and Park (2002), and  $|\langle v, \bar{u}_t \rangle| \leq \|v\| \|\bar{u}_t\|$ , from which it follows that  $\mathbb{E} |\langle v, \bar{u}_t \rangle|^p \leq \|v\|^p \mathbb{E} \|\bar{u}_t\|^p < \infty$ , for  $p > 2$ . Consequently, we may deduce from (33) that

$$T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, \Delta w_t \rangle = T^{-1/2} \sum_{t=1}^{[Tr]} \langle v, \Phi(1) \varepsilon_t \rangle + o_p(1) \tag{34}$$

uniformly in  $r \in [0, 1]$ . Now (30) follows immediately from (31) and (34), as was to be shown to derive (29). Note that  $(\langle v, \Phi(1)\varepsilon_t \rangle)$  is an i.i.d. sequence of random variables with variance  $\langle v, \Phi(1)\Sigma\Phi(1)'v \rangle$ .

To prove (2), we let  $(v_1, v_2) \in H_N \times H_N$  be chosen arbitrarily. Note that

$$\left\langle v_1, \left( T^{-2} \sum_{t=1}^T w_t^N \otimes w_t^N \right) v_2 \right\rangle = T^{-2} \sum_{t=1}^T \langle v_1, w_t^N \rangle \langle v_2, w_t^N \rangle, \quad (35)$$

and that

$$\begin{aligned} T^{-2} \sum_{t=1}^T \langle v_1, w_t \rangle \langle v_2, w_t \rangle &= \int_0^1 \langle v_1, W_T(r) \rangle \langle v_2, W_T(r) \rangle dr + o_p(1) \\ &\rightarrow_d \int_0^1 \langle v_1, W(r) \rangle \langle v_2, W(r) \rangle dr \\ &= \left\langle v_1, \left( \int_0^1 W(r) \otimes W(r) dr \right) v_2 \right\rangle, \end{aligned} \quad (36)$$

for any choice of  $(v_1, v_2) \in H_N \times H_N$ , due to (29) and the continuous mapping theorem. It follows from (35) and (36) that

$$\left\langle v_1, \left( T^{-2} \sum_{t=1}^T w_t^N \otimes w_t^N \right) v_2 \right\rangle \rightarrow_d \left\langle v_1, \left( \int_0^1 W(r) \otimes W(r) dr \right) v_2 \right\rangle,$$

for any choice of  $(v_1, v_2) \in H_N \times H_N$ . This was to be shown.

Next, to deduce (3), we simply note that

$$Q_{NN}^T = T^{-1} \sum_{t=1}^T w_t^S \otimes w_t^S = \Pi_S \left( T^{-1} \sum_{t=1}^T \bar{u}_t \otimes \bar{u}_t \right) \Pi_S,$$

and that

$$T^{-1} \sum_{t=1}^T \bar{u}_t \otimes \bar{u}_t \rightarrow_p \mathbb{E} \bar{u}_t \otimes \bar{u}_t = \sum_{i=0}^{\infty} \bar{\Phi}_i \Sigma \bar{\Phi}_i'.$$

Finally, since we have for any  $(v_1, v_2) \in H_N \times H_S$

$$\langle v_1, Q_{NS}^T v_2 \rangle = T^{-1} \sum_{t=1}^T \langle v_1, w_t^N \rangle \langle v_2, w_t^S \rangle = O_p(1)$$

from which (4) readily follows. This completes the proof.  $\square$

**Proof of Proposition 3.2** It can be easily deduced from (1) and Lemma 3.1 that

$$T^{-2} Q^T = Q_{NN}^T + O_p(T^{-1})$$

for all large  $T$ . Obviously, by construction, the  $M$ -leading eigenvectors of  $Q_{NN}^T$  associated with nonzero eigenvalues are obtained in the unit root subspace of  $H_N$  of  $H$ , and the stated result for  $\Pi_N^T$  follows immediately. The result for  $\Pi_S^T$  can be deduced immediately from the fact that  $\Pi_S = 1 - \Pi_N$ . The proof is therefore complete.  $\square$

