Econometric Theory, **24**, 2008, 587–615. Printed in the United States of America.
doi: 10.1017/S0266466608080249

ASYMPTOTICS FOR COINTEGRATED PROCESSES WITH INFREQUENT STOCHASTIC LEVEL SHIFTS AND OUTLIERS

ILIYAN GEORGIEV
Universidade Nova de Lisboa

This is an analytical study of the effect of level-shift and temporary-change components, when present but neglected, on the trace test for cointegration. The contribution is threefold. First, we discuss in a multivariate framework, and jointly, effects that in the previous literature have been discussed in a univariate setting and in isolation. Second, we consider a rather general specification of shifts and outliers with random size, number, and timing and with flexible dynamics. It nests the classical cases of additive shifts, innovational outliers, and additive outliers. Third, as an instrument for this analysis we develop an asymptotic theory for product moment matrices of linear processes with stochastic level-shift components, generalizing results of Leipus and Viano (2003, *Statistics and Probability Letters* 61, 177–190).

1. INTRODUCTION

Structural breaks are a characteristic of economic time series that, in the area of cointegration testing alone, has given rise to a large literature on the case of a single break (Gregory and Hansen, 1996; Inoue, 1999; Saikkonen and Lütkepohl, 2000; Lütkepohl, Saikkonen, and Trenkler, 2004, among others) and to fewer contributions on the more difficult case of multiple breaks (Johansen, Mosconi, and Nielsen, 2000, assuming known location of the breaks). The importance of this literature is due to the fact that structural breaks render standard inference on the cointegration rank unreliable or invalid. Analytical results in this respect are typically univariate (e.g., Perron, 1989, 1990; Leybourne and Newbold, 2000a, 2000b) and assume a single break (exceptions to the latter are Chen and Tiao, 1990; Rappoport and Reichlin, 1989; Ohara, 1999). Cointegration analysis, however, is multivariate by its nature, and furthermore, there is rarely sufficient justification for fixing the number of breaks a priori to one or any other number; cf. Lumsdaine and Papell (1997, p. 218).

I thank Søren Johansen for commenting on and discussing this work at various stages of progress, and I also thank Pentti Saikkonen, Anders Rahbek, and Helmut Lütkepohl for reviewing its thesis-chapter version. Further, I am indebted to the referees for a substantial contribution to the readability and the correctness of the paper. Remaining defects are entirely my fault. Address correspondence to Iliyan Georgiev, Faculdade de Economia, Universidade Nova de Lisboa, PT 1099-032, Portugal; e-mail: igeorgiev@fe.unl.pt.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

With this in mind, we adopt a multivariate framework and study the effect of level shifts and outliers with random location, number, and magnitude on the Johansen (1991) trace test on the cointegration rank. Specifically, we consider the basic cointegrated vector autoregressive (VAR) model

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where X_t and ε_t are random $n \times 1$ vectors, μ is an $n \times 1$ vector, α and β are $n \times r$ matrices of rank r, and T is the sample size (further assumptions will be made in Sect. 3.1). The rank r is interpreted as the number of long-run equilibrium relations among the components of X_t. We wish to examine the behavior of the trace test for inference on r when it is applied not to X_t but to a series obtained from X_t by the superposition of a level-shift process with a flexible transition pattern.

In a search for a statistical model characterized by structural breaks whose timing and magnitude are random, Perron (1989) suggests that econometricians could adopt as break-generating processes linear processes constructed from thinned sequences; see, e.g., Example 8.2(a) and Section 9.3 in Daley and Vere-Jones (1988) for a discussion of thinning. In the univariate case, for two independent sequences of independent and identically distributed (i.i.d.) random variables $\{\eta_t\}_{t=1}^{\infty}$ and $\{\pi_t\}_{t=1}^{\infty}$, the sequence $\{\pi_t \eta_t\}_{t=1}^{\infty}$ is called an independent p-thinning of $\{\eta_t\}_{t=1}^{\infty}$ if π_t have the Bernoulli distribution $P(\pi_t = 1) = p > 0$, $P(\pi_t = 0) = 1 - p > 0$.

Although alternative stochastic models of breaklike behavior have also been put forward (see, e.g., Engle and Smith, 1999), Perron's suggestion seems to have been followed by more researchers, maybe because the thinning probability p can be of direct economic interest; cf. Rappoport and Reichlin (1989, p. 169). Thus, to model level shifts, Chen and Tiao (1990) superimpose a random walk with thinned increments on a stationary autoregressive moving average (ARMA) process and study autoregressive integrated moving average (ARIMA) approximations to the resulting process. Franses and Haldrup (1994) and Haldrup, Montanés, and Sanso (2005) study time-series models in which each nonzero realization of π_t is associated with the occurrence of an outlier. In all of these papers p is fixed as T increases, and thus, the number of shifts or outliers is of the order of magnitude of T almost surely (a.s.). Further, under their assumptions Wiener asymptotics apply to the partial sum processes of both $\{\pi_t \eta_t\}_{t=1}^{\infty}$ and the ordinary innovations of the time series. Thus, the distinction between ordinary stochastic shocks and outliers is lost in the limit.

Closer to the original suggestion of Perron, in which the rarity of breaks is central, is the specification of the thinning probability as decreasing in T. Granger and Hyung (2004) and Diebold and Inoue (2001) specify it, respectively, as $O(T^{-1})$ and $O(T^{-\eta})$, $\eta \in (0, 2)$, and study the implied orders of magnitude of some variances and covariances. Leipus and Viano (2003) allow the thinning probability to vary in the sample and, keeping it infinitesimal, obtain weak con-

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

vergence of the partial sum process of $\{\pi_t \eta_t\}_{t=1}^{\infty}$ to a (time-nonhomogeneous) compound Poisson process. They employ this result to study the asymptotics of rescaled variance (or V/S) statistics in the presence of trend breaks. The asymptotic distribution they obtain is expressed as a functional of an integrated compound Poisson process.

For the purposes of our analysis, we extend Leipus and Viano's result to several dimensions and to linear process. The latter permits us to cover various kinds of outliers and shifts, including additive outliers, additive shifts, and shifts resulting from the cumulation of innovational outliers. It turns out that, for the partial sum process of a linear process constructed from a component-wise thinned sequence of i.i.d. random vectors, functional compound Poisson convergence obtains under a much stricter condition than in the Wiener-limit case. Fortunately, this condition is satisfied by the permanent component in the Beveridge–Nelson decomposition of the linear process, which allows one to study the convergence of product-moment matrices of linear processes using the approach of Phillips (1988) and Phillips and Solo (1992).

Given the importance of linear processes in time-series econometrics, there is a large body of literature on the convergence of their product-moment matrices; see, e.g., Phillips (1988) for limits involving functionals of Brownian motion and Chan and Tran (1989), Phillips (1990), and Caner (1997) for functionals of stable non-Gaussian Lévy motion. We contribute to this literature with a result of convergence, in the presence of stochastic breaks, to functionals of a compound Poisson process. The result seems convenient to have in the tool box, and a version of our Theorem 2 from a draft of this paper has been used in Perron and Qu (2004).

Finally, we use the product-moment convergences to obtain the promised asymptotics for the trace test. We show that level shifts in the cointegrating directions lead to inconsistency of the test. In the case where the cointegrating directions are free of shifts and the test is consistent, its asymptotic distribution is influenced by shifts in the noncointegrating directions and by large transitory changes. We view this asymptotic distribution as a benchmark that an inferential procedure for cointegration in the presence of multiple level shifts should improve upon, if it is based on multivariate Dickey–Fuller critical values (Johansen, 1996, Ch. 15). Although we are unaware of such a procedure at the moment, it represents a relevant direction of research, given that multiple level shifts are among the possible explanations of why data tend to exhibit less cointegration than predicted by economic theory.

The paper is structured as follows. Preliminary asymptotic results for level-shift processes are presented in Section 2. In Section 3 they are applied to cointegrated VAR processes with influential level shifts and outliers. Proofs are collected in the Appendix.

Throughout, the components of vectors are denoted by numerical superscripts. For two column vectors $x = (x^1, \ldots, x^k)'$ and $y = (y^1, \ldots, y^k)'$ of the same dimension, $x \cdot y$ denotes their Hadamard product $(x^1 y^1, \ldots, x^k y^k)'$. The

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

euclidean norm of a vector is denoted by $\|x\| := (x'x)^{1/2}$ and the trace of a matrix by $\mathrm{tr}(\cdot)$. As usual, a summation with no addends equals zero. The function $\lfloor \cdot \rfloor$ gives the integer part of its argument. Bold characters denote continuous-time stochastic processes, and $D_k(I)$ denotes the space of càdlàg functions from $I \subseteq [0,\infty)$ to $\mathbb{R}^k$, endowed with the Skorohod topology.

2. PRELIMINARIES

Here we discuss weak convergence of the product moments of linear processes constructed from a component-wise thinned sequence of i.i.d. random vectors.

2.1. Assumptions

First, in two steps we introduce the main random elements under study.

Assumption 1.

(i) $\{\eta_t\}_{t=1}^{\infty}$ is a $k \times 1$ sequence of i.i.d. random vectors.

(ii) $\{\pi_{T,t}\}_{t,T=1}^{\infty}$ is an array of independent $k \times 1$ random vectors with Bernoulli-distributed components and with the following properties. First, there exist nondecreasing continuous functions $F^i:[0,\infty) \to [0,\infty)$, $i = 1,\dots,k$, none of them identically equal to zero, and such that

$$\sum_{t=1}^{\lfloor Tu \rfloor} P(\pi_{T,t} = e_i) \to F^i(u) \quad \text{as } T \to \infty \quad \text{for all } u \in [0,\infty), \tag{1}$$

where e_i is the ith vector in the canonical base of $\mathbb{R}^k$. Second,

$$\sum_{t=1}^{\lfloor Tu \rfloor} P(\|\pi_{T,t}\| > 1) \to 0 \quad \text{as } T \to \infty \quad \text{for all } u \in [0,\infty). \tag{2}$$

(iii) $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ is a doubly infinite sequence of $n \times 1$ i.i.d. random vectors with mean zero and covariance matrix Ω;

(iv) $\{\varepsilon_t\}_{t=-\infty}^{\infty}$, $\{\eta_t\}_{t=1}^{\infty}$, and $\{\pi_{T,t}\}_{t,T=1}^{\infty}$ are jointly independent.

The sequence $\{\varepsilon_t\}$ will be the innovations sequence of our cointegrated VAR model, and $\{\eta_t\}$ and $\{\pi_{T,t}\}$ will be used to generate level shifts and outliers.

Remark 2.1. Conditions (1) and (2) jointly are equivalent to the component-wise conditions that (a) $\sum_{t=1}^{\lfloor Tu \rfloor} P(\pi_{T,t}^i = 1) \to F^i(u)$ and (b) $\sum_{t=1}^{\lfloor Tu \rfloor} P(\pi_{T,t}^i = \pi_{T,t}^j = 1) \to 0$ for $i,j = 1,\dots,k$, $i \neq j$. Condition (a) reproduces (3.2) of Leipus and Viano (2003). Their condition (3.3) is actually implied by their (3.2), or our (a), and so we do not impose it explicitly. If $\pi_{T,t}^i$ and $\pi_{T,t}^j$ are independent for $i \neq j$, then also (b) is implied by (a). Differently from Leipus and Viano, we do not assume the existence of any moments of η_t.

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Assumption 2.

(i) $\{\mu_{T,\cdot}\}_{T=1}^{\infty}$ is a sequence of random walks such that $\mu_{T,t} = \sum_{s=1}^{t} \delta_{T,s}$, $t = 1,\ldots,$ where $\delta_{T,s} := \pi_{T,s} \cdot \eta_s$;

(ii) $\{\theta_{T,\cdot}\}_{T=1}^{\infty}$ and $\{\phi_{T,\cdot}\}_{T=1}^{\infty}$ are sequences of processes such that $\theta_{T,t} = \sum_{i=0}^{t-1} \Theta_i \mu_{T,t-i}$ and $\phi_{T,t} := \sum_{i=0}^{t-1} \Phi_i \mu_{T,t-i}$, $t = 1,\ldots,$ where $\Theta(z) := \sum_{i=0}^{\infty} \Theta_i z^i$ and $\Phi(z) := \sum_{i=0}^{\infty} \Phi_i z^i$ $(z \in \mathbb{C})$ are an $m_1 \times k$ and an $m_2 \times k$ matrix power series convergent for $|z| < 1 + \delta$, for some $\delta > 0$;

(iii) $\{\psi_t\}_{t=1}^{\infty}$ satisfies $\psi_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}$, $t = 1,\ldots,$ where $\Psi(z) = \sum_{i=0}^{\infty} \Psi_i z^i$ is an $m_3 \times n$ matrix power series with the convergence property of $\Theta(z)$.

Finally, we associate a continuous-time process $\mathbf{J}_T$ to $\mu_{T,\cdot}$ by $\mathbf{J}_T(u) := \mu_{T,\lfloor Tu \rfloor}$, $u \in [0,\infty)$.

Remark 2.2. Under Assumption 1(ii), the expected number of jumps of $\mathbf{J}_T^i$ on $[0,w]$ converges to $F^i(w)$, $i = 1,\ldots,k$. In the asymptotic study of cointegrated processes only the interval $[0,1]$ will matter, and we will then assume that F^i are not identically zero on $[0,1]$.

2.2. Convergence of Product Moments

Under Assumption 2(ii), $\theta_{T,t}$ has the following representation, known as Beveridge–Nelson decomposition (cf. Phillips and Solo, 1992, Lem. 2.1):

$$\theta_{T,t} = \Theta(1)\mu_{T,t} - \sum_{i=0}^{t-1} \Theta_i^* \delta_{T,t-i} = \Theta(1)\mu_{T,t} - \Delta\theta_{T,t}^*, \qquad t = 1,2,\ldots, \tag{3}$$

where $\Theta_i^* := \sum_{j=i+1}^{\infty} \Theta_j$ $(i = 0,1,\ldots)$, $\theta_{T,t}^* := \sum_{i=0}^{t-1} \Theta_i^* \mu_{T,t-i}$ $(t = 0,1,\ldots)$, $\Delta\theta_{T,t}^* := \theta_{T,t}^* - \theta_{T,t-1}^*$ $(t = 1,2,\ldots)$, and $\Theta^*(z) := \sum_{i=0}^{\infty} \Theta_i^* z^i$ is convergent for $|z| < 1 + \delta$ with the same δ as $\Theta(z)$. In (3) the process $\theta_{T,t}$ is decomposed into its permanent (additive-shift) and transitory (temporary-change) components, respectively, $\Theta(1)\mu_{T,t}$ and $-\Delta\theta_{T,t}^*$.

Our first result has three parts, addressing convergence of the permanent component of $\theta_{T,t}$, of its transitory component, and of the process itself. We denote by $\mathrm{diag}(x)$ the diagonal matrix having the vector x on its diagonal.

PROPOSITION 1. *Under Assumptions 1 and 2, it holds that*

(i)
$$\left((\Theta(1)\mathbf{J}_T)', T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} \psi_t' \right)' \xrightarrow{w} ((\Theta(1)\mathbf{J})', (\Psi(1)\mathbf{W})')' \tag{4}$$

in $D_{m_1+m_3}[0,\infty)$, where the components of $\mathbf{J} = (\mathbf{J}^1,\ldots,\mathbf{J}^k)'$ are compound Poisson processes

$$\mathbf{J}^i(u) = \sum_{j=1}^{\mathbf{N}^i(u)} v_j^i, \qquad i = 1,\ldots,k, \qquad u \in [0,\infty),$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$\mathbf{N}^i$ *are jointly independent Poisson processes with intensities* F^i, υ_j^i *are independent across* i,j *and distributed as* η_1^i, *and* $\mathbf{N}^i$ *are independent of* $\{\upsilon_j^i\}$ *for all* i,j. *Further,* $\mathbf{W}$ *is a Wiener process with covariance matrix* Ω, *independent of* $\mathbf{J}$.

(ii) $\Delta\theta_{T,t}^*$ *are uniformly infinitesimal:*

$$\max_{t\le\lfloor Tu\rfloor} P(\|\Delta\theta_{T,t}^*\| > \epsilon) \to 0 \quad \text{as } T\to\infty, \quad \text{for all } u \ge 0 \quad \text{and} \quad \epsilon > 0.$$

The finite-dimensional distributions of the process $\theta_{T,\lfloor T\cdot\rfloor}$ *converge to those of* $\Theta(1)\mathbf{J}$.

