Testing for Stochastic Cointegration and Evidence for Present Value Models

Brendan McCabe
School of Management,
University of Liverpool

Stephen Leybourne
Department of Economics,
University of Nottingham

David Harris
Department of Economics,
University of Melbourne

June 2003

Abstract

Using the stochastic integration/cointegration framework of Harris, McCabe and Leybourne (2002) we revisit the problem of assessing the empirical evidence for or against the present value class of models in the bond and stock markets. This framework allows for volatility in excess of that catered for by the conventional integration/cointegration paradigm by introducing nonlinear heteroscedasticity. We propose a test for stochastic cointegration against the alternative of no cointegration and a secondary test for stationary cointegration against the heteroscedastic alternative. Asymptotic distributions of these tests under their respective null hypotheses are derived and consistency under their respective alternatives is established. In contrast to conventional cointegration tests, which we show via simulation are unreliable in the presence of the kind of volatility typical of financial data, our tests are able to uncover new cointegration evidence in favour of the present value model, particularly in the bond market.

1 Introduction

The cointegration framework of Engle and Granger (1987) (EG) is characterized by two widely held stylized empirical facts. The first is that, of the set of economic time series that exhibit trending behaviour, many are adequately modelled by processes that are integrated, usually of order 1, $I(1)$. The second is that, despite this trending behaviour, such series often tend to co-move over time according to a stationary, or $I(0)$, process i.e. they are cointegrated. Many empirical tests of important theoretical economic hypotheses are carried out within the EG cointegration framework. Two noteworthy examples are the relationship between long run and short run interest rates and the relationship between dividends and stock prices, which belong to the class of hypotheses known as present value models (PVM); see Campbell and Shiller (1987, 1988). However, the EG approach has, perhaps surprisingly, uncovered only very limited empirical evidence in support of the PVM; see Campbell and Shiller (1987), Diba and Grossman (1988), Froot and Obstfeld (1991), Lamont (1998), Balke and Wohar (2001, 2002). An explanation often put forward for this is that bond and stock 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 consistent with $I(1)$ behaviour, particularly when using higher frequency data. When co-movements between series are analyzed, most simply by examining the spreads (the differences between long and short run interest rates, or between dividends and prices), these also tend to display periods of volatility in excess of that which could be associated with stationary behaviour. In the words of Campbell and Shiller (1987), the spreads tend to "move too much".

An example of this type of volatility and co-movement is given in Figure 1. The series are UK long run and short run interest rates, specifically the monthly yield on 10 year Government bonds and the overnight interbank rate from 1978:1 to 2002:12. Both series display clear periods of volatility higher than is consistent with them being $I(1)$ (this is confirmed by statistical testing later) and, whilst the two series do appear to follow a broadly similar underlying path, there are also distinct episodes when they deviate from each other rather more radically than would tend to be the case if the spread was simply stationary. Particularly volatile episodes include the exports/exchange rate led recession of 1979-1982, together with the period 1989-1993 associated with the exit of Sterling from the European Exchange Rate Mechanism and the collapse of confidence after the preceding boom. In the EG framework, according to the PVM hypothesis, the series in Figure 1 should cointegrate, and with a coefficient of unity. Yet the empirical analysis carried out in Section 6 suggests that there is no cointegrating relationship between these series (let alone one with a unit coefficient). We also find the same is true for all similar pairs of interest rate series from several other major economies that we examine. The story relating dividends and prices in the US stock market is similar.

Prompted by the apparent inability of conventional methods to handle excess volatility in important economic variables, this paper takes another look at the evidence for the PVM by assessing if the series involved cointegrate, but in a setting that specifically allows for volatile behaviour in the integration/cointegration paradigm. The key to this is to replace the restrictive stationarity requirement of first differences of individual series and cointegrating error terms of the EG setup with a looser condition that these are simply free of $I(1)$ stochastic trend terms. In the next section, we define what is meant for a series to be *stochastically trendless* and show that the *stochastic integration/cointegration* framework of Harris, McCabe and Leybourne (2002) (HMLa) satisfies this requirement (whilst encompassing the EG setup as a special case). This approach permits a much richer range of behaviour than is possible under EG. Of particular importance in the present context is that it induces a nonlinear form of heteroscedasticity that gives rise to volatile behaviour in the first differences of individual series and in cointegrating error terms that closely mimic those seen in the real data.

Given that the HMLa framework seems to be an appropriate means of analysis for volatile data, 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 heteroscedastic), or not cointegrated. This parallels the EG approach and we suggest a statistic to test the null of stochastic cointegration based on regression residuals. Within stochastic cointegration, we also consider the hypothesis that the cointegration is stationary against the alternative that it is heteroscedastic and we suggest a second residual-based 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 heteroscedastic integration. Both statistics have the advantage of being very simple to construct.

In Section 4 the asymptotic null distributions of these two test statistics are derived under weak regularity conditions. Conveniently, both are shown to have straightforward normal limiting 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. Monte Carlo studies, which examine the finite sample size and power characteristics of the new tests, along with those of their conventional counterparts, in the stochastic cointegration environment are provided in Section 5. These highlight very clearly the benefits to be gained by adopting the new test procedures, together with the shortcomings of conventional ones. Finally, in Section 6 we give a detailed stochastic cointegration analysis of the evidence for and against the cointegration implications of the PVM in the bond and stock market.[1] For several major economies, our new testing framework uncovers evidence in favour of the PVM in the bond market. For the US stock market, evidence supporting the PVM is also found, contingent on the time period under study. In contrast, in neither case do we find that a parallel conventional cointegration analysis provides any evidence in support of the PVM whatsoever. Interestingly, for all the series we consider here, we conclude they are better modelled by heteroscedastically integrated rather than $I(1)$ processes.

2 Stochastic Integration and Cointegration

We first outline the model introduced in HMLa, augmented with a linear trend

$$\begin{aligned} \mathbf{z}_t &= \boldsymbol{\mu} + \boldsymbol{\delta} t + \Pi_t \mathbf{w}_t + \boldsymbol{\varepsilon}_t, \\ \mathbf{w}_t &= \mathbf{w}_{t-1} + \boldsymbol{\eta}_t, \\ \Pi_t &= \Pi + \mathbf{V}_t \end{aligned} \tag{1}$$

for $t = 1, \ldots, T$. Here $\mathbf{z}_t, \boldsymbol{\eta}_t, \mathbf{w}_t, \boldsymbol{\varepsilon}_t, \boldsymbol{\delta}$ and $\boldsymbol{\mu}$ are $m \times 1$ vectors while Π_t, Π and $\mathbf{V}_t$ are $m \times m$ matrices. Only the process $\mathbf{z}_t$ is observed. The disturbances $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\eta}_t$ and $\mathbf{V}_t$ are mean zero stationary processes, which may be correlated with one another, $\mathbf{w}_t$ is a vector integrated process with $\mathbf{w}_0 = \boldsymbol{\eta}_0$ and $\boldsymbol{\mu}$ is a vector of constants. The key feature of the model is that Π_t is random and this introduces shocks nonlinearly into the generating process for $\mathbf{z}_t$. Rewriting $\mathbf{z}_t$ as

$$\mathbf{z}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t + \Pi \mathbf{w}_t + (\boldsymbol{\varepsilon}_t + \mathbf{V}_t \mathbf{w}_t) \tag{2}$$

we see that $\mathbf{z}_t$ consists of an integrated process plus a new shock term. The shock term has a linear component $\boldsymbol{\varepsilon}_t$ and a nonlinear component $\mathbf{V}_t \mathbf{w}_t$ that is heteroscedastic through its dependence on the $I(1)$ process $\mathbf{w}_t$. Notice that $\mathbf{z}_t$ is not difference stationary as

$$\Delta \mathbf{z}_t = \boldsymbol{\delta} + \Pi \boldsymbol{\eta}_t + \Delta \boldsymbol{\varepsilon}_t + \mathbf{w}_{t-1} \Delta \mathbf{V}_t \tag{3}$$

involves the level $\mathbf{w}_{t-1}$. Let $\mathbf{e}_i$ be an $m \times 1$ vector with 1 in its i'th position and 0 elsewhere, so that $\mathbf{e}_i' \mathbf{z}_t = z_{it}$, the i'th element of the vector $\mathbf{z}_t$. Then, from (2), we have

$$z_{it} = \mathbf{e}_i' \boldsymbol{\mu} + \mathbf{e}_i' \boldsymbol{\delta} t + \mathbf{e}_i' \Pi \mathbf{w}_t + \mathbf{e}_i' (\boldsymbol{\varepsilon}_t + \mathbf{V}_t \mathbf{w}_t) \tag{4}$$

and if $\mathbf{e}_i' \Pi \neq \mathbf{0}$ 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 *heteroscedastically integrated* (HI) due to the term $\mathbf{e}_i' \mathbf{V}_t \mathbf{w}_t$; whilst if $\mathbf{e}_i' \mathbf{V}_t = 0$ then z_{it} is simply $I(1)$. So, a stochastically integrated variable encompasses both ordinary and heteroscedastic integration. The essential difference between HI and $I(1)$ is that the variance of a *change* is allowed to vary in the former case whilst it is constant in the latter. Examples of such series are given in Figures 2 and 3 which show, respectively, the first difference of UK long run and short run interest rates of Figure 1. These series appear highly heteroscedastic rather than stationary - our test statistics later confirm this in Section 6.

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

$$\mathbf{c}' \mathbf{z}_t = \mathbf{c}' \boldsymbol{\mu} + \mathbf{c}' \boldsymbol{\delta} t + \mathbf{c}' \Pi \mathbf{w}_t + \mathbf{c}' (\boldsymbol{\varepsilon}_t + \mathbf{V}_t \mathbf{w}_t). \tag{5}$$

If $\mathbf{c}' \Pi = \mathbf{0}$ then the variables of $\mathbf{z}_t$ are said to be *stochastically cointegrated*; otherwise they are not cointegrated.[2] Under stochastic cointegration $\mathbf{c}' \mathbf{z}_t = \mathbf{c}' (\boldsymbol{\mu} + \boldsymbol{\delta} t + \boldsymbol{\varepsilon}_t + \mathbf{V}_t \mathbf{w}_t)$ behaves like a stochastically

[1] For brevity we often refer to evidence for or against the cointegration implications of the PVM simply as evidence for or against the PVM.

[2] This definition of stochastic cointegration follows HMLa, and generalises that of Ogaki and Park (1998) for the EG framework in which $\mathbf{V}_t = 0$ a.s.. An extended definition of Ogaki and Park's concept of deterministic cointegration in our model is $\mathbf{c}' \boldsymbol{\delta} = 0$ in addition to $\mathbf{c}' \Pi = 0$.

integrated process *net* of its integrated component, and we will subsequently show that such a process is *stochastically trendless*, a term formalized below. If $\mathbf{c}'E(\mathbf{V}_t\mathbf{V}_t')\mathbf{c} = 0$, then $\mathbf{c}'\mathbf{z}_t = \mathbf{c}'(\boldsymbol{\mu} + \boldsymbol{\delta}t + \boldsymbol{\varepsilon}_t)$ is trend-stationary. If, in addition, $\mathbf{V}_t = \mathbf{0}$, the variables are all integrated and cointegrated in the *EG* sense. Because of the stationary behaviour of $\mathbf{c}'\mathbf{z}_t$ in either case, we simply refer to this as *stationary cointegration.* If $\mathbf{c}'E(\mathbf{V}_t\mathbf{V}_t')\mathbf{c} > 0$, the variables $\mathbf{z}_t$ are said to be *heteroscedastically cointegrated.* Thus, stochastic cointegration encompasses both stationary cointegration (possibly of the *EG* kind) and heteroscedastic cointegration. Under the PVM of the term structure, the long and short interest rates series in Figure 1 should stochastically cointegrate with a coefficient of unity. The interest spread is shown in Figure 4 and this shows little visual evidence of a stochastic trend, whereas heteroscedasticity remains a distinct possibility. We will see in Section 6 that our new tests indicate that these two series are in fact heteroscedastically cointegrated. In contrast, a conventional analysis finds them not to be cointegrated, despite the visual evidence to the contrary.

It remains to clarify the statistical properties of the error term $\mathbf{c}'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{w}_t)$ that defines stochastic cointegration. To do this we need to be more precise about the statistical properties of the disturbances in the model and so we make extensive use of the following linear process assumption in the remainder of the paper.

Assumption LP.
Let $\boldsymbol{\zeta}_t$ be generated by the vector linear process $\boldsymbol{\zeta}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\xi}_{t-j}$, where

1. $\sum_{j=0}^{\infty} j^2 \left\|\mathbf{C}_j\right\|^2 < \infty$ *with $\mathbf{C}_0$ having full rank.*[3]
2. $\boldsymbol{\xi}_t$ *is a martingale difference sequence with respect to $\Im_t$, the sigma field generated by $\{\boldsymbol{\xi}_s, s \leq t\}$,*
3. $E\left(\boldsymbol{\xi}_t\boldsymbol{\xi}_t' \middle| \Im_{t-1}\right) = \mathbf{I}$.
4. *For all i and t, $E\left[\xi_{it}^{16}\right] < B^*$ i.e. the sixteenth moments of ξ_{it} are uniformly bounded.*

We set $\boldsymbol{\zeta}_t = \left[\mathrm{vec}(\mathbf{V}_t)', \boldsymbol{\eta}_t', \boldsymbol{\varepsilon}_t'\right]'$ so that all the individual disturbances in the model are possibly correlated linear processes.

