

Testing for rational bubbles in a co-explosive vector autoregression¹

TOM ENGSTED² & BENT NIELSEN³

18 November 2011

Summary: Asset bubbles can be described through the rational bubble solution of the standard stock price model linking stock prices and dividends. We show how the hypothesis of a rational bubble can be tested in the context of a bivariate coexplosive vector autoregression. The methodology is illustrated using US stock prices and dividends for the period 1974-2000.

Keywords: *Rational bubbles, Explosiveness and co-explosiveness, Cointegration, Vector autoregression, Likelihood ratio tests.*

1 Introduction

Asset bubbles can be described as an equilibrium phenomenon under rational expectations and informationally efficient capital markets. We show how the hypothesis of a rational bubble can be tested in the context of a coexplosive and cointegrated vector autoregression. The proposed procedure is an extension to the tests for the present value model for stock prices without bubbles in the context of the cointegrated vector autoregression, see Campbell and Shiller (1987, 1988) and Johansen and Swensen (1999, 2004, 2011). The present work allows for bubbles and draws on the coexplosive analysis proposed by Nielsen (2010) and its application to stock prices by Engsted (2006).

The proposed procedure complements the econometric procedures available in the literature. Flood and Garber (1980) and Froot and Obstfeld (1991) were concerned with explosive bubbles which are either deterministic or deterministic functions of functionals. West (1987) and Diba and Grossman (1988) were concerned with testing the hypothesis of no bubbles. The West (1987) test is inconsistent according to West (1985) as a result of the explosiveness under the alternative. Our proposed testing procedure does not face such a problem because the model and the hypothesis

¹Comments from the referees are gratefully acknowledged.

²CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C. E-mail: tengsted@creates.au.dk. CREATES is funded by the Danish National Research Foundation.

³Nuffield College, Oxford OX1 1NF, UK. E-mail: bent.nielsen@nuffield.ox.ac.uk. Financial support from ESRC grant RES-000-27-0179 and the Institute for New Economic Thinking at the Oxford Martin School is gratefully acknowledged.

explicitly include explosive features. The methodology is illustrated using US stock prices and dividends for the period 1974-2000.

2 The model, restrictions, interpretation and estimation

The economic model for stock price determination is formulated. As this is a partial model it is then embedded in a vector autoregressive framework. It is shown how the stock price model can be formulated as a testable hypothesis.

2.1 The model for stock price determination

The standard model for stock price determination involves real stock prices P_t and dividends D_t . The model is given by the equation

$$P_t = \frac{1}{1+R} \mathcal{E}_t(P_{t+1} + D_{t+1}), \quad (2.1)$$

where $R > 0$ is the expected one-period return on the stock, which is assumed constant. More general models allowing for time-varying expected returns are common in empirical finance, see for instance Cochrane (2008), but since the previous empirical bubble literature has assumed R constant, models with time-varying expected returns will not be pursued here. As it stands the model is partial in that the conditional expectations operator $\mathcal{E}_t$ is left unspecified. Later a complete model is presented in which $\mathcal{E}_t$ is precisely defined. The equation (2.1) can be reformulated as

$$\mathcal{E}_{t-1}M_t = 0 \quad \text{where} \quad M_t = P_t + D_t - (1+R)P_{t-1}. \quad (2.2)$$

In other words, M_t is a martingale difference, which is a version of the efficient market hypothesis, see Leroy (1989).

Considerable attention has been given to the situation in which P_t and D_t have random walk features and the “spread” $S_t = P_t - D_t/R$ is a cointegrating relation, see e.g. Campbell and Shiller (1987, 1988) and Johansen and Swensen (1999, 2004). Often cointegration refers to reduction of random walk behaviour to stationarity. In the present context it is useful to use the term cointegration simply for removal of random walk behaviour so that S_t remains a cointegrating relation. To see this add and subtract $\{(1+R)/R\}\Delta_1 D_t$ to M_t , where $\Delta_1 D_t = D_t - D_{t-1}$, to get

$$M_t = \Delta_{1+R}S_t + (1+R^{-1})\Delta_1 D_t, \quad (2.3)$$

where $\Delta_{1+R}S_t \equiv S_t - (1+R)S_{t-1}$. The relation shows that if $\Delta_1 D_t$ is stationary then $\Delta_{1+R}S_t$ is also stationary since M_t is assumed a martingale difference. Since the operator Δ_{1+R} removes exponential features but does not reduce random walks

to stationarity for $R \neq 0$ then S_t could be explosive, but must be free of unit roots even if P_t and D_t have random walk features. Hence S_t is a cointegrating relation.