(iii) For the process $(\theta'_{T,\lfloor T\cdot\rfloor}, T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi'_t)'$ *to converge weakly in* $D_{m_1+m_3}[0,\infty)$, *it is sufficient that* $\Theta(z) = \Theta(1)\mathrm{diag}(z^{j_1},\dots,z^{j_k})$ *for some* $j_i \ge 0$ $(i=1,\dots,k)$, *i.e., at most one matrix among* $\{\Theta_s\}_{s=0}^{\infty}$ *has nonzero ith column. If* $P(\eta_1^i \ne 0) > 0$ *for all* $i=1,\dots,k$, *this condition is also necessary. In the case of weak convergence, the limit is the right-hand-side expression in (4).*

Remark 2.3. Independence of the components of $\mathbf{J}$ is due to the fact that, with probability approaching one, the components of $\mathbf{J}_T$ have no common jump dates and thus possible dependence between the components of η_t is irrelevant in the limit.

Remark 2.4. The condition in item (iii) for weak convergence of $\theta_{T,\lfloor T\cdot\rfloor}$ does not permit the effect of any $\delta_{T,t}^i$ $(i=1,\dots,k)$ on $\theta_{T,\lfloor T\cdot\rfloor}$ to be gradual. The intuition is that, in the case of a gradual effect, a jump of $\mathbf{J}_T$ is associated with a cluster of jumps of $\theta_{T,\lfloor T\cdot\rfloor}$ and as $T\to\infty$, the time distance between the asymptotically nonnegligible jumps in a cluster vanishes. This, however, is not compatible with convergence in the Skorohod topology (cf. Jacod and Shiryaev, 2003, Exp. VI.1.20). Fortunately, to obtain convergence of product moments, weak convergence of $\theta_{T,\lfloor T\cdot\rfloor}$ is not necessary. The derivations, following Phillips and Solo (1992), can be based on decomposition (3), convergence (4), and Proposition 1(ii).

To state the next theorem, we need some notation. The integral $\int_0^u \mathbf{X}(s\text{-})(d\mathbf{Y}(s))'$ is denoted by $\int_0^u \mathbf{X}(d\mathbf{Y})'$ and is understood as a stochastic integral with respect to a semimartingale (it is an Itô integral if $\mathbf{Y}=\mathbf{W}$, and it is an a.s. finite random sum if $\mathbf{Y}=\mathbf{J}$). The abbreviations $\int \mathbf{X}(d\mathbf{Y})' := \int_0^1 \mathbf{X}(d\mathbf{Y})'$ and $\int \mathbf{X} := \int_0^1 \mathbf{X}(u)\,du$ are also used. The quadratic variation $[\mathbf{J}]_1$ of $\mathbf{J}$ at unity is defined as $[\mathbf{J}]_1 := \mathrm{diag}([\mathbf{J}^1]_1,\dots,[\mathbf{J}^k]_1)$, with $[\mathbf{J}^i]_1 := \sum_{j=1}^{\mathbf{N}^i(1)}(\upsilon_j^i)^2$, $i=1,\dots,k$.

THEOREM 2. *Under Assumptions 1 and 2, the following quantities converge jointly:*

(i) $\sum_{t=1}^{T}\Delta\theta_{T,t}\Delta\phi'_{T,t} \xrightarrow{w} \sum_{i=0}^{\infty}\Theta_i[\mathbf{J}]_1\Phi_i'$;

(ii) $\sum_{t=1}^{T}\theta_{T,t}\Delta\phi'_{T,t} \xrightarrow{w} \Theta(1)\int\mathbf{J}(d\mathbf{J})'\Phi(1)' + \sum_{i=0}^{\infty}\sum_{j=0}^{i}\Theta_j[\mathbf{J}]_1\Phi_i'$;

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

(iii) $T^{-1/2}\sum_{t=1}^{T}\theta_{T,t}\psi_t' \xrightarrow{w} \Theta(1)\int \mathbf{J}(d\mathbf{W})'\Psi(1)';$
(iv) $T^{-1/2}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\psi_i)\Delta\theta_{T,t}' \xrightarrow{w} \Psi(1)\int \mathbf{W}(d\mathbf{J})'\Theta(1)';$
(v) $T^{-3/2}\sum_{t=1}^{T}\theta_{T,t}(\sum_{i=1}^{t}\psi_i)' \xrightarrow{w} \Theta(1)\int \mathbf{J}\mathbf{W}'\Psi(1)';$
(vi) $T^{-1}\sum_{t=1}^{T}\theta_{T,t}\phi_{T,t}' \xrightarrow{w} \Theta(1)\int \mathbf{J}\mathbf{J}'\Phi(1)';$
(vii) $T^{-1}\sum_{t=1}^{T}\theta_{T,t} \xrightarrow{w} \Theta(1)\int \mathbf{J};$
(viii) $T^{-1}\sum_{t=1}^{T}(t-1)\Delta\theta_{T,t}' \xrightarrow{w} \int u(d\mathbf{J}(u))'\Theta(1)'.$

3. LEVEL SHIFTS IN COINTEGRATED PROCESSES

Here we put the results from the previous section to work and study some asymptotic properties of cointegrated VAR processes with superimposed influential level shifts and rare transitory changes. In particular, we discuss the implications for the Johansen (1991) trace test on the cointegration rank. This could be considered a generalization of the Leybourne and Newbold (2000a, 2000b) study on the behavior of the Dickey–Fuller test in the presence of a single break in the level of a univariate process.

3.1. The Basic I(1) Model

First, we summarize briefly the facts that we use about the basic cointegrated VAR model

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \varepsilon_t, \qquad t = 1,\ldots,T; \tag{5}$$

see Johansen (1996) for a general theory of these models. In (5), X_t and ε_t are random $n \times 1$ vectors, with ε_t as in Assumption 1(iii). Additionally, we make the following assumptions.

Assumption 3.

(i) μ is an $n \times 1$ vector,[1] α and β are $n \times r$ $(0 < r < n)$ matrices, and the covariance matrix Ω of ε_t is strictly positive definite.
(ii) The eigenvalues of $I + \beta'\alpha$ lie inside the complex unit circle.
(iii) The initial value X_0 is such that $\beta' X_0 = \sum_{i=0}^{\infty}(I + \beta'\alpha)^i\beta'(\varepsilon_{-i} + \mu)$ and CX_0 is an arbitrary constant, where $C := I_n - \alpha(\beta'\alpha)^{-1}\beta'$ is the so-called long-run impact matrix.

Assumption 3(ii) implies that α and β are of full column rank r and $\beta'\alpha$ is nonsingular. We introduce $\bar{\alpha} := \alpha(\alpha'\alpha)^{-1}$ and similarly for other matrices of full column rank. Further, let $\alpha_\perp$ and $\beta_\perp$ be $n \times (n - r)$ matrices of full column rank such that $\alpha_\perp'\alpha = 0$ and $\beta_\perp'\beta = 0$. Assumption 3(ii) implies that $\alpha_\perp'\beta_\perp$ is nonsingular, $\beta_\perp$ and α span $\mathbb{R}^n$, and $\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta' = I_n$. This identity yields $C = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$ and allows us to decompose any vector in $\mathbb{R}^n$ into the sum of two vectors, respectively, in the span of $\beta_\perp$ and α.

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

The corresponding decomposition of X_t under Assumption 3 is given by Granger's representation theorem (Johansen, 1996, Thm. 4.2) as

$$X_t = C\left(\mu t + \sum_{s=1}^{t} \varepsilon_s + X_0\right) + \alpha(\beta'\alpha)^{-1}\beta' \sum_{i=0}^{\infty} (I + \alpha\beta')^i(\varepsilon_{t-i} + \mu). \tag{6}$$

Thus, we have $\beta' X_t = \sum_{i=0}^{\infty}(I + \beta'\alpha)^i\beta'(\varepsilon_{t-i} + \mu)$ and $\alpha_\perp' X_t = \alpha_\perp'\mu t + \alpha_\perp' \sum_{s=1}^{t} \varepsilon_s + \alpha_\perp' X_0$. The process $\beta' X_t$ is stationary, whereas $\alpha_\perp' X_t$ has a deterministic (for $\tau := C\mu \neq 0$) and a stochastic trend. The directions of β and $\alpha_\perp$ are referred to as cointegrating and noncointegrating, respectively. We wish to do inference on $r = \text{rank}(\beta)$, i.e., on the number of linearly independent stationary relations among the components of X_t. Under Gaussianity of ε_t, the likelihood ratio statistic (known as the trace statistic) for the null hypothesis $r = \rho$ against the alternative of a higher rank is given in Theorem 6.1 of Johansen (1996). Let $\bar{\Delta}X_T := T^{-1}\sum_{t=1}^{T} \Delta X_t$, $\bar{X}_T := T^{-1}\sum_{t=1}^{T} X_{t-1}$, and

$$S_{ij}^x := T^{-1} \sum_{t=1}^{T} U_{i,t} U_{j,t}', \qquad i, j \in \{0, 1\}, \tag{7}$$

with $U_{0,t} = \Delta X_t - \bar{\Delta}X_T$, $U_{1,t} = X_{t-1} - \bar{X}_T$. Then, with $\hat{\lambda}_i^x$ $(\rho + 1 \leq i \leq n)$ denoting the $n - \rho$ smallest solutions of the singular-value problem $|S_{11}^x \lambda - S_{10}^x (S_{00}^x)^{-1} S_{01}^x| = 0$, the trace statistic is $\Lambda^x(\rho) := -T\sum_{i=\rho+1}^{n} \ln(1 - \hat{\lambda}_i^x)$. Its limiting distribution depends on whether $\tau = 0$, and we assume that $\tau \neq 0$ in what follows.

3.2. Introducing Level Shifts and Outliers

We assume that we have observations on a variable $Y_{T,t}$ instead of X_t, where $Y_{T,0} = X_0$,

$$Y_{T,t} = X_t + T^{1/2}\theta_{T,t}, \qquad t = 1, \ldots, T, \tag{8}$$

and $\theta_{T,t}$ satisfies Assumption 2(ii). If no transitory component is present in (3), so that $\theta_{T,t} = \Theta(1)\mu_{T,t}$, equation (8) defines an additive-shift model. Alternatively, if $\Theta(1) = 0$, then only transitory changes are superimposed on X_t; e.g., with $\theta_{T,t} = \Theta_0 \delta_{T,t}$, (8) defines an additive-outlier model. With general Θ_i a variety of smooth shifts can be modeled. The normalization of $\theta_{T,t}$ by $T^{1/2}$ renders it asymptotically influential; it has been employed, e.g., in Doornik, Hendry, and Nielsen (1998, p. 537) and Leybourne and Newbold (2000a, 2000b).

In the analysis of the contribution of $\theta_{T,t}$ to $\beta' Y_{T,t}$, the rank of $\beta'\Theta(1)$ plays a central role, and we denote it by s. If $1 \leq s \leq r - 1$, we have $\beta'\Theta(1) = \beta_1\zeta'$ for some matrices β_1 and ζ of dimensions, respectively, $r \times s$ and $k \times s$ and of full column rank s. In this case introduce $\beta_2 := (\beta_1)_\perp$ of dimensions $r \times (r - s)$ and rank $r - s$ such that $\beta_2'\beta_1 = 0$. If $s = r$ (resp. $s = 0$), let $\beta_1 := I_r$ and $\beta_2 := 0$ (resp. $\beta_2 := I_r$ and $\beta_1 := 0$). Further, to distinguish the noncointe-

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

grating directions in which Y_t is not affected by a linear trend, introduce a matrix γ with columns orthogonal to (β, τ), e.g., $\gamma := P_{\beta_\perp}(I - \tau(\tau' P_{\beta_\perp}\tau)^{-1}\tau')P_{\beta_\perp}$, where $P_{\beta_\perp} = \beta_\perp \bar{\beta}_\perp'$. In what follows we write as if β_1, β_2, and γ are nonzero. If any of them is zero, the terms involving it drop from the corresponding expressions.

Note that $C'\gamma$ and $C'\tau$ span the same space as $\alpha_\perp$, so that $C'\gamma$, $C'\tau$, $\beta\beta_1$, and $\beta\beta_2$ span $\mathbb{R}^n$. We discuss next the properties of $Y_{T,t}$ in the directions of each of these matrices. From (6), (8), and (3), using the identities $\beta_2'\beta'\Theta(1) = 0$, $C\alpha = 0$ and $\tau = C\mu$, we get

$$\begin{aligned}
\bar{\beta}_2'\beta' Y_{T,t} &= && \bar{\beta}_2'\beta' X_t - T^{1/2}\bar{\beta}_2'\beta'\Delta\theta^*_{T,t} \\
\bar{\beta}_1'\beta' Y_{T,t} &= T^{1/2}\zeta'\mu_{T,t} && + \bar{\beta}_1'\beta' X_t - T^{1/2}\bar{\beta}_1'\beta'\Delta\theta^*_{T,t} \\
\bar{\gamma}' C Y_{T,t} &= \bar{\gamma}' C \sum_{s=1}^{t}\varepsilon_s + T^{1/2}\bar{\gamma}' C\Theta(1)\mu_{T,t} && - T^{1/2}\bar{\gamma}' C\Delta\theta^*_{T,t} + \bar{\gamma}' C X_0 \\
\bar{\tau}' C Y_{T,t} &= t + \bar{\tau}' C \sum_{s=1}^{t}\varepsilon_s + T^{1/2}\bar{\tau}' C\Theta(1)\mu_{T,t} && - T^{1/2}\bar{\tau}' C\Delta\theta^*_{T,t} + \bar{\tau}' C X_0.
\end{aligned} \tag{9}$$

Thus, $Y_{T,t}$ may have contributions from the transitory process $\Delta\theta^*_{T,t}$ in all directions. In the directions of $\beta\beta_2$, $Y_{T,t}$ is itself transitory. In those of $\beta\beta_1$ the transitory components are dominated by the level shifts originating from $T^{1/2}\zeta'\mu_{T,t}$. They could be interpreted as shifts in the cointegration means. In the directions of $C'\gamma$ the stochastic trend $\gamma' C\sum_{s=1}^{t}\varepsilon_s$ appears and is dominating together with the level shifts. Finally, in the direction of $C'\tau$, $Y_{T,t}$ is dominated by a linear trend. The consequences for the asymptotic behavior of $Y_{T,t}$ are stated in the next lemma, which is a counterpart of Lemma 10.2 in Johansen (1996).

LEMMA 3. *Define the process* $\mathbf{G}_0$ *such that* $\mathbf{G}_0(u) = ((\bar{\gamma}' C(\mathbf{W}(u) + \Theta(1)\mathbf{J}(u)))', (\zeta'\mathbf{J}(u))', u)'$ *for* $u \in [0,\infty)$, *where the processes* $\mathbf{W}$ *and* $\mathbf{J}$ *are as in Proposition 1. Introduce also* $B_T := (C'\bar{\gamma}, \beta\bar{\beta}_1, T^{-1/2}C'\bar{\tau})$. *Then, under Assumptions 1–3, as* $T \to \infty$,

$$T^{-1/2}B_T' Y_{T,\lfloor T\cdot\rfloor} + B_T'\Delta\theta^*_{T,\lfloor T\cdot\rfloor} \xrightarrow{w} \mathbf{G}_0 \quad \text{in } D_{n-r+s}[0,\infty). \tag{10}$$

Further, if $\bar{Y}_T := T^{-1}\sum_{t=1}^{T} Y_{T,t-1}$ *and the process* $\mathbf{G}$ *satisfies* $\mathbf{G}(u) = \mathbf{G}_0(u) - \int \mathbf{G}_0$, *then*

$$T^{-1/2}B_T'(Y_{T,\lfloor T\cdot\rfloor} - \bar{Y}_T) + B_T'\Delta\theta^*_{T,\lfloor T\cdot\rfloor} \xrightarrow{w} \mathbf{G} \quad \text{in } D_{n-r+s}[0,\infty). \tag{11}$$

From Proposition 1(iii) it follows that the process $T^{-1/2}B_T' Y_{T,\lfloor T\cdot\rfloor}$ does not converge weakly in $D_{n-r+s}[0,\infty)$, unless, in the directions of $C'\gamma$ and $\beta\beta_1$, either no or only additive level shifts are present.