2.1 Stochastically Trendless Processes

We make the following definition

Definition 1 *A vector stochastic process, $\mathbf{u}_t$, to 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 definition states that the (MSE) optimal s step ahead forecasts of a stochastically trendless process converge to the unconditional mean of the process as the forecast horizon s increases. That is, behaviour of the process up to time t has negligible effect on its behaviour into the infinite future.[4] The effect of shocks on the level of the process is transitory rather than permanent, although no statement about the higher order moments of the process is made. The presence of a stochastic trend induces long memory in $\mathbf{u}_t$ in the sense that information at time t, $\Im_t$, is useful improving the forecasts of the process at all horizons. Well-known examples of stochastically trendless processes are weakly stationary series such as those in Assumption LP (this includes the stationary $AR(1)$ case as used in the discussion below). The central example of a process with a stochastic trend is an $I(1)$ process such as $\mathbf{w}_t$.

The following result, proved in Section 9, shows that $(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{w}_t)$ is stochastically trendless, despite the presence of the $I(1)$ component $\mathbf{w}_t$.

Proposition 1 *Under Assumption LP, $\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{w}_t$ is stochastically trendless.*

[3] $\|A\| = \sqrt{\mathrm{tr}\,(A'A)}$

[4] Trendlessness is similar to the concept of a mixingale and the associated notion of asymptotic unpredictability. Analogous definitions have also been used in the literature on economic convergence; see Bernard and Durlauf (1996).

That is, the multiplicative combination of a zero-mean stochastically trendless process $\mathbf{V}_t$ and a stochastic trend $\mathbf{w}_t$ is stochastically trendless. This holds even if $\mathbf{V}_t$ is correlated with $\boldsymbol{\eta}_t$, the disturbances of $\mathbf{w}_t$. Therefore, even though the disturbances $\boldsymbol{\eta}_t$ have an infinitely persistent effect on $\mathbf{w}_{t+s}$, their effect on the level of $\mathbf{V}_{t+s}\mathbf{w}_{t+s}$ is transitory. This implies that under stochastic cointegration $\mathbf{c}'\mathbf{z}_t = \mathbf{c}'(\boldsymbol{\varepsilon}_t + \mathbf{V}_t\mathbf{w}_t)$ is stochastically trendless. It is this property that bestows meaning to the concept of co-movement of a heteroscedastic kind. It also follows directly that $\Delta\mathbf{z}_t$ in (3) is stochastically trendless.

The proof of the proposition is somewhat tedious under the generality of Assumption LP, but a simple example illustrates the result. Suppose v_t and w_t are univariate processes generated by

$$\begin{aligned} v_t &= \phi v_{t-1} + \xi_{v,t}, \\ w_t &= w_{t-1} + \eta_t \end{aligned}$$

where $|\phi| < 1$ and

$$\left[\begin{array}{c} \xi_{v,t} \\ \eta_t \end{array}\right] \sim \text{i.i.d.}\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma_v^2 & \sigma_{v\eta} \\ \sigma_{v\eta} & \sigma_\eta^2 \end{array}\right)\right].$$

This captures the essential features of our data generating process: v_t is autocorrelated but stationary, w_t is $I(1)$, and the disturbances driving the two processes are potentially correlated. It is useful to represent the processes as

$$\begin{aligned} v_{t+s} &= \phi^s v_t + \sum_{j=0}^{s-1} \phi^j \xi_{v,t+s-j}, \\ w_{t+s} &= w_t + \sum_{j=0}^{s-1} \eta_{t+s-j}. \end{aligned}$$

From these representations it is simple to deduce that v_t is stochastically trendless. Denoting the conditional expectation by E_t, it follows, since t is fixed, that $E_t(v_{t+s}) = \phi^s v_t \xrightarrow{p} 0$, as $s \to \infty$ and w_t has a stochastic trend since $E_t(w_{t+s}) = w_t$ does not disappear as $s \to \infty$. Of interest here is the multiplicative process $v_t w_t$, for which

$$E_t(v_{t+s}w_{t+s}) = \phi^s v_t w_t + \sigma_{v\eta} \sum_{j=0}^{s-1} \phi^j,$$

and

$$E(v_{t+s}w_{t+s}) = \phi^s E(v_t w_t) + \sigma_{v\eta} \sum_{j=0}^{s-1} \phi^j,$$

so, for fixed t,

$$E_t(v_{t+s}w_{t+s}) - E(v_{t+s}w_{t+s}) = \phi^s (v_t w_t - E(v_t w_t)) \xrightarrow{p} 0 \text{ as } s \to \infty.$$

This shows that the stochastically trendless property of v_t dominates the multiplicative process $v_t w_t$. Note that this is in contrast to the additive process $(v_t + w_t)$, which is dominated by the stochastic trend w_t.

3 Hypothesis Tests and Test Statistics

Our primary goal is to determine if the system is stochastically cointegrated. This null, and alternative of non-cointegration, may be stated as $H^0 : \mathbf{c}'\Pi = \mathbf{0}$ and $H^1 : \mathbf{c}'\Pi \neq \mathbf{0}$. Within stochastic cointegration, we may wish to know whether stationary or heteroscedastic cointegration pertains. The null of stationary cointegration against the heteroscedastic 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 $H_1^0 : \mathbf{c}'E(\mathbf{V}_t\mathbf{V}_t')\mathbf{c} > 0$.

It proves convenient to interpret these hypothesis 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

$$\begin{bmatrix} y_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{bmatrix} + \begin{bmatrix} \delta_y \\ \boldsymbol{\delta}_x \end{bmatrix} t + \begin{bmatrix} \boldsymbol{\pi}'_y \\ \Pi_x \end{bmatrix} \mathbf{w}_t + \begin{bmatrix} \varepsilon_{yt} \\ \boldsymbol{\varepsilon}_{xt} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\nu}'_{yt} \\ \mathbf{V}_{xt} \end{bmatrix} \mathbf{w}_t \tag{6}$$

where y_t, μ_y, δ_y and ε_{yt} are scalars, $\mathbf{x}_t$, $\boldsymbol{\mu}_x$, $\boldsymbol{\delta}_x$ and $\boldsymbol{\varepsilon}_{xt}$ are $(m-1)\times 1$ vectors, $\boldsymbol{\pi}'_y$ and $\boldsymbol{\nu}'_{yt}$ are $1\times m$ vectors while Π_x and $\mathbf{V}_{xt}$ are $(m-1)\times m$ matrices. Letting $\mathbf{c} = [1, -\boldsymbol{\beta}']'$, $\alpha = \mu_y - \boldsymbol{\beta}'\boldsymbol{\mu}_x$, $\kappa = \delta_y - \boldsymbol{\beta}'\boldsymbol{\delta}_x$, $e_t = \varepsilon_{yt} - \boldsymbol{\beta}'\boldsymbol{\varepsilon}_{xt} = \mathbf{c}'\boldsymbol{\varepsilon}_t$, $\mathbf{q}' = \boldsymbol{\pi}'_y - \boldsymbol{\beta}'\Pi_x = \mathbf{c}'\Pi$ and $\boldsymbol{\nu}'_t = \boldsymbol{\nu}'_{yt} - \boldsymbol{\beta}'\mathbf{V}_{xt} = \mathbf{c}'\mathbf{V}_t$, then we have

$$\begin{aligned} y_t &= \alpha + \kappa t + \mathbf{x}'_t\boldsymbol{\beta} + u_t, \\ u_t &= e_t + \mathbf{q}'\mathbf{w}_t + \boldsymbol{\nu}'_t\mathbf{w}_t. \end{aligned} \tag{7}$$

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 heteroscedastic component $\boldsymbol{\nu}'_t\mathbf{w}_t$. 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 is only one cointegrating vector so that $\text{rank}(\Pi_x) = m-1$. The rank condition ensures that further sub-relationships among the $\mathbf{x}_t$ variables in (7) are excluded.[5] The null hypothesis of stochastic cointegration against alternative of non-cointegration can now be expressed via (7) 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 heteroscedastic alternative is $H^0_0 : E(\boldsymbol{\nu}'_t\boldsymbol{\nu}_t) = 0$ and $H^0_1 : E(\boldsymbol{\nu}'_t\boldsymbol{\nu}_t) > 0$.

A difficulty in deriving test statistics for these hypotheses arises because constructing a likelihood for the vector $\{y_t; t = 1, ..., T\}$ in (7) is very complex. The processes involved are unobserved and it is difficult to know just what sorts of dynamics and distributions would be justifiable. Tractability is also an issue because of the multiplicative and additive way in which the constituent random variables appear. Even under the assumption of Gaussianity there is no known form available. Here, we adopt a semi-parametric approach that does not rely on a parametric model for dynamic behaviour or any distributional assumptions. At the outset there is a simplification we can make in that we can deal with the variables u_t, instead of y_t. This is because $\boldsymbol{\beta}$ (a nuisance parameter in this context) may be eliminated by use of regression residuals. Defining $\mathbf{X}_t = (1, t, \mathbf{x}'_t)'$ and $\mathbf{b} = \left(\alpha, \kappa, \boldsymbol{\beta}'\right)'$, we estimate the model by means of the estimator $\hat{\mathbf{b}}_k = \left(\hat{\alpha}_k, \hat{\kappa}_k, \hat{\boldsymbol{\beta}}'_k\right)'$ given by

$$\hat{\mathbf{b}}_k = \left(\sum_{t=k+1}^{T} \mathbf{X}_{t-k}\mathbf{X}'_t\right)^{-1} \sum_{t=k+1}^{T} \mathbf{X}_{t-k}y_t \tag{8}$$

where $k = k(T)$. This estimator, described in HMLa, is called an asymptotic IV estimator (AIV). Under H^0, $\hat{\boldsymbol{\beta}}_k$ is consistent as k and $T \to \infty$ in contrast with OLS which is not consistent under heteroscedastic cointegration unless $\mathbf{x}_t$ consists entirely of $I(1)$ processes. Operationally, then, this means that we will construct test statistics based on the AIV residuals

$$\widehat{u}_t = y_t - \hat{\alpha}_k - \hat{\kappa}_k t - \mathbf{x}'_t\hat{\boldsymbol{\beta}}_k. \tag{9}$$

3.1 Testing H^0 Against H^1

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

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

The null hypothesis is composite here, encompassing both stationary and heteroscedastic cointegration. A consequence of this is the preponderance of nuisance parameters in the distribution of the partial sum process of $\{u_t\}$. Nuisance parameters affect the choice of test statistic as it must have a distribution that is asymptotically free of them. Thus, a statistic is sought which eliminates nuisance parameters and is not

[5] A special case of this model is studied by Hansen (1992a). When $\mathbf{q} = \mathbf{0}$ and $\mathbf{V}_{xt} = \mathbf{0}$, (7) corresponds to a regression model when the regressors variables are all $I(1)$ and the error term is heteroscedastic, so that the regressand and regressors and are treated asymmetrically. Hansen presents a cogent combination of empirical and theoretical evidence which suggests that such models may prove useful in practical situations where conventional cointegrating regressions are too restrictive.

dependent upon specific assumptions made about the distributions of the unobserved variables. By way of motivation, consider

$$S_{nc} = \sum\nolimits_{t=2}^{T} u_t u_{t-1}.$$

In the situation where all the disturbance terms are i.i.d. S_{nc} 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 more general, S_{nc} needs to be modified to eliminate the nuisance parameters that result from the autocorrelation and from the presence of $\boldsymbol{\nu}_t'\mathbf{w}_t$. Thus, we suggest

$$S_{nc} = \sum\nolimits_{t=k+1}^{T} u_t u_{t-k}$$

where the lag k is *allowed to increase with* T. Under the cointegrating null, H^0, the statistic S_{nc} (when standardised 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 below. Because of the linear process representation, letting k become large eliminates any correlation between u_t and u_{t-k} under H^0, while the HAC variance estimator takes care of the term $\boldsymbol{\nu}_t'\mathbf{w}_t$. Under the alternative, because of the $I(1)$ term $\mathbf{q}'\mathbf{w}_t$, allowing k to grow does not eliminate correlation between u_t and u_{t-k}. This distinction is the source of consistency of the test.[6]

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 heteroscedastic alternative, we have

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

where under the null $V(\boldsymbol{\nu}_t) = \mathbf{0}$. To get some idea of what sort of test to use, we temporarily suppose that the unobserved variables are jointly Gaussian. Because of the simple nature of the null hypothesis, $u_t = e_t$, and the fact that we are going to construct a test local in $V(\boldsymbol{\nu}_t)$, we are able to approximate the distribution of the observables with a relatively straightforward parametric form. McCabe and Leybourne (2000) give a general method of approximating the likelihood and of constructing locally most powerful tests in such circumstances. It is shown in Section 9.3 that the statistic is based on

$$S_{hc} = \sum\nolimits_{t=1}^{T} t u_t^2.$$

Once in possession of the structure of the test, we may now abandon the fairly arbitrary assumptions used to derive it. The statistic is standardised and its asymptotic null distribution when calculated from regression residuals is shown, in Theorem 2 below, to be $N(0,1)$ under our weak regularity conditions.

It is also the case that the structure of S_{hc} can be used to test the null of $I(1)$ against the alternative of HI for any given individual series, by simply constructing it using $\Delta y_t - \delta_y$ (or any one $\Delta x_{it} - \delta_{xi}$, $i = 1, ...m-1$) in place of u_t. That is, we calculate $S_{hi} = \sum_{t=1}^{T} t\left(\Delta y_t - \delta_y\right)^2$.

