Econometric Theory, **22**, 2006, 429–456. Printed in the United States of America.
DOI: 10.1017/S026646660606021X

A RESIDUAL-BASED TEST FOR STOCHASTIC COINTEGRATION

Brendan McCabe
University of Liverpool

Stephen Leybourne
University of Nottingham

David Harris
University of Melbourne

We consider the problem of hypothesis testing in a modified version of the stochastic integration and cointegration framework of Harris, McCabe, and Leybourne (2002, *Journal of Econometrics* 111, 363–384). This nonlinear setup allows for volatility in excess of that catered for by the standard integration/cointegration paradigm through the introduction of nonstationary heteroskedasticity. We propose a test for stochastic cointegration against the alternative of no cointegration and a secondary test for stationary cointegration against the heteroskedastic alternative. Asymptotic distributions of these tests under their respective null hypotheses are derived, and consistency under their respective alternatives is established. Monte Carlo evidence suggests that the tests will perform well in practice. An empirical application to the term structure of interest rates is also given.

1. INTRODUCTION

The cointegration framework of Engle and Granger (1987) is characterized by two widely held stylized empirical facts. The first is that, of the set of economic time series that exhibit trending behavior, many are adequately modeled by processes that are integrated, usually of order one, $I(1)$. The second is that, despite this trending behavior, such series often tend to comove over time according to a stationary, or $I(0)$, process; that is, they are cointegrated. Many empirical tests of important economic hypotheses are carried out within the Engle and Granger framework, for example, the relationship between long-run and short-run interest rates—the term structure. The Engle and Granger approach has, perhaps surprisingly, however, uncovered only very limited empirical evidence in support of the term structure (see Campbell and Shiller, 1987). An explanation often put forward for this is that bond market series tend to be too volatile to be compatible with the $I(1)/I(0)$ framework. That is, the individual series often appear visually to be more volatile, or less smooth, than would be

We are most grateful to the Associate Editor and two anonymous referees for providing helpful comments on earlier versions of this paper. Address correspondence to David Harris, e-mail: harrisd@unimelb.edu.au.

 0266-4666/06 $12.00

consistent with $I(1)$ and when comovements between series are analyzed (most simply by examining the spreads) these also tend to display periods of volatility in excess of that typically associated with stationary behavior. In the words of Campbell and Shiller, the spreads tend to "move too much."

One possible approach to dealing with the presence of extra volatility is within the *stochastic integration and cointegration* framework of Harris, McCabe, and Leybourne (2002). Here, the restrictive stationarity requirement of first differences of individual series and cointegrating error terms of the Engle and Granger (1987) setup is replaced with a looser condition that these are *stochastically trendless;* that is, they are simply free of $I(1)$ stochastic trends. This notion, of course, encompasses the Engle and Granger setup as a special case. We outline this framework in Section 2.

In Section 3 we turn to the issue of hypothesis testing in a regression model representation. The central hypothesis of interest is whether series are stochastically cointegrated (either stationary or heteroskedastic), or not cointegrated. We suggest a residual-based statistic to test the null of stochastic cointegration. Within stochastic cointegration, we also consider the hypothesis that the cointegration is stationary against the alternative that it is nonstationary heteroskedastic, and we suggest a second statistic to test this. Moreover, when applied to first differences of an individual series, this same statistic can also be used to test the null of $I(1)$ against heteroskedastic integration. The asymptotic null distributions of these two test statistics are derived under weak regularity conditions. Both are shown to have normal limit distributions that, unlike most cointegration tests, do not depend on the number of regressors involved. Their consistency properties under associated alternative hypotheses are also established.

Some Monte Carlo studies that examine the finite-sample size and power characteristics of the new tests, along with those of their conventional counterparts, are provided in Section 4. These highlight clearly the benefits to be gained by adopting the new test procedures, together with the shortcomings of using conventional ones, in the stochastic cointegration framework. Finally, in Section 5 we apply our tests to bond market data from several major economies. Our new testing framework uncovers supporting evidence in favor of the term structure in the bond market, in the same situation where conventional tests yield inconsistent results. Notably, for all the interest rate series we consider here, we conclude they are better modeled by heteroskedastically integrated, rather than $I(1)$, processes.

2. STOCHASTIC INTEGRATION AND COINTEGRATION

We first consider a variant of the model introduced in Harris et al. (2002):

$$\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{\Pi}\mathbf{w}_t + \boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{h}_t, \tag{1}$$

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \boldsymbol{\eta}_t,$$

$$\mathbf{h}_t = \mathbf{h}_{t-1} + \boldsymbol{\upsilon}_t$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

for $t = 1,\ldots,T$. Here $\mathbf{z}_t$, $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, and, $\boldsymbol{\varepsilon}_{\mathrm{t}}$ are $m \times 1$ vectors; $\mathbf{w}_t$ and $\boldsymbol{\eta}_t$ are $n \times 1$ vectors; $\mathbf{h}_t$ and $\boldsymbol{v}_t$ are $p \times 1$ vectors; $\boldsymbol{\Pi}$ and $\mathbf{V}_t$ are $m \times n$ and $m \times p$ matrices, respectively. Only the process $\mathbf{z}_t$ is observed. The disturbances $\boldsymbol{\varepsilon}_{\mathrm{t}}$, $\boldsymbol{\eta}_t$, $\boldsymbol{v}_t$, and $\mathbf{V}_t$ are mean zero stationary processes, which may be correlated with one another; $\mathbf{w}_t$ and $\mathbf{h}_t$ are vectors of integrated processes. So, apart from deterministics, $\mathbf{z}_t$ consists of an integrated component, $\boldsymbol{\Pi}\mathbf{w}_t$, together with a shock term, $\boldsymbol{\varepsilon}_{\mathrm{t}} + \mathbf{V}_t\mathbf{h}_t$. This latter term has a linear component, $\boldsymbol{\varepsilon}_{\mathrm{t}}$, and a nonlinear component $\mathbf{V}_t\mathbf{h}_t$ that is nonstationary heteroskedastic through its dependence on the $I(1)$ process $\mathbf{h}_t$. Note that it is entirely possible throughout our analysis that $\mathbf{w}_t$ and $\mathbf{h}_t$ contain identical processes, though we do not enforce this restriction.[1]

As regards the statistical properties of the disturbance terms in (1), we make the following linear process assumption. This allows for general forms of serial correlation, cross-correlation, and endogeneity.

Assumption LP. Let $\boldsymbol{\zeta}_t = [\boldsymbol{v}_t', \mathrm{vec}(\mathbf{V}_t)', \boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t']'$ be generated by the vector linear process $\boldsymbol{\zeta}_t = \sum_{j=0}^{\infty}\mathbf{C}_j\boldsymbol{\xi}_{t-j}$, where

1. $\sum_{j=0}^{\infty} j\|\mathbf{C}_j\| < \infty$ with $\mathbf{C}_0$ having full rank.[2]
2. $\boldsymbol{\xi}_t$ is an independent and identically distributed (i.i.d.) sequence.
3. $E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t') = \mathbf{I}$.
4. For all i, $E(\xi_{it}^{16})$ is bounded.

To examine the properties of the model more clearly, we make the temporary simplifying assumption that $\boldsymbol{\mu} = \boldsymbol{\delta} = 0$. Next, let $\mathbf{e}_i$ be an $m \times 1$ vector with 1 in its ith position and 0 elsewhere, so that $\mathbf{e}_i'\mathbf{z}_t = z_{it}$, the ith element of the vector $\mathbf{z}_t$. Then, from (1), we have

$$z_{it} = \mathbf{e}_i'\boldsymbol{\Pi}\mathbf{w}_t + \mathbf{e}_i'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{h}_t),$$

and if $\mathbf{e}_i'\boldsymbol{\Pi} \neq \mathbf{0}$ then z_{it} is said to *stochastically integrated.* If, in addition, $\mathbf{e}_i'E(\mathbf{V}_t\mathbf{V}_t')\mathbf{e}_i > 0$, z_{it} is said to be *heteroskedastically integrated* (*HI*) due to the term $\mathbf{e}_i'\mathbf{V}_t\mathbf{h}_t$, whereas if $\mathbf{e}_i'\mathbf{V}_t = 0$ then z_{it} is simply $I(1)$. So, a stochastically integrated variable encompasses both ordinary and heteroskedastic integration.

To model linear relationships between the variables in $\mathbf{z}_t$, let $\mathbf{c}$ be a nonzero $m \times 1$ vector and consider

$$\mathbf{c}'\mathbf{z}_t = \mathbf{c}'\boldsymbol{\Pi}\mathbf{w}_t + \mathbf{c}'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{h}_t).$$

If $\mathbf{c}'\boldsymbol{\Pi} = \mathbf{0}$ then the variables of $\mathbf{z}_t$ are said to be *stochastically cointegrated*. Under stochastic cointegration $\mathbf{c}'\mathbf{z}_t = \mathbf{c}'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{h}_t)$ behaves like a stochastically integrated process *net* of its stochastic trend component, and we refer to such a process as being *stochastically trendless*.[3] This terminology is adopted because, under Assumption LP, we can show that as $s \to \infty$ (with t fixed)

$$E(\boldsymbol{\varepsilon}_{t+s} + \mathbf{V}_{t+s}\mathbf{h}_{t+s}|\Im_t) - E(\boldsymbol{\varepsilon}_{t+s} + \mathbf{V}_{t+s}\mathbf{h}_{t+s}) \xrightarrow{p} 0.$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

In other words, the behavior of the process up to time t has a negligible effect on its behavior into the infinite future.[4] Therefore, even though the disturbances $\boldsymbol{\upsilon}_t$ have an infinitely persistent effect on $\mathbf{h}_{t+s}$, their effect on the *level* of $\mathbf{V}_{t+s}\mathbf{h}_{t+s}$ is only transitory. This implies that the product process $\mathbf{V}_t\mathbf{h}_t$ is stochastically trendless, even if $\mathbf{V}_t$ is correlated with $\boldsymbol{\upsilon}_t$. Although it is the case that $\mathbf{V}_t\mathbf{h}_t$ is nonstationary heteroskedastic, as it can be shown to exhibit a linear trend in variance, it is the stochastically trendless nature of $\mathbf{c}'\mathbf{z}_t = \mathbf{c}'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{h}_t)$ that bestows meaning to comovement of a nonstationary heteroskedastic kind.

When $\mathbf{c}'E(\mathbf{V}_t\mathbf{V}'_t)\mathbf{c} = 0$, then $\mathbf{c}'\mathbf{z}_t = \mathbf{c}'\boldsymbol{\varepsilon}_t$ is stationary. If, in addition, $\mathbf{V}_t = \mathbf{0}$, the variables are all integrated and cointegrated in the standard Engle and Granger (1987) sense. Because of the stationary behavior of $\mathbf{c}'\mathbf{z}_t$ in either case, we simply refer to this as *stationary cointegration.* When $\mathbf{c}'E(\mathbf{V}_t\mathbf{V}'_t)\mathbf{c} > 0$, the variables $\mathbf{z}_t$ are said to be *heteroskedastically cointegrated*. Thus, stochastic cointegration encompasses both stationary cointegration (possibly of the Engle and Granger kind) and heteroskedastic cointegration.

To further position our concept of heteroskedastic cointegration, note that $I(1)$, *HI*, and the closely related stochastic unit root processes all share the properties of having trends in their variances although not being stochastic trendless.[5] When these models of nonstationarity are extended to the multivariate cointegration setting, standard cointegration implies that a certain linear combination of the series becomes stochastically trendless *and* any trend in variance is removed. Hence, our definition of heteroskedastic cointegration effectively provides a halfway point to standard cointegration, because the linear combination becomes stochastically trendless, yet the trend in variance remains.