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Remark 3.1. In applied work, finite-order VARs with step dummy variables in the cointegration space and with unrestricted impulse dummy variables are often used; see, e.g., Juselius (1998, Table 5). The analogous specification in our framework is

$$\Delta Y_{T,t} = \mu + \alpha(\beta' Y_{T,t-1} - T^{1/2} A \mu_{T,t-1}) + T^{1/2}\Theta_0 \delta_{T,t} + \varepsilon_t \tag{12}$$

for a matrix A of matching dimensions. We check next what restrictions on the coefficients of $\Theta(z)$ are implied by this specification. From (5), (8), and decomposition (3) we obtain the following unrestricted equation for $\Delta Y_{T,t}$:

$$\Delta Y_{T,t} = \mu + \alpha\beta'(Y_{T,t-1} - T^{1/2}\Theta(1)\mu_{T,t-1}) + T^{1/2}(\Delta\theta_{T,t} + \alpha\beta'\Delta\theta^*_{T,t-1}) + \varepsilon_t.$$

The restriction $\Theta_0\delta_{T,t} = \Delta\theta_{T,t} + \alpha\beta'\Delta\theta^*_{T,t-1}$ is satisfied identically in t for Θ_0 arbitrary, Θ_1 satisfying $\alpha'_\perp\Theta_1 = 0$, and $\Theta_i = \alpha(\beta'\alpha)^{-1}(I + \beta'\alpha)^{i-1}\beta'\Theta_1$ $(i \geq 2)$. Then $C\Delta\theta^*_{T,t} = 0$, and from (9) it follows that in the directions of $C'\gamma$ and $C'\tau$ (i.e., in the noncointegrating directions) $Y_{T,t}$ contains no transitory contribution from $\Delta\theta^*_{T,t}$.

Within the preceding restricted model, $\Theta_1 = \alpha\beta'\Theta_0$ must hold for no level shifts to be present in the cointegrating directions. Then $\Theta(1) = C\Theta_0$, and we obtain the innovational-outlier model $\Delta Y_{T,t} = \mu + \alpha\beta' Y_{T,t-1} + T^{1/2}\Theta_0\delta_{T,t} + \varepsilon_t$. Innovational outliers only cumulate in the directions of $\alpha_\perp$: equations (9) reduce to $\beta' Y_{T,t} = \beta' X_t - T^{1/2}\beta'\Delta\theta^*_{T,t}$ and

$$(\bar\gamma, \bar\tau)' C Y_{T,t} = (0, t)' + (\bar\gamma, \bar\tau)' C \left(\sum_{s=1}^{t} \varepsilon_s + T^{1/2}\Theta_0 \mu_{T,t} + X_0 \right).$$

Remark 3.2. For reference in the next section, we formulate a model with no shifts and no large transitory changes in the cointegrating directions and possibly with both shifts and transitory changes in the noncointegrating directions. The model has coefficients satisfying $\beta'\Theta_i = 0$ or, equivalently, $\Theta_i = C\Theta_i$ $(i \geq 0)$. With $\Theta(1) =: C\Theta$, equations (9) become $\beta' Y_{T,t} = \beta' X_t$ and

$$(\bar\gamma, \bar\tau)' C Y_{T,t} = (0, t)' + (\bar\gamma, \bar\tau)' C \left(\sum_{s=1}^{t} \varepsilon_s + T^{1/2}\Theta\mu_{T,t} - T^{1/2}\Delta\theta^*_{T,t} + X_0 \right).$$

3.3. The Trace Test

Lemma 10.3 in Johansen (1996) states the weak limits of S^x_{ij}, $i,j \in \{0,1\}$, introduced in (7). Let S_{ij} denote the counterparts of S^x_{ij} computed from the observable $Y_{T,t}$ instead of X_t and let $\Lambda(\rho)$ be the corresponding trace statistics, i.e., $\Lambda(\rho) := -T\sum_{i=\rho+1}^{n} \ln(1 - \hat\lambda_i)$, where $\hat\lambda_i$ are the $n - \rho$ smallest roots of $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}|$. Next we apply Theorem 2 to characterize the asymptotic behavior of S_{ij}, and subsequently we use it to study that of $\Lambda(\rho)$.

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

LEMMA 4. *Let the processes* $\mathbf{W}$ *and* $\mathbf{J}$ *be as in Proposition 1. Under Assumptions 1–3, and with* $\Sigma_{00} := \mathrm{Cov}(\Delta X_1, \Delta X_1)$, $\Sigma_{\beta\beta} := \mathrm{Cov}(\beta' X_1, \beta' X_1)$, $\Sigma_{\beta 0} := \mathrm{Cov}(\beta' X_0, \Delta X_1)$, $\mathbf{R}_0 := C\mathbf{W} + \Theta(1)\mathbf{J}$, *and* $R_1 := -C\Omega\beta(\alpha'\beta)^{-1}\alpha'$, *it holds that, as* $T \to \infty$,

$$S_{00} \xrightarrow{w} \Sigma_{00} + \sum_{i=0}^{\infty} \Theta_i [\mathbf{J}]_1 \Theta_i' =: \tilde{\Sigma}_{00},$$

$$\bar{\beta}_2' \beta' S_{11} \beta \bar{\beta}_2 \xrightarrow{w} \bar{\beta}_2' \left(\Sigma_{\beta\beta} + \beta' \sum_{i=0}^{\infty} \Theta_i^* [\mathbf{J}]_1 \Theta_i^{*\prime} \beta \right) \bar{\beta}_2 =: \tilde{\Sigma}_{\beta\beta},$$

$$\bar{\beta}_2' \beta' S_{10} \xrightarrow{w} \bar{\beta}_2' \left(\Sigma_{\beta 0} + \beta' \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \Theta_j [\mathbf{J}]_1 \Theta_i' \right) =: \tilde{\Sigma}_{\beta 0},$$

$$B_T' S_{10} \xrightarrow{w} \int \mathbf{G}(d\mathbf{R}_0)' + (C'\bar{\gamma}, \beta\bar{\beta}_1, 0)' \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \Theta_j [\mathbf{J}]_1 \Theta_i' + ((\bar{\gamma}' R_1)', (\bar{\beta}_1' \Sigma_{\beta 0})', 0)' =: \tilde{\Sigma}_{B0},$$

$$B_T' S_{11} \beta \bar{\beta}_2 = O_P(1),$$

$$T^{-1} B_T' S_{11} B_T \xrightarrow{w} \int \mathbf{G}\mathbf{G}'.$$

An important difference from Lemma 10.3 of Johansen (1996) is that, in the s linearly independent directions of $\beta\beta_1$, we need the same normalization as in those of $C'\gamma$. The reason is that $\bar{\beta}_1' \beta' Y_{T,t}$ is influenced by $T^{1/2} \zeta' \mu_{T,t}$ (see (9)), which is of the same magnitude order as the stochastic trend $\gamma' C \sum_{s=1}^{t} \varepsilon_s$. The next theorem states that, as a consequence, the usual sequential testing procedure on the cointegration rank (see Johansen, 1996, Sect. 12.2) does not reject with probability approaching one the ranks from $r - s$ to $r - 1$ (at least at some significance levels), but only the ranks not exceeding $r - s - 1$. In the case $s = 0$ ranks lower than r are rejected with probability approaching one, but the process $\mathbf{J}$ enters the limiting distribution of $\Lambda(r)$ and hence affects the asymptotic critical values of the trace test.

THEOREM 5. *Let the distribution of* η_1 *be absolutely continuous and consider convergence under Assumptions 1–3 and under the hypothesis that* $\mathrm{rank}(\beta) = r$ *and* $\mathrm{rank}(\beta'\Theta(1)) = s$.

(i) It holds, for $r \geq 1$, *that*

$$P(\Lambda(0) > q_0, \ldots, \Lambda(r - s - 1) > q_{r-s-1}) \to 1 \quad \text{as } T \to \infty \tag{13}$$

for every choice of the constants q_i $(0 \leq i \leq r - s - 1)$. *Further, if* $s > 0$ *and the functions* F^i *from (1) are not identically zero on* $[0,1]$, *then there exist* $\vartheta > 0$ *and* $q_{r-s} < \infty$ *such that*

$$\limsup P(\Lambda(r - s) \geq q_{r-s}) \leq 1 - \vartheta \quad \text{as } T \to \infty.$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

(ii) In the case $s = 0$ the limiting distribution of $\Lambda(r)$ can be expressed through the limits from Lemma 4 as follows:

$$\Lambda(r) \xrightarrow{w} \operatorname{tr}\left\{\widetilde{N}^{1/2}\widetilde{\Sigma}'_{B0}\left[\int \mathbf{G}\mathbf{G}'\right]^{-1}\widetilde{\Sigma}_{B0}\widetilde{N}^{1/2}\right\} \quad \text{as } T \to \infty, \tag{14}$$

where $\widetilde{N} := \widetilde{\Sigma}_{00}^{-1} - \widetilde{\Sigma}_{00}^{-1}\widetilde{\Sigma}'_{\beta 0}[\widetilde{\Sigma}_{\beta 0}\widetilde{\Sigma}_{00}^{-1}\widetilde{\Sigma}'_{\beta 0}]^{-1}\widetilde{\Sigma}_{\beta 0}\widetilde{\Sigma}_{00}^{-1}$.

Theorem 5 is better understood in light of decomposition (3) of $\theta_{T,t}$.

For $s > 0$, failure of consistency in Theorem 5(i) is exclusively due to the additive-shift component in decomposition (3). A similar effect arises in the work of Leybourne and Newbold (2000b), who study a stationary univariate process with a superimposed deterministic level shift of the order of $T^{1/2}$ (i.e., $n = r = s = 1$ and no linear trend is present). Their conclusion is that the Dickey–Fuller unit root test is no longer consistent.

Distribution (14) in general is influenced both by the additive-shift and the transitory component of $\theta_{T,t}$. The literature so far has focused on univariate special cases where only one of these effects is present. For comparison, in Corollaries 6 and 7 we consider the multivariate innovational-outlier and additive-outlier models, for which (14) is influenced only by the additive-shift, respectively, by the transitory component of $\theta_{T,t}$. Then in Corollary 8 we discuss the relatively tractable specification of Remark 3.2, where both effects are present.

COROLLARY 6. *(Innovational outliers). Let the assumptions of Theorem 5 hold, Θ_0 be any $n \times k$ matrix, and $\Theta_i = \alpha(I + \beta'\alpha)^{i-1}\beta'\Theta_0$ $(i \geq 1)$. Then $s = 0$ and*

$$\Lambda(r) \xrightarrow{w} \operatorname{tr}\left\{\int (d\mathbf{Z})\mathbf{V}'\left[\int \mathbf{V}\mathbf{V}'\right]^{-1}\int \mathbf{V}(d\mathbf{Z})'\right\} \quad \text{as } T \to \infty, \tag{15}$$

where $\mathbf{Z} = (I_{n-r} + K\Theta_0[\mathbf{J}]_1\Theta_0'K')^{-1/2}(\mathbf{B} + K\Theta_0\mathbf{J})$ *is an* $(n - r) \times 1$ *stochastic process,* $\mathbf{B} = (\mathbf{B}_1', \mathbf{B}_2)'$ *is a standard Brownian motion in* $n - r$ *dimensions, independent of* $\mathbf{J}$*, and is partitioned into the* $n - r - 1$ *dimensional and the scalar Brownian motions* $\mathbf{B}_1$ *and* $\mathbf{B}_2$*,* $K = (K_1', K_2')'$ *is an* $(n - r) \times n$ *matrix with blocks* $K_1 = (\bar{\gamma}'C\Omega C'\bar{\gamma})^{-1/2}\bar{\gamma}'C$ *and* $K_2 = (\mu'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\mu)^{-1/2}\mu'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'$*,* $\mathbf{V} = (\mathbf{V}_1', \mathbf{V}_2)'$*, with* $\mathbf{V}_1 = \mathbf{B}_1 - \int \mathbf{B}_1 + K_1\Theta_0(\mathbf{J} - \int \mathbf{J})$ *and* $\mathbf{V}_2(u) = u - \frac{1}{2}$.

Distribution (15) is expressed through the same functional as the asymptotic distribution in the absence of outliers (see Johansen, 1996, Thm. 11.1). The difference is that here this functional is evaluated at $\mathbf{Z}$ and $\mathbf{V}$ instead of $\mathbf{B}$ and $\mathbf{F} := ((\mathbf{B}_1 - \int \mathbf{B}_1)', u - \frac{1}{2})'$. A similar univariate result is contained, in a disguised form, in Theorem 1 of Leybourne and Newbold (2000a).

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Remark 3.3. Distribution (15) and that in Theorem 11.1 of Johansen (1996) coincide if and only if $\alpha_\perp' \Theta_0 = 0$, so that $K\Theta_0 = 0$. In this case $\alpha_\perp' Y_{T,t} = \alpha_\perp' X_t$, i.e., in the noncointegrating directions $Y_{T,t}$ equals X_t, whereas in the directions of β there can be outliers. For $\alpha_\perp' \Theta_0 \neq 0$ the additive level shifts in $\alpha_\perp' Y_{T,t}$ give rise to the jump component in the process $\mathbf{V}$ and $\mathbf{Z}$.

Remark 3.4. Recall from Remark 3.1 that, within the basic model (5), innovational outliers give rise to a process $\theta_{T,t}$ that, in the noncointegrating directions, has no transitory component. Moreover, the innovational-outlier specification with $\beta'\Theta_0 = 0$ is equivalent to an additive-shift specification with shifts restricted to the noncointegrating directions. If lagged differences of X_t are added to the basic model, then under the innovational-outlier specification $\theta_{T,t}$ may have a transitory component also in the noncointegrating directions, which changes the functional form of the asymptotic distribution similarly to the following corollaries.

We consider next a simple specification in which $\theta_{T,t}$ has only a transitory component.

COROLLARY 7. *(Additive outliers in the noncointegrating directions). Under the assumptions of Theorem 5, let* $\theta_{T,t} = \Theta\delta_{T,t}$, *i.e.,* $\Theta_0 = \Theta$, $\Theta_1 = -\Theta$, *and* $\Theta_i = 0$ $(i \geq 2)$. *Let further* $\Theta = C\Theta$. *Then* $s = 0$ *and*

$$\Lambda(r) \xrightarrow{w} \operatorname{tr}\left\{K_3^{-1/2}\left[K_4' + \int (d\mathbf{B})\mathbf{F}'\right]\left[\int \mathbf{F}\mathbf{F}'\right]^{-1}\left[\int \mathbf{F}(d\mathbf{B})' + K_4\right]K_3^{-1/2}\right\}$$

as $T \to \infty$, *where* $K_3 = I_{n-r} + 2K\Theta[\mathbf{J}]_1\Theta'K'$, $K_4 = -(\bar{\gamma},0)'\Theta[\mathbf{J}]_1\Theta'K'$, *and the rest of the notation is as in Corollary 6 and its discussion.*

Integrals in the preceding limiting distribution involve the processes $\mathbf{B}$ and $\mathbf{F}$, as when no shifts and outliers are present. However, in the presence of outliers in the noncointegration directions, K_4 introduces "noncentrality" in the limiting distribution. A similar phenomenon in a univariate setting with (frequent instead of large) additive outliers was established by Franses and Haldrup (1994).

Finally, we consider the specification from Remark 3.2, where $\theta_{T,t}$ may have both an additive-shift and a transitory component in the noncointegrating directions.

COROLLARY 8. *Under the assumptions of Theorem 5, let* $\Theta_i = C\Theta_i$ $(i \geq 0)$ *and* $\Theta(1) = C\Theta$. *Then* $s = 0$ *and*

$$\Lambda(r) \xrightarrow{w} \operatorname{tr}\left\{\tilde{K}_3^{-1/2}\left[\tilde{K}_4' + \int (d\tilde{\mathbf{Z}})\tilde{\mathbf{V}}'\right]\left[\int \tilde{\mathbf{V}}\tilde{\mathbf{V}}'\right]^{-1}\left[\int \tilde{\mathbf{V}}(d\tilde{\mathbf{Z}})' + \tilde{K}_4\right]\tilde{K}_3^{-1/2}\right\}$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

as $T \to \infty$, *where* $\tilde{K}_3 = I_{n-r} + K(\sum_{i=0}^{\infty} \Theta_i [\mathbf{J}]_1 \Theta_i')K'$, $\tilde{K}_4 = (\bar{\gamma}, 0)' \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \Theta_j [\mathbf{J}]_1 \Theta_i' K'$, $\tilde{\mathbf{Z}} = \mathbf{B} + K\Theta\mathbf{J}$, $\tilde{\mathbf{V}} = (\tilde{\mathbf{V}}_1', \mathbf{V}_2)'$ *with* $\tilde{\mathbf{V}}_1 = \mathbf{B}_1 - \int \mathbf{B}_1 + K_1 \Theta(\mathbf{J} - \int \mathbf{J})$, *and the rest of the notation is as in the previous corollaries.*

Thus, for $\alpha_\perp' \Theta \neq 0$ the effect of level shifts in the nonstationary directions manifests itself in $\tilde{K}_3^{-1/2} \tilde{\mathbf{Z}}$ and $\tilde{\mathbf{V}}$ as in Corollary 6. Additionally, the large transitory changes in $\alpha_\perp' Y_{T,t}$ (see Remark 3.2) introduce in the asymptotic distribution "noncentrality" as in Corollary 7.