4 Asymptotic Distribution of the Test Statistics

As noted above, to implement the two tests we replace u_t with $\widehat{u}_t$ defined in (9). The statistics also use HAC variance estimators (see Andrews (1991)) and the following notation is adopted. Define the lag covariances for any process $\{a_t\}$ by

$$\hat{\gamma}_j\left(a_t\right) = T^{-1} \sum_{s=j+1}^{T} a_s a_{s-j}$$

[6] Whilst it is tempting to consider using cointegrating versions of stationarity tests, e.g. Shin (1994), such procedures suffer from the fact that it is difficult to remove the effects of nuisance parameters under heteroscedastic cointegration and so the tests are unusable.

and define a HAC estimator of the long run variance by

$$\hat{\omega}^2(a_t) = \hat{\gamma}_0(a_t) + 2\sum_{j=1}^{l} \lambda\left(\frac{j}{l}\right)\hat{\gamma}_j(a_t) \tag{10}$$

where $\lambda(.)$ is a window with lag truncation parameter l. We assume that the following assumption holds.

Assumption KN *(Kernel and Lag Length)*

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

4.1 Asymptotic Distribution of S_{nc}

This section derives the asymptotic distribution of the statistic that tests for no cointegration. The distribution of $\widehat{S}_{nc}$ (S_{nc} calculated using the residuals $\widehat{u}_t$ and studentised) is required under the composite null of either stationary or heteroscedastic cointegration. Notwithstanding the fact that quantities like $\widehat{u}_t$ are of different orders of magnitude in these two situations, it is still the case that $\widehat{S}_{nc}$ is asymptotically normal under the composite null.

Theorem 1 *Assume the model (7), Assumption LP and Assumption KN hold. If $k = O(T^{1/2})$ then*
(i) under H^0,

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

(ii) under H^1, the distribution of $\left|\widehat{S}_{nc}\right|$ diverges as $T \to \infty$.
Here $\widehat{u}_t$ is defined in (9) using (8) with $k = O(T^{1/2})$; $\widehat{\omega}(.)$ and l are defined in (10).

The first part of this proposition states that a properly standardised statistic, $\widehat{S}_{nc}$, is asymptotically normal under stationary cointegration (which includes EG) and also under heteroscedastic cointegration; the second part shows that the test is consistent under H^1. The same results arise if linear trends are excluded from (7) and the fitted model.

4.2 Asymptotic Distribution of S_{hc}

In contrast to $\widehat{S}_{nc}$, whose limit distribution is invariant to the type of cointegration present, the statistic $\widehat{S}_{hc}$ is designed to distinguish between stationary and heteroscedastic cointegration in those cases where it is known that stochastic cointegration is present. It therefore exploits the difference between the orders of magnitude of $\widehat{u}_t$ that exists in the two types of cointegration. The null hypothesis of this model, H_0^0, is stationary cointegration.

Theorem 2 *Assume the model (7), Assumptions LP and KN hold, then*
(i) under H_0^0,

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

(ii) under H_1^0, the distribution of $\left|\widehat{S}_{hc}\right|$ diverges as $T \to \infty$.
Here $\hat{\sigma}_u^2 = T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2$; $\widehat{\omega}(.)$ and l are defined in (10).

As a corollary to this result, if $\widehat{S}_{hc}$ is calculated using $\Delta y_t - \overline{\Delta y}$ in place of $\widehat{u}_t$ then, denoting this statistic $\widehat{S}_{hi}$, $\widehat{S}_{hi} \xrightarrow{d} N(0,1)$ if y_t is $I(1)$ and $\left|\widehat{S}_{hi}\right|$ diverges if y_t is HI. The same results arise if linear trends are excluded from (7) and the fitted model.

4.3 Computational Details

To carry out a stochastic cointegration analysis in practice, we need to choose values for the various constants associated with the AIV estimator and the new tests. In what follows, for the AIV estimator we set $k = [T^{1/2}]$ ([.] denoting the integer part of). This same value of k is also used to construct the test $\widehat{S}_{nc}$. All three tests $\widehat{S}_{nc}$, $\widehat{S}_{hc}$ and $\widehat{S}_{hi}$ require the use of a kernel and a lag truncation parameter for their variance estimator. For each we choose the simple Bartlett kernel for $\lambda(.)$ and set the lag truncation parameter $l = [12(T/100)^{1/4}]$; this choice for l being fairly standard in the literature.

When we perform an *EG* analysis we use an efficient OLS estimator in which $[T^{1/4}]$ 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 test for integration and cointegration by adopting the null of stationarity. Here we apply the KPSS test (Kwiatkowski *et al.* (1992), denoted K_s) for the null of $I(0)$ against the alternative of $I(1)$ to the individual series and the residual test of Shin (1994) (denoted K_c) for the null of cointegration between $I(1)$ series against the alternative of no cointegration. The tests K_s and K_c also require the use of a kernel and a lag truncation parameter and for both we again use the Bartlett kernel with $l = 12(T/100)^{1/4}$.

5 Simulation Results

In this section we investigate, via Monte Carlo simulation, the finite sample behaviour of the tests $\widehat{S}_{nc}$ and $\widehat{S}_{hc}$, comparing these with the tests for the conventional paradigm. We considered the model (6) with $m = 2$. Specifically, our DGP was

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & d_1 \end{bmatrix} \mathbf{w}_t + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix} + \begin{bmatrix} \nu_{yt} & 0 \\ \nu_{xt} & 0 \end{bmatrix} \mathbf{w}_t \tag{11}$$

In regression format (11) can be written

$$\begin{aligned} y_t &= \beta x_t + u_t, \\ u_t &= e_t + q w_{2t} + v_t w_{1t} \end{aligned}$$

where

$$\begin{aligned} \beta &= 1, \quad e_t = \varepsilon_{yt} - \varepsilon_{xt}, \\ q &= -d_1, \quad v_t = \nu_{yt} - \nu_{xt}. \end{aligned}$$

The stochastic processes of (11) were generated according to

$$\begin{aligned} \varepsilon_{yt} &= 0.5\varepsilon_{yt-1} + \epsilon_{1t}, \quad \varepsilon_{xt} = -0.5\varepsilon_{xt-1} + \epsilon_{2t}, \\ \nu_{yt} &= -0.8\nu_{yt-1} + d_2\sqrt{0.10}\epsilon_{3t}, \quad \nu_{xt} = 0.8\nu_{xt-1} + d_3\sqrt{0.05}\epsilon_{4t}, \\ \Delta w_{1t} &= \epsilon_{5t}, \quad \Delta w_{2t} = \epsilon_{6t} \end{aligned}$$

with $(\epsilon_{1t}, \epsilon_{2t}, \epsilon_{3t}, \epsilon_{4t}, \epsilon_{5t}, \epsilon_{6t})'$ a multivariate standard normal white noise process. To induce some contemporaneous correlation into the system under heteroscedastic cointegration we imposed $\text{cor}(\epsilon_{2t}, \epsilon_{4t}) = \text{cor}(\epsilon_{5t}, \epsilon_{4t}) = 0.5$.

Here the d_i, $i = 1, 2, 3$ are scalar 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, whilst if $d_1 \neq 0$, H^1 is true and y_t and x_t are not cointegrated in any sense (whatever 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 heteroscedastic 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 $d_3 \neq 0$).

Our simulations considered sample sizes of $T = 200, 400, 600$ and the number of replications for all experiments was $10,000$. The entries in Tables 1 represent empirical rejection frequencies of the various tests, based on regressions allowing constants but not trends, at the nominal asymptotic 0.05-level distribution (two tailed tests in the case of $\widehat{S}_{nc}$, $\widehat{S}_{hc}$ and $\widehat{S}_{hi}$). We denote the efficient OLS estimator by $\hat{\beta}_0$ and the entries in the table for this and $\hat{\beta}_k$ are the empirical means. In Table 1 (a), we have $d_1 = d_2 = d_3 = 0$, so that H_0^0 is true - in this situation there is stationary cointegration between two $I(1)$ series. The KPSS test, K_s, clearly indicates rejection of $I(0)$ in favour of $I(1)$ for both series and the $\widehat{S}_{hi}$ has near nominal size, indicating $I(1)$ rather than HI for both series. The OLS and AIV estimators both have means very close to the actual value of β $(= 1)$. The KPSS cointegration test, K_c, together with $\widehat{S}_{nc}$ and $\widehat{S}_{hc}$, all have sizes close to nominal. Thus, a conventional cointegration analysis and a stochastic one are in agreement here in that stationary cointegration between y_t and x_t holds. In Table 1 (b), we have $d_1 = 1$, $d_2 = d_3 = 0$, so that H^1 is true - here that cointegration between two $I(1)$ series fails to hold. The behaviour of K_s and $\widehat{S}_{hi}$ is the very similar to that in Table 1 (a) since the properties of the individual series y_t and x_t have not changed. Now, however, neither $\hat{\beta}_0$ nor $\hat{\beta}_k$ are consistent estimators of β. The test $\widehat{S}_{nc}$ strongly rejects the null of cointegration and consistency of this test is clearly evident. Interestingly, outside of the smallest sample size considered, the empirical power of $\widehat{S}_{nc}$ exceeds that of K_c. The test $\widehat{S}_{hc}$ also rejects its null, increasingly so with increasing sample size, but has rather less "power" than $\widehat{S}_{nc}$ (in practical terms, therefore, under the current DGP we would not necessarily expect observed rejections by $\widehat{S}_{nc}$ to coincide with rejections by $\widehat{S}_{hc}$). Thus, the results of Tables 1 (a) and (b) lead us to conclude that our new test $\widehat{S}_{nc}$ has considerable ability to detect the presence, or absence, of cointegration within the conventional framework, performing at least comparably to a test designed for this scenario.

We now proceed, in Table 1 (c), to examine the behaviour of the procedures under the weaker paradigm of heteroscedastic cointegration. We set $d_1 = 0$, $d_2 = d_3 = 1$ representing H_1^0, in this case heteroscedastic cointegration between two HI series. The KPSS test, K_s, continues to reject that the series are $I(0)$; as we would expect since they both still contain an $I(1)$ influence. In this respect, then, K_s yields the appropriate inference. However, the OLS estimator of β is now inconsistent and the KPSS cointegration test, K_c, strongly rejects conventional cointegration in favour of non-cointegration. That is, on the basis of this test we would be led to spuriously conclude that the series "differ" by an $I(1)$ component. This clearly emphasizes the point that conventional procedures for testing for cointegration are not just inappropriate but also highly misleading if applied in this weakened context. At a practical level, since the $\widehat{S}_{hi}$ test now indicates the presence of HI rather than $I(1)$ for both series, this could be taken as a prior warning *not* to proceed with the conventional procedures. In sharp contrast to the OLS estimator, it is clear that the AIV estimator remains consistent under heteroscedastic cointegration. The test $\widehat{S}_{nc}$ has reasonably close to nominal size in this case (it is a little under sized in the smaller samples), whilst $\widehat{S}_{hc}$ consistently indicates the presence of heteroscedastic cointegration. Both these outcomes are in line with our theoretical results.

Finally, in Tables 1 (d) and (e) we look at the special cases of "unbalanced" heteroscedastic cointegration. In Table 1 (d), y_t is $I(1)$ and x_t is HI. In Table 1 (e) this situation is reversed. The pattern in Table 1 (d) is very similar to that in Table 1 (c) for both the conventional and new tests. That is, conventional procedures indicate no cointegration whilst the new tests are well able to detect the kind of cointegration present. In Table 1 (e), whilst the results are comparable to Table 1 (d) as regards the new tests, the behaviour of the KPSS cointegration test is markedly different and now indicates the presence of cointegration. The difference is, of course, that OLS estimation can be shown to be consistent under heteroscedastic cointegration only when x_t is $I(1)$. In the unbalanced case the ordering of the variables therefore becomes of crucial importance for OLS-based conventional procedures. On the other hand, the results from our new tests are virtually unaffected by this ordering.

In summary then, the above analysis would seem to fully validate our new test procedures. If the conventional cointegration paradigm holds they compare favourably to standard tests. Under our more general notion of stochastic cointegration they have considerable ability to detect the kind of cointegration present, when, in the same situations, the standard tests can produce badly misleading inference. It is also very important to remember that when applying our new tests in practice, we never actually need to distinguish between $I(1)$ and HI series. That is, we would never need to calculate the test $\widehat{S}_{hi}$ for individual series. The only rationale for calculating $\widehat{S}_{hi}$ is that it provide early warning of situations where standard cointegration tests are likely to be unreliable.

6 Testing for Rational Expectations PVMs

6.1 The Bond Market

A necessary empirical condition for the PVM in the bond market (the expectations theory of the term structure) is that long run and short run interest rates cointegrate with a coefficient of unity. A conventional EG analysis, assuming the variables to be $I(1)$, proceeds by regressing the long run rate (L_t) on a constant and a short run rate (S_t) using (efficient) OLS to estimate a model of the form

$$L_t = \alpha + \beta S_t + u_t. \tag{12}$$

The residuals from the fitted model are then used to test whether u_t is $I(0)$. We will not impose the restriction that $\beta = 1$, since this forms part of the PVM being tested. In our generalized context, the yields are allowed to be stochastically integrated and the cointegration stochastic, encompassing the EG formulation. We carry out an EG cointegration exercise along side a stochastic one to assess the empirical evidence for or against the PVM. To expedite comparison with our simulation results, we calculate exactly the same array of statistics as in Section 5, where the regressions include constants but not trends.