3. HYPOTHESIS TESTS AND TEST STATISTICS

Our primary goal is to determine if the system is stochastically cointegrated. This null, and the alternative of noncointegration, may be stated as $H^0: \mathbf{c}'\boldsymbol{\Pi} = \mathbf{0}$ and $H^1: \mathbf{c}'\boldsymbol{\Pi} \neq \mathbf{0}$. Within stochastic cointegration, we may wish to know whether stationary or heteroskedastic cointegration pertains. The null of stationary cointegration against the heteroskedastic alternative may be tested by partitioning H^0 as $H^0_0: \mathbf{c}'E(\mathbf{V}_t\mathbf{V}'_t)\mathbf{c} = 0$ and $H^0_1: \mathbf{c}'E(\mathbf{V}_t\mathbf{V}'_t)\mathbf{c} > 0$.

It proves convenient to interpret these hypotheses within a regression model. Partition $\mathbf{z}_t$ into a scalar y_t and an $(m-1)\times 1$ vector $\mathbf{x}_t$ as $\mathbf{z}_t = [y_t, \mathbf{x}'_t]'$. Then partitioning (1) conformably, and rearranging, we obtain

$$\begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{bmatrix} + \begin{bmatrix} \delta_y \\ \boldsymbol{\delta}_x \end{bmatrix} t + \begin{bmatrix} \boldsymbol{\pi}'_y \\ \boldsymbol{\Pi}_x \end{bmatrix} \mathbf{w}_t + \begin{bmatrix} \varepsilon_{yt} \\ \boldsymbol{\varepsilon}_{xt} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}'_{yt} \\ \mathbf{V}_{xt} \end{bmatrix} \mathbf{h}_t, \tag{2}$$

where y_t, μ_y, δ_y, and ε_{yt} are scalars, $\mathbf{x}_t$, $\boldsymbol{\mu}_x$, $\boldsymbol{\delta}_x$, and $\boldsymbol{\varepsilon}_{\mathrm{xt}}$ are $(m-1)\times 1$ vectors, and $\boldsymbol{\pi}'_y$ and $\boldsymbol{\nu}'_{yt}$ are $1\times n$ and $1\times p$ vectors, respectively, whereas $\boldsymbol{\Pi}_x$ and $\mathbf{V}_{xt}$

are $(m-1)\times n$ and $(m-1)\times p$ matrices. Letting $\mathbf{c}=[1,-\boldsymbol{\beta}']'$, $\alpha=\mu_y-\boldsymbol{\beta}'\boldsymbol{\mu}_x$, $\kappa=\delta_y-\boldsymbol{\beta}'\boldsymbol{\delta}_x$, $e_t=\varepsilon_{yt}-\boldsymbol{\beta}'\boldsymbol{\varepsilon}_{xt}=\mathbf{c}'\boldsymbol{\varepsilon}_t$, $\mathbf{q}'=\boldsymbol{\pi}_y'-\boldsymbol{\beta}'\boldsymbol{\Pi}_x=\mathbf{c}'\boldsymbol{\Pi}$, and $\boldsymbol{\nu}_t'=\boldsymbol{\nu}_{yt}'-\boldsymbol{\beta}'\mathbf{V}_{xt}=\mathbf{c}'\mathbf{V}_t$, then we have

$$y_t=\alpha+\kappa t+\mathbf{x}_t'\boldsymbol{\beta}+u_t, \tag{3}$$

$$u_t=e_t+\mathbf{q}'\mathbf{w}_t+\boldsymbol{\nu}_t'\mathbf{h}_t. \tag{4}$$

Thus, the regression error term u_t is composed of the stationary term e_t, the integrated term $\mathbf{q}'\mathbf{w}_t$, and the heteroskedastic component $\boldsymbol{\nu}_t'\mathbf{h}_t$. Note that u_t need not have zero mean, so that α is not an intercept in the usual sense. In the regression framework we assume that there is only one cointegrating vector, so that $\text{rank}(\boldsymbol{\Pi}_x)=m-1$, which imposes the restriction that $n\geq m-1$. This implies that further subrelationships among the $\mathbf{x}_t$ variables in (3) are excluded.[6] The null hypothesis of stochastic cointegration against the alternative of noncointegration can now be expressed via (3) as $H^0:\mathbf{q}=\mathbf{0}$ and $H^1:\mathbf{q}\neq\mathbf{0}$. Within H^0, the null hypothesis of stationary cointegration against the heteroskedastic alternative is $H_0^0:E(\boldsymbol{\nu}_t'\boldsymbol{\nu}_t)=0$ against $H_1^0:E(\boldsymbol{\nu}_t'\boldsymbol{\nu}_t)>0$.

For later use, we also define the lag covariances for an arbitrary process $\{a_t\}$ by

$$\gamma_j(a_t)=T^{-1}\sum_{s=j+1}^{T}a_s a_{s-j}$$

and define a heteroskedasticity and autocorrelation consistent (HAC) estimator of the long-run variance (LRV) by

$$\omega^2(a_t)=\gamma_0(a_t)+2\sum_{j=1}^{l}\lambda(j/l)\gamma_j(a_t), \tag{5}$$

where $\lambda(.)$ is a window with lag truncation parameter l. We also assume that Assumption KN, which follows, holds.

Assumption KN (Kernel and lag length).

1. $\lambda(0)=1$.
2. $0\leq\lambda(x)\leq 1$ for $0\leq x<1$.
3. $\lambda(x)$ is continuous and of bounded variation on $[0,1]$.
4. $l\to\infty$ as $T\to\infty$.

3.1. Testing H^0 against H^1

To test stochastic cointegration against noncointegration we need to test whether $\mathbf{q}=\mathbf{0}$ in

$$u_t=e_t+\mathbf{q}'\mathbf{w}_t+\boldsymbol{\nu}_t'\mathbf{h}_t.$$

Here, the null hypothesis is composite, encompassing both stationary and heteroskedastic cointegration; whereas the alternative is $I(1)$ or heteroskedastic integration. Because of the level of generality being entertained it is, however, not clear as to how to construct an optimal test statistic with a tractable limit distribution (even if we restrict ourselves to making Gaussian i.i.d. assumptions about the distributions of the unobserved variables). These complications lead us to examine instead a simple statistic for which we can at least determine a limiting null distribution free of nuisance parameters and also establish consistency. To this end, we consider

$$S_{nc} = \sum_{t=k+1}^{T} u_t u_{t-k}. \tag{6}$$

In the situation where all the disturbance terms are i.i.d., S_{nc} with $k = 1$ would test for zero autocorrelation in u_t against the correlation induced by the $I(1)$ term $\mathbf{q}'\mathbf{w}_t$. When the disturbance terms are not i.i.d., S_{nc} needs to be modified to eliminate nuisance parameter dependence resulting from autocorrelation and also from the presence of $\boldsymbol{\nu}_t'\mathbf{h}_t$. This is accomplished by allowing k to increase *with T*.[7] Under the cointegrating null, H^0, the statistic S_{nc} (when standardized with a HAC variance estimator) is asymptotically $N(0,1)$ and is consistent under the alternative of no cointegration, H^1. This is the content of Theorem 1, which follows. Because of the linear process representation, letting k become large eliminates correlation between u_t and u_{t-k} under H^0, whereas the HAC variance estimator takes care of the term $\boldsymbol{\nu}_t'\mathbf{h}_t$.[8] Under H^1, because of the presence of the $I(1)$ term $\mathbf{q}'\mathbf{w}_t$, letting k grow does not eliminate correlation between u_t and u_{t-k}. This distinction is the source of consistency of the test.

Because y_t and $\mathbf{x}_t$ are observed, we estimate $\mathbf{b} = [\alpha, \kappa, \boldsymbol{\beta}']'$ of (3) by means of the estimator $\hat{\mathbf{b}}_k = [\hat{\alpha}_k, \hat{\kappa}_k, \hat{\boldsymbol{\beta}}_k']'$ given by

$$\hat{\mathbf{b}}_k = \left(\sum_{t=k+1}^{T} \mathbf{X}_{t-k}\mathbf{X}_t' \right)^{-1} \sum_{t=k+1}^{T} \mathbf{X}_{t-k} y_t, \tag{7}$$

where $\mathbf{X}_t = [1, t, \mathbf{x}_t']'$. This estimator, described in Harris et al. (2002), is called an asymptotic instrumental variables estimator (AIV). Under H^0, a minor modification of the proof of Harris et al. (2002) shows that $\hat{\boldsymbol{\beta}}_k$ is consistent as k and $T \to \infty$, in contrast to the ordinary least squares (OLS) estimator, which is not consistent under heteroskedastic cointegration unless $\mathbf{x}_t$ consists entirely of $I(1)$ processes. We now construct (6) using the AIV residuals:

$$\hat{u}_t = y_t - \hat{\alpha}_k - \hat{\kappa}_k t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_k. \tag{8}$$

We then have the following result.

THEOREM 1. *Assume that the model (3), Assumption LP, and Assumption KN hold. If* $k = O(T^{1/2})$, $l = o(k)$, *and* $l < k$, *then* *(i) under H^0;*

$$\hat{S}_{nc} = \frac{T^{-1/2} \sum_{t=k+1}^{T} \hat{u}_t \hat{u}_{t-k}}{\sqrt{\omega^2(\hat{u}_t \hat{u}_{t-k})}} \xrightarrow{d} N(0,1);$$

(ii) under H^1, the distribution of $|\hat{S}_{nc}|$ diverges as $T \to \infty$.

Here $\hat{u}_t$ is defined in (8) using (7); $\omega^2(.)$ is defined in (5).

The first part of this theorem states that a properly standardized statistic, $\hat{S}_{nc}$, is asymptotically normal under stationary cointegration (which includes Engle and Granger cointegration) and also under heteroskedastic cointegration; the second part shows that the test is consistent under H^1. The same results arise if linear trends are excluded from (3) and the fitted model.

3.2. Testing H_0^0 against H_1^0

In decomposing the composite hypothesis H^0 into the null of stationary cointegration against the heteroskedastic alternative, we need to test whether $E(\boldsymbol{\nu}_t' \boldsymbol{\nu}_t) = 0$ in (4), maintaining $\mathbf{q} = \mathbf{0}$. Under the temporary assumption that e_t, $\boldsymbol{\nu}_t$, $\boldsymbol{\eta}_t$, and $\boldsymbol{v}_t$, are all jointly Gaussian i.i.d. and uncorrelated with each other, it follows from a straightforward application of McCabe and Leybourne (2000) that a locally most powerful test of H_0^0 against H_1^0 is given by

$$S_{hc} = \sum_{t=1}^{T} t u_t^2. \quad \textbf{(9)}$$

We then have the following result.

THEOREM 2. *Under the conditions of Theorem 1,*

(i) under H_0^0,

$$\hat{S}_{hc} = (1/12)^{1/2} \frac{T^{-3/2} \sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2)}{\sqrt{\omega^2(\hat{u}_t^2 - \hat{\sigma}_u^2)}} \xrightarrow{d} N(0,1);$$

(ii) under H_1^0, the distribution of $|\hat{S}_{hc}|$ diverges as $T \to \infty$.

Here $\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$; $\omega^2(.)$ is defined in (5).

Notice that $\hat{S}_{hc}$ is calculated using $\hat{u}_t^2 - \hat{\sigma}_u^2$, rather than simply $\hat{u}_t^2$, as (9) might suggest. This alteration is needed to center the statistic and render it invariant to the variance of u_t under H_0^0.