NOTE

1. We use similar notation for the parameter μ and the process $\mu_{T,t}$ because both are related to the level of X_t.

REFERENCES

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Caner, M. (1997) Weak convergence to a matrix stochastic integral with stable processes. *Econometric Theory* 13, 506–528.

Chan, N.H. & L.T. Tran (1989) On the first-order autoregressive process with infinite variance. *Econometric Theory* 5, 354–362.

Chen, C. & G. Tiao (1990) Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of Business & Economic Statistics* 8, 83–97.

Daley, D.J & D. Vere-Jones (1988) *An Introduction to the Theory of Point Processes*. Springer-Verlag.

Diebold, F. & A. Inoue (2001) Long memory and regime-switching. *Journal of Econometrics* 105, 131–159.

Doornik, J., D. Hendry, & B. Nielsen (1998) Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys* 12, 533–572.

Engle, R. & A. Smith (1999) Stochastic permanent breaks. *Review of Economics and Statistics* 81, 553–574.

Franses, P.H. & N. Haldrup (1994) The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business & Economic Statistics* 12, 471–478.

Granger, C. & N. Hyung (2004) Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance* 11, 399–421.

Gregory, A. & B. Hansen (1996) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99–126.

Haldrup, N., A. Montanés, & A. Sanso (2005) Measurement errors and outliers in seasonal unit root testing. *Journal of Econometrics* 127, 103–128.

Ibragimov, R. & P.C.B. Phillips (2004) Regression Asymptotics Using Martingale Convergence Methods. Cowles Foundation Discussion paper 1473 (forthcoming in *Econometric Theory*).

Inoue, A. (1999) Tests of cointegrating rank with a trend-break. *Journal of Econometrics* 90, 215–237.

Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes*. Springer-Verlag.

Johansen, S. (1991) Estimation and hypothesis testing on cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johansen, S., R. Mosconi, & B. Nielsen (2000) Cointegration analysis in the presence of structural breaks of the deterministic trend. *Econometrics Journal* 3, 216–249.

Juselius, K. (1998) A structured VAR for Denmark under changing monetary regimes. *Journal of Business & Economic Statistics* 16, 400–411.

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Kurtz, T. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.
Leipus, R. & M.-C. Viano (2003) Long memory and stochastic trend. *Statistics and Probability Letters* 61, 177–190.
Leybourne, S. & P. Newbold (2000a) Behaviour of the standard and symmetric Dickey–Fuller type tests when there is a break under the null hypothesis. *Econometrics Journal* 3, 1–15.
Leybourne, S. & P. Newbold (2000b) Behavior of Dickey–Fuller *t*-tests when there is a break under the alternative hypothesis. *Econometric Theory* 16, 779–789.
Lumsdaine, R. & D. Papell (1997) Multiple trend breaks and the unit-root hypothesis. *Review of Economics and Statistics* 79, 212–218.
Lütkepohl, H., P. Saikkonen, & C. Trenkler (2004) Testing for the cointegration rank of a VAR process with level shift at unknown time. *Econometrica* 72, 647–662.
Ohara, H. (1999) A unit-root test with multiple trend breaks: A theory and an application to US and Japanese macroeconomic time-series. *Japanese Economic Review* 50, 266–290.
Perron, P. (1989) The Great Crash, the oil price shock, and the unit root hypotheses. *Econometrica* 57, 1361–1401.
Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics* 8, 153–162.
Perron, P. & Z. Qu (2004) An Analytical Evaluation of the Log-Periodogram Estimate in the Presence of Level Shifts and Its Implications for Stock Returns Volatility. Manuscript, Department of Economics, Boston University.
Phillips, P.C.B. (1988) Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528–533.
Phillips, P.C.B. (1990) Time series regression with a unit root and infinite-variance errors. *Econometric Theory* 6, 44–62.
Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Rappoport, P. & L. Reichlin (1989) Segmented trends and non-stationary time series. *Economic Journal* 99, 168–177.
Saikkonen, P. & H. Lütkepohl (2000) Testing for the cointegrating rank of a VAR process with structural shifts. *Journal of Business & Economic Statistics* 18, 451–464.

APPENDIX A

For ease in evaluations, we work with the euclidean norm for both vectors and matrices: $\|x\| := (x'x)^{1/2}$ for $x \in \mathbb{R}^n$ and $\|A\| := [\mathrm{tr}(A'A)]^{1/2}$ for $A \in \mathbb{R}^{n\times k}$.

A.1. Functional and Product-Moments Poisson Convergence. Throughout, $D_p[0,\infty)$ is endowed with the Skorohod topology (see Jacod and Shiryaev, 2003, Thm. VI.1.14) and with the induced Borel σ-algebra. This topology can be metrized (as suggested by Prohorov) such that a complete and separable metric space obtains. We specify a metric explicitly only in contexts not regarding completeness, and so we use the Skorohod metric, equivalent to Prohorov's but simpler to handle (Jacod and Shiryaev, 2003, Rmk. VI.1.27).

Lemma A.1 prepares convergence (4). The proof is based on Theorem VII.3.4 of Jacod and Shiryaev (2003). Ibragimov and Phillips (2004) provide a discussion of related methods of proof.

LEMMA A.1. *For a $p \times k$ matrix Q_1 and a $p \times n$ matrix Q_2, let $\mathbf{U}_T := Q_1\mathbf{J}_T + T^{-1/2}Q_2\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t$ and $\mathbf{U} := Q_1\mathbf{J} + Q_2\mathbf{W}$, where the processes $\mathbf{J}$ and $\mathbf{W}$ are defined in Proposition 1. Under Assumptions 1 and 2(i), it holds that $\mathbf{U}_T \xrightarrow{w} \mathbf{U}$ in $D_p[0,\infty)$.*

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Proof. By Donsker's theorem and Assumption 1(iii), $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t \xrightarrow{w} \mathbf{W}$ in $D_n[0,\infty)$. If we show that $\mathbf{J}_T \xrightarrow{w} \mathbf{J}$ in $D_k[0,\infty)$, it will follow from the independence of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t$ and $\mathbf{J}_T$ (Assumption 1(iv)), and from the continuity of $\mathbf{W}$, that $(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t', \mathbf{J}_T')' \xrightarrow{w} (\mathbf{W}', \mathbf{J}')'$ in $D_{n+k}[0,\infty)$, which together with the continuous mapping theorem (CMT) implies the lemma.

By Theorem VII.3.4 of Jacod and Shiryaev (2003), the convergence $\mathbf{J}_T \xrightarrow{w} \mathbf{J}$ in $D_k[0,\infty)$ can be obtained from conditions [Sup-β_3], [$\gamma_3 - [0,\infty)$], and [$\delta_{3,1} - [0,\infty)$] of Jacod and Shiryaev, which here specialize to

$$\sup_{w\in[0,u]}\left\|\sum_{t=1}^{\lfloor Tw\rfloor}E[h(\delta_{T,t})] - \sum_{i=1}^{k}F^i(w)E[h(\eta_1^i e_i)]\right\| \to 0,$$

$$\sum_{t=1}^{\lfloor Tu\rfloor}(E\{h(\delta_{T,t})[h(\delta_{T,t})]'\} - E[h(\delta_{T,t})]E[h(\delta_{T,t})]') \to \sum_{i=1}^{k}F^i(u)E\{h(\eta_1^i e_i)[h(\eta_1^i e_i)]'\},$$

$$\sum_{t=1}^{\lfloor Tu\rfloor}E[g(\delta_{T,t})] \to \sum_{i=1}^{k}F^i(u)E[g(\eta_1^i e_i)]$$

for all $u \in [0,\infty)$, where h is a (continuous) truncation function (Jacod and Shiryaev, 2003, Def. II.2.3) and $g \in C_1(\mathbb{R}^k)$, a subclass of the nonnegative continuous bounded functions $\mathbb{R}^k \to \mathbb{R}$ that are 0 around 0 and whose further properties are not used here (see Jacod and Shiryaev, 2003, Def. VII.2.7). The preceding form of the conditions follows from equations (II.3.14) and (VII.2.49) of Jacod and Shiryaev for the characteristics $\mathbf{J}_T$ and from the expression defining the characteristic function of $\mathbf{J}(u)$, $u \in [0,\infty)$:

$$E\exp(\mathbf{i}v'\mathbf{J}(u)) = \prod_{i=1}^{k}\exp\bigl(F^i(u)(\mathbf{i}v'E[h(\eta_1^i e_i)] + E[e^{\mathbf{i}v'(\eta_1^i e_i)} - 1 - \mathbf{i}v'h(\eta_1^i e_i)])\bigr),$$

where $\mathbf{i}$ is the imaginary unit and $v \in \mathbb{R}^k$, for the characteristics of $\mathbf{J}$; cf. equation (II.4.16) in Jacod and Shiryaev.

Now we check the validity of the conditions. Let $p_{T,t}^i := P(\pi_t = e_i)$ and $p_{T,t}^{>} := P(\|\pi_t\| > 1)$. For any vector or matrix-valued function f on $\mathbb{R}^k$ with $f(0) = 0$, it holds that

$$\|E[f(\delta_{T,t})] - \sum_{i=1}^{k}p_{T,t}^i E[f(\eta_1^i e_i)]\| \le K_f p_{T,t}^{>}, \tag{A.1}$$

where $K_f := \sup_{x\in\mathbb{R}^k}\|f(x)\|$. From here and (2), for $K_f < \infty$ we find

$$\sup_{w\in[0,u]}\left\|\sum_{t=1}^{\lfloor Tw\rfloor}\left\{E[f(\delta_{T,t})] - \sum_{i=1}^{k}p_{T,t}^i E[f(\eta_1^i e_i)]\right\}\right\| \le K_f\sum_{t=1}^{\lfloor Tu\rfloor}p_{T,t}^{>} \to 0$$

as $T \to \infty$. Combining this convergence with the fact that (1) is uniform on bounded intervals (for it involves nondecreasing functions and the limiting function is continuous), we find that $\sum_{t=1}^{\lfloor Tu\rfloor}E[f(\delta_{T,t})] \to \sum_{i=1}^{k}F^i(u)E[f(\eta_1^i e_i)]$ uniformly on bounded intervals. The choices $f = h$ and $f = g$ yield [Sup-β_3] and [$\delta_{3,1} - [0,\infty)$], whereas $f = hh'$ together with $\sum_{t=1}^{\lfloor Tu\rfloor}E[h(\delta_{T,t})]E[h(\delta_{T,t})]' \to 0$ yields [$\gamma_3 - [0,\infty)$]. For the latter convergence, note that

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\left\| \sum_{t=1}^{\lfloor Tu \rfloor} E[h(\delta_{T,t})]E[h(\delta_{T,t})]' \right\| \leq \max_{t \leq \lfloor Tu \rfloor} \|E[h(\delta_{T,t})]\| \sum_{t=1}^{\lfloor Tu \rfloor} \|E[h(\delta_{T,t})]\|,$$

where, first, $\sum_{t=1}^{\lfloor Tu \rfloor} \|E[h(\delta_{T,t})]\| \leq K_h \sum_{t=1}^{\lfloor Tu \rfloor} P(\pi_{T,t} \neq 0) = K_h \sum_{t=1}^{\lfloor Tu \rfloor} (\sum_{i=1}^{k} p_{T,t}^{i} + p_{T,t}^{>})$, which is bounded in view of (1) and (2), and second,

$$\max_{t \leq \lfloor Tu \rfloor} \|E[h(\delta_{T,t})]\| \leq K_h \max_{t \leq \lfloor Tu \rfloor} P(\pi_{T,t} \neq 0) \leq K_h \left(\sum_{i=1}^{k} \max_{t \leq \lfloor Tu \rfloor} p_{T,t}^{i} + \sum_{t=1}^{\lfloor Tu \rfloor} p_{T,t}^{>} \right),$$

which converges to zero in view of (2) and the relation

$$\max_{t \leq \lfloor Tu \rfloor} p_{T,t}^{i} \leq \max_{t \leq \lfloor Tu \rfloor} |F^i(t/T) - F^i((t-1)/T)| + 2 \max_{t \leq \lfloor Tu \rfloor} \left| \sum_{s=1}^{t} p_{T,s}^{i} - F^i(t/T) \right| \to 0, \quad \textbf{(A.2)}$$

in its turn implied by the uniform continuity of F^i and the uniformity of (1) on $[0,u]$. ■

The suggestions of a referee were useful in reducing several remarks to the following lemma.

LEMMA A.2. *Under Assumptions 1 and 2, it holds that*

(i) $\max_{t \leq \lfloor Tu \rfloor} P(\|\Delta\theta_{T,t}\| > \varepsilon) \to 0$ *for all* $\varepsilon > 0$ *and* $u \geq 0$*;*
(ii) $\sum_{t=1}^{T} \Delta\theta_{T,t} \psi_t' = O_P(1)$.

Proof. Note first that $\|\delta_{T,t}\| \leq \mathbf{1}_{\{\pi_{T,t} \neq 0\}} \|\eta_t\|$, with $\pi_{T,t}$ independent of η_t, so that

$$\max_{t \leq \lfloor Tu \rfloor} \|\delta_{T,t}\| \leq \sum_{t=1}^{\lfloor Tu \rfloor} \|\delta_{T,t}\| \leq \sum_{t=1}^{\lfloor Tu \rfloor} \mathbf{1}_{\{\pi_{T,t} \neq 0\}} \|\eta_t\| = O_P(1), \quad \textbf{(A.3)}$$

because the array $\{\mathbf{1}_{\{\pi_{T,t} \neq 0\}}\}$ satisfies the one-dimensional version of (1) with $F = \sum_{i=1}^{k} F^i$ on the right-hand side, and by Lemma A.1 and the CMT, $\sum_{t=1}^{\lfloor Tu \rfloor} \mathbf{1}_{\{\pi_{T,t} \neq 0\}} \|\eta_t\|$ converges in distribution to a compound Poisson variable for every $u \geq 0$. Further, for $t \leq \lfloor Tu \rfloor$,

$$\|\Delta\theta_{T,t}\| = \left\| \sum_{i=0}^{t-1} \Theta_i \delta_{T,t-i} \right\| \leq \left\| \sum_{i=0}^{(t-1) \wedge \lfloor \ln T \rfloor} \Theta_i \delta_{T,t-i} \right\| + \left(\max_{s \leq \lfloor Tu \rfloor} \|\delta_{T,s}\| \right) \sum_{i=\lfloor \ln T \rfloor}^{\infty} \|\Theta_i\|,$$

and because $\sum_{i=\lfloor \ln T \rfloor}^{\infty} \|\Theta_i\| = o(1)$ as a tail of a convergent series, we find using (A.3) that

$$\max_{t \leq \lfloor Tu \rfloor} P(\|\Delta\theta_{T,t}\| > \varepsilon) \leq \max_{t \leq \lfloor Tu \rfloor} P\left(\left\| \sum_{i=0}^{(t-1) \wedge \lfloor \ln T \rfloor} \Theta_i \delta_{T,t-i} \right\| > \varepsilon/2 \right) + P(o_P(1) > \varepsilon/2) \quad \textbf{(A.4)}$$

for all $\varepsilon > 0$. On the other hand, $\{\sum_{i=0}^{(t-1) \wedge \lfloor \ln T \rfloor} \Theta_i \delta_{T,t-i}\| > \varepsilon/2\} \subseteq \cup_{i=0}^{(t-1) \wedge \lfloor \ln T \rfloor} \{\pi_{T,t-i} \neq 0\}$, and $P(\pi_{T,t-i} \neq 0) = \sum_{j=1}^{k} p_{T,t-i}^{j} + p_{T,t-i}^{>}$ with the notation of the previous proof, so that

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\max_{t\le\lfloor Tu\rfloor} P\left(\left\|\sum_{i=0}^{(t-1)\wedge\lfloor \ln T\rfloor}\Theta_i\delta_{T,t-i}\right\|>\varepsilon/2\right)\le\sum_{j=1}^{k}\left(\max_{\lfloor \ln T\rfloor<t\le\lfloor Tu\rfloor}\sum_{i=0}^{\lfloor \ln T\rfloor}p^{j}_{T,t-i}\right)+\sum_{t=1}^{\lfloor Tu\rfloor}p^{>}_{T,t}\to 0$$

by an evaluation similar to (A.2) for the maxima on the right-hand side and by (2). Combined with (A.4), this completes the proof of (i).