A consequence of the stock price determination model is that the “spread” can be written as a present value of expected future dividends:

$$S_t = P_t - \frac{1}{R}D_t = \frac{1+R}{R} \sum_{r=1}^{\infty} \frac{\mathcal{E}_t(\Delta_1 D_{t+r})}{(1+R)^r} + bB_t. \quad (2.4)$$

for $b \in \mathbb{R}$ and where B_t obeys, see Diba and Grossman (1988) and Engsted (2006),

$$B_{t+1} = (1+R)B_t + \xi_{t+1}, \quad (2.5)$$

where $\mathcal{E}_t \xi_{t+1} = 0$. The variable B_t is called a rational bubble, the component of stock prices that reflects self-fulfilling rational expectations of future price increases independently of fundamentals $\mathcal{E}_t D_{t+s}$. To see that (2.4) and (2.5) solve (2.1), solve for P_t , and insert the expression on both sides of (2.1). If $\Delta_1 D_t$ is stationary then the present value component on the right hand side of (2.4) is well-defined. In other words, if D_t is integrated of order one then the “spread” S_t is a cointegrating relation in that it has no unit roots. At the same time S_t evolves around the explosive component B_t . These features can be described through the co-explosive model.

2.2 The vector autoregressive model

The partial economic model (2.1) is completed by assuming that $X_t = (P_t, D_t)$ is vector autoregressive in line with Campbell and Shiller (1988) and Johansen and Swensen (1999, 2004, 2011). The vector autoregressive model of order k is given by

$$X_t = \sum_{j=1}^k A_j X_{t-j} + \mu + \varepsilon_t, \quad (2.6)$$

where $A_j \in \mathbb{R}^{2 \times 2}$ and $\mu \in \mathbb{R}^2$. The errors, ε_t , are independent $\mathbf{N}_2(0, \Omega)$ -distributed, or more generally a martingale difference sequence. The completed model implies a natural filtration $\mathcal{F}_t = \sigma(X_s; 1 - k \leq s \leq t)$ and an expectations operator $\mathbf{E}$. Within this model the efficient marked hypothesis amounts to $\mathbf{E}(M_t | \mathcal{F}_{t-1}) = 0$.

2.3 The coexplosive model and its interpretation

The coexplosive model arises as a restriction to the vector autoregressive model (2.6). It allows for common random walk trend and a common explosive stochastic component with explosive root $\rho > 1$. To facilitate the analysis the model is reparametrised in equilibrium correction form, see Nielsen (2010), as

$$\Delta_1 \Delta_\rho X_t = \Pi_1 \Delta_\rho X_{t-1} + \Pi_\rho \Delta_1 X_{t-1} + \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho X_{t-j} + \mu + \varepsilon_t, \quad (2.7)$$

where $\Delta_\rho X_t = X_t - \rho X_{t-1}$ and $\Delta_1 X_t = X_t - X_{t-1}$. The parameters satisfy $\Pi_1, \Pi_\rho, \Phi_j \in \mathbb{R}^{2 \times 2}$, $\mu \in \mathbb{R}^2$, and $\rho \in \mathbb{R}$.

The additional assumptions that X_t has one unit root and one explosive root are accommodated by reduced rank restrictions so

$$\mathbf{H}_1 : \quad (\Pi_1, \mu) = \alpha_1(\beta'_1, \zeta_1), \quad \Pi_\rho = \alpha_\rho \beta'_\rho,$$

where $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}^2$, $\zeta_1 \in \mathbb{R}$. The model $\mathbf{M}$ restricted by $\mathbf{H}_1$ is denoted $\mathbf{M}_1$.

The process can be interpreted through its Granger-Johansen representation. Such a representation was given in Nielsen (2010, Theorem 1). This shows that $\beta'_1 \Delta_\rho X_t$, $\beta'_\rho \Delta_1 X_t$ and $\Delta_1 \Delta_\rho X_t$ can be given a stationary distribution. The vectors β_1 and β_ρ are therefore referred to as the cointegrating vector and the co-explosive vector. In the context of the stock price model it is useful to restate the result with detailed expressions for the initial values of the process which will be used for the computation of Figure 1. For a matrix α with full column rank $\alpha_\perp$ denotes an orthogonal complement of a matrix α so $\alpha'_\perp \alpha = 0$ and $(\alpha, \alpha_\perp)$ is invertible while $\bar{\alpha} = \alpha(\alpha' \alpha)^{-1}$.