**Proof of Theorem 3.3** To prove the first part, we invoke the well known Skorokhod representation theorem in, e.g., Pollard (1984, Theorem IV.13), and assume that  $Q_{NN}^T \rightarrow_{a.s.} Q_{NN}$ . Obviously, the infinite dimensionality of  $H$  does not give any complications, since the unit root subspace  $H_N$ , on which  $Q_{NN}^T$  and  $Q_{NN}$  are defined, is finite dimensional. Here we note that the eigensubspaces spanned by  $v_i(Q_{NN})$  are one dimensional a.s., and define  $v_i^T(Q_{NN}) = \text{sgn}\langle v_i(Q_{NN}), v_i(Q^T) \rangle v_i(Q_{NN})$ ,  $i = 1, \dots, M$ . The introduction of  $(v_i^T(Q_{NN}))$  is necessary for the definitiveness of eigenvectors, since both  $(v_i(Q_{NN}))$  and  $(-v_i(Q_{NN}))$  are eigenvectors corresponding to  $(\lambda_i(Q_{NN}))$ ,  $i = 1, \dots, M$ . In what follows, however, we write  $(v_i^T(Q_{NN}))$  simply as  $(v_i(Q_{NN}))$ ,  $i = 1, \dots, M$ , and this should cause no confusion.

It follows from the proofs of Lemma 3.1 and Proposition 3.2 and Lemma 4.3 of Bosq (2000) that

$$\begin{aligned} |T^{-2}\lambda_i(Q^T) - \lambda_i(Q_{NN})| &\leq \|T^{-2}Q^T - Q_{NN}\| \\ &\leq \|T^{-2}Q^T - Q_{NN}^T\| + \|Q_{NN}^T - Q_{NN}\| \\ &= O_p(T^{-1}) + O_p(T^{1-p/2}) \end{aligned}$$

and

$$\begin{aligned} \|v_i(Q^T) - v_i(Q_{NN})\| &= O_p(1) \|T^{-2}Q^T - Q_{NN}\| \\ &\leq O_p(1) (\|T^{-2}Q^T - Q_{NN}^T\| + \|Q_{NN}^T - Q_{NN}\|) \\ &= O_p(T^{-1}) + O_p(T^{1-p/2}) \end{aligned}$$

for  $i = 1, \dots, M$ . Note that  $\lambda_i(Q_{NN}) = O_p(1)$  for  $i = 1, \dots, M$ . Consequently, we have

$$(T^{-2}\lambda_i(Q^T), v_i(Q^T)) \rightarrow_p (\lambda_i(Q_{NN}), v_i(Q_{NN}))$$

for  $i = 1, \dots, M$ , from which the stated result follows immediately.

For the second part, we write

$$\Pi_S^T Q^T \Pi_S^T = T^2 \Pi_S^T Q_{NN}^T \Pi_S^T + T \Pi_S^T Q_{NS}^T \Pi_S^T + T \Pi_S^T Q_{SN}^T \Pi_S^T + T \Pi_S^T Q_{SS}^T \Pi_S^T, \quad (37)$$

where  $Q_{NN}^T, Q_{NS}^T, Q_{SN}^T$  and  $Q_{SS}^T$  are defined as in (1), and note that

$$\Pi_S^T \Pi_N = \Pi_S \Pi_N + (\Pi_S^T - \Pi_S) \Pi_N = O_p(T^{-1}) \quad (38)$$

and

$$\Pi_S^T \Pi_S = \Pi_S \Pi_S + (\Pi_S^T - \Pi_S) \Pi_S = \Pi_S + O_p(T^{-1}) \quad (39)$$

due to Lemma 3.1. Clearly, we have  $\Pi_S \Pi_N = 0$  and  $\Pi_S \Pi_S = \Pi_S$ .