Next, as $\|\delta_{T,t}\|=\mathbf{1}_{\{\pi_{T,t}\neq 0\}}\|\delta_{T,t}\|$,

$$\left\|\sum_{t=1}^{T}\Delta\theta_{T,t}\psi_t'\right\|\le\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\|\Theta_i\|\,\|\delta_{T,t-i}\|\right)\|\psi_t\|\le\left(\max_{t\le T}\|\delta_{T,t}\|\right)\sigma_T,$$

where $\max_{t\le T}\|\delta_{T,t}\|=O_P(1)$ by (A.3) and $\sigma_T:=\sum_{t=1}^{T}(\sum_{i=0}^{t-1}\|\Theta_i\|\mathbf{1}_{\{\pi_{T,t-i}\neq 0\}})\|\psi_t\|=O_P(1)$ by Markov's inequality and the evaluation

$$E\sigma_T=E\|\psi_1\|\sum_{t=1}^{T}\sum_{i=0}^{t-1}\|\Theta_i\|P(\pi_{T,t-i}\neq 0)\le E\|\psi_1\|\left(\sum_{i=0}^{\infty}\|\Theta_i\|\right)\sum_{t=1}^{T}P(\pi_{T,t}\neq 0)=O(1),$$

the equality because $\{\psi_t\}$ is stationary and independent of $\{\pi_{T,t}\}$, and boundedness by (1) and (2) and because $E\|\psi_1\|<\infty$. This proves (ii). ■

For reference in what follows, we note that, similarly to (3), the following Beveridge–Nelson decompositions hold for the processes $\phi_{T,t}$ and ψ_t satisfying Assumption 2(ii) and (iii):

$$\phi_{T,t}=\Phi(1)\mu_{T,t}-\Delta\phi^{*}_{T,t}\quad\text{and}\quad\psi_t=\Psi(1)\varepsilon_t-\Delta\psi^{*}_{t},\qquad t=1,2,\ldots \tag{A.5}$$

with $\phi^{*}_{T,t}:=\sum_{i=0}^{t-1}\Phi^{*}_i\mu_{T,t-i}$, $\psi^{*}_t:=\sum_{i=0}^{\infty}\Psi^{*}_i\varepsilon_{t-i}$ $(t=0,1,\ldots)$, $\Phi^{*}_i:=\sum_{j=i+1}^{\infty}\Phi_j$, and $\Psi^{*}_i:=\sum_{j=i+1}^{\infty}\Psi_j$ $(i=0,1,\ldots)$.

Proof of Proposition 1. By (A.5), $\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi_t=\Psi(1)\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t-\psi^{*}_{\lfloor T\cdot\rfloor}+\psi^{*}_0$. Thus, item (i) obtains from Lemma A.1 with $Q_1=(\Theta(1)',0_{k\times m_3})'$ and $Q_2=(0_{n\times m_1},\Psi(1)')'$ and from the relation $\max_{0\le t\le\lfloor Tw\rfloor}\|T^{-1/2}\psi^{*}_t\|=o_P(1)$ for all $w\ge 0$ (see, e.g., Johansen, 1996, p. 251, eqn. (B.17)). The first part of item (ii) reproduces Lemma A.2(i) for $\Delta\theta^{*}_{T,t}$. It implies the finite-dimensional convergence of $(0_{1\times m_1},\Delta\theta^{*\prime}_{T,\lfloor T\cdot\rfloor})'$ to $0\in\mathbb{R}^{2m_1}$, which together with (i) yields finite-dimensional convergence of $((\Theta(1)\mathbf{J}_T)',\theta'_{T,\lfloor T\cdot\rfloor})'=(\Theta(1)',\Theta(1)')'\mathbf{J}_T-(0_{1\times m_1},\Delta\theta^{*\prime}_{T,\lfloor T\cdot\rfloor})'$ to $(\Theta(1)',\Theta(1)')'\mathbf{J}$.

For the sufficiency part of item (iii), assuming $\Theta(z)=\Theta(1)\mathrm{diag}(z^{j_1},\ldots,z^{j_k})$, we first show that $\theta_{T,\lfloor T\cdot\rfloor}\xrightarrow{w}\Theta(1)\mathbf{J}$ in $D_{m_1}[0,\infty)$. Specifically, fix an arbitrary $w\in[0,\infty)$ and let δ_w be the Skorohod distance in $D_{m_1}[0,w]$ between the restrictions of $\theta_{T,\lfloor T\cdot\rfloor}$ and $\Theta(1)\mathbf{J}_T$ to $[0,w]$ (see Billingsley, 1999, p. 124, eqns. (12.11)–(12.13)) with $x,y\in D_{m_1}[0,w]$ and sup over $t\in[0,w]$. We argue that $\delta_w=o_P(1)$, which together with (4) and with the almost sure continuity of $\mathbf{J}$ at w implies that $\theta_{T,\lfloor T\cdot\rfloor}\xrightarrow{w}\Theta(1)\mathbf{J}$ according to Theorems 16.7 and 3.1 of Billingsley (1999).

Define $j:=\max_{i=1,\ldots,k}\{j_i\}$. Let $v_1<\cdots<v_{n(T)}$ be the jump times of $\mathbf{J}_T$ on $[0,w]$, with $n(T)\le\lfloor Tw\rfloor$ random, and let $v_0:=0$ and $v_{n(T)+1}:=w$. Consider the event $\mathcal{B}_T$ that $v_{s+1}-v_s\ge(j+2)/T$, $s=0,\ldots,n(T)$, and none among $\{\pi_{T,t}\}_{t=1}^{\lfloor Tw\rfloor}$ has more than one nonzero component; then $P(\mathcal{B}_T)\to 1$ under Assumption 1. For outcomes in $\mathcal{B}_T$, let $l_s:=j_i$ $(s=1,\ldots,n(T))$ if at time v_s a jump in component i of $\mathbf{J}_T$ occurs and note that the intervals $(v_s-1/T,v_s+(l_s+1)/T)$, $s=1,\ldots,n(T)$, are disjoint. Define λ as the function on $[0,w]$ that interpolates linearly between $\lambda(0)=0$, $\lambda(v_s-1/T)=v_s-1/T$,

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$\lambda(v_s) = v_s + l_s/T$, $\lambda(v_s + (l_s+1)/T) = v_s + (l_s+1)/T$ $(s = 1, \ldots, n(T))$, and $\lambda(w) = w$. Then λ is a strictly increasing continuous map of $[0,w]$ onto $[0,w]$, $\sup_{u\in[0,w]} |\lambda(u) - u| \le j/T$ and, for outcomes in $\mathcal{B}_T$, $\sup_{u\in[0,w]} \|\theta_{T,\lfloor T\lambda(u)\rfloor} - \Theta(1)\mathbf{J}_T(u)\| = 0$. Thus, for such outcomes the definition of the Skorohod distance on $D_{m_1}[0,w]$ gives $\delta_w \le j/T$. Because $P(\mathcal{B}_T) \to 1$, we conclude that $\delta_w = o_P(1)$, which completes the proof of the convergence $\theta_{T,\lfloor T\cdot\rfloor} \xrightarrow{w} \Theta(1)\mathbf{J}$ in $D_{m_1}[0,\infty)$.

Together with this convergence we have that $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi_t \xrightarrow{w} \Psi(1)\mathbf{W}$. As $\theta_{T,\lfloor T\cdot\rfloor}$ is independent of $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi_t$, and $\Psi(1)\mathbf{W} \in C_{m_3}[0,\infty)$, it follows that $(\theta'_{T,\lfloor T\cdot\rfloor}, T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi'_t)'$ converges weakly to $((\Theta(1)\mathbf{J})', (\Psi(1)\mathbf{W})')'$ in $D_{m_1+m_3}[0,\infty)$.

Finally, we turn to the necessity part of (iii). Take an $i \in \{1,\ldots,k\}$ and a $w_i \in (0,\infty)$ such that $F^i(w_i/2) > 0$; it exists by Assumption 1. We argue first that if $\theta_{T,\lfloor T\cdot\rfloor}$ converges weakly in $D_{m_1}[0,\infty)$, then, with $\tilde{}$ denoting the restriction of a process to $[0,w_i]$,

$$((\Theta(1)\tilde{\mathbf{J}}_T)', \tilde{\theta}'_{T,\lfloor T\cdot\rfloor})' \xrightarrow{w} (\Theta(1)', \Theta(1)')'\tilde{\mathbf{J}} \quad \text{in } D_{m_1}[0,w_i] \times D_{m_1}[0,w_i]. \tag{A.6}$$

As the Skorohod topology on $D_{m_1}[0,\infty)$ can be metrized such that a complete and separable metric space obtains, we can invoke Prohorov's theorem (e.g., Jacod and Shiryaev, 2003, Thm. VI.3.5), by which, first, if $\theta_{T,\lfloor T\cdot\rfloor}$ converges weakly in $D_{m_1}[0,\infty)$, then the sequence of induced probability measures is tight, and second, because $\Theta(1)\mathbf{J}_T$ converges weakly in $D_{m_1}[0,\infty)$ by (4), its induced sequence of probability measures is tight too. But then the sequence of measures induced by $((\Theta(1)\mathbf{J}_T)', \theta'_{T,\lfloor T\cdot\rfloor})'$ on $D_{m_1}[0,\infty) \times D_{m_1}[0,\infty)$ is also tight, for in product spaces coordinate-wise tightness implies tightness. Hence, by the converse part of Prohorov's theorem, this sequence is relatively compact. This fact in conjunction with the finite-dimensional convergence of $((\Theta(1)\mathbf{J}_T)', \theta'_{T,\lfloor T\cdot\rfloor})'$ to $(\Theta(1)', \Theta(1)')'\mathbf{J}$ implies that $((\Theta(1)\mathbf{J}_T)', \theta'_{T,\lfloor T\cdot\rfloor})' \xrightarrow{w} (\Theta(1)', \Theta(1)')'\mathbf{J}$ in $D_{m_1}[0,\infty) \times D_{m_1}[0,\infty)$. Because $\mathbf{J}$ is a.s. continuous at w_i, (A.6) follows from the last convergence and Theorem 16.7 in Billingsley (1999).

We do not use completeness in what follows, and we return to the Skorohod metric d on $D_{m_1}[0,w_i]$. As a metric, d is continuous on its domain, and in view of (A.6) and the CMT, $d(\Theta(1)\tilde{\mathbf{J}}_T, \tilde{\theta}_{T,\lfloor T\cdot\rfloor}) \xrightarrow{w} d(\Theta(1)\tilde{\mathbf{J}}, \Theta(1)\tilde{\mathbf{J}}) = 0$, and the same convergence holds in probability.

The implication for the form of $\Theta(z)$ is discussed next. If all Θ_s $(s = 0,1,\ldots)$ have ith column $\Theta_{s,\cdot i} = 0$, define $j_i := 0$. Otherwise, let $j_i := \min\{s \in \mathbb{N} \cup \{0\} : \Theta_{s,\cdot i} \ne 0\}$ and assume that $\Theta_{s,\cdot i} \ne 0$ for some $s > j_i$, so that $l := \min\{s > j_i : \Theta_{s,\cdot i} \ne 0\}$ is defined over a nonempty set. Consider also the event $\mathcal{C}_T := \{\tilde{\mathbf{J}}_T$ has one jump, it occurs at a point in $(0, w_i/2]$ and takes place in component i only$\}$. For outcomes in $\mathcal{C}_T$, denote the relative time of the single jump by v and its size by $\eta = \eta^i e_i$. It follows from (3) that, for $u \in [0,w_i]$,

$$\tilde{\theta}_{T,\lfloor Tu\rfloor} = \mathbf{1}_{\{u\ge v\}}(\Theta(1) - \Theta^*_{\lfloor Tu\rfloor - \lfloor Tv\rfloor})\eta \quad \text{and} \quad \Theta(1)\tilde{\mathbf{J}}_T(u) = \mathbf{1}_{\{u\ge v\}}\Theta(1)\eta.$$

Thus, if λ is any continuous strictly increasing map of $[0,w_i]$ onto $[0,w_i]$, we get for $u \ge v$ that

$$\Theta(1)\tilde{\mathbf{J}}_T(\lambda(u)) - \tilde{\theta}_{T,\lfloor Tu\rfloor} = \mathbf{1}_{\{\lambda(u)\ge v\}}\Theta^*_{\lfloor Tu\rfloor - \lfloor Tv\rfloor}\eta + \mathbf{1}_{\{\lambda(u)<v\}}(\Theta^*_{\lfloor Tu\rfloor - \lfloor Tv\rfloor} - \Theta(1))\eta. \tag{A.7}$$

For $\lfloor Tw_i/2\rfloor \ge l$ and outcomes in $\mathcal{C}_T$, $v + l/T \le w_i$ holds, and hence, $\lambda(v + s/T)$ is well defined for $s \in \{j_i, l-1, l\}$. Either $\lambda(v + j_i/T) < v$ or $\lambda(v + j_i/T) \ge v$. In the former case

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\sup_{u\in[0,w_i]}\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(u))-\tilde{\theta}_{T,\lfloor Tu\rfloor}\|\geq\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(v+j_i/T))-\tilde{\theta}_{T,\lfloor Tv+j_i\rfloor}\|$$

$$=\|(\Theta^*_{j_i}-\Theta(1))\eta\|$$

$$=|\eta^i|\left\|\sum_{s=0}^{j_i}\Theta_{s,\cdot i}\right\|=|\eta^i|\|\Theta_{j_i,\cdot i}\|$$

by (A.7), the definition of $\Theta^*_{j_i}$, and the choice of j_i. If $\lambda(v+j_i/T)\geq v$, then $\lambda(v+s/T)\geq v$ for $s\in\{l-1,l\}$ because $l-1\geq j_i$ and λ is increasing. From here and (A.7), in this case

$$\sup_{u\in[0,w_i]}\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(u))-\tilde{\theta}_{T,\lfloor Tu\rfloor}\|\geq\max_{s\in\{l-1,l\}}\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(v+s/T))-\tilde{\theta}_{T,\lfloor Tv+s\rfloor}\|$$

$$=|\eta^i|\max\{\|\Theta^*_{l-1,\cdot i}\|,\|\Theta^*_{l,\cdot i}\|\}.$$

By combining the cases $\lambda(v+j_i/T)<v$ and $\lambda(v+j_i/T)\geq v$, it follows that

$$\sup_{u\in[0,w_i]}\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(u))-\tilde{\theta}_{T,\lfloor Tu\rfloor}\|\geq|\eta^i|\min\{\|\Theta_{j_i,\cdot i}\|,\max\{\|\Theta^*_{l-1,\cdot i}\|,\|\Theta^*_{l,\cdot i}\|\}\}$$

for outcomes in $\mathcal{C}_T$. As $\Theta_{j_i,\cdot i}\neq 0$, $\Theta^*_{l-1,\cdot i}=\Theta_{l,\cdot i}+\Theta^*_{j_i,\cdot i}$ and $\Theta_{l,\cdot i}\neq 0$, the preceding minimum is positive. Further, assuming that $P(|\eta^i|>0)>0$, by continuity of P from above there exists an $\epsilon>0$ such that $P(|\eta^i|>\epsilon)>0$. For outcomes in $\mathcal{C}_T\cap\{|\eta^i|>\epsilon\}$ we get that $\sup_{u\in[0,w_i]}\|\Theta(1)\tilde{\mathbf{J}}_T(\lambda(u))-\tilde{\theta}_{T,\lfloor Tu\rfloor}\|$ is bounded from below by a positive constant independent of λ and hence for such outcomes $d(\Theta(1)\tilde{\mathbf{J}}_T,\tilde{\theta}_{T,\lfloor T\cdot\rfloor})$ is bounded from below by the same constant. Because $d(\Theta(1)\tilde{\mathbf{J}}_T,\tilde{\theta}_{T,\lfloor T\cdot\rfloor})\xrightarrow{P}0$, this is possible only if $P(\mathcal{C}_T\cap\{|\eta^i|>\epsilon\})\to 0$, which is not the case. Indeed, $P(\mathcal{C}_T\cap\{|\eta^i|>\epsilon\})=P(\mathcal{C}_T)P(|\eta^i|>\epsilon)$ using the independence of jump dates and jump sizes, $P(\mathcal{C}_T)\to F^i(w_i/2)\exp(-\sum_{j=1}^k F^j(w_i))>0$ by the choice of w_i, and $P(|\eta^i|>\epsilon)$ was chosen positive. Therefore, the assumption that $\Theta_{s,\cdot i}\neq 0$ for some $s>j_i$ is wrong, and by the arbitrariness of i, $\Theta(z)=\Theta(1)\mathrm{diag}(z^{j_1},\ldots,z^{j_k})$ as asserted. ■

Proof of Theorem 2. In this proof, for arrays the dependence on T is subsumed.