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

The structure of S_{hc} can also be used to test the null of $I(1)$ against the alternative of *HI* for any given individual series by simply constructing $\hat{S}_{hc}$ by redefining $\hat{u}_t$ as $\hat{u}_t = \Delta y_t - \hat{\delta}_y$ where $\hat{\delta}_y$ is an estimator of the trend coefficient δ_y given by $\hat{\delta}_y = T^{-1}\sum_{t=1}^{T}\Delta y_t$. We denote this statistic $\hat{S}_{hi}$. It is a straightforward special case of our results to show that $\hat{S}_{hi} \xrightarrow{d} N(0,1)$ if y_t is $I(1)$ and $|\hat{S}_{hi}|$ diverges if y_t is *HI*. The same results arise if linear trends are excluded from (3), in which case $\hat{u}_t = \Delta y_t$.[9]

4. SIMULATION RESULTS

In this section we investigate, via Monte Carlo simulation, the finite-sample behavior of our new tests, comparing these with tests applied assuming the conventional paradigm. To test for the null of conventional cointegration we apply the Shin (1994) adaptation of the Kwiatkowski et al. (1992) (KPSS) stationarity test. This test uses an efficient OLS estimator in which $[T^{1/4}]$ ($[.]$ denoting the integer part) lead and lag terms in $\Delta\mathbf{x}_t$ are added into the regression equation of y_t on $\mathbf{x}_t$; see Saikkonen (1991) for details. We denote this test K_c. The tests $\hat{S}_{nc}$, $\hat{S}_{hc}$, $\hat{S}_{hi}$, and K_c all require the use of a kernel and a lag truncation parameter in their respective variance estimators. For all tests we use the Bartlett kernel for $\lambda(.)$. As regards choice of l, we allow two schemes. The first simply fixes $l = [12(T/100)^{1/4}]$, which is a fairly mainstream choice in the literature, whereas the second is the automatic data-dependent selection method of Newey and West (1994).[10] Here we enforce the restriction that, for our new tests, $l < k$ when l is chosen automatically. Regarding the choice of k, we wish to avoid choosing k in a data-dependent manner as the $O(T^{1/2})$ rate is designed to deal with *all* processes covered by Assumption LP. Of course, $O(T^{1/2})$ is not uniquely defined, and so different possibilities need to be considered. Here we examine three candidates. These are $k = [0.75T^{1/2}]$, $[T^{1/2}]$, and $[1.25T^{1/2}]$. Although not exhaustive, these choices nonetheless prove sufficient for us to gauge the finite-sample influence of different values of k and also for us to recommend a value for use in practice.

The simulation model we examine is (2) with $m = n = p = 2$. Specifically, our data-generating process is

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & d_1 \end{bmatrix}\begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix} + \begin{bmatrix} \nu_{yt} & 0 \\ 0 & \nu_{xt} \end{bmatrix}\begin{bmatrix} h_{1t} \\ h_{2t} \end{bmatrix}, \qquad \textbf{(10)}$$

and the stochastic processes of (10) are generated according to

$$\varepsilon_{yt} = \phi_{\varepsilon,y}\varepsilon_{yt-1} + \epsilon_{1t}, \qquad \varepsilon_{xt} = \phi_{\varepsilon,x}\varepsilon_{xt-1} + \epsilon_{2t},$$

$$\nu_{yt} = \phi_{v,y}\nu_{yt-1} + d_2\epsilon_{3t}, \qquad \nu_{xt} = \phi_{v,x}\nu_{xt-1} + d_3\epsilon_{4t},$$

$$\Delta w_{1t} = \epsilon_{5t}, \qquad \Delta w_{2t} = \epsilon_{6t},$$

$$\Delta h_{1t} = \epsilon_{7t}, \qquad \Delta h_{2t} = \epsilon_{8t}$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

with $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t}, \epsilon_{5t}, \epsilon_{6t}, \epsilon_{7t}, \epsilon_{8t})'$ a multivariate standard normal white noise process. Here the d_i, $i = 1,2,3$ are constants. Within this setup, if $d_1 = d_2 = d_3 = 0$, then H_0^0 is true and stationary cointegration between two $I(1)$ series pertains, whereas if $d_1 \neq 0$, H^1 is true and y_t and x_t are not cointegrated in any sense (irrespective of the status of d_2 and d_3). If $d_1 = 0$ with $d_2 \neq 0$ and/or $d_3 \neq 0$, there is heteroskedastic cointegration. This may exist either between two *HI* series ($d_2 \neq 0$ and $d_3 \neq 0$) or between an $I(1)$ and *HI* series (e.g., $d_2 = 0$ and $d_3 \neq 0$). The model is generated over $t = -99,\dots,0,1,\dots,T$, with the first 100 startup values discarded. We consider sample sizes of $T = 200,400,600$, and the number of replications for all experiments is 10,000. Table entries represent empirical rejection frequencies of the various tests, based on regressions allowing constants but not trends, at the nominal asymptotic 0.05 level (these being two-tailed tests in the case of $\hat{S}_{hi}$, $\hat{S}_{nc}$, and $\hat{S}_{hc}$). For brevity, we only report results for the $\hat{S}_{hi}$ tests applied to y_t. In terms of notation in the tables, if $\phi_{i,j}$ is not explicitly given, its value is set to zero. Variants of the tests based on the automatic lag selection are superscripted with an *a*.

In Table 1 we have $d_1 = d_2 = d_3 = 0$ throughout, so that H_0^0 is true—stationary cointegration between two $I(1)$ series. The $\hat{S}_{hi}$ test has near nominal size, indicating that $I(1)$ rather than *HI* series are present, and any additional serial correlation in the form of nonzero values of $\phi_{\varepsilon,y}$ clearly has little effect on its size. As regards the test $\hat{S}_{nc}$, its size is well controlled apart from when $\phi_{\varepsilon,y} = 0.9$ and $\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = 0.9$. Here, when $k = [0.75T^{1/2}]$ it is moderately oversized and thus too frequently indicates absence of cointegration. However, setting $k = [T^{1/2}]$ or $k = [1.25T^{1/2}]$ virtually removes the oversizing problems, especially if the automated variants are considered. When we examine the test $\hat{S}_{hc}$, we find that the choice of k has far less effect on the size. For $\phi_{\varepsilon,y} = 0.9$ and $\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = 0.9$, all three choices (whether based on automated variants or not) produce oversized tests and thus indicate spurious heteroskedastic cointegration, although the degree of oversizing is not particularly serious and is mostly ameliorated as the sample size increases. On the basis of these results then, specifically those pertaining to $\hat{S}_{nc}$, we would conclude that setting $k = [0.75T^{1/2}]$ is realistically too low to maintain reliable finite-sample size. Notice that the nonautomated KPSS cointegration test, K_c, is quite badly oversized when $\phi_{\varepsilon,y} = 0.9$ and $\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = 0.9$, and automating the lag choice struggles to correct this to a satisfactory degree. Interestingly, the automated K_c test can be badly oversized in the presence of negative autocorrelation, unless the sample size is large. None of the other tests, however, appear to be adversely affected by negative autocorrelation.

Table 2 examines the size and power of the tests under six different models of heteroskedastic cointegration, H_1^0. In the first four, both y_t and x_t are *HI* ($d_2 \neq 0$ and $d_3 \neq 0$); in the fifth y_t is *HI* ($d_2 \neq 0$) and x_t is $I(1)$ ($d_3 = 0$), with these roles being reversed in the sixth model. The size issue relates to $\hat{S}_{nc}$, and it is clear that the test does not appear particularly sensitive to k, with size being controlled reasonably well for all choices, across all model specifica-

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

Table 1. Size of the tests under stationary cointegration: $d_1 = d_2 = d_3 = 0$

T	$\hat{S}_{hi}$ (y_t)	$\hat{S}_{hi}^a$	$k=[0.75T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$k=[T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$k=[1.25T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	K_c	K_c^a
	$\phi_{\varepsilon,y} = 0.5$															
200	0.049	0.046	0.038	0.047	0.052	0.057	0.037	0.044	0.056	0.059	0.030	0.033	0.061	0.061	0.056	0.056
400	0.051	0.045	0.044	0.048	0.048	0.051	0.049	0.051	0.049	0.052	0.044	0.045	0.049	0.051	0.053	0.051
600	0.052	0.048	0.049	0.051	0.052	0.053	0.048	0.048	0.051	0.051	0.046	0.049	0.052	0.052	0.059	0.064
	$\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = 0.5$															
200	0.049	0.049	0.044	0.060	0.053	0.060	0.040	0.048	0.054	0.059	0.038	0.047	0.066	0.071	0.063	0.064
400	0.051	0.045	0.044	0.051	0.047	0.051	0.049	0.052	0.051	0.054	0.047	0.052	0.051	0.057	0.053	0.053
600	0.050	0.048	0.047	0.052	0.054	0.057	0.047	0.050	0.050	0.056	0.053	0.052	0.057	0.056	0.062	0.060
	$\phi_{\varepsilon,y} = -0.7$															
200	0.060	0.054	0.038	0.049	0.055	0.060	0.038	0.044	0.056	0.059	0.038	0.038	0.059	0.061	0.057	0.157
400	0.049	0.052	0.050	0.056	0.052	0.054	0.045	0.045	0.051	0.052	0.046	0.050	0.052	0.055	0.052	0.063
600	0.048	0.052	0.052	0.056	0.050	0.053	0.053	0.053	0.050	0.052	0.046	0.048	0.049	0.052	0.049	0.051
	$\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = -0.7$															
200	0.058	0.057	0.040	0.055	0.054	0.062	0.041	0.046	0.054	0.061	0.043	0.046	0.052	0.061	0.054	0.307
400	0.048	0.051	0.045	0.052	0.052	0.054	0.046	0.050	0.052	0.051	0.044	0.047	0.051	0.052	0.048	0.190
600	0.047	0.051	0.056	0.058	0.051	0.054	0.054	0.053	0.051	0.052	0.046	0.046	0.052	0.053	0.047	0.052
	$\phi_{\varepsilon,y} = 0.9$															
200	0.049	0.046	0.121	0.170	0.088	0.102	0.053	0.055	0.092	0.088	0.044	0.035	0.105	0.092	0.171	0.119
400	0.048	0.046	0.130	0.137	0.085	0.089	0.065	0.056	0.087	0.076	0.054	0.039	0.094	0.075	0.143	0.085
600	0.054	0.053	0.134	0.134	0.082	0.082	0.066	0.057	0.082	0.068	0.073	0.056	0.083	0.068	0.154	0.090
	$\phi_{\varepsilon,y} = \phi_{\varepsilon,x} = 0.9$															
200	0.048	0.046	0.138	0.193	0.100	0.115	0.054	0.055	0.098	0.098	0.044	0.034	0.116	0.100	0.238	0.150
400	0.048	0.047	0.145	0.155	0.081	0.081	0.065	0.054	0.083	0.072	0.051	0.037	0.090	0.081	0.215	0.126
600	0.058	0.055	0.136	0.136	0.074	0.076	0.070	0.058	0.077	0.067	0.065	0.053	0.081	0.068	0.218	0.113