Assumption A Consider the model $\mathbf{M}$ and the hypothesis $\mathbf{H}_1$. Suppose

- (i) The vectors $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}$ are non-zero;
- (ii) The nonstationary characteristic roots of X_t are at 1 or $\rho > 0$;
- (iii) $\det(\alpha'_{1\perp} \Psi_1 \beta_{1\perp}) \neq 0$ and $\det(\alpha'_{\rho\perp} \Psi_\rho \beta_{\rho\perp}) \neq 0$ where

$$\Psi_1 = I_2 + \frac{\alpha_\rho \beta'_\rho}{\rho - 1} - \sum_{j=1}^{k-2} \Phi_j, \quad \Psi_\rho = I_2 + \frac{\alpha_1 \beta'_1}{1 - \rho} - \sum_{j=1}^{k-2} \rho^{-j} \Phi_j.$$

Theorem 2.1 Suppose Assumption A holds. Then it holds

$$X_t = \frac{1}{1 - \rho} C_1 (A_1 + \sum_{s=1}^t \varepsilon_s) + \frac{1}{\rho - 1} C_\rho \rho^t (A_\rho + \sum_{s=1}^t \rho^{-s} \varepsilon_s) + Y_t - \frac{\bar{\beta}_1 \zeta_1}{1 - \rho},$$

where Y_t is a stationary process. The impact matrices and initial values are⁴

$$\begin{aligned} C_x &= x \beta_{x\perp} (\alpha'_{x\perp} \Psi_x \beta_{x\perp})^{-1} \alpha'_{x\perp}, \\ A_x &= \Psi_x \Delta_y X_0 - \frac{\alpha_y \beta'_y}{y - x} \Delta_x X_0 + \sum_{j=1}^{k-2} \Phi_j \sum_{h=0}^{j-1} x^{h-j} \Delta_x \Delta_y X_{-h}, \end{aligned}$$

where (x, y) is $(1, \rho)$ or $(\rho, 1)$. The stationary process Y_t is given by $Y_t = \theta_U U_t$ where U_t has zero mean stationary and is defined in terms of $\tilde{X}_t = X_t - \bar{\beta}_1 \zeta_1 / (1 - \rho)$ as

$$U_{t-1} = \{(\beta'_1 \Delta_\rho \tilde{X}_{t-1})', (\beta'_\rho \Delta_1 \tilde{X}_{t-1})', \Delta_1 \Delta_\rho \tilde{X}_{t-1}, \dots, \Delta_1 \Delta_\rho \tilde{X}_{t-k+2}\}',$$

⁴The definition of C_x includes a factor x which is erroneously left out in Nielsen (2010).

while the parameter θ_U is defined through

$$\begin{aligned}\theta_U &= (G_{1,\rho}, G_{\rho,1}, H_{1,\rho,1} + H_{\rho,1,1}, \dots, H_{1,\rho,k-2} + H_{\rho,1,k-2}), \\ G_{x,y} &= -\frac{C_y \alpha_x}{(y-x)^2} - \frac{C_x \Psi_x \bar{\beta}_x}{x-y} + \frac{x \bar{\beta}_x}{x-y}, \quad H_{x,y,n} = \frac{C_x x^{n-1}}{y-x} \sum_{j=n}^{k-2} \Phi_j x^{-j}.\end{aligned}$$

The cointegration rank, that is the rank of the matrix Π_1 , can be determined with Johansen's likelihood test based procedure. In particular, the likelihood is maximised using reduced rank regression of $\Delta_1 X_t$ on $(X'_{t-1}, 1)'$ correcting for $\Delta_1 X_{t-1}, \dots, \Delta_1 X_{t-k+1}$; see also Nielsen (2010, §3.3) for a discussion.

2.4 Testing that dividends are non-explosive

In the stock price model the dividends are assumed to have stationary differences. In other words the dividends are non-explosive. This can be tested through the simple hypothesis on the coexplosive vector that

$$\mathbf{H}_D : \quad \beta_\rho = (0, 1)'$$

Under the hypothesis $\mathbf{H}_D$ the model equation of $\mathbf{M}_1$ reduces to

$$\mathbf{M}_{1D}: \quad \Delta_1 \Delta_\rho X_t = \alpha_1 \beta_1^{*'} \Delta_\rho X_{t-1}^* + \alpha_\rho \Delta_1 D_{t-1} + \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho X_{t-j} + \varepsilon_t, \quad (2.8)$$

where $\beta_1^* = (\beta_1', \zeta_1')'$ and $\Delta_\rho X_{t-1}^* = (\Delta_\rho X_{t-1}, 1)$. For a given value of ρ the likelihood is maximised by reduced rank regression of $\Delta_1 \Delta_\rho X_t$ on $\Delta_\rho X_{t-1}$ correcting for lagged dividend growth $\Delta_1 D_{t-1}$ and lagged differences $\Delta_1 \Delta_\rho X_{t-j}$; see Nielsen (2010, §3.3)1.