Due to (38) and (39), we have

$$\begin{aligned} \Pi_S^T Q_{NN}^T \Pi_S^T &= \Pi_S^T \Pi_N \left( \frac{1}{T^2} \sum_{t=1}^T w_t \otimes w_t \right) \Pi_N \Pi_S^T = O_p(T^{-2}) \\ \Pi_S^T Q_{SN}^T \Pi_S^T &= \Pi_S^T \left( \frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t \right) \Pi_N \Pi_S^T = O_p(T^{-1}) \\ \Pi_S^T Q_{SS}^T \Pi_S^T &= \Pi_S^T \left( \frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t^S \right) \Pi_S^T = \frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t^S + O_p(T^{-1}), \end{aligned}$$



from which it follows that

$$T^{-1}\Pi_S^T Q^T \Pi_S^T = \Pi_S^T Q_{SS}^T \Pi_S^T + O_p(T^{-1}) = \frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t^S + O_p(T^{-1}). \quad (40)$$

Moreover, we have

$$\frac{1}{T} \sum_{t=1}^T w_t^S \otimes w_t^S = Q_{SS} + O_p(T^{-1/2}) \quad (41)$$

for all large  $T$ .

Now we may deduce from (40) and (41) that

$$\left| T^{-1}\lambda_i(\Pi_S^T Q^T \Pi_S^T) - \lambda_i(Q_{SS}) \right| \leq \|T^{-1}\Pi_S^T Q^T \Pi_S^T - Q_{SS}\| = O_p(T^{-1/2}).$$

Moreover, if we use the convention made in the proof of the first part and redefine  $v_i(Q_{SS})$  as  $\text{sgn}\langle v_i(Q_{SS}), v_i(\Pi_S^T Q^T \Pi_S^T) \rangle v_i(Q_{SS})$ ,  $i = 1, 2, \dots$ , then it follows from Lemma 4.3 of Bosq (2000) that and

$$\|v_i(\Pi_S^T Q^T \Pi_S^T) - v_i(Q_{SS})\| \leq O(1) \|T^{-1}\Pi_S^T Q^T \Pi_S^T - Q_{SS}\| = O_p(T^{-1/2})$$

for  $i = 1, 2, \dots$ . To complete the proof of the second part, note that  $\lambda_{M+i}(Q^T) = \lambda_i(\Pi_S^T Q^T \Pi_S^T)$  and  $v_{M+i}(Q^T) = v_i(\Pi_S^T Q^T \Pi_S^T)$  for  $i = 1, 2, \dots$   $\square$

**Proof of Corollary 3.4** The first part follows immediately from

$$\begin{aligned} |\langle \Pi_N^T v, w_t \rangle - \langle \Pi_N v, w_t \rangle| &= |\langle (\Pi_N^T - \Pi_N) v, w_t \rangle| \\ &\leq \|(\Pi_N^T - \Pi_N) v\| \|w_t\| \\ &\leq \|(\Pi_N^T - \Pi_N)\| \left( \max_{1 \leq t \leq T} \|w_t\| \right) \\ &= O_p(T^{-1}) O_p(T^{1/2}). \end{aligned}$$

Similarly, the second part can be deduced from

$$\begin{aligned} |\langle \Pi_N^T v, \Delta w_t \rangle - \langle \Pi_N v, \Delta w_t \rangle| &= |\langle (\Pi_N^T - \Pi_N) v, \Delta w_t \rangle| \\ &\leq \|(\Pi_N^T - \Pi_N) v\| \|\Delta w_t\| \\ &\leq \|(\Pi_N^T - \Pi_N)\| \left( \max_{1 \leq t \leq T} \|\Delta w_t\| \right) \\ &= O_p(T^{-1}) O_p(T^{1/p}). \end{aligned}$$

upon noticing that

$$\max_{1 \leq t \leq T} \|\Delta w_t\| = O_p(T^{1/p}).$$

The proof is therefore complete.  $\square$

**Proof of Theorem 3.5** As noted, we may let

$$z_t = (\langle \Pi_N v_1(Q^T), w_t \rangle, \dots, \langle \Pi_N v_M(Q^T), w_t \rangle)'$$

without loss of generality. Therefore, it follows immediately from Corollary 3.4 that

$$\max_{1 \leq t \leq T} \|\tilde{z}_t - z_t\| = O_p(T^{-1/2}) \quad (42)$$

$$\max_{1 \leq t \leq T} \|\Delta \tilde{z}_t - \Delta z_t\| = O_p(T^{-1+1/p}) \quad (43)$$