TABLE 2. Size/power of the tests under heteroskedastic cointegration

T	$\hat{S}_{hi}$	$\hat{S}_{hi}^a$ (y_t)	$k=[0.75T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$k=[T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$k=[1.25T^{1/2}]$ $\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	K_c	K_c^a
	$d_1=0;\ d_2=d_3=0.1^{1/2};\ \phi_{v,y}=0.8$															
200	0.461	0.424	0.022	0.037	0.400	0.442	0.028	0.026	0.412	0.418	0.026	0.026	0.388	0.385	0.123	0.107
400	0.593	0.536	0.042	0.049	0.598	0.603	0.039	0.042	0.595	0.577	0.044	0.038	0.588	0.558	0.117	0.086
600	0.661	0.563	0.048	0.049	0.698	0.698	0.044	0.043	0.692	0.670	0.045	0.035	0.690	0.645	0.119	0.108
	$d_1=0;\ d_2=d_3=0.1^{1/2};\ \phi_{v,x}=0.8$															
200	0.444	0.416	0.068	0.081	0.402	0.441	0.049	0.051	0.381	0.388	0.038	0.033	0.374	0.360	0.461	0.378
400	0.578	0.527	0.055	0.061	0.570	0.575	0.051	0.046	0.557	0.547	0.053	0.047	0.547	0.514	0.601	0.478
600	0.647	0.557	0.063	0.034	0.664	0.664	0.056	0.053	0.660	0.635	0.061	0.057	0.638	0.590	0.694	0.560
	$d_1=0;\ d_2=d_3=0.1^{1/2};\ \phi_{v,y}=0.6,\ \phi_{v,x}=-0.6;\ \phi_{\varepsilon,y}=0.6, \phi_{\varepsilon,x}=-0.6$															
200	0.452	0.409	0.033	0.043	0.460	0.491	0.030	0.033	0.447	0.456	0.029	0.031	0.432	0.427	0.083	0.083
400	0.600	0.545	0.040	0.045	0.638	0.648	0.040	0.041	0.632	0.620	0.038	0.037	0.617	0.595	0.075	0.070
600	0.662	0.551	0.047	0.049	0.723	0.726	0.044	0.042	0.720	0.698	0.039	0.039	0.715	0.678	0.078	0.077
	$d_1=0;\ d_2=d_3=0.1^{1/2};\ \phi_{v,y}=-0.6,\ \phi_{v,x}=0.6;\ \phi_{\varepsilon,y}=-0.6, \phi_{\varepsilon,x}=0.6$															
200	0.402	0.383	0.034	0.043	0.432	0.465	0.030	0.034	0.430	0.459	0.030	0.032	0.419	0.416	0.365	0.334
400	0.536	0.512	0.038	0.042	0.642	0.649	0.040	0.041	0.635	0.628	0.040	0.040	0.620	0.596	0.457	0.383
600	0.604	0.558	0.045	0.046	0.715	0.717	0.047	0.046	0.715	0.704	0.044	0.043	0.701	0.668	0.522	0.430
	$d_1=0;\ d_2=0.1^{1/2};\ d_3=0;\ \phi_{v,y}=0.8$															
200	0.447	0.416	0.030	0.041	0.405	0.440	0.030	0.036	0.393	0.400	0.028	0.026	0.373	0.386	0.089	0.077
400	0.585	0.535	0.040	0.044	0.526	0.525	0.049	0.047	0.512	0.500	0.039	0.035	0.501	0.477	0.090	0.073
600	0.670	0.577	0.046	0.047	0.620	0.620	0.047	0.046	0.614	0.606	0.043	0.038	0.596	0.547	0.090	0.070
	$d_1=0;\ d_2=0;\ d_3=0.1^{1/2};\ \phi_{v,x}=0.8$															
200	0.045	0.044	0.077	0.086	0.377	0.367	0.054	0.058	0.356	0.367	0.041	0.036	0.348	0.338	0.506	0.405
400	0.045	0.047	0.069	0.074	0.488	0.492	0.062	0.059	0.483	0.473	0.061	0.051	0.464	0.447	0.667	0.511
600	0.050	0.052	0.062	0.063	0.600	0.587	0.060	0.057	0.588	0.568	0.063	0.055	0.574	0.533	0.739	0.566

tions. If anything, setting $k = [0.75T^{1/2}]$ sometimes leads to slight oversizing; setting $k = [1.25T^{1/2}]$ occasionally yields slight undersizing. When considering the power of $\hat{S}_{hc}$, both fixed and automated variants exhibit consistency. The power does not appear to change particularly dramatically across model specifications either. Power does tend to decrease monotonically as k increases, although the rate of decrease is fairly low. The test $\hat{S}_{hi}$ is also seen to be consistent (aside obviously from when y_t is $I(1)$). The behavior of the K_c test is much less predictable, however. This is because, as mentioned in Section 3, the distribution of K_c in the *HI* case depends on nuisance parameters. This test can have very low or reasonably high power to reject its null of stationary cointegration, depending on the nature of the heteroskedastic cointegration. For example, if x_t is $I(1)$ as in the fifth case, its power is trivial. If, on the other hand, if x_t is *HI* and ν_{xt} is persistent, as in the second or sixth case, it can reject stationary cointegration very frequently. This differing behavior is due to the inconsistency of the OLS estimator of β $(= 1)$ whenever x_t is *HI*.[11]

In Table 3, we examine the power of the tests under the case of no cointegration, H^1, here between two $I(1)$ series ($\hat{S}_{hi}$ is not included now). Consistency of $\hat{S}_{nc}$ is clearly evident, as is the role of k in determining its power. The power is seen to fall fairly rapidly with increasing k for both fixed and automated variants.[12] Notice also that power of $\hat{S}_{nc}$ often exceeds that of K_c. There is no contradiction here, however: the optimality properties associated with the raw form of the KPSS statistic, on which K_c is based, do not necessarily carry over to the current empirical version of the statistic, which needs to be robustified both to serial correlation and to endogeneity. It is also apparent that the power of K_c drops quite sharply when moving from the fixed to automated lag selection.

In unreported simulations, we also examined the properties of the tests when some endogeneity is introduced. The first case revisited H_0^0, stationary cointegration between two $I(1)$ series, where we set $\text{cor}(\epsilon_1, \epsilon_5) = -0.7$ and $\text{cor}(\epsilon_2, \epsilon_5) = 0.7$, such that the increment processes of ε_{yt} and ε_{xt} are correlated with that of the random walk w_{1t}. The sizes of $\hat{S}_{nc}$ and $\hat{S}_{hc}$ were largely unaffected by introducing such correlation. A second case revisited H_1^0, heteroskedastic cointegration between two *HI* series. Here we made w_{1t} and h_{1t} identical random walks, so that the $I(1)$ process driving part of the heteroskedasticity also drove the level of the processes. In addition, we set $\text{cor}(\epsilon_4, \epsilon_8) = 0.7$, such that the increment process of ν_{xt} was correlated with that of the random walk h_{2t} it multiplies into. Again, the size of $\hat{S}_{nc}$ remained reasonably accurate, and consistency of $\hat{S}_{hc}$ (and $\hat{S}_{hi}$) appeared unaffected. Full details of these simulations are available upon request.

All the preceding simulation results concerning $\hat{S}_{nc}$, $\hat{S}_{hc}$, and $\hat{S}_{hi}$ are pretty much in line with what we would expect given our theoretical results of Section 3 regarding asymptotic normality of the tests, their robustness to serial correlation and endogeneity, and their consistency. They all detect the appropriate departures from their respective null hypotheses. The choice of k remains

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

TABLE 3. Power of the tests under no cointegration

T	$k=[0.75T^{1/2}]$				$k=[T^{1/2}]$				$k=[1.25T^{1/2}]$					
	$\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	$\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	K_c	K_c^a
	$d_1=0.2;\ d_2=d_3=0$													
200	0.330	0.389	0.126	0.141	0.229	0.238	0.130	0.134	0.166	0.150	0.145	0.138	0.502	0.393
400	0.775	0.788	0.231	0.242	0.656	0.623	0.241	0.223	0.530	0.447	0.233	0.212	0.733	0.522
600	0.931	0.932	0.319	0.319	0.877	0.853	0.324	0.286	0.774	0.700	0.336	0.265	0.836	0.538
	$d_1=0.4;\ d_2=d_3=0$													
200	0.543	0.614	0.186	0.228	0.360	0.363	0.192	0.194	0.247	0.204	0.203	0.183	0.547	0.391
400	0.914	0.919	0.297	0.305	0.778	0.741	0.307	0.270	0.635	0.532	0.318	0.243	0.764	0.509
600	0.984	0.984	0.352	0.352	0.937	0.918	0.366	0.311	0.847	0.755	0.367	0.277	0.848	0.532
	$d_1=0.6;\ d_2=d_3=0$													
200	0.605	0.693	0.301	0.245	0.382	0.383	0.218	0.219	0.266	0.218	0.231	0.202	0.546	0.375
400	0.935	0.938	0.308	0.319	0.809	0.762	0.321	0.274	0.652	0.549	0.334	0.252	0.767	0.514
600	0.989	0.989	0.376	0.376	0.945	0.925	0.382	0.328	0.859	0.770	0.388	0.295	0.855	0.511

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

an issue, however. Predominantly led by the behavior of $\hat{S}_{nc}$, the facts are that setting k too low can, in certain situations, induce size distortions (cf. Table 1), whereas setting k too high leads to a loss of power (cf. Table 3). Moreover, it seems rather unlikely that such a trade-off can be entirely avoided however k is chosen. A reasonable compromise would appear to be the middle value of the three we have considered, and so we recommend setting $k = [T^{1/2}]$ as a matter of practice. Whether l is selected using a fixed or an automated method does not appear particularly crucial to our test's performance, and we would not favor one approach over the other.

Our results also highlight the problems of using OLS-based procedures such as K_c to test for cointegration. Inconsistency of the OLS estimator whenever the heteroskedastic cointegration involves x_t that is HI causes the test to reject, so that K_c is unable to discern between this situation (i.e., when series "differ" by a heteroskedastic but stochastically trendless term) and noncointegration (i.e., when series "differ" by a stochastic trend term). Of course, we may take the view that because neither situation represents a stationary cointegrating relation, a rejection of the null of stationary cointegration is an appropriate outcome. However, if the heteroskedastic cointegration involves x_t that is $I(1)$, the same test tends to no longer reject this null, which clearly cannot also represent an appropriate outcome. This of course means that the inference drawn can become crucially dependent on the ordering of the $I(1)$ and HI variables, even asymptotically. Such considerations do not apply to our new tests as their asymptotic distributions are free of nuisance parameters. It is also important to remember that when applying $\hat{S}_{nc}$ and $\hat{S}_{hc}$, we never actually need to *distinguish* between which series are $I(1)$ and HI. That is, we do not need to calculate the test $\hat{S}_{hi}$ for individual series. Perhaps the only rationale for calculating $\hat{S}_{hi}$ is that it may provide early warning of situations where it would be unwise to apply conventional cointegration tests.

5. AN EMPIRICAL EXAMPLE: THE TERM STRUCTURE OF INTEREST RATES

A necessary empirical condition for the expectations theory of the term structure of interest rates is that long-run and short-run interest rates cointegrate. We test this empirically using monthly data from the United States, Canada, the United Kingdom, and Japan, taken from the OECD/MEI database. A single long-run interest rate, L_t, and a variety of short-run rates, S_{it}, are used for each country, and we consider bivariate regressions of L_t on S_{it} and also the reordered regression of S_{it} on L_t.[13] We calculate the same array of statistics as in Section 4, where again the regressions include constants but not trends. We also calculate the standard KPSS stationarity test allowing a constant (denoted K_s), to test individual series for stationarity. Both fixed and automatic lag selection procedures are employed for all tests, and, in view of the results of the previous section, we set $k = [T^{1/2}]$.

The results are given in Table 4, where the entries are p-values of the tests based on the asymptotic distribution. Bold print indicates a p-value of 0.05 or less, and in the current context we will consider this to represent a rejection of the associated null hypothesis. As regards the individual series, we first note that the KPSS test, K_s, indicates rejection of $I(0)$ for every one of the 17 individual interest rate series considered. In addition, the $\hat{S}_{hi}$ test shows that all of these interest rate series appears to be HI rather than $I(1)$, so that excess volatility would certainly appear to be an issue for this data set.[14]

Turning now to the bivariate regression results, first L_t on S_{it}, we see that according to $\hat{S}_{nc}$, stochastic cointegration is not rejected for eight of the 13 pairs. In both Canada and the United Kingdom, the nonrejection is unambiguous. In the case of the United States the evidence is mixed; rejections are found for two of the four pairs considered. No evidence of stochastic cointegration at all is found for Japan, though the peculiar nature of Japanese short-run interest rates in recent times (being effectively zero) may partly explain this finding. According to the $\hat{S}_{hc}$ test of the eight pairwise regressions that do not reject stochastic cointegration, five represent stationary cointegration between HI series (three for Canada, two for the United Kingdom) and three represent heteroskedastic cointegration between HI series (two for the United States, one for the United Kingdom). This pattern of results is the same whether the lag selection is fixed or automated. When we consider the regressions of S_{it} on L_t, qualitatively, the results for Canada, the United Kingdom, and Japan are unchanged. The United States now shows no rejections of stochastic cointegration, with one of the four being stationary cointegration, one being heteroskedastic cointegration, and two being indeterminate. This makes the total of nonrejections now 10 out of the 13 pairs. Thus, there is certainly a reasonable consensus of support for the term structure of interest rates in these data, particularly if the somewhat anomalous case of Japan is excluded from consideration.

A less coherent picture emerges if we examine the outcomes from the OLS-based KPSS cointegration test, K_c. For regressions of L_t on S_{it}, conventional cointegration is rejected for every one of the 13 pairs of long- and short-run rates if a fixed lag selection is used (this drops to four rejections if lag selection is automated, though as shown earlier the power of this test can be a good deal lower than that of the fixed lag test). However, no rejections at all are obtained for the reordered regressions of S_{it} on L_t. Hence, the differing degrees of excess volatility of long- and short-run interest rate data appear to exert a substantial influence on the outcomes for conventional OLS-based cointegration tests, to the extent that inference can be crucially dependent on variable ordering. By way of a contrast, the new procedures we have proposed in this paper are designed to provide inference that is rather more robust when analyzing this sort of data.