We use Theorem 2.7 of Kurtz and Protter (1991), which provides a sufficient condition for convergence of integrals whose integrators (like $\mathbf{J}_T$) are constant except for finitely many discontinuities on bounded intervals. First, the theorem requires, for every $u>0$, the number of jumps of $\mathbf{J}_T$ on $[0,u]$ to be stochastically bounded. Here this number is at most $\sum_{i=1}^k\sum_{t=1}^{\lfloor Tu\rfloor}\pi^i_t$, which converges weakly to a Poisson random variable. The second requirement in our context is the convergence $(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi'_t,\mathbf{J}'_T,\mathbf{J}'_T,\mathbf{J}'_T)'\xrightarrow{w}((\Psi(1)\mathbf{W})',\mathbf{J}',\mathbf{J}',\mathbf{J}')'$ in $D_{n+3k}[0,\infty)$, and it is fulfilled by Lemma A.1. Thus, by the theorem of Kurtz and Protter,

$$\int_0^{(\cdot)}\left(T^{-1/2}\sum_{t=1}^{\lfloor Ts-\rfloor}\psi'_t,\mathbf{J}_T(s-)'\right)'(d\mathbf{J}_T(s))'\xrightarrow{w}\int_0^{(\cdot)}((\Psi(1)\mathbf{W})',\mathbf{J}')'(d\mathbf{J})'$$

jointly with $(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\psi'_t,\mathbf{J}'_T)'\xrightarrow{w}((\Psi(1)\mathbf{W})',\mathbf{J}')'$. The process $\int_0^{(\cdot)}((\Psi(1)\mathbf{W})',\mathbf{J}')'(d\mathbf{J})'$ is a.s. continuous at 1, and from Theorem VI.2.3 in Jacod and Shiryaev (2003) and the CMT it follows that

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\int_0^1 \left(T^{-1/2} \sum_{t=1}^{\lfloor Ts- \rfloor} \psi_t', \mathbf{J}_T(s-)' \right)' (d\mathbf{J}_T(s))' \xrightarrow{w} \int_0^1 ((\Psi(1)\mathbf{W})', \mathbf{J}')'(d\mathbf{J})', \tag{A.8}$$

still jointly with $(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot \rfloor}\psi_t', \mathbf{J}_T')' \xrightarrow{w} ((\Psi(1)\mathbf{W})', \mathbf{J}')'$. As the left-hand sides of convergences (i) to (vi) are, up to $o_P(1)$ terms, continuous functions of $\int_0^1 (T^{-1/2}\sum_{t=1}^{\lfloor Ts- \rfloor}\psi_t', \mathbf{J}_T(s-)')'(d\mathbf{J}_T(s))'$, and $(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot \rfloor}\psi_t', \mathbf{J}_T')'$, these convergences are also joint. The details follow.

To show convergence (i), note that $\sum_{t=1}^T \Delta\theta_t \Delta\phi_t' = \sum_{t=1}^T \sum_{i=0}^{t-1} \Theta_i \delta_{t-i}(\Phi_i \delta_{t-i})' + o_P(1)$. Here $o_P(1)$ stands for $\sum_{t=1}^T \sum_{0\le i\ne j\le t-1} \Theta_i \delta_{t-i}(\Phi_j \delta_{t-j})'$, as its norm does not exceed

$$\sum_{t=1}^T \sum_{0\le i\ne j\le t-1} \|\Theta_i\| \|\Phi_j\| \|\delta_{t-i}\| \|\delta_{t-j}\| \le \left(\max_{t\le T} \|\delta_t\| \right)^2 \xi_T,$$

where $\max_{t\le T}\|\delta_t\| = O_P(1)$ by (A.3) and $\xi_T := \sum_{t=1}^T \sum_{0\le i\ne j\le t-1} \|\Theta_i\| \|\Phi_j\| \mathbf{1}_{\{\pi_{t-i}\ne 0\}} \mathbf{1}_{\{\pi_{t-j}\ne 0\}}$ is $o_P(1)$ by Markov's inequality and the evaluation

$$E\xi_T \le \max_{t\le T} P(\pi_t \ne 0) \sum_{t=1}^T P(\pi_t \ne 0) \sum_{i=0}^\infty \|\Theta_i\| \sum_{j=0}^\infty \|\Phi_j\| \to 0,$$

the inequality using the independence of π_{t-i} and π_{t-j} for $i \ne j$, and the convergence from (A.2), (1), and (2). Further, we have the identity

$$\sum_{t=1}^T \sum_{i=0}^{t-1} \Theta_i \delta_{t-i}(\Phi_i \delta_{t-i})' = \sum_{i=0}^\infty \Theta_i \left(\sum_{t=1}^T \delta_t \delta_t' \right) \Phi_i' - \sum_{t=0}^{T-1} \sum_{i=t+1}^\infty \Theta_i \delta_{T-t}(\Phi_i \delta_{T-t})'.$$

Similarly to the previous argument,

$$\left\| \sum_{t=0}^{T-1} \sum_{i=t+1}^\infty \Theta_i \delta_{T-t}(\Phi_i \delta_{T-t})' \right\| \le \left(\max_{t\le T} \|\delta_t\| \right)^2 \sum_{t=0}^{T-1} \mathbf{1}_{\{\pi_{T-t}\ne 0\}} \sum_{i=t+1}^\infty \|\Theta_i\| \|\Phi_i\| = o_P(1)$$

because $\max_{t\le T}\|\delta_t\| = O_P(1)$ and

$$E\left[\sum_{t=0}^{T-1} \mathbf{1}_{\{\pi_{T-t}\ne 0\}} \sum_{i=t+1}^\infty \|\Theta_i\| \|\Phi_i\| \right] \le \max_{t\le T} P(\pi_t \ne 0) \sum_{t=0}^\infty \sum_{i=t+1}^\infty \|\Theta_i\| \|\Phi_i\| \to 0.$$

We can conclude that $\sum_{t=1}^T \Delta\theta_t \Delta\phi_t' = \sum_{i=0}^\infty \Theta_i (\sum_{t=1}^T \delta_t \delta_t') \Phi_i' + o_P(1)$, which converges as asserted in view of the representation

$$\sum_{t=1}^T \delta_t \delta_t' = \mathbf{J}_T(1)\mathbf{J}_T(1)' - \int \mathbf{J}_T (d\mathbf{J}_T)' - \int (d\mathbf{J}_T)\mathbf{J}_T' \xrightarrow{w} \mathbf{J}(1)\mathbf{J}(1)' - \int \mathbf{J}(d\mathbf{J})' - \int (d\mathbf{J})\mathbf{J}',$$

the convergence of $\mathbf{J}_T(1)$ by Proposition 1(i) and the almost sure continuity of $\mathbf{J}$ at 1, and that of the integrals, by (A.8). Because the components of $\mathbf{J}$ a.s. do not have common jump times, the limit in the last display a.s. equals $[\mathbf{J}]_1 = \mathrm{diag}([\mathbf{J}^1]_1, \ldots, [\mathbf{J}^k]_1)$.

Using (3), the left-hand side of convergence (ii) can be written as

$$\sum_{t=1}^T \theta_t \Delta\phi_t' = \Theta(1) \sum_{t=1}^T \mu_t \Delta\phi_t' - \sum_{t=1}^T \Delta\theta_t^* \Delta\phi_t'. \tag{A.9}$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Summing partially and applying (A.5) gives for the first addend

$$\sum_{t=1}^{T} \mu_t \Delta\phi_t' = \mu_T \phi_T' - \sum_{t=1}^{T} \delta_t \phi_{t-1}'$$

$$= \mu_T \mu_T' \Phi(1)' - \mu_T \Delta\phi_T^{*\prime} - \left(\sum_{t=1}^{T} \delta_t \mu_{t-1}' \right) \Phi(1)' + \sum_{t=1}^{T} \delta_t \Delta\phi_{t-1}^{*\prime}.$$

Here $\mu_T \xrightarrow{w} \mathbf{J}(1)$ by Proposition 1(i) and the almost sure continuity of $\mathbf{J}$ at 1, $\Delta\phi_T^* = o_P(1)$ by Lemma A.2(i), and the summations converge, respectively, to $\int (d\mathbf{J})\mathbf{J}'$ and 0 by (A.8) and Theorem 2(i). It follows that $\sum_{t=1}^{T} \mu_t \Delta\phi_t' \xrightarrow{w} [\mathbf{J}(1)\mathbf{J}(1)' - \int (d\mathbf{J})\mathbf{J}']\Phi(1)'$. For the second addend on the right-hand side of (A.9), Theorem 2(i) yields $\sum_{t=1}^{T} \Delta\theta_t^* \Delta\phi_t' \xrightarrow{w} \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \Theta_j [\mathbf{J}]_1 \Phi_i'$, where $\Theta_i^* = \sum_{j=i+1}^{\infty} \Theta_j$ has been inserted. Combining the limits of the two addends gives that

$$\sum_{t=1}^{T} \theta_t \Delta\phi_t' \xrightarrow{w} \Theta(1)\mathbf{J}(1)\mathbf{J}(1)'\Phi(1)' - \Theta(1) \int (d\mathbf{J})\mathbf{J}'\Phi(1)' - \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \Theta_j [\mathbf{J}]_1 \Phi_i'.$$

To obtain the expression in (ii), $\mathbf{J}(1)\mathbf{J}(1)' - \int (d\mathbf{J})\mathbf{J}'$ could be substituted by $[\mathbf{J}]_1 + \int \mathbf{J}(d\mathbf{J})'$, and then $\Theta(1)[\mathbf{J}]_1\Phi(1)' - \sum_{i=0}^{\infty} \sum_{j=i+1}^{\infty} \Theta_j [\mathbf{J}]_1 \Phi_i'$ by $\sum_{i=0}^{\infty} \sum_{j=0}^{i} \Theta_j [\mathbf{J}]_1 \Phi_i'$.

The validity of item (iii) can be shown departing from

$$\sum_{t=1}^{T} \theta_t \psi_t' = \Theta(1) \sum_{t=1}^{T} \mu_t \psi_t' - \sum_{t=1}^{T} \Delta\theta_t^* \psi_t' = \Theta(1) \sum_{t=1}^{T} \mu_t \psi_t' + O_P(1),$$

which holds in view of (3) and Lemma A.2(ii). By partial summation,

$$T^{-1/2} \sum_{t=1}^{T} \mu_t \psi_t' = \mu_T \left(T^{-1/2} \sum_{t=1}^{T} \psi_t' \right) - \sum_{t=1}^{T} \delta_t \left(T^{-1/2} \sum_{i=1}^{t-1} \psi_i' \right)$$

$$\xrightarrow{w} \left[\mathbf{J}(1)\mathbf{W}(1)' - \int d\mathbf{J}(\mathbf{W})' \right] \Psi(1)',$$

the convergences by Proposition 1(i) and the almost sure continuity of $\mathbf{J}$ and $\mathbf{W}$ at 1, and from (A.8). Integration by parts in the last limit gives (iii).

Similarly, (3) yields the decomposition

$$\sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \psi_i \right) \Delta\theta_t' = \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \psi_i \right) \delta_t' \Theta(1)' - \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \psi_i \right) (\Delta^2\theta_t^*)',$$

where $\Delta^2\theta_t^* := \Delta\theta_t^* - \Delta\theta_{t-1}^*$ $(t > 1)$ and $\Delta^2\theta_1^* := \Delta\theta_1^*$. Here $T^{-1/2} \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \psi_i)\delta_t' \xrightarrow{w} \Psi(1) \int \mathbf{W}(d\mathbf{J})'$ by (A.8), whereas

$$\sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \psi_i \right) (\Delta^2\theta_t^*)' = \left(\sum_{t=1}^{T} \psi_t \right) \Delta\theta_T^* - \sum_{t=1}^{T} \psi_t (\Delta\theta_t^*)' = o_P(T^{1/2})$$

by Proposition 1(i) and (ii) and Lemma A.2(ii). This shows (iv).

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Again using (3), the left-hand side in statement (v) can be written as

$$T^{-3/2}\sum_{t=1}^{T}\theta_t\left(\sum_{i=1}^{t}\psi_i\right)' = T^{-3/2}\Theta(1)\sum_{t=1}^{T}\mu_t\left(\sum_{i=1}^{t}\psi_i\right)' - T^{-3/2}\sum_{t=1}^{T}\Delta\theta_t^*\left(\sum_{i=1}^{t}\psi_i\right)',$$

where the first term converges, according to Proposition 1(i) and the CMT, to the asserted limit in (v), whereas the second term is $o_P(1)$ by item (iv).

The validity of (vi) can be shown by using (3) and (A.5) to write $\sum_{t=1}^{T}\theta_t\phi_t'$ as

$$\Theta(1)\left(\sum_{t=1}^{T}\mu_t\mu_t'\right)\Phi(1)' - \left(\sum_{t=1}^{T}\Delta\theta_t^*\mu_t'\right)\Phi(1)' - \sum_{t=1}^{T}\theta_t\Delta\phi_t^{*\prime}$$

$$= \Theta(1)\left(\sum_{t=1}^{T}\mu_t\mu_t'\right)\Phi(1)' + O_P(1),$$

the equality by item (ii). Now item (vi) follows from Proposition 1(i) and the CMT.

According to (3), $\sum_{t=1}^{T}\theta_t = \Theta(1)\sum_{t=1}^{T}\mu_t - \sum_{t=1}^{T}\Delta\theta_t^* = \Theta(1)\sum_{t=1}^{T}\mu_t - \theta_T^*$. By (4) and the CMT, $T^{-1}\sum_{t=1}^{T}\mu_t \xrightarrow{w} \int\mathbf{J}$, whereas by Proposition 1(ii), $T^{-1}\theta_T^* = o_P(1)$; hence (vii). Finally, by partial summation

$$T^{-1}\sum_{t=1}^{T}(t-1)\Delta\theta_t' = \theta_T' - T^{-1}\sum_{t=1}^{T}\theta_t' \xrightarrow{w} \left(\mathbf{J}(1) - \int\mathbf{J}\right)'\Theta(1)' = \int u(d\mathbf{J}(u))'\Theta(1)',$$

the convergence by Proposition 1(i) and (ii) and Theorem 2(vii); hence (viii). ■

A.2. Cointegrated Processes with Level Shifts.