Monthly data from the United States, Canada, United Kingdom and Japan are used. These are taken from the OECD/MEI database. A single long run interest rate and a variety of short run rates are used for each country - see Section 8 for details of the data. The results are given in Table 2, where bold print indicates a rejection of the corresponding null hypothesis at the 0.05 significance level.

With regard to a conventional cointegration analysis we first note that the KPSS test, K_s, indicates rejection of $I(0)$ in favour of $I(1)$, at the 0.05 level for every one of the 17 individual interest rate series considered. However, according to the KPSS cointegration test, K_c, conventional cointegration is rejected in favour of non-cointegration for *every* one of the 13 pairs of long and short run rates. Thus, a conventional analysis fails to find any support for the PVM in the bond market. Conversely, when we examine the outcomes from the new analysis, we find that, according to the $\widehat{S}_{nc}$ test, stochastic cointegration is not rejected for 8 of the 13 pairwise regressions. In both Canada and the United Kingdom, the non-rejection is unambiguous (evidence in favour of PVM is found). In the case of the United States the evidence is mixed; rejections are found for 2 of the 4 pairs considered. No evidence of cointegration at all is found for Japan, though the peculiar nature of Japanese short run interest rates in recent times (being effectively zero) may explain this finding. Of the 8 pairwise regressions that do not reject stochastic cointegration, according to the $\widehat{S}_{hc}$ test 5 represent stationary cointegration and 3 heteroscedastic. It is also informative to compare the estimates of β from the two analyses; when cointegration is indicated the AIV estimates are typically much closer to the theoretical value of unity than is OLS. Hence, in contrast to the conventional approach, the new analysis uncovers rather more evidence in favour of the bond market PVM - at least as indicated by the presence of the cointegration implications.

A plausible explanation as to why the EG and new analyses often yield such different conclusions arises from examining the $\hat{S}_{hi}$ test. This shows that all of the interest rate series appears to be HI rather than $I(1)$.[7] Recalling our simulation results in Section 5, if two HI series are cointegrated then conventional tests tend to indicate non-cointegration and it is therefore entirely possible that we are witnessing the empirical counterpart of this behaviour. We conjecture that it is the failure of EG analyses to account properly for the prevalence of HI series in the bond market that is an important factor underpinning their difficulties in finding empirical support for the PVM.

6.2 The Stock Market

Let D_t denote the natural logarithm of the dividend payment and P_t the logarithm of stock prices. According to Campbell and Shiller (1988) the stock market PVM implies

$$D_t = \alpha + \beta P_t + u_t \tag{13}$$

where if D_t and P_t are $I(1)$, then $\beta = 1$ and u_t is $I(0)$. We perform conventional and stochastic cointegration analyses of this model using United States monthly data on S&P composite dividends and stock prices for the period 1871:1 to 2001:12. These are taken from Robert Shiller's website (http://aida.econ.yale.edu/~shiller). The data are graphed in Figure 5. Since there is some possibility of deterministic trending behaviour in the data here, we perform two sets of analyses; one using (13) and one which incorporates an additional linear trend in (13). Moreover, since it is also evident from visual inspection of the data that the behaviour of (and relationship between) dividends and stock prices may well have altered after World War II, we also perform subperiod analyses based on the two periods 1871:1-1946:12 and 1947.1-2001:12. The results

[7] This finding adds to the growing body of evidence that many economic and financial time series previously considered to be $I(1)$ are more appropriately modelled as HI (or the closely related stochastic unit root) processes. See, *inter alia,* Hansen (1992a), Leybourne McCabe and Tremayne (1996), Granger and Swanson (1997), Wu and Chen (1997) and Psaradakis *et al* (2001).

are given in Table 3. For the full sample period the K_s test indicates rejection of $I(0)$ in favour of $I(1)$ (possibly with drift) for either series. Examining K_c and $\widehat{S}_{nc}$, we find that both the conventional and stochastic cointegration analyses emphatically reject cointegration between dividends and prices, irrespective of whether linear trends are incorporated or not. Note that it is difficult, however, to have a great deal of confidence in the conventional analysis since the series are found to be HI rather than $I(1)$ according to the $\hat{S}_{hi}$ test. As regards the subperiod analyses, for the first subperiod the conventional analysis continues to reject cointegration, with or without linear trends, whereas the stochastic approach finds stationary cointegration when a constant alone is included, and heteroscedastic cointegration when a linear trend is included. Again, both series appear to be HI so we conjecture once more that this is the reason why the conventional approach to modelling cointegration is unable to yield any empirical support for the PVM. For the second subperiod neither approach finds any evidence of cointegration. This is, however, quite plausible since any relationship has probably been weakened in recent times by the behaviour of the price series and the tendency of firms to retain profits and pay less in dividends.

7 References

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, 59, 817–58.

Balke, N.S. and M.E. Wohar (2001) Explaining stock price movements: is there a case for fundamentals? *Federal Reserve Bank of Dallas Review*, Third Quarter, 22 - 34.

Balke, N.S. and M.E. Wohar (2002), Low frequency movements in stock prices: a state-space decomposition, *Review of Economics and Statistics*, (forthcoming).

Campbell, J.Y. and R.J. Shiller (1987) Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062-88.

Campbell, J.Y. and R.J. Shiller (1988) The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies*, 1, 195-227.

Diba, B.T. and Grossman, H.I. (1988) Explosive rational bubbles in stock prices, *American Economic Review*, 78, 520-30.

Bernard, A.B. and S.N. Durlauf (1996) Interpreting tests of the convergence hypothesis, *Journal of Econometrics*, 71, 161–73.

Engle, R.F. and C.W.J. Granger (1987) Cointegration and error correction: representation, estimation and testing, *Econometrica*, 55, 251–76.

Froot, K and M. Obstfeld (1991) Intrinsic bubbles: The case of stock prices, *American Economic Review*, 81, 1189-1214.

Granger, C.W.J and 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, 139-58.

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

Harris, D., B.P.M. McCabe and S.J. Leybourne (2002) Stochastic cointegration: estimation and inference, *Journal of Econometrics,* 111, 2, 363-84.

Harris, D., B.P.M. McCabe and S.J. Leybourne (2003) Some limit theory for infinite order autocovariances, *Econometric Theory, 19, 829-64.*

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and 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, 159-78.

Lamont, O. (1998) Earnings and expected returns, *Journal of Finance*, 53, 1563-1587.

Leybourne, S.J., B.P.M. McCabe and A.R. Tremayne (1996) Can economic time series be differenced to stationarity?, *Journal of Business and Economic Statistics*, 14, 435–46.

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

M. Ogaki and J.Y. Park (1998) A cointegration approach to estimating preference parameters, *Journal of Econometrics,* 82, 107-134.

Psaradakis, Z, M. Sola and F. Spagnolo (2001) A simple procedure for detecting periodically collapsing rational bubbles, *Economics Letters*, 72, 317-23.

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

Table 1. Size and Power of the Tests.

(a) $d_1 = 0, d_2 = 0, d_3 = 0$

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				$T = 200$			
y_t	0.723			0.044			
x_t	0.726	0.999	0.055	0.050	1.005	0.041	0.049
				$T = 400$			
y_t	0.866			0.050			
x_t	0.867	1.000	0.056	0.051	0.999	0.047	0.053
				$T = 600$			
y_t	0.923			0.048			
x_t	0.923	1.000	0.056	0.053	1.000	0.047	0.052

(b) $d_1 = 1, d_2 = 0, d_3 = 0$

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				$T = 200$			
y_t	0.723			0.044			
x_t	0.716	0.433	0.521	0.048	0.430	0.388	0.217
				$T = 400$			
y_t	0.866			0.050			
x_t	0.861	0.434	0.756	0.049	0.434	0.802	0.321
				$T = 600$			
y_t	0.923			0.048			
x_t	0.919	0.434	0.862	0.050	0.428	0.941	0.366

(c) $d_1 = 0, d_2 = 1, d_3 = 1$

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				$T = 200$			
y_t	0.718			0.377			
x_t	0.676	0.843	0.303	0.354	1.091	0.022	0.439
				$T = 400$			
y_t	0.862			0.520			
x_t	0.831	0.854	0.453	0.529	1.048	0.036	0.567
				$T = 600$			
y_t	0.918			0.583			
x_t	0.902	0.850	0.570	0.608	1.031	0.043	0.626

(d) $d_1 = 0, d_2 = 0, d_3 = 1$

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				$T = 200$			
y_t	0.720			0.044			
x_t	0.676	0.843	0.317	0.354	1.080	0.025	0.352
				$T = 400$			
y_t	0.864			0.048			
x_t	0.831	0.854	0.471	0.533	1.047	0.039	0.508
				$T = 600$			
y_t	0.918			0.049			
x_t	0.896	0.850	0.585	0.608	1.033	0.043	0.585

(e) $d_1 = 0, d_2 = 1, d_3 = 0$

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				$T = 200$			
y_t	0.721			0.381			
x_t	0.726	0.999	0.055	0.050	0.998	0.026	0.404
				$T = 400$			
y_t	0.864			0.528			
x_t	0.867	1.000	0.045	0.049	1.000	0.036	0.541
				$T = 600$			
y_t	0.921			0.588			
x_t	0.923	1.000	0.042	0.053	0.999	0.045	0.602

Table 2. Application to Bond Market Data.

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
United States							
L_t	**1.465**			**2.732**			
S_{1t}	**1.327**	0.560	**0.403**	**2.075**	0.731	1.794	**2.162**
S_{2t}	**1.072**	0.595	**0.625**	**2.022**	0.834	**2.002**	**1.989**
S_{3t}	**1.333**	0.581	**0.401**	**2.195**	0.756	1.920	**2.140**
S_{4t}	**1.368**	0.537	**0.391**	**2.212**	0.686	**2.003**	**2.289**
Canada							
L_t	**1.508**			**3.112**			
S_{1t}	**1.266**	0.675	**0.364**	**2.380**	0.805	1.300	0.648
S_{2t}	**1.279**	0.657	**0.359**	**2.910**	0.790	1.323	1.147
S_{3t}	**1.141**	0.687	**0.508**	**2.954**	0.877	1.532	1.316
United Kingdom							
L_t	**1.741**			**3.289**			
S_{1t}	**1.280**	0.743	**0.501**	**3.337**	0.934	0.989	1.726
S_{2t}	**1.231**	0.728	**0.547**	**3.405**	0.931	0.609	**2.166**
S_{3t}	**1.307**	0.738	**0.471**	**3.694**	0.908	1.000	1.440
Japan							
L_t	**1.193**			**2.176**			
S_{1t}	**0.994**	0.940	**0.369**	**2.878**	0.992	**2.314**	0.778
S_{2t}	**1.049**	0.668	**0.335**	**2.327**	0.714	**2.206**	0.826
S_{3t}	**1.047**	0.665	**0.344**	**2.765**	0.732	**2.291**	1.095

Table 3. Application to Stock Market Data.

	K_s	$\hat{\beta}_0$	K_c	$\hat{S}_{hi}$	$\hat{\beta}_k$	$\hat{S}_{nc}$	$\hat{S}_{hc}$
				constant			
1871:1-2001:12							
D_t	**6.087**			1.327			
P_t	**5.876**	0.850	**0.533**	0.090	0.848	**3.635**	1.628
1871:1-1944:12							
D_t	**3.363**			**1.702**			
P_t	**3.250**	0.932	**0.627**	**2.401**	1.176	1.427	0.576
1945:1-2001:12							
D_t	**3.413**			**2.670**			
P_t	**3.174**	0.756	**0.570**	0.147	0.739	**3.885**	1.248
				constant and trend			
1871:1-2001:12							
D_t	**1.206**			1.539			
P_t	**1.237**	0.659	**0.218**	0.035	0.672	**3.453**	0.385
1871:1-1944:12							
D_t	**0.335**			1.648			
P_t	**0.140**	0.672	**0.429**	**2.392**	2.403	0.758	**2.339**
1945:1-2001:12							
D_t	**0.190**			**2.817**			
P_t	**0.418**	0.054	**0.154**	0.025	-.052	**2.811**	**2.607**

8 Bond Market Data

The data used in Table 2 is 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} = US 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.

9 Proofs

9.1 Notation and Conventions

For the model specified by equations (1), (6) and (7), define the $\begin{pmatrix} m^2 \times 1 & m \times 1 & m \times 1 \end{pmatrix}'$ vector

$$\boldsymbol{\zeta}_t = \begin{pmatrix} \operatorname{vec}(\mathbf{V}_t) \\ \boldsymbol{\eta}_t \\ \boldsymbol{\varepsilon}_t \end{pmatrix},$$

and let $\boldsymbol{\zeta}_t$ satisfy Assumption LP with coefficients $\mathbf{C}_j$ and disturbances $\boldsymbol{\xi}_t$. 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 $\Omega_{11} = \mathbf{CC}'$ and $\Omega_{22} = \sum_{j=-\infty}^{\infty} \mathbf{G}_j \mathbf{G}_j'$. Also define $\mathbf{S}_t$ to be the partial sum of the $\boldsymbol{\zeta}_t$ i.e. $\Delta \mathbf{S}_t = \boldsymbol{\zeta}_t$. Selector matrices $\mathbf{R}_\eta$, $\mathbf{R}_\varepsilon$ and $\mathbf{R}_\nu$ can be defined such that $\boldsymbol{\eta}_t = \mathbf{R}_\eta' \boldsymbol{\zeta}_t$, $\boldsymbol{\varepsilon}_t = \mathbf{R}_\varepsilon' \boldsymbol{\zeta}_t$ and $\boldsymbol{\nu}_t = \mathbf{R}_\nu' \boldsymbol{\zeta}_t$. The forms of $\mathbf{R}_\eta$ and $\mathbf{R}_\varepsilon$ are obvious and, since $\boldsymbol{\nu}_t = \mathbf{V}_t' \mathbf{c} = (\mathbf{I}_m \otimes \mathbf{c}') \operatorname{vec}(\mathbf{V}_t)$,

$$\mathbf{R}_\nu = \begin{pmatrix} (\mathbf{I}_m \otimes \mathbf{c}) \\ \mathbf{0}_m \\ \mathbf{0}_m \end{pmatrix}.$$

In manipulating expressions involving kernels we adopt the notation $\lambda^+(j/l) = 2\lambda(j/l)$, $j > 0$, $\lambda^+(0) = 1$. Constants whose precise values have no significance, are denoted by the generic term *const*. Often these arise by the use of Assumption LP.4. When taking expectations through an infinite summation sign, we generally do not remark on the operation when obviously square summable linear processes are involved and we use elementary properties of norms (for sums and products) extensively without comment.