TABLE 4. Application to bond market data: p-values of tests

	K_s	K_s^a	$\hat{S}_{hi}$	$\hat{S}_{hi}^a$	$\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	K_c	K_c^a	$\hat{S}_{nc}$	$\hat{S}_{nc}^a$	$\hat{S}_{hc}$	$\hat{S}_{hc}^a$	K_c	K_c^a
					L_t on S_{it} regression						S_{it} on L_t regression					
United States																
L_t	**0.00**	**0.01**	**0.01**	**0.01**												
S_{1t}	**0.00**	**0.01**	**0.04**	**0.04**	0.08	0.08	**0.03**	**0.04**	**0.02**	0.08	0.11	0.13	**0.05**	0.07	0.14	0.28
S_{2t}	**0.00**	**0.02**	**0.04**	**0.05**	**0.05**	**0.05**	**0.04**	**0.05**	**0.00**	**0.03**	0.12	0.13	0.06	0.07	0.13	0.28
S_{3t}	**0.00**	**0.01**	**0.03**	**0.03**	0.06	0.07	**0.03**	**0.04**	**0.02**	0.09	0.09	0.11	**0.05**	0.07	0.14	0.28
S_{4t}	**0.00**	**0.01**	**0.03**	**0.03**	**0.05**	**0.05**	**0.02**	**0.03**	**0.02**	0.09	0.08	0.10	**0.04**	**0.04**	0.11	0.23
Canada																
L_t	**0.00**	**0.01**	**0.00**	**0.00**												
S_{1t}	**0.00**	**0.01**	**0.02**	**0.02**	0.19	0.19	0.52	0.52	**0.03**	0.08	0.25	0.25	0.75	0.75	0.37	0.52
S_{2t}	**0.00**	**0.01**	**0.00**	**0.00**	0.19	0.19	0.25	0.25	**0.03**	0.08	0.25	0.25	0.36	0.36	0.39	0.55
S_{3t}	**0.00**	**0.01**	**0.00**	**0.00**	0.13	0.13	0.19	0.19	**0.01**	**0.04**	0.22	0.22	0.40	0.40	0.30	0.47
United Kingdom																
L_t	**0.00**	**0.01**	**0.00**	**0.00**												
S_{1t}	**0.00**	**0.01**	**0.00**	**0.00**	0.32	0.34	0.08	0.09	**0.01**	**0.05**	0.36	0.37	0.13	0.14	0.42	0.61
S_{2t}	**0.00**	**0.01**	**0.00**	**0.00**	0.54	0.54	**0.03**	**0.03**	**0.01**	**0.03**	0.58	0.58	**0.04**	**0.04**	0.44	0.61
S_{3t}	**0.00**	**0.01**	**0.00**	**0.00**	0.32	0.33	0.15	0.16	**0.01**	0.06	0.36	0.37	0.21	0.22	0.44	0.64
Japan																
L_t	**0.00**	**0.01**	**0.03**	**0.03**												
S_{1t}	**0.00**	**0.02**	**0.00**	**0.00**	**0.02**	**0.02**	0.44	0.44	**0.03**	0.12	**0.02**	**0.02**	0.16	0.15	0.12	0.22
S_{2t}	**0.00**	**0.01**	**0.02**	**0.02**	**0.03**	**0.02**	0.41	0.41	**0.04**	0.10	**0.02**	**0.02**	0.12	0.12	0.10	0.19
S_{3t}	**0.00**	**0.01**	**0.01**	**0.00**	**0.02**	**0.02**	0.27	0.26	**0.04**	0.11	**0.03**	**0.02**	0.27	0.31	0.11	0.20

Note: The data used in Table 4 are defined as follows.
United States (1978:1–2002:12): L_t = government composite bond yield (>10 years); S_{1t} = federal funds rate; S_{2t} = prime rate; S_{3t} = rate on certificates of deposit; S_{4t} = U.S. dollar in London, 3-month deposit rate. Canada (1982:6–2002:12): L_t = benchmark bond yield (10 years); S_{1t} = official discount rate; S_{2t} = overnight money market rate; S_{3t} = rate on 90-day deposits. United Kingdom (1978:1–2002:12): L_t = yield on 10-year government bonds; S_{1t} = London clearing banks rate; S_{2t} = overnight interbank rate; S_{3t} = rate on 3-month interbank loans. Japan (1989:1–2002:12): L_t = yield on interest bearing government bonds (10 years); S_{1t} = official discount rate; S_{2t} = uncollateralized overnight rate; S_{3t} = rate on 90-day certificates of deposit.

NOTES

1. In Harris et al. (2002), $\mathbf{h}_t = \mathbf{w}_t$. Here if any element of $\mathbf{w}_t$ is identical to that of $\mathbf{h}_t$ we would simply delete the corresponding element of $\boldsymbol{v}_t$ from Assumption LP.

2. Here and throughout $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}'\mathbf{A})}$.

3. More formally, a vector stochastic process, $\mathbf{u}_t$, is said to be stochastically trendless if, as $s \to \infty$ (t fixed),

$$E(\mathbf{u}_{t+s}|\Im_t) - E(\mathbf{u}_{t+s}) \xrightarrow{p} 0,$$

where $\Im_t$ is the sigma field of information of all the elements in the vector up to time t. This implies that the mean square error optimal s step ahead forecasts of a stochastically trendless process converge to the unconditional mean of the process as the forecast horizon s increases. Following the Beveridge and Nelson (1981) definition, such a process has no stochastic trend (or permanent component), hence the terminology "stochastically trendless." An analogous definition has also been used in the literature on economic convergence; see Bernard and Durlauf (1996). Trendlessness is similar to the concept of a mixingale and the associated notion of asymptotic unpredictability, with the minor difference, in practical terms, that the convergence of the conditional expectation in our definition is in probability rather than in an L_p norm.

4. A proof of this result is available upon request.

5. There is a growing body of evidence that many economic and financial time series previously considered $I(1)$ are more appropriately modeled as *HI* or stochastic unit root processes. See the results in Section 5 of this paper and, inter alia, Hansen (1992a), Leybourne, McCabe, and Tremayne (1996), Granger and Swanson (1997), Wu and Chen (1997), and Psaradakis, Sola, and Spagnolo (2001).

6. A special case of this model is studied by Hansen (1992a). When $\mathbf{q} = \mathbf{0}$ and $\mathbf{V}_{xt} = \mathbf{0}$, (3) corresponds to a regression model when the regressors variables are all $I(1)$ and the error term is heteroskedastic, so that the regressand and regressors are treated asymmetrically.

7. The form of the statistic S_{nc} was earlier considered by Harris et al. (2003) in the context of stationarity testing in a deterministic regression.

8. Cointegrating versions of KPSS stationarity tests, such as that of Shin (1994), suffer from the fact that it is not possible to remove the effects of nuisance parameters in the partial sum process of u_t under the null of heteroskedastic cointegration, leading to incorrect size. The simulation studies of Section 4 confirm this.

9. Analogous statistics can of course be constructed for each element of the vector $\mathbf{x}_t$.

10. In the context of stationarity testing, this has been demonstrated by Hobijn, Franses, and Ooms (1998) to remove many of the well-documented oversizing problems associated with KPSS tests.

11. Busetti and Taylor (2003) demonstrate that the KPSS tests applied to an individual series with heteroskedastic errors can overreject the null of stationarity. In the current context, the cointegrating KPSS statistic actually diverges because of the inconsistency of the ordinary least squares estimator when x_t is *HI*.

12. These observations also apply to $\hat{S}_{hc}$, though it has rather less power than $\hat{S}_{nc}$ because it is not constructed to detect this alternative.

13. See the note to Table 4 for a full description of the data.

14. It is easily shown that the KPSS stationarity test is consistent when the alternative is *HI*.

REFERENCES

Bernard, A.B. & S.N. Durlauf (1996) Interpreting tests of the convergence hypothesis. *Journal of Econometrics* 71, 161–173.

Beveridge, S. & C. R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics* 7, 151–174.

Busetti, F. & A.M.R. Taylor (2003) Variance shifts, structural breaks and stationarity tests. *Journal of Business & Economic Statistics* 21, 510–531.

Campbell, J.Y. & R.J. Shiller (1987) Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1088.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Granger, C.W.J. & N.R. Swanson (1997) An introduction to stochastic unit root processes. *Journal of Econometrics* 80, 35–62.

Hansen, B.E. (1992a) Heteroscedastic cointegration. *Journal of Econometrics* 54, 139–158.

Hansen, B.E. (1992b) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Harris, D., B.P.M. McCabe, & S.J. Leybourne (2002) Stochastic cointegration: Estimation and inference. *Journal of Econometrics* 111, 363–384.

Harris, D., B.P.M. McCabe, & S.J. Leybourne (2003) Some limit theory for infinite order autocovariances. *Econometric Theory* 19, 829–864.

Hobijn, B., P.H. Franses, & M. Ooms (1998) Generalizations of the KPSS-Test for Stationarity. Econometric Institute Working paper, Erasmus University, Rotterdam.

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

Leybourne, S.J., B.P.M. McCabe, & A.R. Tremayne (1996) Can economic time series be differenced to stationarity? *Journal of Business & Economic Statistics* 14, 435–446.

McCabe, B.P.M. & S.J. Leybourne (2000) A general method of testing for random parameter variation in statistical models. In R.D.H Heijmans, D.S.G. Pollock, & A. Satorra (eds.), *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker*, pp. 75–85. Kluwer.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance estimation. *Review of Economic Studies* 61, 631–654.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Psaradakis, Z., M. Sola, & F. Spagnolo (2001) A simple procedure for detecting periodically collapsing rational bubbles. *Economics Letters* 72, 317–323.

Saikkonen, P. (1991) Asymptotically efficient estimation of cointegration regressions. *Econometric Theory* 7, 1–21.

Shin, Y. (1994) A residual based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.

Wu, J.-L. & S.-L. Chen (1997) Can nominal exchange rates be differenced to stationarity? *Economics Letters* 55, 397–402.

APPENDIX: Proofs

Notation and Conventions. In what follows we assume that Assumptions LP and KN, the model (3), and $k = O(T^{1/2})$ hold. For the model specified by equations (1)–(3), with $\boldsymbol{\zeta}_t = [\boldsymbol{v}_t', \text{vec}(\mathbf{V}_t)', \boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t']'$, let $\mathbf{C} = \sum_{j=0}^{\infty} \mathbf{C}_j$ and $\mathbf{G}_j = \sum_{i=j\vee 0}^{\infty} (\mathbf{C}_{i-j} \otimes \mathbf{C}_i)$ and define covariance matrices $\boldsymbol{\Omega}_{11} = \mathbf{C}\mathbf{C}'$ and $\boldsymbol{\Omega}_{22} = \sum_{j=-\infty}^{\infty} \mathbf{G}_j \mathbf{G}_j'$. Also define $\mathbf{S}_t$ to be the partial sum of the $\boldsymbol{\zeta}_t$, that is, $\Delta \mathbf{S}_t = \boldsymbol{\zeta}_t$. Selector matrices $\mathbf{R}_{\boldsymbol{v}}$, $\mathbf{R}_{\boldsymbol{\nu}}$, $\mathbf{R}_{\eta}$, and $\mathbf{R}_{\boldsymbol{\varepsilon}}$ are defined implicitly such that $\boldsymbol{v}_t = \mathbf{R}_{\boldsymbol{v}}' \boldsymbol{\zeta}_t$, $\boldsymbol{\nu}_t = \mathbf{R}_{\boldsymbol{\nu}}' \boldsymbol{\zeta}_t$, $\boldsymbol{\eta}_t = \mathbf{R}_{\boldsymbol{\eta}}' \boldsymbol{\zeta}_t$, and $\boldsymbol{\varepsilon}_{\text{t}} = \mathbf{R}_{\boldsymbol{\varepsilon}}' \boldsymbol{\zeta}_t$. When taking expectations through an infinite summation sign, we generally do not remark on the operation when obviously square summable linear processes are involved.