We may easily deduce from (42) that

$$\begin{aligned} \left\| \sum_{t=1}^T \tilde{z}_t \tilde{z}_t' - \sum_{t=1}^T z_t z_t' \right\| &\leq 2 \left( \max_{1 \leq t \leq T} \|\tilde{z}_t - z_t\| \right) \sum_{t=1}^T \|z_t\| + T \left( \max_{1 \leq t \leq T} \|\tilde{z}_t - z_t\| \right)^2 \\ &= O_p(T^{-1/2}) O_p(T^{3/2}) + T O_p(T^{-1}) = O_p(T), \end{aligned}$$

which implies that

$$\tilde{Q}_M^T = Q_M^T + O_p(T) \quad (44)$$

for all large  $T$ .

Moreover, we have from (43) that

$$\begin{aligned} \left| \tilde{\Gamma}_T(i) - \Gamma_T(i) \right| &\leq \left( \max_{1 \leq t \leq T} \|\Delta \tilde{z}_t - \Delta z_t\| \right) \frac{1}{T} \sum_{t=1}^T \|\Delta z_t\| + \left( \max_{1 \leq t \leq T} \|\Delta \tilde{z}_{t-i} - \Delta z_{t-i}\| \right)^2 \\ &= O_p(T^{-1+1/p}) O_p(1) + O_p(T^{-2+2/p}) = O_p(T^{-1+1/p}), \end{aligned}$$

and therefore,

$$\begin{aligned} \left| \tilde{\Omega}_M^T - \Omega_M^T \right| &\leq \sum_{|i| \leq \ell} |\varpi_\ell(i)| \left| \tilde{\Gamma}_T(i) - \Gamma_T(i) \right| \\ &= O_p(\ell) O_p(T^{-1+1/p}) = O_p(\ell T^{-1+1/p}) = o_p(1). \end{aligned} \quad (45)$$

Now it can be easily deduced from (44) and (45) that

$$\lambda_i(\tilde{Q}_M^T, \tilde{\Omega}_M^T) = \lambda_i(Q_M^T, \Omega_M^T) + o_p(1)$$

from which the first part follows immediately.

To establish the second part, we may simply consider the asymptotic behavior of  $\lambda_M(Q_M^T, \Omega_M^T)$ . Note that the asymptotic behavior of the newly defined statistic  $\tau_M^T$  is different from that of  $\sigma_M^T$  also under the alternative hypothesis (8). Momentarily, we let  $M = 0$  and  $\Gamma(i) = 0$  for all  $|i| > \ell$ . In this simple case, we have  $\Omega_M = 0$  and

$$\Omega_M^{T-1} = O_p(T^{1/2})$$

with the choice of  $\varpi_\ell(i) \equiv 1$  for all  $|i| \leq \ell$ . The order of  $\Omega_M^{T-1}$  is in general smaller than  $O_p(T^{1/2})$  if  $\Gamma(i) \neq 0$  for infinitely many  $i$ 's and we have to increase  $\ell$  as  $T \rightarrow \infty$  with an appropriate choice of  $\varpi_\ell$ . Under the alternative hypothesis (8),  $\Omega_M$  becomes singular and we may now easily see

$$\lambda_M(Q_M^T, \Omega_M^T) = O_p(T^{3/2}),$$

which implies that  $\tau_M^T = O_p(T^{-1/2})$ . This was to be shown.  $\square$

**Proof of Lemma 4.1** We write

$$\hat{w}_t = \hat{f}_t - T^{-1} \sum_{t=1}^T \hat{f}_t = \underline{w}_t + \left( \Delta_t - T^{-1} \sum_{t=1}^T \Delta_t \right),$$

where  $\underline{w}_t = w_t - T^{-1} \sum_{t=1}^T w_t$ , and define

$$\underline{Q}_{NN}^T = \frac{1}{T^2} \Pi_N \left( \sum_{t=1}^T \underline{w}_t \otimes \underline{w}_t \right) \Pi_N.$$