Proof of Lemma 3. Let $\mathbf{V}_T := C(T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t + \Theta(1)\mathbf{J}_T)$; it follows from Lemma A.1 that $\mathbf{V}_T \xrightarrow{w} \mathbf{V} := C(\mathbf{W} + \Theta(1)\mathbf{J})$. From (9) we obtain that $B_T'(T^{-1/2}Y_{T,\lfloor T\cdot\rfloor} + \Delta\theta_{T,\lfloor T\cdot\rfloor}^*)$ equals

$$((\bar\gamma'\mathbf{V}_T)', (\bar\beta_1'\beta'\Theta(1)\mathbf{J}_T)', \lfloor T\cdot\rfloor/T)' + T^{-1/2}(0_{1\times(n-r-1)}, (\bar\beta_1'\beta'X_{\lfloor T\cdot\rfloor})', (\bar\tau'\mathbf{V}_T)')' + o(1),$$

where $o(1)$ stands for $T^{-1/2}B_T'CX_0$. The first addend in the preceding expression converges weakly to $\mathbf{G}_0$ in $D_{n-r+s}[0,\infty)$ by Lemma A.1 and by the uniform convergence of $\lfloor T\cdot\rfloor/T$ to the identity function on $[0,\infty)$. The second addend is $o_P(1)$ in $D_{n-r+s}[0,\infty)$ because, for every fixed $u\in[0,\infty)$, first, $\max_{s\in[0,u]}\|\mathbf{V}_T\| \xrightarrow{w} \max_{s\in[0,u]}\|C(\mathbf{W}+\Theta(1)\mathbf{J})\|$ by Lemma A.1 and the CMT, and second, $\max_{s\in[0,u]}\|T^{-1/2}\beta'X_{\lfloor T\cdot\rfloor}\| = o_P(1)$ by, e.g., Johansen (1996, p. 251, eqn. (B.17)). Hence, convergence (10). Further, the left-hand side of (11) equals

$$B_T'(T^{-1/2}Y_{T,\lfloor T\cdot\rfloor} + \Delta\theta_{T,\lfloor T\cdot\rfloor}^*) - T^{-1}B_T'\sum_{s=1}^{T-1}(T^{-1/2}Y_{T,s} + \Delta\theta_{T,s}^*) + o_P(1),$$

where $o_P(1)$ stands for $T^{-1}B_T'\sum_{s=1}^{T-1}\Delta\theta_{T,s}^* = T^{-1}B_T'\theta_{T,T-1}^*$ in agreement with Proposition 1(ii) and weak convergence to $\mathbf{G}$ follows from (10) and the CMT. ■

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Proof of Lemma 4. Recall S_{ij}^x $(i,j \in \{0,1\})$ defined in (7). From $Y_{T,t} = X_t + T^{1/2}\theta_{T,t}$ we obtain the following decomposition of the sample covariances S_{ij}:

$$S_{ij} = S_{ij}^x + (M_{ij}^{\theta\theta} - M_i^\theta (M_j^\theta)') + (M_{ij}^{x\theta} - M_i^x (M_j^\theta)') + (M_{ij}^{\theta x} - M_i^\theta (M_j^x)'), \tag{A.10}$$

where, with $U_{0,t}^x := \Delta X_t$, $U_{0,t}^\theta := T^{1/2}\Delta\theta_{T,t}$, $U_{1,t}^x := X_{t-1}$, and $U_{1,t}^\theta := T^{1/2}\theta_{T,t-1}$, we define

$$M_i^a := T^{-1}\sum_{t=1}^T U_{i,t}^a \quad \text{and} \quad M_{ij}^{ab} := T^{-1}\sum_{t=1}^T U_{i,t}^a [U_{j,t}^b]', \qquad i,j \in \{0,1\}, \qquad a,b \in \{x,\theta\}.$$

Also let $M_0^{\theta^*} := T^{-1/2}\sum_{t=1}^{T-1}\Delta\theta_{T,t}^* = T^{-1/2}\theta_{T,T-1}^*$. The following convergences will be useful:

$$M_0^x \xrightarrow{P} E(\Delta X_1) = \tau, \qquad \beta' M_1^x \xrightarrow{P} E(\beta' X_1), \tag{A.11}$$

by a law of large numbers (LLN) for covariance stationary sequences, and

$$T^{1/2}M_0^\theta = \theta_{T,T} \xrightarrow{w} \Theta(1)\mathbf{J}(1), \quad T^{1/2}M_0^{\theta^*} \xrightarrow{w} \Theta^*(1)\mathbf{J}(1), \quad T^{-1/2}M_1^\theta \xrightarrow{w} \Theta(1)\int \mathbf{J}, \tag{A.12}$$

the first two by Proposition 1(i) and (ii) and the last one by Theorem 2(vii).

Consider now the convergences in the lemma. We find that $S_{00} = S_{00}^x + M_{00}^{\theta\theta} + o_P(1)$ because, first, $M_0^\theta = o_P(1)$ by (A.12) and, second,

$$M_{00}^{x\theta} - M_0^x (M_0^\theta)' = T^{-1/2}\sum_{t=1}^T (\Delta X_t - \tau)\Delta\theta_{T,t}' - (M_0^x - \tau)(M_0^\theta)' = o_P(1) \tag{A.13}$$

by Lemma A.2(ii), (A.11), and (A.12). Further, $S_{00}^x \xrightarrow{P} \Sigma_{00}$ by an LLN, and $M_{00}^{\theta\theta} = \sum_{t=1}^T \Delta\theta_{T,t}\Delta\theta_{T,t}' \xrightarrow{w} \sum_{i=0}^\infty \Theta_i [\mathbf{J}]_1 \Theta_i'$ by Theorem 2(i), from where the limit of S_{00} follows.

To obtain the limit of $\bar\beta_2'\beta' S_{11}\beta\bar\beta_2$, recall that $\bar\beta_2'\beta'\Theta(1) = 0$ by the choice of β_2, so that, also using (3), we find $\bar\beta_2'\beta'\theta_{T,t} = -\bar\beta_2'\beta'\Delta\theta_{T,t}^*$ and $\bar\beta_2'\beta' M_1^\theta = -\bar\beta_2'\beta' M_0^{\theta^*}$. Thus,

$$\beta' M_{11}^{x\theta}\beta\bar\beta_2 - \beta' M_1^x (M_1^\theta)'\beta\bar\beta_2 = -T^{-1/2}\sum_{t=1}^T (\beta' X_{t-1} - E(\beta' X_1))(\Delta\theta_{T,t-1}^{*\prime}\beta\bar\beta_2)$$
$$+ (\beta' M_1^x - E(\beta' X_1))(M_0^{\theta^*})'\beta\bar\beta_2 = o_P(1)$$

by Lemma A.2(ii), (A.11), and (A.12). In conjunction with (A.10), this gives

$$\bar\beta_2'\beta' S_{11}\beta\bar\beta_2 = \bar\beta_2'\beta' S_{11}^x \beta\bar\beta_2 + \bar\beta_2'\beta'[M_{11}^{\theta\theta} - M_1^\theta (M_1^\theta)']\beta\bar\beta_2 + o_P(1).$$

The preceding terms converge as follows: $\bar\beta_2'\beta' S_{11}^x\beta\bar\beta_2 \xrightarrow{P} \bar\beta_2'\Sigma_{\beta\beta}\bar\beta_2$ by an LLN,

$$\bar\beta_2'\beta' M_{11}^{\theta\theta}\beta\bar\beta_2 = \bar\beta_2'\beta'\sum_{t=1}^T \Delta\theta_{T,t-1}^*\Delta\theta_{T,t-1}^{*\prime}\beta\bar\beta_2 \xrightarrow{w} \bar\beta_2'\beta'\sum_{i=0}^\infty \Theta_i^* [\mathbf{J}]_1 \Theta_i^{*\prime}\beta\bar\beta_2$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

by the identity $\bar{\beta}_2'\beta'\theta_{T,t} = -\bar{\beta}_2'\beta'\Delta\theta^*_{T,t}$ and Theorem 2(i), and $\bar{\beta}_2'\beta' M_1^\theta = -\bar{\beta}_2'\beta' M_0^{\theta^*} = o_P(1)$ by (A.12). Hence, the asserted limit of $\bar{\beta}_2'\beta' S_{11}\beta\bar{\beta}_2$. Convergence of $\bar{\beta}_2'\beta' S_{10}$ follows analogously, and the proof is omitted.

The limit of $B_T' S_{10}$ is obtained next. First, $B_T' M_{10}^{\theta x} - B_T' M_1^\theta (M_0^x)'$ equals

$$T^{-1/2}B_T'\sum_{t=1}^{T}\theta_{T,t-1}(\Delta X_t - \tau)' - (T^{-1/2}B_T' M_1^\theta)T^{-1/2}\sum_{t=1}^{T}(\Delta X_t - \tau)'$$

$$\xrightarrow{w} (C'\bar{\gamma}, \beta\bar{\beta}_1, 0)'\Theta(1)\left(\int \mathbf{J}(d\mathbf{W})' - \left(\int \mathbf{J}\right)\mathbf{W}(1)'\right)C' \qquad \textbf{(A.14)}$$

by Theorem 2(iii), (A.12), and (4) with $\psi_t = \Delta X_t - \tau$. Second, with $\bar{Y}_T := M_1^x + M_1^\theta$,

$$B_T'(M_{10}^{x\theta} + M_{10}^{\theta\theta}) - (B_T'\bar{Y}_T)(M_0^\theta)' = T^{-1/2}B_T'\sum_{t=1}^{T} Y_{T,t-1}\Delta\theta_{T,t}' - (T^{-1/2}B_T'\bar{Y}_T)(T^{1/2}M_0^\theta)', \qquad \textbf{(A.15)}$$

where the components of $T^{-1/2}B_T'\sum_{t=1}^{T} Y_{T,t-1}\Delta\theta_{T,t}'$ converge as follows. On the one hand,

$$T^{-1/2}(C'\bar{\gamma}, \beta\bar{\beta}_1)'\sum_{t=1}^{T} Y_{T,t-1}\Delta\theta_{T,t}'$$

$$= T^{-1/2}(C'\bar{\gamma}, 0)'\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\varepsilon_s\right)\Delta\theta_{T,t}' + (C'\bar{\gamma}, \beta\bar{\beta}_1)'\sum_{t=1}^{T}\theta_{T,t-1}\Delta\theta_{T,t}' + o_P(1)$$

$$\xrightarrow{w} \int [(C'\bar{\gamma}, 0)'\mathbf{W} + (C'\bar{\gamma}, \beta\bar{\beta}_1)'\Theta(1)\mathbf{J}](d\mathbf{J})'\Theta(1)'$$

$$+ (C'\bar{\gamma}, \beta\bar{\beta}_1)'\sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\Theta_j[\mathbf{J}]_1\Theta_i'$$

by (9) (recall that $(\bar{\gamma}, \beta\bar{\beta}_1)'\tau = 0$) and Theorem 2(ii) and (iv), with $o_P(1)$ standing for

$$T^{-1/2}\sum_{t=1}^{T}(0_{r\times(n-r-1)}, \bar{\beta}_1)'[\beta' X_{t-1} - E(\beta' X_1)]\Delta\theta_{T,t}' + ((\bar{\gamma}' C X_0)', (\bar{\beta}_1' E(\beta' X_1))')'(M_0^\theta)'$$

in agreement with Lemma A.2(ii) and (A.12). On the other hand, again using (9),

$$T^{-1}\bar{\tau}' C\sum_{t=1}^{T} Y_{T,t-1}\Delta\theta_{T,t}' = T^{-1}\sum_{t=1}^{T}(t-1)\Delta\theta_{T,t}' + o_P(1) \xrightarrow{w} \int u(d\mathbf{J}(u))'\Theta(1)'$$

by Theorem 2(viii), with $o_P(1)$ standing for $T^{-1}\bar{\tau}' C\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s + T^{1/2}\theta_{T,t-1} + X_0)\Delta\theta_{T,t}'$ in agreement with Theorem 2(ii) and (iv) and Proposition 1(ii). Returning to (A.15), weak convergence of $T^{-1/2}B_T'\bar{Y}_T$ to $\int \mathbf{G}_0$ follows from its representation as $T^{-1}B_T'\sum_{s=1}^{T-1}(T^{-1/2}Y_{T,s} + \Delta\theta^*_{T,s}) - T^{-1}B_T'\theta^*_{T,T-1}$ by applying Lemma 3, the CMT, and Proposition 1(ii), whereas the convergence of $T^{1/2}M_0^\theta$ is given in (A.12). Thus, the weak limit of (A.15) is

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\int \mathbf{G}_0(d\mathbf{J})'\Theta(1)' - \left(\int \mathbf{G}_0\right)\mathbf{J}(1)'\Theta(1)' + (C'\bar{\gamma}, \beta\bar{\beta}_1, 0)' \sum_{i=1}^{\infty}\sum_{j=0}^{i-1} \Theta_j[\mathbf{J}]_1\Theta_i'.$$

Third, from equation (B.20) and the proof of Lemma 10.3, both in Johansen (1996), we have

$$B_T'S_{10}^x \xrightarrow{w} \left(\left(C\int\left(\mathbf{W} - \int \mathbf{W}\right)(d\mathbf{W})'C' + R_1\right)'\bar{\gamma}, (\bar{\beta}_1'\Sigma_{\beta 0})', \left(\int\left(u - \frac{1}{2}\right)(d\mathbf{W})'C'\right)'\right)',$$

where $R_1 := \sum_{h=1}^{\infty}\mathrm{Cov}(C\Delta X_t, \Delta X_{t+h}) = \sum_{h=1}^{\infty} C\Omega\beta(I+\alpha'\beta)^{h-1}\alpha' = -C\Omega\beta(\alpha'\beta)^{-1}\alpha'$. The limit of $B_T'S_{10}$ obtains from (A.10) by combining the preceding results.

Next, the terms in the decomposition of $B_T'S_{11}\beta\bar{\beta}_2$ are $B_T'S_{11}^x\beta\bar{\beta}_2 = O_P(1)$ by Lemma 10.3 of Johansen (1996); $B_T'M_{11}^{\theta x}\beta\bar{\beta}_2 - B_T'M_1^{\theta}(M_1^x)'\beta\bar{\beta}_2 = O_P(1)$ similarly to (A.14); $B_T'(M_{11}^{x\theta} + M_{11}^{\theta\theta})\beta\bar{\beta}_2 - B_T'\bar{Y}_T(M_1^{\theta})'\beta\bar{\beta}_2 = O_P(1)$ similarly to (A.15) and using the identity $\bar{\beta}_2'\beta'M_1^{\theta} = -\bar{\beta}_2'\beta'M_0^{\theta^*}$ justified earlier. We omit the details.

Finally, let $\tilde{Y}_{T,t} := B_T'[T^{-1/2}(Y_{T,t} - \bar{Y}_T) + \Delta\theta_{T,t}^*]$, so that $B_T'S_{11}B_T = \sum_{t=1}^{T-1}\tilde{Y}_{T,t}\tilde{Y}_{T,t}' + B_T'\sum_{t=1}^{T-1}\Delta\theta_{T,t}^*\Delta\theta_{T,t}^{*\prime}B_T - \sum_{t=1}^{T-1}\tilde{Y}_{T,t}\Delta\theta_{T,t}^{*\prime}B_T - B_T'\sum_{t=1}^{T-1}\Delta\theta_{T,t}^*\tilde{Y}_{T,t}'$. Upon normalization by T^{-1}, the first term converges weakly to $\int \mathbf{G}\mathbf{G}'$ in view of (11) and the CMT, whereas the remaining terms vanish. This yields the weak limit of $T^{-1}B_T'S_{11}B_T$. ■

Proof of Theorem 5. The argument resembles that for Theorem 11.1 of Johansen (1996), and so we only present the modifications. With $S(\lambda) = \lambda S_{11} - S_{10}S_{00}^{-1}S_{01}$, similarly to Johansen's equation (11.16), we get from Lemma 4 that $|\bar{\beta}_2'\beta'S(\lambda)\beta\bar{\beta}_2| \xrightarrow{w} |\lambda\tilde{\Sigma}_{\beta\beta} - \tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{00}^{-1}\tilde{\Sigma}_{0\beta}|$, where $\tilde{\Sigma}_{0\beta} := \tilde{\Sigma}_{\beta 0}'$. With Ω strictly positive definite and $\beta\beta_2$ of full column rank, the matrices $\Sigma_{00} = \Omega + \alpha\Sigma_{\beta\beta}\alpha'$ and $\beta_2'\Sigma_{\beta\beta}\beta_2 = \beta_2'(\beta'\Omega\beta + (I+\beta'\alpha)\Sigma_{\beta\beta}(I+\alpha'\beta))\beta_2$ are strictly positive definite, and hence so are $\tilde{\Sigma}_{00}$ and $\tilde{\Sigma}_{\beta\beta}$. Thus, if $\tilde{\Sigma}_{\beta 0}$ a.s. has full row rank $r-s$, the polynomial $|\lambda\tilde{\Sigma}_{\beta\beta} - \tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{00}^{-1}\tilde{\Sigma}_{0\beta}|$ will have $r-s$ a.s. nonzero roots. To evaluate the rank of $\tilde{\Sigma}_{\beta 0}$, we depart from the inequality

$$P(\mathrm{rank}(\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{0\beta}) < r-s) \le \sum_{\varphi\subseteq\{1,\ldots,k\}} P(\mathrm{rank}(\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{0\beta}) < r-s | \{\mathbf{N}^j(1) \ge 1 \text{ iff } j\in\varphi\}).$$

(A.16)

The condition corresponding to $\varphi = \varnothing$ is $\{\mathbf{N}(1) = 0\}$. It is contained in the set $\{[\mathbf{J}]_1 = 0\}$, on which we have $\tilde{\Sigma}_{\beta 0} = \bar{\beta}_2'\Sigma_{\beta 0}$, so that $\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{0\beta}$ equals $\bar{\beta}_2'\Sigma_{\beta\beta}(\alpha'\alpha)\Sigma_{\beta\beta}\bar{\beta}_2$ by Lemma 10.1 of Johansen (1996) and has full rank $r-s$. Therefore, in (A.16) we need to consider only $\varphi \ne \varnothing$. For such φ, let $\mathbf{J}_\varphi$ be the process (and $\Theta_{i,\varphi}$, $i \ge 0$, the matrix) that obtains from $\mathbf{J}$ (resp. from Θ_i) by keeping only the components $\mathbf{J}^j$ (resp. the columns $\Theta_{i,\cdot j}$) with $j\in\varphi$. Introduce also $p_\varphi(x) := \beta_2'(\Sigma_{\beta 0} + \beta'\sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\Theta_{j,\varphi}\,\mathrm{diag}(x)\Theta_{i,\varphi}')$ and, denoting by $\#(\varphi)$ the cardinality of φ, $\mathcal{D}_\varphi := \{x\in\mathbb{R}^{\#(\varphi)} : |p_\varphi(x)p_\varphi'(x)| = 0\}$. Using the expression for $\tilde{\Sigma}_{\beta 0}$ from Lemma 4, the independence of the components of $\mathbf{J}$, and the diagonal structure of $[\mathbf{J}]_1$, we can write (A.16) as

$$P(\mathrm{rank}(\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{0\beta}) < r-s) \le \sum_{\varnothing\ne\varphi\subseteq\{1,\ldots,k\}} P(\mathrm{dg}[\mathbf{J}_\varphi]_1 \in \mathcal{D}_\varphi | \{\mathbf{N}^j(1)\ge 1 \text{ iff } j\in\varphi\}),$$

(A.17)

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

where $\mathrm{dg}[\mathbf{J}_\varphi]_1$ denotes the diagonal of $[\mathbf{J}_\varphi]_1$. The set $\mathcal{D}_\varphi$, as the zero set of a nonzero polynomial, has Lebesgue measure 0 in $\mathbb{R}^{\#(\varphi)}$. The measures $P(\mathrm{dg}[\mathbf{J}_\varphi]_1 \in (\cdot) | \{\mathbf{N}^j(1) \geq 1 \text{ iff } j \in \varphi\})$, $\varnothing \neq \varphi \subseteq \{1, \ldots, k\}$, are absolutely continuous with respect to Lebesgue measure in $\mathbb{R}^{\#(\varphi)}$ because the distribution of η_1 is assumed such with respect to Lebesgue measure in $\mathbb{R}^k$, and thus the summation in (A.17) is zero. This confirms that $\mathrm{rank}(\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{0\beta}) = r - s$ (a.s.), and hence the $r - s$ roots of $|\lambda\tilde{\Sigma}_{\beta\beta} - \tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{00}^{-1}\tilde{\Sigma}_{0\beta}|$ are a.s. nonzero.