9.2 Proof of Proposition 1

Writing $\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \mathbf{V}_t \mathbf{w}_t$, showing $\mathbf{u}_t$ is stochastically trendless requires consideration of

$$\left\| E\left(\mathbf{u}_{t+s} \middle| \Im_t \right) - E\left(\mathbf{u}_{t+s} \right) \right\| \leq \left\| E\left(\boldsymbol{\varepsilon}_{t+s} \middle| \Im_t \right) \right\| + \left\| E\left(\mathbf{V}_{t+s} \mathbf{w}_{t+s} \middle| \Im_t \right) - E\left(\mathbf{V}_{t+s} \mathbf{w}_{t+s} \right) \right\|. \tag{14}$$

Then

$$\begin{aligned} E\left\| E\left(\boldsymbol{\varepsilon}_{t+s} \middle| \Im_t \right) \right\| &\leq \left\| \mathbf{R}_\varepsilon \right\| E \left\| E\left(\boldsymbol{\zeta}_{t+s} \middle| \Im_t \right) \right\| \\ &= \left\| \mathbf{R}_\varepsilon \right\| E \left\| \sum_{j=0}^{\infty} \mathbf{C}_{j+s} \boldsymbol{\xi}_{t-j} \right\| \\ &\leq \left\| \mathbf{R}_\varepsilon \right\| E \left(\sum_{j=0}^{\infty} \left\| \mathbf{C}_{j+s} \boldsymbol{\xi}_{t-j} \right\| \right) \\ &\leq const. \left\| \mathbf{R}_\varepsilon \right\| \sum_{j=s}^{\infty} \left\| \mathbf{C}_j \right\| \\ &= o\left(s^{-1/2} \right) \end{aligned}$$

using Assumption LP and the monotone convergence theorem. The order follows from Assumption LP by noting that LP.1 implies $\sum_{j=0}^{\infty} j^{1/2} \left\| \mathbf{C}_j \right\| < \infty$ and hence $\sum_{j=s}^{\infty} j^{1/2} \left\| \mathbf{C}_j \right\| = o(1)$ as $s \to \infty$, so

$$\sum_{j=s}^{\infty} \left\| \mathbf{C}_j \right\| = \sum_{j=s}^{\infty} j^{-1/2} j^{1/2} \left\| \mathbf{C}_j \right\| \leq s^{-1/2} \sum_{j=s}^{\infty} j^{1/2} \left\| \mathbf{C}_j \right\| = o\left(s^{-1/2} \right).$$

This shows the first term of (14) converges to zero in probability.

To deal with the second term of (14), write $\mathbf{V}_t$ in terms of its rows as

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{V}_{1,t}' \\ \vdots \\ \mathbf{V}_{m,t}' \end{pmatrix}.$$

The selector matrices $\mathbf{R}_{\mathbf{V},j}$, $j = 1, \ldots, m$, can then be defined so that $\mathbf{R}'_{\mathbf{V},j}\boldsymbol{\zeta}_t = \mathbf{V}_{j,t}$, along with $\mathbf{R}_{\eta}$ such that $\mathbf{R}'_{\eta}\boldsymbol{\zeta}_t = \boldsymbol{\eta}_t$. Thus the j^{th} element of the $m \times 1$ vector $\mathbf{V}_t\mathbf{w}_t$ can be written as

$$\mathbf{V}'_{j,t}\mathbf{w}_t = \boldsymbol{\zeta}'_t\mathbf{R}_{\mathbf{V},j}\mathbf{R}'_{\boldsymbol{\eta}}\mathbf{S}_t = \operatorname{tr}\mathbf{R}_{\mathbf{V},j}\mathbf{R}'_{\boldsymbol{\eta}}\mathbf{S}_t\boldsymbol{\zeta}'_t$$

where $\Delta\mathbf{S}_t = \boldsymbol{\zeta}_t$. Now

$$\begin{aligned} & \left|E\left(\mathbf{V}'_{j,t+s}\mathbf{w}_{t+s}\right|\Im_t\right) - E\left(\mathbf{V}'_{j,t+s}\mathbf{w}_{t+s}\right)\right| \\ = & \left|\operatorname{tr}\mathbf{R}_{\mathbf{V},j}\mathbf{R}'_{\boldsymbol{\eta}}\left(E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right|\Im_t\right) - E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right)\right)\right| \\ \leq & \left\|\mathbf{R}_{\mathbf{V},j}\right\|\left\|\mathbf{R}_{\boldsymbol{\eta}}\right\|\left\|E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right|\Im_t\right) - E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right)\right\|, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\zeta}_{t+s} &= \sum_{r=-s}^{-1}\mathbf{C}_{r+s}\boldsymbol{\xi}_{t-r} + \sum_{r=0}^{\infty}\mathbf{C}_{r+s}\boldsymbol{\xi}_{t-r}, \\ \mathbf{S}_{t+s} &= \sum_{j=-t+1}^{s}\sum_{i=-j}^{\infty}\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i} = \sum_{j=-t+1}^{s}\sum_{i=-j}^{-1}\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i} + \sum_{j=-t+1}^{s}\sum_{i=0}^{\infty}\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i} \end{aligned}$$

Thus

$$\begin{aligned} E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right) &= \sum_{j=-t+1}^{s}\sum_{r=-s}^{-1}\mathbf{C}_{r+j}\mathbf{C}'_{r+s} + \sum_{r=0}^{\infty}\sum_{j=-t+1}^{s}\mathbf{C}_{r+j}\mathbf{C}'_{r+s}, \\ E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right|\Im_t\right) &= \sum_{j=-t+1}^{s}\sum_{r=-s}^{-1}\mathbf{C}_{r+j}\mathbf{C}'_{r+s} + \sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=-t+1}^{s}\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i}\mathbf{C}'_{r+s}\boldsymbol{\xi}'_{t-r}, \end{aligned}$$

and so using the monotone convergence theorem again,

$$\begin{aligned} & E\left\|E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right|\Im_t\right) - E\left(\mathbf{S}_{t+s}\boldsymbol{\zeta}'_{t+s}\right)\right\| \\ = & E\left\|\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=-t+1}^{s}\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i}\mathbf{C}'_{r+s}\boldsymbol{\xi}'_{t-r} - \sum_{r=0}^{\infty}\sum_{j=-t+1}^{s}\mathbf{C}_{r+j}\mathbf{C}'_{r+s}\right\| \\ \leq & E\left(\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=-t+1}^{s}\left\|\mathbf{C}_{i+j}\boldsymbol{\xi}_{t-i}\boldsymbol{\xi}'_{t-r}\mathbf{C}'_{r+s}\right\|\right) + \sum_{r=0}^{\infty}\sum_{j=-t+1}^{s}\left\|\mathbf{C}_{r+j}\mathbf{C}'_{r+s}\right\| \\ \leq & \, const.\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=-t+1}^{s}\left\|\mathbf{C}_{i+j}\right\|\left\|\mathbf{C}_{r+s}\right\| + \sum_{r=0}^{\infty}\sum_{j=-t+1}^{s}\left\|\mathbf{C}_{r+j}\right\|\left\|\mathbf{C}_{r+s}\right\| \\ \leq & \, const.\sum_{j=1}^{s+t}\sum_{i=j}^{\infty}\left\|\mathbf{C}_i\right\|\sum_{r=s}^{\infty}\left\|\mathbf{C}_r\right\| + \sum_{r=s}^{\infty}\left\|\mathbf{C}_r\right\|\sum_{j=0}^{\infty}\left\|\mathbf{C}_j\right\| \\ = & \, o\left(s^{1/2}\right)\cdot o\left(s^{-1/2}\right) + o\left(s^{-1/2}\right) \\ = & \, o(1). \end{aligned}$$

This shows that $E\left(\mathbf{V}'_{j,t+s}\mathbf{w}_{t+s}\right|\Im_t\right) - E\left(\mathbf{V}'_{j,t+s}\mathbf{w}_{t+s}\right) \xrightarrow{p} 0$ for $j = 1, \ldots, m$, and hence that

$$\left\|E\left(\mathbf{V}_{t+s}\mathbf{w}_{t+s}\right|\Im_t\right) - E\left(\mathbf{V}_{t+s}\mathbf{w}_{t+s}\right)\right\| \xrightarrow{p} 0$$

in (14).

9.3 Derivation of S_{hc}

We assume e_t, $\boldsymbol{\nu}_t$ and $\boldsymbol{\eta}_t$ are all jointly Gaussian and uncorrelated with each other. We also assume that e_t is white noise, $\boldsymbol{\nu}_t$ has a diagonal covariance matrix with elements $\lambda_i = E\left(v_{it}^2\right), i = 1, ..., k$ and there are no restrictions on Σ_η, the covariance matrix of $\boldsymbol{\eta}_t$. These assumptions are to ensure that the likelihood is relatively tractable and could be relaxed at the expense of a more complicated derivation. There is no compelling reason to do so, however, as we effectively treat the approximation to the joint density of the u_t as a pseudo-likelihood and the validity of the asymptotic distribution of the test is not dependent on any of the simplifying assumptions used in deriving the statistic itself. Define $\mathbf{V} = [\boldsymbol{\nu}_1, ..., \boldsymbol{\nu}_T]$, $\boldsymbol{\nu} = \mathrm{vec}(\mathbf{V})$ and $\Sigma_{\boldsymbol{\nu}}(\boldsymbol{\lambda}) = E(\boldsymbol{\nu}\boldsymbol{\nu}') = \mathbf{I}_T \otimes E(\boldsymbol{\nu}_t\boldsymbol{\nu}_t') = \mathbf{I}_T \otimes \mathrm{diag}(\lambda_i)$. Let $\ell(\mathbf{u}|\boldsymbol{\nu})$ be the log-likelihood of $\mathbf{u}$ given $\boldsymbol{\nu}$. According to McCabe and Leybourne (2000), a locally most powerful test of H_0^0 against H_1^0 is given by

$$S_{hc} = \mathrm{tr}\left(\mathbf{M}|_{\{\mathbf{v}=\mathbf{0}\}} \cdot \sum\nolimits_{i=1}^{k} \left.\frac{\partial \Sigma_\nu}{\partial \lambda_i}\right|_{\{\boldsymbol{\lambda}=\mathbf{0}\}} \right),$$

where $\mathbf{M} = \frac{\partial \ell}{\partial \boldsymbol{\nu}} \cdot \frac{\partial \ell}{\partial \boldsymbol{\nu}'} + \frac{\partial^2 \ell}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}'}$, $\ell(\mathbf{u}|\boldsymbol{\nu}) = -(1/2)\mathbf{u}'\mathbf{G}(\boldsymbol{\nu})^{-1}\mathbf{u} + const$ and $\mathbf{G}(\boldsymbol{\nu}) = \mathbf{V}'\Sigma_\eta \mathbf{V} \odot \mathbf{N} + \sigma_e^2 \mathbf{I}_T$. The constant depends on $\boldsymbol{\nu}$ but is not random so that, effectively, it may be ignored it in the construction of the test. The matrix $\mathbf{N}$ is the usual $\min(i,j)$ matrix and $\odot$ denotes the Hadamard product operator.