Similarly as in the proof of Lemma 3.1, we may readily deduce that

$$\langle v_1, \underline{Q}_{NN}^T v_2 \rangle \rightarrow_d \langle v_1, \underline{Q}_{NN} v_2 \rangle$$

for  $(v_1, v_2) \in H_N \times H_N$  chosen arbitrarily, which implies that

$$\underline{Q}_{NN}^T \rightarrow_d \underline{Q}_{NN}. \quad (46)$$

Moreover, if we define  $\underline{w}_t^N = \Pi_N \underline{w}_t$ , it follows after some tedious algebra that

$$\hat{Q}_{NN}^T - \underline{Q}_{NN}^T = \frac{1}{T^2} \sum_{t=1}^T \hat{w}_t^N \otimes \hat{w}_t^N - \frac{1}{T^2} \sum_{t=1}^T \underline{w}_t^N \otimes \underline{w}_t^N = o_p(1).$$

It is also rather straightforward to show that

$$\hat{Q}_{SS}^T = Q_{SS}^T + o_p(1), \quad (47)$$

once we write

$$\hat{Q}_{SS}^T - Q_{SS}^T = \left( \hat{Q}_{SS}^T - \underline{Q}_{SS}^T \right) + \left( \underline{Q}_{SS}^T - Q_{SS}^T \right),$$

where

$$\underline{Q}_{SS}^T = \frac{1}{T} \sum_{t=1}^T \underline{w}_t^S \otimes \underline{w}_t^S$$

with  $\underline{w}_t^S = \Pi_S \underline{w}_t$ . The stated results follow immediately from (46) and (47). The details of the proof are omitted to save the space.  $\square$

**Proof of Theorem 4.2** The proofs of the first and second parts are analogous respectively to the proofs of Proposition 3.2 and Theorem 3.3.  $\square$

**Proof of Theorem 4.3** Let  $\underline{Q}_M^T$  be defined from  $(\underline{z}_t)$ , where

$$\underline{z}_t = \left( \langle v_1(Q^T), \underline{w}_t \rangle, \dots, \langle v_M(Q^T), \underline{w}_t \rangle \right)',$$

analogously as  $\hat{Q}_M^T$  in (22) defined from  $(\hat{z}_t)$  in (21), where  $(\underline{w}_t)$  is introduced in the proof of Lemma 4.1. After tedious, but straightforward, algebra, we may readily deduce that

$$\hat{Q}_M^T = \underline{Q}_M^T + o_p(T^{3/2}) \quad \text{and} \quad \hat{\Omega}_M^T = \Omega_M^T + o_p(1),$$

from which the stated results follow immediately.  $\square$

## References

- Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent matrix estimation. *Econometrica* 59, 817–858.
- Autor, D., Katz, L., Kearney, M., 2008. Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics* 90, 300–323.
- Barth, E., Bryson, A., Davis, J., Freeman, R., 2010. The contribution of dispersion across plants to the increase in US earnings dispersion, Working paper, London School of Economics.
- Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics 110, 2nd Edition. Springer, New York.
- Bosq, D., 2000. *Linear Processes in Function Spaces*, Lecture Notes in Statistics 149. Springer, New York.
- Chang, Y., Park, J., 2002. On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–448.
- Dynan, K., Elmendorf, D., Sichel, D., 2012. The evolution of household income volatility. *B.E. Journal of Economic Analysis & Policy* 12-2.
- Hansen, B., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.
- Honda, T., 2009. Nonparametric density estimation for linear processes with infinite variance. *Annals of the Institute of Statistical Mathematics* 61, 413–439.
- Lemieux, T., 2006. Increased residual wage inequality: Composition effects, noisy data, or rising demand for skill. *American Economic Review* 96, 461–498.
- Li, Q., Racine, J. S., 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.
- Liu, L., 2011. On hourly wages and weekly earnings in the Current Population Survey. *Economics Letters* 105, 113–116.
- Park, J., Qian, J., 2012. Functional regression of continuous state distributions. *Journal of Econometrics* 167, 397–412.
- Phillips, P., Solo, V., 1992. Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
- Shin, D., Solon, G., 2011. Trends in men’s earnings volatility: What does the Panel Study of Income Dynamics show? *Journal of Public Economics* 95, 973–982.
- Stock, J., 1994. Unit roots, structural breaks and trends. In: Engle, R. F., McFadden, D. L. (Eds.), *Handbook of Econometrics*. Vol. 4. Elsevier, Ch. 46, pp. 2740–2841.