As the ordered roots of $|\bar{\beta}_2'\beta' S(\lambda)\beta\bar{\beta}_2|$ are continuous functions of the coefficient matrices, we conclude from the previous analysis that they converge weakly to a.s. nonzero random variables. Further, as these roots are lower bounds for the $r - s$ largest roots of $|S(\lambda)|$, the latter are bounded away from zero in probability. This implies that $\Lambda(\rho) = -T\sum_{i=\rho+1}^{n} \ln(1 - \hat{\lambda}_i) \xrightarrow{P} \infty$ whenever $\rho < r - s$, and hence (13).

We discuss next the $n - r + s$ smallest roots of $|S(\lambda)|$. For $s > 0$ the matrix $\int \mathbf{G}\mathbf{G}'$ is singular with positive probability, and for problematic outcomes we modify the equation under study. For a fixed $\epsilon > 0$, define the function $f_\epsilon^n : \mathbb{R}^{n\times n} \times \mathbb{R}^{k\times k} \to \mathbb{R}^{n\times n}$ by $f_\epsilon^n(A, B) = A\mathbf{1}_{\mathcal{P}(B)} + I_n(1 - \mathbf{1}_{\mathcal{P}(B)})$, where $\mathcal{P}(B) := \{\lambda_{\min}(\zeta' B\zeta) > \epsilon\}$, and consider the equation

$$\left| \lambda f_\epsilon^n \left(A_T' S_{11} A_T, \int \mathbf{J}_T \mathbf{J}_T' \right) - A_T' S_{10} S_{00}^{-1} S_{01} A_T \right| = 0, \tag{A.18}$$

where $A_T = (\beta\bar{\beta}_2, T^{-1/2}B_T)$. For outcomes in $\mathcal{P}(\int \mathbf{J}_T\mathbf{J}_T')$, this is the equation $|A_T' S(\lambda) A_T| = 0$. As $P(\lambda_{\min}(\zeta' \int \mathbf{J}\mathbf{J}'\zeta) = \epsilon) = 0$, from Lemma 4 and the CMT it follows that $f_\epsilon^n(A_T' S_{11} A_T, \int \mathbf{J}_T\mathbf{J}_T') \xrightarrow{w} f_\epsilon^n(\mathrm{diag}(\tilde{\Sigma}_{\beta\beta}, \int \mathbf{G}\mathbf{G}'), \int \mathbf{J}\mathbf{J}')$, which is a.s. nonsingular. Imitating the proof of Theorem 11.1 of Johansen (1996), we conclude that the $n - r + s$ smallest solutions of (A.18) multiplied by T converge weakly to the solutions of

$$\left| \omega f_\epsilon^{n-r+s} \left(\int \mathbf{G}\mathbf{G}', \int \mathbf{J}\mathbf{J}' \right) - \tilde{\Sigma}_{B0} \tilde{N} \tilde{\Sigma}_{B0}' \right| = 0, \tag{A.19}$$

where $\tilde{\Sigma}_{B0}$ is as in Lemma 4 and $\tilde{N} := \tilde{\Sigma}_{00}^{-1} - \tilde{\Sigma}_{00}^{-1}\tilde{\Sigma}_{0\beta}[\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{00}^{-1}\tilde{\Sigma}_{0\beta}]^{-1}\tilde{\Sigma}_{\beta 0}\tilde{\Sigma}_{00}^{-1}$ can be written as $\tilde{N} = (\tilde{\Sigma}_{0\beta})_\perp[(\tilde{\Sigma}_{0\beta})_\perp'\tilde{\Sigma}_{00}(\tilde{\Sigma}_{0\beta})_\perp]^{-1}(\tilde{\Sigma}_{0\beta})_\perp'$, with $(\tilde{\Sigma}_{0\beta})_\perp$ standing for an $n \times (n - r + s)$ matrix of full column rank such that $\tilde{\Sigma}_{\beta 0}(\tilde{\Sigma}_{0\beta})_\perp = 0$.

Denote by $\hat{\lambda}_{i,\epsilon}$ $(i = 1, \ldots, n)$ and ω_i $(i = 1, \ldots, n - r + s)$ the solutions of, respectively, (A.18) and (A.19) in decreasing order. Then for $q > 0$ it holds that

$$\begin{aligned}
P(\Lambda(r-s) < q) &\geq P\left(\{\Lambda(r-s) < q\} \cap \mathcal{P}\left(\int \mathbf{J}_T\mathbf{J}_T' \right) \right) \\
&= P\left(\left\{ -T \sum_{i=r-s+1}^{n} \ln(1 - \hat{\lambda}_{i,\epsilon}) < q \right\} \cap \mathcal{P}\left(\int \mathbf{J}_T\mathbf{J}_T' \right) \right) \\
&\to P\left(\left\{ \sum_{i=1}^{n-r+s} \omega_i < q \right\} \cap \mathcal{P}\left(\int \mathbf{J}\mathbf{J}' \right) \right) \\
&\geq P\left(\{\omega_1 < q/(n-r+s)\} \cap \mathcal{P}\left(\int \mathbf{J}\mathbf{J}' \right) \right).
\end{aligned}$$

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

Because F^i are not identically zero on $[0,1]$, ϵ can be chosen such that $P(\mathcal{P}(\int \mathbf{JJ}')) > 0$, and because ω_1 is a.s. finite, continuity of P from below guarantees that q can be chosen large enough for the last probability in the preceding display to be positive. This proves part (i) of the theorem.

Consider next the case $s = 0$, so that we can choose $\beta_2 = I_r$, giving $\beta\beta_2 = \beta$. Now $\int \mathbf{GG}'$ is a.s. nonsingular, and no modification of the equation $|S(\lambda)| = 0$ is needed. From Johansen's argument it follows that $T\hat{\lambda}_i$ $(r+1 \le i \le n)$ converge weakly to the ordered solutions of the equation $|\omega \int \mathbf{GG}' - \tilde{\Sigma}_{B0}\tilde{N}\tilde{\Sigma}'_{B0}| = 0$, say, $\omega_1 \ge \cdots \ge \omega_{n-r}$. Thus, for $\rho \ge r$, we have $\Lambda(\rho) = -T\sum_{i=\rho+1}^{n} \ln(1-\hat{\lambda}_i) = T\sum_{i=\rho+1}^{n} \hat{\lambda}_i + o_P(1) \xrightarrow{w} \sum_{i=\rho-r+1}^{n-r} \omega_i$. In particular,

$$\Lambda(r) \to \operatorname{tr}\left\{\left[\int \mathbf{GG}'\right]^{-1} \tilde{\Sigma}_{B0}\tilde{N}\tilde{\Sigma}'_{B0}\right\}, \tag{A.20}$$

which is equivalent to (14). ■

Proof of Corollary 6. Because $\beta'\Theta(1) = \beta' C\Theta_0 = 0$, we have $s = 0$. Using the expression for $\tilde{\Sigma}_{\beta 0}$ from Lemma 4 and the relation $\Sigma_{\beta 0} = \Sigma_{\beta\beta}\alpha'$ from Lemma 10.1 of Johansen (1996), it is checked directly that, for the considered Θ_i, we can choose $(\tilde{\Sigma}_{0\beta})_\perp = \alpha_\perp$. Thus, from Lemma 4, $\tilde{\Sigma}_{B0}\alpha_\perp = \int \mathbf{G}(d\mathbf{R}_0)'\alpha_\perp$ ($R_1\alpha_\perp = 0$; $\sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\Theta_j[\mathbf{J}]_1\Theta_i'\alpha_\perp = 0$ because $\Theta_i'\alpha_\perp = 0$ for $i \ge 1$; the terms with β_1 are absent because $s = 0$). Further, $\alpha_\perp'\Theta(1) = \alpha_\perp'\Theta_0$, so that $\alpha_\perp'\mathbf{R}_0 = \alpha_\perp'(\mathbf{W} + \Theta_0\mathbf{J})$. Finally,

$$(\tilde{\Sigma}_{0\beta})'_\perp \tilde{\Sigma}_{00}(\tilde{\Sigma}_{0\beta})_\perp = \alpha_\perp'\left(\Sigma_{00} + \sum_{i=0}^{\infty}\Theta_i[\mathbf{J}]_1\Theta_i'\right)\alpha_\perp = \alpha_\perp'(\Omega + \Theta_0[\mathbf{J}]_1\Theta_0')\alpha_\perp$$

by Lemma 10.1 in Johansen (1996) for $\alpha_\perp'\Sigma_{00}\alpha_\perp$ and by the choice of Θ_i. Thus, the limit in (A.20) specializes to

$$\operatorname{tr}\left\{\left[\int \mathbf{GG}'\right]^{-1} \int \mathbf{G}d(\mathbf{W} + \Theta_0\mathbf{J})'\alpha_\perp[\alpha_\perp'(\mathbf{W} + \Theta_0\mathbf{J})]_1^{-1}\alpha_\perp' \int d(\mathbf{W} + \Theta_0\mathbf{J})\mathbf{G}'\right\}.$$

The process $\mathbf{Z}_0 = (\mathbf{Z}_{01}', \mathbf{Z}_{02}')'$ defined by $\mathbf{Z}_{0i} := K_i(\mathbf{W} + \Theta_0\mathbf{J}) = \mathbf{B}_i + K_i\Theta_0\mathbf{J}$ $(i = 1, 2)$ can be substituted in place of $\alpha_\perp'(\mathbf{W} + \Theta_0\mathbf{J})$ because it is a nonsingular linear transformation of it. Then $\mathbf{Z}$ obtains by standardizing $\mathbf{Z}_0$. Similarly, $\mathbf{V}$ can be substituted for $\mathbf{G}$. Further linear transformations would need to be orthonormal to keep the Brownian motion standard, but no such transformation can eliminate the nuisance parameters from the distribution. ■

Proof of Corollary 7. With the given Θ_i we find $s = 0$, $\tilde{\Sigma}_{00} = \Sigma_{00} + 2\Theta[\mathbf{J}]_1\Theta'$, $\tilde{\Sigma}_{\beta 0} = \Sigma_{\beta 0}$ (because $\beta'\Theta = \beta' C\Theta = 0$), and

$$\tilde{\Sigma}_{B0} = \left[\int \mathbf{G}(d\mathbf{W})' - (\bar{\gamma}, 0)'\Theta[\mathbf{J}]_1\Theta'\right]C' = \left[\int \mathbf{G}(d\mathbf{B})' + K_4\right](\bar{\alpha}_\perp'K')^{-1}(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'$$

(because $C\Theta = \Theta$ and $\alpha_\perp\bar{\alpha}_\perp'K' = K'$). As in the proof of Corollary 6, we can choose $(\tilde{\Sigma}_{0\beta})_\perp = \alpha_\perp$, so that $(\tilde{\Sigma}_{0\beta})'_\perp\tilde{\Sigma}_{00}(\tilde{\Sigma}_{0\beta})_\perp = \alpha_\perp'(\Omega + 2\Theta[\mathbf{J}]_1\Theta')\alpha_\perp$, and the limit in (A.20) becomes

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press

$$\operatorname{tr}\left\{\left[\int \mathbf{G}\mathbf{G}'\right]^{-1}\left[\int \mathbf{G}(d\mathbf{B})' + K_4\right](K\bar{\alpha}_\perp\alpha_\perp'(\Omega + 2\Theta[\mathbf{J}]_1\Theta')\alpha_\perp\bar{\alpha}_\perp'K')^{-1}\right.$$
$$\left.\times\left[K_4' + \int (d\mathbf{B})\mathbf{G}'\right]\right\}.$$

Because $\Theta(1) = 0$, $\mathbf{G}$ contains no contribution from $\mathbf{J}$, and $\mathbf{F}$ can be inserted instead of $\mathbf{G}$ as a nonsingular transformation of it. Finally, as $K\bar{\alpha}_\perp\alpha_\perp'\Omega\alpha_\perp\bar{\alpha}_\perp'K' = I_{n-r}$ and $K\bar{\alpha}_\perp\alpha_\perp' = K$, we obtain the expression for K_3. ■

Proof of Corollary 8. Because $\beta'\Theta_i = 0$ $(i \geq 0)$, we have $\tilde{\Sigma}_{\beta 0} = \Sigma_{\beta 0}$. As before, $(\tilde{\Sigma}_{0\beta})_\perp = \alpha_\perp$ is a valid choice, and by Lemma 4,

$$\tilde{\Sigma}_{B0}\alpha_\perp = \int \mathbf{G}(d\mathbf{R}_0)'\alpha_\perp + (\bar{\gamma},0)'\sum_{i=1}^{\infty}\sum_{j=0}^{i-1}\Theta_j[\mathbf{J}]_1\Theta_i'\alpha_\perp = \left[\int \mathbf{G}(d\tilde{\mathbf{Z}})' + \tilde{K}_4\right](\bar{\alpha}_\perp'K')^{-1},$$

the equality because, first, $\Theta(1) = C\Theta$ and $\mathbf{R}_0'\alpha_\perp = (\mathbf{W} + \Theta\mathbf{J})'C'\alpha_\perp = (\mathbf{W} + \Theta\mathbf{J})'\alpha_\perp$, and second, $C\Theta_j = \Theta_j$ and $\alpha_\perp\bar{\alpha}_\perp'K' = K'$. Thus, the limit in (A.20) becomes

$$\operatorname{tr}\left\{\left[\int \mathbf{G}\mathbf{G}'\right]^{-1}\left[\int \mathbf{G}(d\tilde{\mathbf{Z}})' + \tilde{K}_4\right](K\bar{\alpha}_\perp\alpha_\perp'\tilde{\Sigma}_{00}\alpha_\perp\bar{\alpha}_\perp'K')^{-1}\left[\tilde{K}_4' + \int (d\tilde{\mathbf{Z}})\mathbf{G}'\right]\right\}.$$

Using again the identities $\alpha_\perp\bar{\alpha}_\perp'K' = K'$ and $\alpha_\perp'\Sigma_{00}\alpha_\perp = \alpha_\perp'\Omega\alpha_\perp$, we find $K\bar{\alpha}_\perp\alpha_\perp'\tilde{\Sigma}_{00}\alpha_\perp\bar{\alpha}_\perp'K' = K(\Omega + \sum_{i=0}^{\infty}\Theta_i[\mathbf{J}]_1\Theta_i')K' = \tilde{K}_3$, and to complete the proof it remains to note that $\tilde{\mathbf{V}}$ is a linear transformation of $\mathbf{G}$. ■

https://doi.org/10.1017/S0266466608080249 Published online by Cambridge University Press