It is simple to verify that $\sum_{i=1}^{k} \partial \Sigma_\nu / \partial \boldsymbol{\lambda}_i|_{\{\boldsymbol{\lambda}=\mathbf{0}\}} = \mathbf{I}_{Tk}$. So,

$$\begin{aligned} S_{hc} &= \mathrm{tr}(\mathbf{M})|_{\{\boldsymbol{\nu}=\mathbf{0}\}} \\ &= \sum_{t=1}^{T} \mathrm{tr}\left.\left(\frac{\partial \ell}{\partial \boldsymbol{\nu}_t} \cdot \frac{\partial \ell}{\partial \boldsymbol{\nu}_t'} \right)\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}} + \sum_{t=1}^{T} \mathrm{tr} \left.\frac{\partial^2 \ell}{\partial \boldsymbol{\nu}_t \partial \boldsymbol{\nu}_t'}\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}}. \end{aligned}$$

Now,

$$\frac{\partial \ell}{\partial v_{it}} = -(1/2)\mathbf{u}'\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial v_{it}}\mathbf{G}^{-1}\mathbf{u}$$

for $i = 1, ..., k$ and since $\mathbf{G}|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = \sigma_e^2 \mathbf{I}_T$ and $\partial \mathbf{G}/\partial v_{it}|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = \mathbf{0}$, it follows that $\partial \ell / \partial v_{it}|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = 0$ and so S_{hc} simplifies to

$$S_{hc} = \sum_{t=1}^{T} \mathrm{tr} \left.\frac{\partial^2 \ell}{\partial \boldsymbol{\nu}_t \partial \boldsymbol{\nu}_t'}\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = \sum_{t=1}^{T}\sum_{i=1}^{k} \left.\frac{\partial^2 \ell}{\partial v_{it}^2}\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}}.$$

Next,

$$\frac{\partial^2 \ell}{\partial v_{it}^2} = -\frac{1}{2}\mathbf{u}'\left\{ -2\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial v_{it}}\mathbf{G}^{-1}\frac{\partial \mathbf{G}}{\partial v_{it}}\mathbf{G}^{-1} + \mathbf{G}^{-1}\frac{\partial^2 \mathbf{G}}{\partial v_{it}^2}\mathbf{G}^{-1} \right\}\mathbf{u}$$

and

$$\left.\frac{\partial^2 \ell}{\partial v_{it}^2}\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = -\frac{1}{2}\mathbf{u}'\mathbf{G}^{-1}\frac{\partial^2 \mathbf{G}}{\partial v_{it}^2}\mathbf{G}^{-1}\mathbf{u}\Big|_{\{\boldsymbol{\nu}=\mathbf{0}\}}.$$

It is then easy to show that $\partial^2 \mathbf{G} / \partial v_{it}^2$ is a $T \times T$ matrix with $2t\sigma_{\eta ii}$ in its t, t'th position and zeroes elsewhere. Thus,

$$\left.\frac{\partial^2 \ell}{\partial v_{it}^2}\right|_{\{\boldsymbol{\nu}=\mathbf{0}\}} = -\,\sigma_e^{-4}\sigma_{\eta ii} t u_t^2$$

and hence

$$S_{hc} = -\sigma_e^{-4}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{k}\sigma_{\eta ii} t u_t^2 = -\sigma_e^{-4}\,\mathrm{tr}\left(\Sigma_\eta\right)\sum\nolimits_{t=1}^{T} t u_t^2.$$

9.4 Proof of Theorems

In what follows we assume that Assumptions LP and KN, the model (7) and $k = O\left(T^{1/2}\right)$ hold. To avoid unnecessary complexity we analyse the regression model without a time trend included, though our results can be shown to extend to the trend case. We also make repeated use of the following representations:

$$\begin{aligned} \hat{u}_t &= u_t - (\hat{\alpha}_k - \alpha) - \left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right)' \mathbf{x}_t && (15) \\ &= u_t - \left(\hat{\mathbf{b}}_k - \mathbf{b}\right)' X_t && (16) \end{aligned}$$

where $X_t = (1, \mathbf{x}_t')'$ and $\left(\hat{\mathbf{b}}_k - \mathbf{b}\right) = \left((\hat{\alpha}_k - \alpha), \left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right)'\right)'$. Also

$$\begin{aligned} \hat{u}_t\hat{u}_{t-k} &= u_t u_{t-k} - \left(\hat{\mathbf{b}}_k - \mathbf{b}\right)' X_t u_{t-k} - u_t X_{t-k}' \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) + \left(\hat{\mathbf{b}}_k - \mathbf{b}\right)' X_t X_{t-k}' \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \\ &= u_t u_{t-k} + z_{k,t} && (17) \end{aligned}$$

where the $z_{k,t}$ are defined implicitly. On several occasions we consider terms similar to (17) written in a more generic notation as

$$\hat{z}_t \hat{z}_{t-k} = z_t^* z_{t-k}^* + z_{k,t}^+$$

As a preliminary to the main proofs we establish some results on estimating long run variances.

9.4.1 Long Run Variance Estimators

A typical long run variance estimator calculation involves showing that the variance when computed from a residual, say, has the same distribution as when computed from the true disturbance. The asymptotic distribution theory is then established using the disturbance expression. For some $\delta > 0$ consider

$$\begin{aligned} & \left\| T^{-(\delta-1)} \{\hat{\omega}^2 \left(\hat{z}_t \hat{z}_{t-k}\right) - \hat{\omega}^2 \left(z_t^* z_{t-k}^*\right)\} \right\| && (18) \\ \leq & \left\| \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) T^{-\delta} \sum_{t=k+j+1}^{T} z_t^* z_{t-k}^* z_{k,t-j}^+ \right\| && (19) \\ & + \left\| \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) T^{-\delta} \sum_{t=k+j+1}^{T} z_{t-k}^* z_{t-j-k}^* z_{k,t}^+ \right\| && (20) \\ & + \left\| \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) T^{-\delta} \sum_{t=k+j+1}^{T} z_{k,t} z_{k,t-j}^+ \right\| . && (21) \end{aligned}$$

Dealing with terms like (19) we use the C-S inequality to get

$$\begin{aligned} & \left\| \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) T^{-\delta} \sum_{t=k+j+1}^{T} z_t^* z_{t-k}^* z_{k,t-j}^+ \right\| \\ \leq & \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) \cdot \sqrt{T^{-\delta} \sum_{t=k+1}^{T} (z_t^* z_{t-k}^*)^2} \cdot \sqrt{T^{-\delta} \sum_{t=k+1}^{T} z_{k,t}^{+2}} \\ \leq & \sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) \cdot \left(T^{-\delta} \sum_{t=1}^{T} z_t^{*4}\right) \cdot \sqrt{T^{-\delta} \sum_{t=k+1}^{T} z_{k,t}^{+2}} && (22) \end{aligned}$$

Typically for $z_{k,t}$ in (17) the linear terms are of the same order and the quadratic is of lower order. This is demonstrated by showing

$$\begin{aligned}
&\left\| \sum_{t=k+1}^{T} \left(\hat{\mathbf{b}}_k - \mathbf{b}\right)' X_t u_{t-k}^2 X_t' \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \right\| \\
\leq\ & \left\| \mathbf{D}_T \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \right\|^2 \left\| \sum_{t=k+1}^{T} \mathbf{D}_T^{-1} X_t X_t' \mathbf{D}_T^{-1} u_{t-k}^2 \right\| \\
\leq\ & \left\| \mathbf{D}_T \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \right\|^2 \sqrt{\sum_{t=1}^{T} \left\| \mathbf{D}_T^{-1} X_t X_t' \mathbf{D}_T^{-1} \right\|^2} . \sqrt{\sum_{t=1}^{T} u_t^4}. \qquad (23)
\end{aligned}$$

where $\mathbf{D}_T = \operatorname{diag}\left(1, \sqrt{T}\mathbf{I}_m\right)$. Based on these inequalities, there now follows a set of lemmata that describe the behaviour of variance estimators under the various sub-hypotheses of the stochastic cointegration framework.

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

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

Proof of Lemma 1. To deal with $\hat{\omega}^2\left(\hat{u}_t \hat{u}_{t-k}\right)$ we use (17) and (18) with $\delta = 1$. Since $T^{1/2}\mathbf{D}_T\left(\hat{\mathbf{b}}_k - \mathbf{b}\right)$ is $O_p(1)$ it is clear that the quadratic form in $\left(\hat{\mathbf{b}}_k - \mathbf{b}\right)$ is of a lower order than the two linear terms in $\left(\hat{\mathbf{b}}_k - \mathbf{b}\right)$ in the expression for $z_{t,k}$ in (17). The linear terms are of the same order. So, by (18) we have that $\left\|\hat{\omega}^2\left(\hat{u}_t \hat{u}_{t-k}\right) - \hat{\omega}^2\left(e_t e_{t-k}\right)\right\|$ is bounded by (19) to (21) with $\delta = 1$.

The first of these (19) ($\delta = 1$) is bounded, using (22) ($\delta = 1$), by

$$\sum_{j=0}^{l} \lambda^+ \left(\frac{j}{l}\right) . T^{-1} \sum_{t=1}^{T} u_t^4 \sqrt{T^{-1} \sum_{t=k+1}^{T} z_{k,t}^2} \qquad (24)$$

where the order of the first term is $O(l)$ (Assumption KN.2) and since $u_t = e_t$ here, the second term is $O_p(1)$, independent of k, by Markov's inequality and Assumption LP. Now, $T^{-1}\sum_{t=k+1}^{T} z_{k,t}^2$, just like $z_{k,t}$, has two dominant terms of the same order and the first, which is bounded by (23) ($\delta = 1$), is

$$\begin{aligned}
&\left\| T^{-1} \sum_{t=k+1}^{T} \left(\hat{\mathbf{b}}_k - \mathbf{b}\right)' X_t u_{t-k}^2 X_t' \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \right\| \\
\leq\ & T^{-1} \left\| T^{1/2}\mathbf{D}_T \left(\hat{\mathbf{b}}_k - \mathbf{b}\right) \right\|^2 \sqrt{T^{-1}\sum_{t=1}^{T} \left\| \mathbf{D}_T^{-1} X_t X_t' \mathbf{D}_T^{-1} \right\|^2} . \sqrt{T^{-1}\sum_{t=1}^{T} u_t^4} \\
=\ & T^{-1}.O_p(1).O_p(1).O_p(1).
\end{aligned}$$

Thus

$$T^{-1} \sum_{t=k+1}^{T} z_{k,t}^2 = O_p(T^{-1}). \qquad (25)$$

Hence, (19) ($\delta = 1$) is bounded by an $O_p(lT^{-1/2})$ random variable.

Now (20) ($\delta = 1$) is identical in structure to (19) ($\delta = 1$) and therefore a similar approach establishes that it too is bounded to the same order. The second summation in (21) ($\delta = 1$) is bounded, using C-S, by means of (25) and so (21) is also bounded by an order $O_p(lT^{-1/2})$ variable. Thus,

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

and the regression effect is eliminated. Applying Theorem LRV of Harris, Leybourne and McCabe (2003) (HMLb) (with $n=1$, $\alpha=2$ and $\boldsymbol{\mu}=0$) then shows that

$$\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right) \xrightarrow{p} \omega_{e1}^2. \tag{26}$$

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

Proof of Lemma 2. Here H_1^0 is true so $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{w}_t$. Using (17) and, with $\delta=3$, (18) to (21) again, $\left\|T^{-2}\{\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right) - \hat{\omega}^2\left(e_te_{t-k}\right)\}\right\|$ is bounded by three terms with (19) bounded by

$$\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right).T^{-3}\sum_{t=1}^{T}u_t^4.\sqrt{T^{-3}\sum_{t=k+1}^{T}z_{k,t}^2}.$$

Now, $T^{-3}\sum_{t=1}^{T}u_t^4 = T^{-3}\sum_{t=1}^{T}\left(\mathbf{v}_t'\mathbf{w}_t+e_t\right)^4 = O_p(1)$. The dominant term of $T^{-3}\sum_{t=k+1}^{T}z_{k,t}^2$, by (23) ($\delta=3$), is

$$\begin{aligned}
&\left\|T^{-3}\sum_{t=k+1}^{T}\left(\hat{\mathbf{b}}_k-\mathbf{b}\right)'X_tu_{t-k}^2X_t'\left(\hat{\mathbf{b}}_k-\mathbf{b}\right)\right\| \\
\leq\;& T^{-1}\left\|\mathbf{D}_T\left(\hat{\mathbf{b}}_k-\mathbf{b}\right)\right\|^2\sqrt{T^{-1}\sum_{t=1}^{T}\left\|\mathbf{D}_T^{-1}X_tX_t'\mathbf{D}_T^{-1}\right\|^2}.\sqrt{T^{-3}\sum_{t=1}^{T}u_t^4} \\
=\;& T^{-1}O_p(1).O_p(1).O_p(1).
\end{aligned}$$

Thus

$$T^{-3}\sum_{t=k+1}^{T}z_{k,t}^2 = O_p(T^{-1}). \tag{27}$$

Hence, (19) ($\delta=3$) is bounded by a $O_p(lT^{-1/2})$ variable. Now, (20) ($\delta=3$) has exactly the same structure and so the order is the same. As before (27) shows that (21) ($\delta=3$) is $O_p(lT^{-1/2})$. Combining these results gives

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

The regression effect is therefore removed from the problem.