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

For transparency, we analyze the regression model without a time trend included, though all our results can be shown to extend to the trend case. We also make repeated use of the following representations:

$$\hat{u}_t = u_t - (\hat{\mathbf{b}}_k - \mathbf{b})'\mathbf{X}_t, \tag{A.1}$$

$$\begin{aligned}\hat{u}_t\hat{u}_{t-k} &= u_t u_{t-k} - (\hat{\mathbf{b}}_k - \mathbf{b})'\mathbf{X}_t u_{t-k} - u_t \mathbf{X}'_{t-k}(\hat{\mathbf{b}}_k - \mathbf{b}) + (\hat{\mathbf{b}}_k - \mathbf{b})'\mathbf{X}_t\mathbf{X}'_{t-k}(\hat{\mathbf{b}}_k - \mathbf{b})\\ &= u_t u_{t-k} + z_{k,t}\end{aligned} \tag{A.2}$$

with $\mathbf{X}_t = [1, \mathbf{x}'_t]'$ and $\hat{\mathbf{b}}_k - \mathbf{b} = [(\hat{\alpha}_k - \alpha), (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})']'$ and where $z_{k,t}$ is defined implicitly.

When dealing with LRV terms it is convenient to utilize the following results. First, in manipulating expressions involving kernels we adopt the notation $\lambda^+(j/l) = 2\lambda(j/l)$, $j > 0$, $\lambda^+(0) = 1$. Next, for any sequences $\{a_t\}$ and $\{b_t\}$ define

$$\gamma_j(a,b) = T^{-1}\sum_{s=j+1}^{T} a_s b_{s-j}.$$

We use the convention that $\gamma_j(a) = \gamma_j(a,a)$. Then for the sequence $\{a_t + b_t\}$ we have

$$\gamma_j(a+b) = \gamma_j(a) + \gamma_j(a,b) + \gamma_j(b,a) + \gamma_j(b).$$

Also define for any sequences $\{a_t\}$ and $\{b_t\}$

$$\omega(a,b) = \sum_{j=0}^{l} \lambda^+(j/l)\gamma_j(a,b),$$

again with the convention that $\omega^2(a) = \omega(a,a)$. So, we have for the sequence $\{a_t + b_t\}$,

$$\omega^2(a+b) = \omega^2(a) + \omega(a,b) + \omega(b,a) + \omega^2(b).$$

Thus, for $\delta > 0$ we can write

$$|T^{-\delta}\{\omega^2(a+b) - \omega^2(a)\}| \le |T^{-\delta}\omega(a,b)| + |T^{-\delta}\omega(b,a)| + T^{-\delta}\omega^2(b). \tag{A.3}$$

Note too that

$$|T^{-\delta}\omega(a,b)| \le \sum_{j=0}^{l}\lambda^+(j/l)\cdot\sqrt{T^{-(\delta+1)}\sum_{t=1}^{T}a_t^2}\cdot\sqrt{T^{-(\delta+1)}\sum_{t=1}^{T}b_t^2} \tag{A.4}$$

with the obvious modification for $a = b$.

In our applications a_t is often a product sequence, $a_t = c_t c_{t-k}$, say. The summation in s starts at $k + j + 1$ and in t starts at $t = k + 1$. Then, (A.4) yields

$$\begin{aligned}|T^{-\delta}\omega(a,b)| &\le \sum_{j=0}^{l}\lambda^+(j/l)\cdot\sqrt{T^{-(\delta+1)}\sum_{t=k+1}^{T}(c_t c_{t-k})^2}\cdot\sqrt{T^{-(\delta+1)}\sum_{t=1}^{T}b_t^2}\\ &\le \sum_{j=0}^{l}\lambda^+(j/l)\cdot\sqrt{T^{-(\delta+1)}\sum_{t=1}^{T}c_t^4}\cdot\sqrt{T^{-(\delta+1)}\sum_{t=1}^{T}b_t^2}.\end{aligned} \tag{A.5}$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

Proof of Theorems. We also use the following lemmas in establishing the results of Theorems 1 and 2.

LEMMA 1. *Under* H_0^0, $\omega^2(\hat{u}_t\hat{u}_{t-k}) \xrightarrow{p} \omega_{e1}^2$ *where*

$$\omega_{e1}^2 = \lim_{T\to\infty} T^{-1}\,\mathrm{var}\left(\sum_{t=k+1}^{T} e_t e_{t-k}\right) = \mathbf{c}'\mathbf{R}_\varepsilon'\mathbf{\Omega}_{22}\mathbf{R}_\varepsilon\mathbf{c}.$$

Proof. In this case $u_t = e_t$ and $\hat{u}_t\hat{u}_{t-k} = e_t e_{t-k} + z_{k,t}$. Setting $\delta = 0$, $a_t = e_t e_{t-k}$, and $b_t = z_{k,t}$ we have that $|\omega^2(\hat{u}_t\hat{u}_{t-k}) - \omega^2(e_t e_{t-k})|$ is bounded by (A.3). The first term in (A.3) is bounded by (A.5). That is

$$|\omega(e_t e_{t-k}, z_{k,t})| \le \sum_{j=0}^{l} \lambda^+(j/l)\cdot\sqrt{T^{-1}\sum_{t=1}^{T} e_t^4}\cdot\sqrt{T^{-1}\sum_{t=k+1}^{T} z_{k,t}^2},$$

where the order of the first right-hand side term is $O(l)$ (Assumption KN.2) and the second term is $O_p(1)$, independent of k, by Markov's inequality and Assumption LP. As for the third term, recalling the expression for $z_{k,t}$ in (A.2), note that $T^{1/2}\mathbf{D}_T(\hat{\mathbf{b}}_k - \mathbf{b})$ is $O_p(1)$ where $\mathbf{D}_T = \mathrm{diag}[1, \sqrt{T}\mathbf{I}_m]$ as follows from Harris et al. (2002). Thus, in (A.2), the quadratic form in $(\hat{\mathbf{b}}_k - \mathbf{b})$ is of a lower order than the two linear terms in $(\hat{\mathbf{b}}_k - \mathbf{b})$. The linear terms are of the same order. So the two dominant terms in $T^{-1}\sum_{t=k+1}^{T} z_{k,t}^2$ are $T^{-1}\sum_{t=k+1}^{T}(\hat{\mathbf{b}}_k - \mathbf{b})'X_t u_{t-k}^2 X_t'(\hat{\mathbf{b}}_k - \mathbf{b})$ and $T^{-1}\sum_{t=k+1}^{T}(\hat{\mathbf{b}}_k - \mathbf{b})'X_{t-k} u_t^2 X_{t-k}'(\hat{\mathbf{b}}_k - \mathbf{b})$. But

$$\left\| T^{-1}\sum_{t=k+1}^{T}(\hat{\mathbf{b}}_k - \mathbf{b})'X_t u_{t-k}^2 X_t'(\hat{\mathbf{b}}_k - \mathbf{b})\right\|$$

$$\le T^{-1}\|T^{1/2}\mathbf{D}_T(\hat{\mathbf{b}}_k - \mathbf{b})\|^2\sqrt{T^{-1}\sum_{t=1}^{T}\|\mathbf{D}_T^{-1}X_t X_t'\mathbf{D}_T^{-1}\|^2}\cdot\sqrt{T^{-1}\sum_{t=1}^{T} u_t^4}$$

$$= T^{-1}.O_p(1).O_p(1).O_p(1),$$

and it is clear that the second dominant term is of the same order. So, $\sqrt{T^{-1}\sum_{t=k+1}^{T} z_{k,t}^2} = O_p(T^{-1/2})$. Hence

$$|\omega(e_t e_{t-k}, z_{k,t})| = O_p(lT^{-1/2}).$$

The same method of proof shows that $|\omega(z_{k,t}, e_t e_{t-k})|$ and $\omega^2(z_{k,t})$ are also $O_p(lT^{-1/2})$.
Thus,

$$|\omega^2(\hat{u}_t\hat{u}_{t-k}) - \omega^2(e_t e_{t-k})| = O_p(lT^{-1/2}).$$

Applying Theorem LRV of Harris, McCabe, and Leybourne (2003) (with $n = 1$, $\alpha = 2$, and $\boldsymbol{\mu} = 0$) then shows that $\omega^2(e_t e_{t-k}) \xrightarrow{p} \omega_{e1}^2$. ■

LEMMA 2. *Under* H_1^0, $T^{-2}\omega^2(\hat{u}_t\hat{u}_{t-k}) = T^{-2}\omega^2(\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k}) + O_p(lT^{-1/2})$.

Proof. Now $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{h}_t$ and $\hat{u}_t\hat{u}_{t-k} = u_t u_{t-k} + z_{k,t}$. Setting $\delta = 2$, $a_t = u_t u_{t-k}$, and $b_t = z_{k,t}$ we have that $|T^{-2}\{\omega^2(\hat{u}_t\hat{u}_{t-k}) - \omega^2(u_t u_{t-k})\}|$ is bounded by (A.3). The first term in (A.3) is bounded by (A.5). That is,

$$|T^{-2}\omega(u_t u_{t-k}, z_{k,t})| \le \sum_{j=0}^{l} \lambda^+(j/l)\cdot\sqrt{T^{-3}\sum_{t=1}^{T} u_t^4}\cdot\sqrt{T^{-3}\sum_{t=k+1}^{T} z_{k,t}^2},$$

where the first right-hand side term is $O(l)$ and the second $O_p(1)$. The dominant term of $T^{-3}\sum_{t=k+1}^{T} z_{k,t}^2$ is

$$\left\| T^{-3}\sum_{t=k+1}^{T}(\hat{\mathbf{b}}_k - \mathbf{b})'X_t u_{t-k}^2 X_t'(\hat{\mathbf{b}}_k - \mathbf{b})\right\|$$

$$\le T^{-1}\|\mathbf{D}_T(\hat{\mathbf{b}}_k - \mathbf{b})\|^2\sqrt{T^{-1}\sum_{t=1}^{T}\|\mathbf{D}_T^{-1}X_t X_t'\mathbf{D}_T^{-1}\|^2}\cdot\sqrt{T^{-3}\sum_{t=1}^{T}u_t^4}$$

$$= T^{-1}O_p(1).O_p(1).O_p(1),$$

where the first two $O_p(1)$ results can be shown to hold via a simple modification of the approach of Harris et al. (2002). Thus

$$T^{-3}\sum_{t=k+1}^{T} z_{k,t}^2 = O_p(T^{-1}),$$

and so $|T^{-2}\omega(u_t u_{t-k}, z_{k,t})|$ is bounded by an $O_p(lT^{-1/2})$ variable. That $|T^{-2}\omega(z_{k,t}, u_t u_{t-k})|$ and $T^{-2}\omega^2(z_{k,t})$ are also bounded by an $O_p(lT^{-1/2})$ variable follows similarly. Combining these results gives

$$|T^{-2}\{\omega^2(\hat{u}_t\hat{u}_{t-k}) - \omega^2(u_t u_{t-k})\}| = O_p(lT^{-1/2}).$$

Because e_t is of a lower order of magnitude than $\boldsymbol{\nu}_t'\mathbf{h}_t$ it follows by similar arguments that

$$|T^{-2}\{\omega^2(u_t u_{t-k}) - \omega^2(\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k})\}| = O_p(lT^{-1/2}). \qquad \blacksquare$$