2.5 Testing the bubble model

The bubble model imposes two restrictions on the model $\mathbf{M}_{1D}$. The first restriction is that the “spread” $S_t = P_t - D_t/R$ is a cointegrating relation so that the coefficient R is linked to the explosive root through $\rho = 1 + R$. This gives the hypothesis

$$\mathbf{H}_S : \quad \beta_1 = (1, -1/R)', \quad \text{where } R = \rho - 1. \quad (2.9)$$

The second restriction is the martingale restriction (2.2). To formulate this hypothesis pre-multiply the equation (2.8) by the vector $\iota' = (1, 1)$ to get

$$\Delta_1 \Delta_\rho (P_t + D_t) = \iota' \alpha_1 \Delta_\rho S_{t-1} + \iota' \alpha_\rho \Delta_1 D_{t-1} + \iota' \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho X_{t-j} + \iota' \alpha_1 \zeta_1 + \iota' \varepsilon_t.$$

This equation reduces to the equation $M_t = \iota' \varepsilon_t$, which can be denoted $\varepsilon_{M,t}$, when

$$\mathbf{H}_B: \quad \iota' \alpha_1 = -1, \quad \iota' \alpha_\rho = -(1+R)^2/R, \quad \iota' \Phi_j = 0, \quad \zeta_1 = 0, \quad (2.10)$$

for $j = 1, \dots, k$. The model $\mathbf{M}_{1D}$ restricted by $\mathbf{H}_S$ and $\mathbf{H}_B$ is denoted $\mathbf{M}_{1DSB}$.

It is convenient to reparametrise the restricted model $\mathbf{M}_{1DSB}$. To deal with the correlation structure for the errors, let ω denote the population regression coefficient of $\varepsilon_{D,t} = (0, 1)\varepsilon_t$ on $\varepsilon_{M,t} = (1, 1)\varepsilon_t$ and rewrite the model in terms of the marginal equation for M_t and the conditional equation for $\Delta_1 D_t$ so

$$\mathbf{M}_{1DSB}: \quad M_t = \varepsilon_{M,t}, \quad (2.11)$$

$$\begin{aligned} \Delta_1 D_t &= \alpha_{1,D} \Delta_\rho S_{t-1} + (\alpha_{\rho,D} + \rho) \Delta_1 D_{t-1} \\ &\quad + \sum_{j=1}^{k-2} \Phi_{j,D} \Delta_1 \Delta_\rho X_{t-j} + \omega M_t + \varepsilon_{D \cdot M,t}, \end{aligned} \quad (2.12)$$

where the errors $\varepsilon_{M,t}$ and $\varepsilon_{D \cdot M,t} = \varepsilon_{D,t} - \omega \varepsilon_{M,t}$ are uncorrelated. The likelihood implied by these equations is maximised through a profile argument using that for a known R the regressions (2.11) and (2.12) are unrelated. The likelihood is then maximised by maximising over R .

3 Asymptotic analysis

The proposal is to test the hypotheses $\mathbf{H}_1, \mathbf{H}_D, \mathbf{H}_S, \mathbf{H}_B$ through a sequence of likelihood ratio test statistics. It is shown that these have standard distributions.

The result is proved under the assumption that the sequence of innovations ε_t is an $\mathcal{F}_t$ -martingale difference sequence. This is in line with the stock price model in which the rational expectation essentially amounts to a martingale difference assumption. This is remarkable as the asymptotic distribution theory of the explosive root estimator itself relies on a normality assumption. Two additional conditions are needed.

Assumption B *For some $\gamma > 2$ it holds $\sup_t \mathbf{E}\{(\varepsilon'_t \varepsilon_t)^{(2+\gamma)/2} | \mathcal{F}_{t-1}\} < \infty$ a.s.*

Assumption C *$\mathbf{E}(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}) = \Omega$ a.s. where Ω is positive definite.*

The purpose of Assumption B is to give a bound to the fluctuations of the paths while Assumption C makes the conditional variances time invariant. While allowing for a degree of temporal dependence each of the conditions rule out autoregressive conditional heteroscedasticity (ARCH).