In $u_t = \mathbf{v}_t'\mathbf{w}_t + e_t$, e_t is of a lower order of magnitude than $\mathbf{v}_t'\mathbf{w}_t$ and hence, in expressions involving u_t, terms containing e_t are negligible. More formally we may write, in an obvious notation,

$$\begin{aligned}
u_tu_{t-k} &= \boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k} + \widetilde{z}_{k,t} \\
&= z_t^*z_{t-k}^* + z_{k,t}^+
\end{aligned}$$

and hence $\left\|\hat{\omega}^2\left(u_tu_{t-k}\right) - \hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right)\right\|$ may be bounded by (18) to (21), with $\delta=1$. Now each of these terms may be treated exactly the same way as before to establish that

$$\left\|\hat{\omega}^2\left(u_tu_{t-k}\right) - \hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right)\right\| = O_p\left(lT^{-1/2}\right).$$

Lemma 3 *Under H_1^0, $T^{-2}\hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right) \xrightarrow{d} \int_0^1\left(W\otimes W\right)'\Omega_{PP}\left(W\otimes W\right)$ where*

$$\Omega_{PP}=\left(\mathbf{R}_\nu\otimes\mathbf{R}_\nu\right)'\Omega_{22}\left(\mathbf{R}_\nu\otimes\mathbf{R}_\nu\right), W=\mathbf{R}_\eta'B_1.$$

Proof of Lemma 3. Write

$$\begin{aligned}
&T^{-2}\hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right) \\
=\;& \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_t\right)'\operatorname{vec}\left(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}'\right)\operatorname{vec}\left(\boldsymbol{\nu}_{t-j}\boldsymbol{\nu}_{t-k-j}'\right)'\left(\mathbf{w}_{t-k-j}\otimes\mathbf{w}_{t-j}\right) \\
=\;& \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}\left(\mathbf{S}_{t-k}\otimes\mathbf{S}_t\right)'\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right)\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)'\left(\operatorname{vec}\left(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}'\right)\operatorname{vec}\left(\boldsymbol{\zeta}_{t-j}\boldsymbol{\zeta}_{t-k-j}'\right)'\right) \\
&\qquad\times\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right)'\left(\mathbf{S}_{t-k-j}\otimes\mathbf{S}_{t-j}\right).
\end{aligned} \tag{28}$$

The basis of proof is that $\left(\operatorname{vec}\left(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}'\right)\operatorname{vec}\left(\boldsymbol{\zeta}_{t-j}\boldsymbol{\zeta}_{t-k-j}'\right)'\right)$ can be replaced by its expectation and hence that (28) can be written as

$$\begin{aligned}
& T^{-3}\sum_{t=l+1}^{T-k}\left(\mathbf{S}_t\otimes\mathbf{S}_t\right)'\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right) \\
&\qquad\times\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)'\left[\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)E\left(\operatorname{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\right)E\left(\operatorname{vec}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-j}')\right)'\right]\left(\mathbf{R}_v\otimes\mathbf{R}_v\right) \\
&\qquad\times\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right)'\left(\mathbf{S}_t\otimes\mathbf{S}_t\right)+o_p\left(1\right) \\
& \xrightarrow{d}\int_0^1\left(B_2\otimes B_2\right)'\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right)\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)'\Omega_{22}\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)\left(\mathbf{R}_\eta\otimes\mathbf{R}_\eta\right)'\left(B_2\otimes B_2\right) \\
= & \int_0^1\left(W\otimes W\right)'\Omega_{PP}\left(W\otimes W\right).
\end{aligned}$$

The convergence of the expression in square brackets to Ω_{22} follows because it can be shown to be a consistent estimate of the long run variance of $\operatorname{vec}\left(\zeta_t\zeta_{t-k}'\right)$, which is the definition of Ω_{22}, i.e. $\Omega_{22} = \lim_{T\to\infty}\operatorname{var}\left(T^{-1/2}\sum_{t=k+1}^{T}\operatorname{vec}\left(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_{t-k}'\right)\right)$. Then $\Omega_{PP}=\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)'\Omega_{22}\left(\mathbf{R}_v\otimes\mathbf{R}_v\right)$ by definition.

The proof proceeds by showing (using the scalar analogue of (28)) that

$$\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}S_tS_{t-k}S_{t-j}S_{t-k-j}\zeta_t\zeta_{t-k}\zeta_{t-j}\zeta_{t-j-k}\xrightarrow{d}\omega_2^2\int B_2^4.$$

It involves a lengthy sequence of approximants to the LHS above with the limit theory being derived from the final expression. Specifically, we write

$$\begin{aligned}
& \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}S_tS_{t-j}S_{t-k}S_{t-k-j}\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j} \\
= & \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}S_{t-k}^4\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j}+o_p\left(1\right) && (29)\\
= & \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}S_{t-k}^4E_{t-k}\left(\zeta_t\zeta_{t-j}\zeta_{t-k}\zeta_{t-k-j}\right)+o_p\left(1\right) && (30)\\
= & \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=j+1}^{T-k}S_t^4E\left(\zeta_t\zeta_{t-j}\right)^2+o_p\left(1\right) && (31)\\
= & \left[\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)E\left(\zeta_t\zeta_{t-j}\right)^2\right]T^{-3}\sum_{t=l+1}^{T-k}S_t^4+o_p\left(1\right) && (32)\\
& \xrightarrow{d}\omega_2^2\int B_2^4 && (33)
\end{aligned}$$

The proofs of the approximations (29) to (33) are extremely lengthy and are available from the authors on request.

Lemma 4 *Under* H^1, $\hat{\omega}^2\left(u_tu_{t-k}\right)=O_p\left(lT^2\right)$.

Proof of Lemma 4. Consider

$$
\begin{aligned}
T^{-2}\hat{\omega}^2\left(u_t u_{t-k}\right) &= \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}u_t u_{t-k}u_{t-j}u_{t-k-j} \\
&= \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}\mathbf{w}_t'\mathbf{q}_t\mathbf{q}_{t-k}'\mathbf{w}_{t-k}\mathbf{w}_{t-j}'\mathbf{q}_{t-j}\mathbf{q}_{t-k-j}'\mathbf{w}_{t-k-j}+o_p\left(1\right) \\
&= \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_{t-k}\right)'\boldsymbol{\tau}_{t,k}\boldsymbol{\tau}_{t-j,k}'\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_{t-k}\right)+o_p\left(1\right) \quad (34)
\end{aligned}
$$

where

$$
\boldsymbol{\tau}_{t,k}=\mathrm{vec}(\mathbf{q}_t\mathbf{q}_{t-k}')=\mathrm{vec}\left(\mathbf{q}\mathbf{q}'+\mathbf{q}\boldsymbol{\nu}_{t-k}'+\boldsymbol{\nu}_t\mathbf{q}'+\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}'\right), \quad (35)
$$

and $\mathbf{q}_t=\mathbf{q}+\boldsymbol{\nu}_t$. The details of these approximations are tedious but, briefly, the $o_p\left(1\right)$ term in the second line above arises from neglecting terms involving e_t which are clearly dominated by those involving $\mathbf{q}_t'\mathbf{w}_t$, while the $o_p\left(1\right)$ term in the third line arises from replacing each of w_t, w_{t-j} and w_{t-j-k} by w_{t-k}. Substituting the first term in $\tau_{t,k}$ into (34) gives

$$
\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-3}\sum_{t=k+j+1}^{T}\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_{t-k}\right)'\mathrm{vec}\left(\mathbf{q}\mathbf{q}'\right)\mathrm{vec}\left(\mathbf{q}\mathbf{q}'\right)'\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_{t-k}\right)=O_p\left(l\right).
$$

Substituting the other terms from (35) into (34) gives lower orders than $O_p\left(l\right)$. Substituting vec $\left(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}'\right)$ into (34) also gives $O_p\left(1\right)$. Substituting $\mathbf{q}\boldsymbol{\nu}_{t-k}'$ and $\boldsymbol{\nu}_t\mathbf{q}'$ also gives $O_p\left(1\right)$. Therefore from (34) we conclude that $\hat{\omega}^2\left(u_t u_{t-k}\right)=O_p\left(lT^2\right)$.

Lemma 5 *Under H_0^0, $\hat{\omega}^2\left(\hat{u}_t^2-\hat{\sigma}_u^2\right)\xrightarrow{p}\omega_{e2}^2$ where*

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

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

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

Lemma 6 *Under H^1, $\hat{\omega}^2\left(u_t^2-\sigma_u^2\right)=O_p\left(lT^2\right)$.*

Proof of Lemma 6. Now,

$$
\begin{aligned}
\hat{\omega}^2\left(u_t^2-\sigma_u^2\right) &= \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\sum_{t=j+1}^{T}\left(u_t^2-T^{-1}\sum_{t=1}^{T}u_t^2\right)\left(u_{t-j}^2-T^{-1}\sum_{t=1}^{T}u_t^2\right) \\
&= \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\sum_{t=j+1}^{T}u_t^2u_{t-j}^2-\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\left(T-j\right)\left(T^{-1}\sum_{t=1}^{T}u_t^2\right)^2.
\end{aligned}
$$

On substituting $u_t=e_t+\boldsymbol{\nu}_t'\mathbf{w}_t$ the asymptotically dominant terms are those arising from $\boldsymbol{\nu}_t'\mathbf{w}_t$. So, on defining $\left\|Q\right\|_p=E\left(Q^p\right)^{1/p}$ and using Holders inequality,

$$
\begin{aligned}
\left\|\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\sum_{t=j+1}^{T}u_t^2u_{t-j}^2\right\|_1 &= \left\|\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\sum_{t=j+1}^{T}\left(\boldsymbol{\nu}_t'\mathbf{w}_t\right)^2\left(\boldsymbol{\nu}_{t-j}'\mathbf{w}_{t-j}\right)^2\right\|_1 \quad (36) \\
&\leq \sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right)T^{-1}\sum_{t=j+1}^{T}\left\|\mathbf{w}_t\right\|_8^2\left\|\boldsymbol{\nu}_t\right\|_8^2\left\|\mathbf{w}_{t-j}\right\|_8^2\left\|\boldsymbol{\nu}_{t-j}\right\|_8^2 \\
&\leq \mathit{const.}\sum_{j=0}^{l}\lambda^+\left(\frac{j}{l}\right).T^{-1}\sum_{t=1}^{T}t^2 \\
&= O\left(lT^2\right),
\end{aligned}
$$

using the fact that $\|\mathbf{w}_t\|_p = const.t^{1/2}$ for $p \leq 16$ under Assumption LP.4. Similarly,

$$\begin{aligned}
\left\| \sum_{j=0}^{l} \lambda^{+}\left(\frac{j}{l}\right) T^{-1}(T-j)\left(T^{-1}\sum_{t=1}^{T} u_t^2\right)^2 \right\|_1 &\leq \sum_{j=0}^{l} \lambda^{+}\left(\frac{j}{l}\right) . \left\| \left(T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{\nu}_t'\mathbf{w}_t\right)^2\right) \right\|_2^2 \\
&\leq \sum_{j=0}^{l} \lambda^{+}\left(\frac{j}{l}\right) . \left(T^{-1}\sum_{t=1}^{T} \|\mathbf{w}_t\|_8^2 \|\boldsymbol{\nu}_t\|_8^2\right)^2 \\
&\leq const. \sum_{j=0}^{l} \lambda^{+}\left(\frac{j}{l}\right) . \left(T^{-1}\sum_{t=1}^{T} t\right)^2 \\
&= O\left(lT^2\right),
\end{aligned}$$

which shows that $\hat{\omega}^2\left(u_t^2 - \sigma_u^2\right) = O_p\left(lT^2\right)$.

9.4.2 Proof of Theorem 1

Before we proceed we need a preliminary result on convergence to a stochastic integral.

Lemma 7 *Let $\boldsymbol{\zeta}_t$ satisfy Assumption LP and let $k = O(T^{1/2})$. Then, as $T \to \infty$,*

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

where $W = \mathbf{R}_\eta' B_1$ and $P = (\mathbf{R}_\nu \otimes \mathbf{R}_\nu)' B_2$.

Proof of Lemma 7 First rewrite using $\Delta \mathbf{S}_t = \boldsymbol{\zeta}_t$, so that

$$\begin{aligned}
&T^{-3/2} \sum \left(\mathbf{w}_{t-k} \otimes \mathbf{w}_t\right)' \operatorname{vec}\left(\boldsymbol{\nu}_t \boldsymbol{\nu}_{t-k}'\right) \\
= \ &T^{-3/2} \sum \left(\mathbf{R}_\eta' \mathbf{S}_{t-k} \otimes \mathbf{R}_\eta' \mathbf{S}_{t-1}\right)' \operatorname{vec}\left(\mathbf{R}_\nu' \boldsymbol{\zeta}_t \boldsymbol{\zeta}_{t-k}' \mathbf{R}_\nu\right) + T^{-3/2} \sum \mathbf{w}_{t-k}' \boldsymbol{\nu}_{t-k} \boldsymbol{\nu}_t' \boldsymbol{\eta}_t.
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{m}_{k,t} &= \sum_{j=1}^{\infty} \mathbf{G}_{k,k-j} \operatorname{vec}\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-j}'\right) \\
\mathbf{r}_{k,t} &= \sum_{i=-\infty}^{\infty} \mathbf{F}_{k,k-i}(L) \operatorname{vec}\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-i}'\right) \\
\tilde{\mathbf{r}}_{k,t} &= \sum_{i=1}^{\infty} \tilde{\mathbf{G}}_{k,k-i}(L) \operatorname{vec}\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t-i}'\right)
\end{aligned}$$

and the coefficients are defined by

$$\begin{aligned}
\mathbf{G}_{k,r}(L) &= \sum_{j=r\vee 0}^{k-1} (\mathbf{C}_{j-r}\otimes\mathbf{C}_j) L^j \\
\mathbf{F}_{k,r}(L) &= \sum_{j=r\vee k}^{\infty} (\mathbf{C}_{j-r}\otimes\mathbf{C}_j) L^j \\
\tilde{\mathbf{G}}_{k,k-i}(L) &= \sum_{j=0}^{k-2} \tilde{\mathbf{G}}_{k,k-i,j} L^j, \quad \tilde{\mathbf{G}}_{k,k-i,j} = \sum_{r=(k-i)\vee(j+1)}^{k-1} \mathbf{C}_{r-k+i}\otimes\mathbf{C}_r.
\end{aligned}$$