LEMMA 3. *Under* H_1^0, $T^{-2}\omega^2(\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k}) \xrightarrow{d} \int_0^1 (W\otimes W)'\boldsymbol{\Omega}_{PP}(W\otimes W)$ *where*

$$\boldsymbol{\Omega}_{PP} = (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'\boldsymbol{\Omega}_{22}(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}}), \qquad W = \mathbf{R}_{\boldsymbol{\nu}}'B_1$$

with B_1 *a Brownian motion process.*

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

Proof. Write

$$
\begin{aligned}
&T^{-2}\omega^2(\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k})\\
&\quad=\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=k+j+1}^{T}(\mathbf{h}_{t-k}\otimes\mathbf{h}_t)'\,\mathrm{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}')\mathrm{vec}(\boldsymbol{\nu}_{t-j}\boldsymbol{\nu}_{t-k-j}')'(\mathbf{h}_{t-k-j}\otimes\mathbf{h}_{t-j})\\
&\quad=\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=k+j+1}^{T}(\mathbf{S}_{t-k}\otimes\mathbf{S}_t)'(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' \qquad \mathbf{(A.6)}\\
&\qquad\times\{\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}')\mathrm{vec}(\boldsymbol{\zeta}_{t-j}\boldsymbol{\zeta}_{t-k-j}')'\}(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'(\mathbf{S}_{t-k-j}\otimes\mathbf{S}_{t-j})\\
&\quad=T^{-3}\sum_{t=l+1}^{T-k}(\mathbf{S}_t\otimes\mathbf{S}_t)'(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})\\
&\qquad\times(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'\left[\sum_{j=0}^{l}\lambda^+(j/l)E\{\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\}E\{\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\}'\right](\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})\\
&\qquad\times(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'(\mathbf{S}_t\otimes\mathbf{S}_t)+o_p(1) \qquad \mathbf{(A.7)}\\
&\quad\xrightarrow{d}\int_0^1(B_1\otimes B_1)'(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'\boldsymbol{\Omega}_{22}(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'(B_1\otimes B_1)\\
&\quad=\int_0^1(W\otimes W)'\boldsymbol{\Omega}_{PP}(W\otimes W).
\end{aligned}
$$

The key to the proof lies in replacing $\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}')\mathrm{vec}(\boldsymbol{\zeta}_{t-j}\boldsymbol{\zeta}_{t-k-j}')'$ in (A.6) by $E\{\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\}E\{\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\}'$ in (A.7). This means that the convergence in square brackets is nonstochastic and thus the continuous mapping theorem (CMT) is sufficient to deduce the asymptotic distribution. Also the quantity in square brackets converges to $\boldsymbol{\Omega}_{22}$ because it can be shown to be a consistent estimate of the long-run variance of $\mathrm{vec}(\zeta_t\zeta_{t-k}')$, which is the definition of $\boldsymbol{\Omega}_{22}$, that is, $\boldsymbol{\Omega}_{22}=\lim_{T\to\infty}\mathrm{var}\{T^{-1/2}\sum_{t=k+1}^{T}\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}')\}$. Then $\boldsymbol{\Omega}_{PP}=(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'\boldsymbol{\Omega}_{22}(\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})$ by definition.

The validity of replacing $\mathrm{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}')\mathrm{vec}(\boldsymbol{\zeta}_{t-j}\boldsymbol{\zeta}_{t-k-j}')'$ by the double expectation involves establishing the following sequence of results (expressed in the scalar case for simplicity). That is,

$$
\begin{aligned}
&\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=k+j+1}^{T}S_tS_{t-j}S_{t-k}S_{t-k-j}\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j}\\
&\quad=\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=k+j+1}^{T}S_{t-k}^4\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j}+o_p(1)\\
&\quad=\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=k+j+1}^{T}S_{t-k}^4E_{t-k}(\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j})+o_p(1)\\
&\quad=\sum_{j=0}^{l}\lambda^+(j/l)T^{-3}\sum_{t=j+1}^{T-k}S_t^4\{E(\zeta_t\zeta_{t-j})\}^2+o_p(1)\\
&\quad=\left[\sum_{j=0}^{l}\lambda^+(j/l)\{E(\zeta_t\zeta_{t-j})\}^2\right]T^{-3}\sum_{t=1}^{T-k-l}S_t^4+o_p(1)\\
&\quad\xrightarrow{d}\omega_2^2\int_0^1B_1^4.
\end{aligned}
$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

The complete proofs of these steps are available from the authors on request. Notice that the last equality shows the virtue of using the expectation device as the CMT and then delivers the result in a very straightforward way. ■

LEMMA 4. *Under* H_0^0, $\omega^2(\hat{u}_t^2 - \hat{\sigma}_u^2) \xrightarrow{p} \omega_{e2}^2$ *where*

$$\omega_{e2}^2 = \lim_{T\to\infty} T^{-1} \operatorname{var}\left\{\sum_{t=1}^{T}(e_t^2 - \sigma_e^2)\right\}$$

and $\sigma_e^2 = E(e_t^2)$.

The proof is similar to that of Lemma 1 and is thus omitted.

LEMMA 5. *Let* $\boldsymbol{\zeta}_t$ *satisfy Assumption LP and let* $k = O(T^{1/2})$. *Then, as* $T \to \infty$,

$$\left[T^{-3/2}\sum_{t=k+1}^{T}(\mathbf{h}_{t-k}\otimes\mathbf{h}_t)'\operatorname{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}'),\, T^{-1}\mathbf{h}_{[Ts]-k}\otimes\mathbf{h}_{[Ts]-1},\, T^{-1/2}\sum_{t=k+1}^{[Ts]}\operatorname{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}')\right]$$
$$\Rightarrow\left[\int_0^1 (W\otimes W)'\,dP, W\otimes W, P\right],$$

where $W = \mathbf{R}_{\boldsymbol{\nu}}'B_1$ *and* $P = (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})'B_2$ *where* B_1 *and* B_2 *are independent Brownian motion processes.*

Proof. First rewrite using $\Delta\mathbf{S}_t = \boldsymbol{\zeta}_t$, so that

$$T^{-3/2}\sum(\mathbf{h}_{t-k}\otimes\mathbf{h}_t)'\operatorname{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}')$$
$$= T^{-3/2}\sum(\mathbf{R}_{\boldsymbol{\nu}}'\mathbf{S}_{t-k}\otimes\mathbf{R}_{\boldsymbol{\nu}}'\mathbf{S}_{t-1})'\operatorname{vec}(\mathbf{R}_{\boldsymbol{\nu}}'\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}'\mathbf{R}_{\boldsymbol{\nu}}) + T^{-3/2}\sum\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k}\boldsymbol{\nu}_t'\boldsymbol{\upsilon}_t.$$

The proof proceeds by applying the Beveridge–Nelson decomposition to the first term and showing that the second term is asymptotically negligible. We use the notation

$$\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k} = \mathbf{m}_{k,t} + \mathbf{r}_{k,t} - \Delta\tilde{\mathbf{r}}_{k,t},$$

where

$$\mathbf{m}_{k,t} = \sum_{j=1}^{\infty}\mathbf{G}_{k,k-j}\operatorname{vec}(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-j}'),\qquad \mathbf{r}_{k,t} = \sum_{i=-\infty}^{\infty}\mathbf{F}_{k,k-i}(L)\operatorname{vec}(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-i}'),$$

$$\tilde{\mathbf{r}}_{k,t} = \sum_{i=1}^{\infty}\tilde{\mathbf{G}}_{k,k-i}(L)\operatorname{vec}(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t-i}'),$$

and the coefficients are defined by

$$\mathbf{G}_{k,r}(L) = \sum_{j=r\vee 0}^{k-1}(\mathbf{C}_{j-r}\otimes\mathbf{C}_j)L^j,\qquad \mathbf{F}_{k,r}(L) = \sum_{j=r\vee k}^{\infty}(\mathbf{C}_{j-r}\otimes\mathbf{C}_j)L^j,$$

$$\tilde{\mathbf{G}}_{k,k-i}(L) = \sum_{j=0}^{k-2}\tilde{\mathbf{G}}_{k,k-i,j}L^j,\qquad \tilde{\mathbf{G}}_{k,k-i,j} = \sum_{r=(k-i)\vee(j+1)}^{k-1}\mathbf{C}_{r-k+i}\otimes\mathbf{C}_r.$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

Apply Theorem BN of Harris et al. (2003) to $\mathrm{vec}(\boldsymbol{\zeta}_t \boldsymbol{\zeta}'_{t-k})$ to get a martingale approximation, $\mathbf{m}_{k,t}$, a remainder term $\mathbf{r}_{k,t}$, and an overdifferenced factor $\Delta\tilde{\mathbf{r}}_{k,t}$. The idea is that the martingale term is dominant and that the dependence on k is absorbed into its variance. In this way the proof of convergence to a stochastic integral can be treated by conventional methods of analysis. Thus,

$$T^{-3/2} \sum (\mathbf{h}_{t-k} \otimes \mathbf{h}_t)' \, \mathrm{vec}(\boldsymbol{\nu}_t \boldsymbol{\nu}'_{t-k})$$

$$= T^{-3/2} \sum (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-1})' (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' \mathbf{m}_{k,t}$$

$$+ T^{-3/2} \sum (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-1})' (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' (\mathbf{r}_{k,t} - \Delta\tilde{\mathbf{r}}_{k,t})$$

$$+ T^{-3/2} \sum \mathbf{h}'_{t-k} \boldsymbol{\nu}_{t-k} \boldsymbol{\nu}'_t \boldsymbol{v}_t.$$

We find

$$T^{-3/2} \sum (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-1})' (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' \mathbf{r}_{k,t} \xrightarrow{p} \mathbf{0},$$

$$T^{-3/2} \sum (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-1})' (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' \Delta\tilde{\mathbf{r}}_{k,t} \xrightarrow{p} \mathbf{0},$$

$$T^{-3/2} \sum \mathbf{h}'_{t-k} \boldsymbol{\nu}_{t-k} \boldsymbol{\nu}'_t \boldsymbol{v}_t \xrightarrow{p} \mathbf{0}.$$

The first result follows directly from Theorem SI of Harris et al. (2003), and the second is established along very similar lines. The last follows by writing

$$T^{-3/2} \sum \mathbf{h}'_{t-k} \boldsymbol{\nu}_{t-k} \boldsymbol{\nu}'_t \boldsymbol{v}_t = T^{-3/2} \sum \mathbf{h}'_{t-k} \{\mathbf{a}_t - E_{t-k}(\mathbf{a}_t)\} + T^{-3/2} \sum \mathbf{h}'_{t-k} E_{t-k}(\mathbf{a}_t),$$

where $\mathbf{a}_t = \boldsymbol{\nu}_{t-k} \boldsymbol{\nu}'_t \boldsymbol{v}_t$. The first term can be shown to disappear on exploiting the properties of the increment process, that is, that $E_{t-k}\{\mathbf{a}_t - E_{t-k}(\mathbf{a}_t)\} = \mathbf{0}$; the second term disappears by applying Theorem 3.3 of Hansen (1992b).

Thus,

$$T^{-3/2} \sum (\mathbf{h}_{t-k} \otimes \mathbf{h}_t)' \, \mathrm{vec}(\boldsymbol{\nu}_t \boldsymbol{\nu}'_{t-k})$$

$$= T^{-3/2} \sum (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{t-1})' (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' \mathbf{m}_{k,t} + o_p(1).$$

Now, because $k = o(T)$, it follows from Theorem FCLT of Harris et al. (2003) that

$$\mathbf{U}_{T,[Ts]} \equiv T^{-1} (\mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{[Ts]-k} \otimes \mathbf{R}'_{\boldsymbol{\nu}} \mathbf{S}_{[Ts]-1}) \Rightarrow \mathbf{R}'_{\boldsymbol{\nu}} B_1 \otimes \mathbf{R}'_{\boldsymbol{\nu}} B_1,$$

jointly with

$$\mathbf{N}_{T,[Ts]} = T^{-1/2} \sum^{[Ts]} \mathrm{vec}(\boldsymbol{\nu}_t \boldsymbol{\nu}'_{t-k}) = \mathbf{M}_{T,[Ts]} + o_p(1) \Rightarrow (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})' B_2,$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

where $\mathbf{M}_{T,[Ts]} = T^{-1/2}\sum^{[Ts]}\mathbf{m}_{k,t}$. Thus Theorem SI of Harris et al. (2003) applies, and setting $B_Q \equiv (\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})'B_2 = P$ and $U \equiv \mathbf{R}_{\boldsymbol{\nu}}'B_1 \otimes \mathbf{R}_{\boldsymbol{\nu}}'B_1 = W \otimes W$ we have that

$$\left[T^{-1/2}\sum T^{-1}(\mathbf{R}_{\boldsymbol{\nu}}'\mathbf{S}_{t-k} \otimes \mathbf{R}_{\boldsymbol{\nu}}'\mathbf{S}_{t-1})'(\mathbf{R}_{\boldsymbol{\nu}} \otimes \mathbf{R}_{\boldsymbol{\nu}})'\mathbf{m}_{k,t}, \mathbf{U}_{T,[Ts]}, \mathbf{M}_{T,[Ts]}\right]$$

$$\Rightarrow \left[\int_0^1 (W \otimes W)'\,dP, U, P\right]. \qquad \blacksquare$$

Proof of Theorem 1.

Part (i) (Null distribution). Sections (a) and (b) derive the asymptotic null distribution of $\hat{S}_{nc}$ under H_0^0 and H_1^0, respectively.

(a) Under H_0^0, $u_t = e_t$ and from Harris et al. (2002), $(\hat{\mathbf{b}}_k - \mathbf{b})' = [O_p(T^{-1/2}), O_p(T^{-1})]$ and

$$T^{-2}\sum_{t=k+1}^{T}\mathbf{x}_t\mathbf{x}_{t-k}', \qquad T^{-1/2}\sum_{t=k+1}^{T}e_t, \qquad T^{-1}\sum_{t=k+1}^{T}\mathbf{x}_{t-k}e_t,$$

$$T^{-1}\sum_{t=k+1}^{T}\mathbf{x}_t e_{t-k}, \qquad T^{-3/2}\sum_{t=k+1}^{T}\mathbf{x}_{t-k}$$

are all $O_p(1)$. Consequently, using (A.2) we find

$$T^{-1/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k} = T^{-1/2}\sum_{t=k+1}^{T}e_t e_{t-k} + O_p(T^{-1/2}).$$

Because $e_t = \mathbf{c}'\boldsymbol{\varepsilon}_t$ is a linear combination of a vector linear process, it follows from an application of Theorem FCLT of Harris et al. (2003) that

$$T^{-1/2}\sum_{t=k+1}^{T}e_t e_{t-k} \xrightarrow{d} N(0, \omega_{e1}^2),$$

where by Lemma 1, $\omega^2(\hat{u}_t\hat{u}_{t-k}) \xrightarrow{p} \omega_{e1}^2$. Thus,

$$\hat{S}_{nc} = \frac{T^{-1/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k}}{\omega(\hat{u}_t\hat{u}_{t-k})} \xrightarrow{d} N(0,1).$$

(b) Under H_1^0, $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{h}_t$, and from a minor modification to the results of Harris et al. (2002), $(\hat{\mathbf{b}}_k - \mathbf{b})' = [O_p(1), O_p(T^{-1/2})]$, $T^{-1}\sum_{t=k+1}^{T}u_t$, and $T^{-3/2}\sum_{t=k+1}^{T}\mathbf{x}_t u_{t-k}$ are $O_p(1)$. Hence, using (A.2) we find

$$T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k} = T^{-3/2}\sum_{t=k+1}^{T}u_t u_{t-k} + O_p(T^{-1/2}).$$

Now, substituting $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{h}_t$, we can write

$$T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k} = T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k} + T^{-3/2}\sum_{t=k+1}^{T}e_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k}$$

$$+ T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{h}_t e_{t-k} + T^{-3/2}\sum_{t=k+1}^{T}e_t e_{t-k} + O_p(T^{-1/2})$$

$$= T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{h}_t\mathbf{h}_{t-k}'\boldsymbol{\nu}_{t-k} + O_p(T^{-1/2})$$

$$= T^{-3/2}\sum_{t=k+1}^{T}(\mathbf{h}_{t-k}\otimes\mathbf{h}_t)'\,\mathrm{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}') + O_p(T^{-1/2})$$

$$\xrightarrow{d}\int_0^1 (W\otimes W)'\,dP,$$

where $W = \mathbf{R}_{\boldsymbol{\nu}}'B_1$ and $P = (\mathbf{R}_\nu \otimes \mathbf{R}_\nu)'B_2$ and B_1 and B_2 are independent Brownian motions with covariance matrices $\boldsymbol{\Omega}_{11}$ and $\boldsymbol{\Omega}_{22}$. The weak convergence follows from Lemma 5. The covariance matrix of P is $\boldsymbol{\Omega}_{PP} = (\mathbf{R}_\nu\otimes\mathbf{R}_\nu)'\boldsymbol{\Omega}_{22}(\mathbf{R}_\nu\otimes\mathbf{R}_\nu)$.

Combining the results of Lemmas 2 and 3 shows that

$$T^{-2}\omega^2(\hat{u}_t\hat{u}_{t-k})\xrightarrow{d}\int_0^1 (W\otimes W)'\boldsymbol{\Omega}_{PP}(W\otimes W).$$

We now require the distribution of the ratio of $T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k}$ to $T^{-1}\omega(\hat{u}_t\hat{u}_{t-k})$. As shown in Lemma 5,

$$\left[T^{-3/2}\sum_{t=k+1}^{T}(\mathbf{h}_{t-k}\otimes\mathbf{h}_t)'\,\mathrm{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}'), T^{-1}\mathbf{h}_{[Ts]-k}\otimes\mathbf{h}_{[Ts]-1}, T^{-1/2}\sum_{t=k+1}^{[Ts]}\mathrm{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}')\right]'$$

$$\Rightarrow\left[\int_0^1 (W\otimes W)'\,dP, W\otimes W, P\right].$$

Next the CMT, with the preceding vector as argument and the ratio as the map, applies to conclude that

$$\frac{T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k}}{T^{-1}\omega(\hat{u}_t\hat{u}_{t-k})}\xrightarrow{d}\frac{\int_0^1 (W\otimes W)'\,dP}{\sqrt{\int_0^1 (W\otimes W)'\boldsymbol{\Omega}_{PP}(W\otimes W)}}. \tag{A.8}$$

As

$$\int_0^1 (W\otimes W)'\,dP \sim N\left(0, \int_0^1 (W\otimes W)'\boldsymbol{\Omega}_{PP}(W\otimes W)\right)$$

conditional on W, the distribution in (A.8) is unconditionally $N(0,1)$.

Part (ii) (Consistency). Under H^1, $u_t = e_t + \mathbf{q}'\mathbf{w}_t + \boldsymbol{\nu}_t'\mathbf{h}_t$ where $\mathbf{q} \neq 0$. Here, it is easy to show that $\hat{\alpha}_k - \alpha = O_p(T^{1/2})$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p(1)$, and, using (A.2), this implies that $\hat{u}_t\hat{u}_{t-k}$ is of the same order in probability as $u_t u_{t-k}$. It is then straightforward to deduce that

$$T^{-1/2} \sum_{t=k+1}^{T} \hat{u}_t \hat{u}_{t-k} = O_p(T^{3/2}).$$

Now we require a bound for the order of probability of $\omega^2(\hat{u}_t\hat{u}_{t-k})$, which again is the same as the order of probability of $\omega^2(u_t u_{t-k})$. Setting $a = b = u_t u_{t-k}$ and $\delta = 2$ in (A.5) yields

$$\begin{aligned} l^{-1}T^{-2}\omega^2(u_t u_{t-k}) &\leq l^{-1} \sum_{j=0}^{l} \lambda^{+}(j/l).T^{-3} \sum_{t=1}^{T} u_t^4 \\ &= O(1).O_p(1). \end{aligned}$$

Thus we conclude that $\omega^2(\hat{u}_t\hat{u}_{t-k}) = O_p(lT^2)$ at most. Hence the distribution of $|\hat{S}_{nc}|$ diverges at least as fast as $O_p(\sqrt{T/l})$. ■

Proof of Theorem 2.

Part (i) (Null distribution). Under H_0^0, $u_t = e_t$ we have $\hat{\alpha}_k - \alpha = O_p(T^{-1/2})$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p(T^{-1})$. Then, it follows from (A.1) that

$$T^{-3/2} \sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2) = T^{-1/2} \sum_{t=1}^{T} \left(\frac{t}{T} - \frac{1}{2} - \frac{1}{2T} \right)(e_t^2 - \sigma_e^2) + O_p(T^{-1/2}),$$

where $\sigma_e^2 = E(e_t^2)$. Write

$$T^{-1/2} \sum_{t=1}^{T} \left(\frac{t}{T} - \frac{1}{2} - \frac{1}{2T} \right)(e_t^2 - \sigma_e^2) = \int_0^1 \left(s - \frac{1}{2} \right) dF_T(s) + o_p(1).$$

Here $F_T(s)$ is the partial sum process of $\{e_t^2 - \sigma_e^2\}$ that weakly converges to $F(s)$ by Theorem 3.8 of Phillips and Solo (1992). Then, noting by integration by parts that $\int_0^1 (s - \frac{1}{2})\, dF_T(s) = \frac{1}{2}F_T(1) - \int_0^1 F_T(s)\, ds$, we can use the CMT to deduce

$$T^{-3/2} \sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2) \xrightarrow{d} \int_0^1 \left(s - \frac{1}{2} \right) dF(s),$$

where $F(s)$ is a Brownian motion with variance ω_{e2}^2, as defined in Lemma 4. Hence, $\int_0^1 (s - \frac{1}{2})\, dF(s)$ is normally distributed with mean zero and variance

$$\omega_{e2}^2 \int_0^1 \left(s - \frac{1}{2} \right)^2 ds = 12\omega_{e2}^2,$$

which shows

$$T^{-3/2} \sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2) \xrightarrow{d} N(0, 12\omega_{e2}^2).$$

https://doi.org/10.1017/S026646660606021X Published online by Cambridge University Press

From Lemma 4, $\omega^2(\hat{u}_t^2 - \hat{\sigma}_u^2) \xrightarrow{p} \omega_{e2}^2$, and so the result follows.

Part (ii) (Consistency). Under H_1^0, $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{h}_t$ we have $\hat{\alpha}_k - \alpha = O_p(1)$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p(T^{-1/2})$. We may write

$$T^{-3/2}\sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2) = T^{-1/2}\sum_{t=1}^{T}\left(\frac{t}{T} - \frac{1}{2} - \frac{1}{2T}\right)\hat{u}_t^2.$$

From (A.1), $\hat{u}_t$ is of the same order in probability as u_t, and it is then straightforward to show that

$$T^{-1/2}\sum_{t=1}^{T}\left(\frac{t}{T} - \frac{1}{2} - \frac{1}{2T}\right)\hat{u}_t^2 = O_p(T^{3/2})$$

and hence

$$T^{-3/2}\sum_{t=1}^{T} t(\hat{u}_t^2 - \hat{\sigma}_u^2) = O_p(T^{3/2}).$$

In the denominator, $\omega^2(\hat{u}_t^2 - \hat{\sigma}_u^2)$ and $\omega^2(u_t^2 - \bar{\sigma}_u^2)$ (where $\bar{\sigma}_u^2 = T^{-1}\sum_{t=1}^{T} u_t^2$) are of the same order in probability. Setting $a = b = u_t^2 - \bar{\sigma}_u^2$ and $\delta = 2$ in (A.4) yields

$$\begin{aligned} l^{-1}T^{-2}\omega^2(u_t^2 - \bar{\sigma}_u^2) &\leq l^{-1}\sum_{j=0}^{l}\lambda^+(j/l).T^{-3}\sum_{t=1}^{T}(u_t^2 - \bar{\sigma}_u^2)^2 \\ &= O(1).\left\{T^{-3}\sum_{t=1}^{T} u_t^4 - \left(T^{-2}\sum_{t=1}^{T} u_t^2\right)^2\right\}. \end{aligned}$$

It is easily shown that both $T^{-3}\sum_{t=1}^{T} u_t^4$ and $T^{-2}\sum_{t=1}^{T} u_t^2$ are $O_p(1)$. Hence $\omega^2(u_t^2 - \bar{\sigma}_u^2)$ and consequently $\omega^2(\hat{u}_t^2 - \hat{\sigma}_u^2)$ are $O_p(lT^2)$ at most. So, the distribution of $|\hat{S}_{hc}|$ diverges at least as fast as $O_p(\sqrt{T/l})$. ■