Apply Theorem BN of HMLb to $\mathrm{vec}\left(\boldsymbol{\zeta}_t\boldsymbol{\zeta}'_{t-k}\right)$ to get a martingale approximation, $\mathbf{m}_{k,t}$, a remainder term $\mathbf{r}_{k,t}$ and an over-differenced 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,

$$\begin{aligned}
& T^{-3/2}\sum (\mathbf{w}_{t-k}\otimes\mathbf{w}_t)' \,\mathrm{vec}\left(\boldsymbol{\nu}_t\boldsymbol{\nu}'_{t-k}\right) \\
= \;& T^{-3/2}\sum \left(\mathbf{R}'_\eta\mathbf{S}_{t-k}\otimes\mathbf{R}'_\eta\mathbf{S}_{t-1}\right)' (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' \mathbf{m}_{k,t} + T^{-3/2}\sum \left(\mathbf{R}'_\eta\mathbf{S}_{t-k}\otimes\mathbf{R}'_\eta\mathbf{S}_{t-1}\right)' (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' (\mathbf{r}_{k,t}-\Delta\tilde{\mathbf{r}}_{k,t}) \\
& + T^{-3/2}\sum \mathbf{w}'_{t-k}\boldsymbol{\nu}_{t-k}\boldsymbol{\nu}'_t\boldsymbol{\eta}_t.
\end{aligned}$$

We find

$$\begin{aligned}
& T^{-3/2}\sum \left(\mathbf{R}'_\eta\mathbf{S}_{t-k}\otimes\mathbf{R}'_\eta\mathbf{S}_{t-1}\right)' (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' \mathbf{r}_{k,t} \xrightarrow{p} \mathbf{0}, \\
& T^{-3/2}\sum \left(\mathbf{R}'_\eta\mathbf{S}_{t-k}\otimes\mathbf{R}'_\eta\mathbf{S}_{t-1}\right)' (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' \Delta\tilde{\mathbf{r}}_{k,t} \xrightarrow{p} \mathbf{0}, \\
& T^{-3/2}\sum \mathbf{w}'_{t-k}\boldsymbol{\nu}_{t-k}\boldsymbol{\nu}'_t\boldsymbol{\eta}_t \xrightarrow{p} \mathbf{0}.
\end{aligned}$$

The first result follows directly from Theorem SI of HMLb and the second is established along very similar lines. The last follows by writing

$$T^{-3/2}\sum \mathbf{w}'_{t-k}\boldsymbol{\nu}_{t-k}\boldsymbol{\nu}'_t\boldsymbol{\eta}_t = T^{-3/2}\sum \mathbf{w}'_{t-k}[\mathbf{a}_t - E_{t-k}(\mathbf{a}_t)] + T^{-3/2}\sum \mathbf{w}'_{t-k}E_{t-k}(\mathbf{a}_t)$$

where $\mathbf{a}_t = \boldsymbol{\nu}_{t-k}\boldsymbol{\nu}'_t\boldsymbol{\eta}_t$. The first term can be shown to disappear on exploiting the properties of the increment process i.e. 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,

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

Now, since $k = o(T)$, it follows from the FCLT that

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

and from Theorem FCLT of HMLb this convergence is joint with $\mathbf{N}_{T,[Ts]} = T^{-1/2}\sum^{[Ts]}\mathrm{vec}\left(\boldsymbol{\nu}_t\boldsymbol{\nu}'_{t-k}\right) = \mathbf{M}_{T,[Ts]} + o_p(1)$ where $\mathbf{M}_{T,[Ts]} = T^{-1/2}\sum^{[Ts]}\mathbf{m}_{k,t}$. Thus Theorem SI of HMLb applies and setting $B_Q \equiv (\mathbf{R}_{\boldsymbol{\nu}}\otimes\mathbf{R}_{\boldsymbol{\nu}})' B_2 = P$ and $U \equiv \mathbf{R}'_\eta B_1 \otimes \mathbf{R}'_\eta B_1 = W \otimes W$ we have that

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

converges weakly to

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

Part (i) The Null.

(a) This part of the proof deals with the sub-hypothesis that u_t is stationary, H_0^0. The first thing to note are the orders of the regression moments. Under H_0^0, $u_t = e_t$ and from HMLa we have that $\hat{\alpha}_k - \alpha = O_p\left(T^{-1/2}\right)$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p\left(T^{-1}\right)$ and hence

$$T^{-2}\sum_{t=k+1}^{T}\mathbf{x}_t\mathbf{x}_{t-k}',\ T^{-1/2}\sum_{t=k+1}^{T}e_t,\ T^{-1}\sum_{t=k+1}^{T}\mathbf{x}_{t-k}e_t,\ T^{-1}\sum_{t=k+1}^{T}\mathbf{x}_te_{t-k},\ T^{-3/2}\sum_{t=k+1}^{T}\mathbf{x}_{t-k}$$

are all $O_p(1)$. There are three further steps required to find the asymptotic distribution of the statistic $\widehat{S}_{nc}$ of Theorem 1; a step for each of the numerator, denominator and their ratio.

(a.1) The numerator. Since $e_t = \mathbf{c}'\boldsymbol{\varepsilon}_t$ is a linear combination of a vector linear process, it follows from (17) and by an application of Theorem FCLT of HMLb that

$$\begin{aligned} T^{-1/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k} &= T^{-1/2}\sum_{t=k+1}^{T}e_te_{t-k} + O_p\left(T^{-1/2}\right) \\ &\xrightarrow{d} N\left(0,\omega_{e1}^2\right). \end{aligned} \tag{37}$$

(a.2) The denominator. By Lemma 1, $\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right) \xrightarrow{p} \omega_{e1}^2$

(a.3) The ratio. Since the convergence in (26) is to a constant the continuous mapping theorem (CMT) combines (37) and (26) to show

$$\widehat{S}_{nc} = \frac{T^{-1/2}\sum_{t=k+1}^{T}\widehat{u}_t\widehat{u}_{t-k}}{\widehat{\omega}(\widehat{u}_t\widehat{u}_{t-k})} \xrightarrow{d} N\left(0,\omega_{e1}^2\right)$$

under H_0^0.

(b) This part of the proof deals with the heteroscedastic sub-hypothesis of the null, H_1^0. As before we require the orders of the regression moments and the three steps to deal with the numerator, denominator and the ratio. Under H_1^0, $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{w}_t$ and from HMLa, $\hat{\alpha}_k - \alpha = O_p(1)$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p\left(T^{-1/2}\right)$. We find

$$T^{-3/2}\sum_{t=k+1}^{T}\left(\hat{\alpha}_k - \alpha\right)^2 = O_p\left(T^{-1/2}\right),$$

$$T^{-1/2}.T^{1/2}\left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right)'T^{-2}\sum_{t=k+1}^{T}\mathbf{x}_t\mathbf{x}_{t-k}'T^{1/2}\left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right) = O_p\left(T^{-1/2}\right),$$

$$T^{-1/2}\left(\hat{\alpha}_k - \alpha\right)T^{-1}\sum_{t=k+1}^{T}u_t = O_p\left(T^{-1/2}\right),$$

$$T^{-1/2}.T^{1/2}\left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right)'T^{-3/2}\sum_{t=k+1}^{T}\mathbf{x}_tu_{t-k} = O_p\left(T^{-1/2}\right),$$

$$T^{-1/2}\left(\hat{\alpha}_k - \alpha\right)T^{1/2}\left(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\right)'T^{-3/2}\sum_{t=k+1}^{T}\mathbf{x}_t = O_p\left(T^{-1/2}\right).$$

(b.1) The numerator. It follows from (17) that the numerator is

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

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

$$\begin{aligned}
T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k} &= T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}+T^{-3/2}\sum_{t=k+1}^{T}e_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k} \\
&\quad +T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{w}_te_{t-k}+T^{-3/2}\sum_{t=k+1}^{T}e_te_{t-k}+O_p\left(T^{-1/2}\right) \\
&= T^{-3/2}\sum_{t=k+1}^{T}\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}+O_p\left(T^{-1/2}\right) \\
&= T^{-3/2}\sum_{t=k+1}^{T}\left(\mathbf{w}_{t-k}\otimes\mathbf{w}_t\right)'\operatorname{vec}(\boldsymbol{\nu}_t\boldsymbol{\nu}_{t-k}')+O_p\left(T^{-1/2}\right) \\
&\xrightarrow{d}\int_0^1\left(W\otimes W\right)'dP
\end{aligned}$$

where $W=\mathbf{R}_\eta'B_1$ and $P=(\mathbf{R}_\nu\otimes\mathbf{R}_\nu)'B_2$ and B_1 and B_2 are independent Brownian motions with covariance matrices Ω_{11} and Ω_{22}. The weak convergence follows from Lemma 7. The covariance matrix of P is $\Omega_{PP}=\left(\mathbf{R}_\nu\otimes\mathbf{R}_\nu\right)'\Omega_{22}\left(\mathbf{R}_\nu\otimes\mathbf{R}_\upsilon\right)$.

(b.2) The denominator. Lemma 2 gives

$$T^{-2}\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right)=T^{-2}\hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right)+O_p\left(lT^{-1/2}\right)$$

and Lemma 3 shows that

$$T^{-2}\hat{\omega}^2\left(\boldsymbol{\nu}_t'\mathbf{w}_t\mathbf{w}_{t-k}'\boldsymbol{\nu}_{t-k}\right)\xrightarrow{d}\int_0^1\left(W\otimes W\right)'\Omega_{PP}\left(W\otimes W\right).$$

Thus

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

Note by the continuity of the sample paths of the Brownian motion that

$$\Pr\left(\int_0^1\left(W\otimes W\right)'\Omega_{PP}\left(W\otimes W\right)=0\right)=0.$$

(b.3) The ratio. 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 $\sqrt{T^{-2}\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right)}$. The vector

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

converges weakly to

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

as shown in Lemma 7. Next the mapping theorem, with the above vector as argument and the ratio as the map, applies to conclude that

$$\frac{T^{-3/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k}}{\sqrt{T^{-2}\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right)}}\xrightarrow{d}\frac{\int_0^1 (W\otimes W)'\,dP}{\sqrt{\int_0^1 (W\otimes W)'\,\Omega_{PP}\,(W\otimes W)}} \tag{38}$$

since the set of discontinuities of the map has probability zero with respect to the limit measure. Conditional on W,

$$\int_0^1 (W\otimes W)'\,dP\sim N\left(0,\int_0^1 (W\otimes W)'\,\Omega_{PP}\,(W\otimes W)\right) \tag{39}$$

so the distribution in (38) is unconditionally standard normal as required.

Part (ii) The Alternative.
Under H^1, $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{w}_t + \mathbf{q}'\mathbf{w}_t$ where $\mathbf{q}\neq 0$. Here, it is easy to show that $\hat{\alpha}_k-\alpha = O_p\left(T^{1/2}\right)$ and $\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}=O_p\left(1\right)$ and, using (17), this implies that $\hat{u}_t\hat{u}_{t-k}$ is of the same order in probability as u_tu_{t-k}. It is then straightforward to deduce that

$$T^{-1/2}\sum_{t=k+1}^{T}\hat{u}_t\hat{u}_{t-k}=O_p\left(T^{3/2}\right). \tag{40}$$

Now we require a bound for the order of probability of $\hat{\omega}^2\left(\hat{u}_t\hat{u}_{t-k}\right)$, which again is the same as the order of probability of $\hat{\omega}^2\left(u_tu_{t-k}\right)$. By Lemma 4, $\hat{\omega}^2\left(u_tu_{t-k}\right)=O_p\left(lT^2\right)$ and hence the distribution of $\left|\widehat{S}_{nc}\right|$ diverges at least as fast as $O_p(\sqrt{T/l})$.

9.4.3 Proof of Theorem 2

Part (i) The Null.
Under H_0^0, $u_t=e_t$ and from HMLa, we have that $\hat{\alpha}_k-\alpha=O_p\left(T^{-1/2}\right)$ and $\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}=O_p\left(T^{-1}\right)$. Then, it follows from (15) that

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

where $\sigma_e^2=E(e_t^2)$. An application of a CLT shows the right hand side to be asymptotically normal. A quick route to the variance of the limiting normal distribution is to use integral notation and write

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

here $F_T(s)$ is the partial sum process of $\{e_t^2-\sigma_e^2\}$. Thus,

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

where $F\left(s\right)$ is a Brownian motion with variance ω_{e2}^2. Hence, $\int_0^1\left(s-\frac{1}{2}\right)dF\left(s\right)$ 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

$$T^{-3/2}\sum_{t=1}^{T}t\left(\hat{u}_t^2-\hat{\sigma}_u^2\right)\xrightarrow{d}N(0,12\omega_{e2}^2) \tag{41}$$

Using Lemma 5

$$\hat{\omega}^2 \left(\hat{u}_t^2 - \hat{\sigma}_u^2\right) \xrightarrow{p} \omega_{e2}^2 \tag{42}$$

and since the limit in (42) is a constant, the CMT with (41) gives the result.

Part (ii) The Alternative.
Under H_1^0, $u_t = e_t + \boldsymbol{\nu}_t'\mathbf{w}_t$ and from HMLa, $\hat{\alpha}_k - \alpha = O_p(1)$ and $\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta} = O_p\left(T^{-1/2}\right)$. We may write

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

From (15), $\hat{u}_t$ is of the same order in probability as u_t and it is then straightforward to show that

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

and hence

$$T^{-3/2} \sum_{t=1}^{T} t \left(\hat{u}_t^2 - \hat{\sigma}_u^2\right) = O_p\left(T^{3/2}\right). \tag{43}$$

In the denominator, $\hat{\omega}^2 \left(\hat{u}_t^2 - \hat{\sigma}_u^2\right)$ and $\hat{\omega}^2 \left(u_t^2 - \sigma_u^2\right)$ are of the same order (where $\sigma_u^2 = T^{-1} \sum_{t=1}^{T} u_t^2$) and hence, by Lemma 6, the distribution of $\left|\widehat{S}_{hc}\right|$ diverges at least as fast as $O_p(\sqrt{T/l})$.