A range of specification tests are available for the vector autoregression $\mathbf{M}$ in the presence of an explosive root. Nielsen (2006a,b) has shown that likelihood-based

procedures for lag-length determination apply as opposed to procedures based on the Yule-Walker equations. Engler and Nielsen (2009) show that plots of empirical quantiles of the residuals against normal quantiles are valid. Nielsen and Sohkanen (2011) show that cumulative sums of squares tests for constant variance of ε_t can be applied.

As a next step, the test for cointegration rank, that is the likelihood ratio test for H_1 within M , has standard critical values in the presence of an explosive root. This has previously been proved in the univariate case by Nielsen (2001) and for the model with a restricted linear trend by Nielsen (2010, Theorem 2).

Theorem 3.1 *Suppose model M_1 and Assumptions A, B, C hold with $\gamma > 0$ only. Then $LR(M_1|M)$ has the same limiting distribution as in Johansen (1995, §6),*

$$LR(M_1|M) \xrightarrow{D} \text{tr}\left\{\int_0^1 dB_u F'_u \left(\int_0^1 F_u F'_u du\right)^{-1} \int_0^1 F_u dB'_u\right\},$$

where $F_u = (B_u, 1)'$ with B_u being a standard Brownian motion of dimension 1.

Thirdly, the likelihood ratio test statistic for H_D within M_1 is asymptotically χ^2 . As this statistic is concerned exclusively with the explosive terms normality is required. Some discussion of the normality assumption is given in Nielsen (2010, §4.3). The argument is similar to the proof of Theorem 4 and Corollary 1 in that paper. For the formulation of the result define $\tau_\perp = \Psi_\rho \beta_{\rho\perp}$ and $\tau = (I_2 - \tau_\perp \tau'_\perp) \alpha_\rho$.

Theorem 3.2 *Suppose model M_{1D} with $\rho \geq \varrho$ for some $\varrho > 1$ and Assumptions A, B, C hold, and that $\tau' \varepsilon_t$ is independent $N(0, \tau' \Omega \tau)$ -distributed. Then $LR(M_{1D}|M_1)$ is asymptotically $\chi^2(1)$.*

Finally, the main result of this paper concerns the likelihood ratio test for the joint hypothesis H_S, H_B within the model M_{1D} . The test statistic is asymptotically $\chi^2(2k)$. The hypotheses H_S and H_B link the cointegrating vector, the explosive root, the dynamic adjustment parameters and the intercept in a non-linear fashion and it is not immediately obvious that a χ^2 result should hold, but it actually does under martingale difference assumptions.

Theorem 3.3 *Suppose model M_{1DSB} and Assumptions A, B, C are satisfied. Then the test statistic $LR(M_{1DSB}|M_{1D})$ is asymptotically $\chi^2(2k)$.*

The proof is given in the appendix. It is related to the analysis in Nielsen (2010). Because the model M_{1DSB} is analysed by regression the likelihood function can be analysed in a different way that gives rise to stronger results than in the above paper.

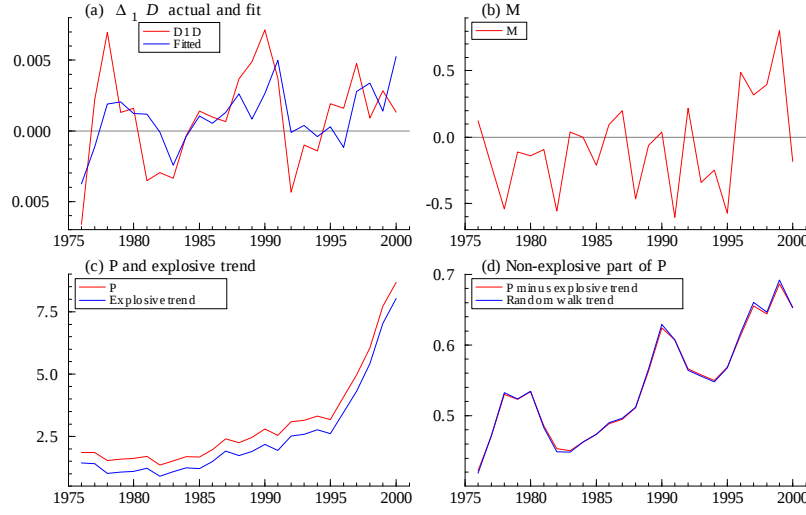


Figure 1: Results from M_{1DSB} applied to sub-sample. (a) shows $\Delta_1 D_t$ and the fit. (b) shows M_t . (c) shows P_t and the estimated explosive trend. (d) shows P_t minus explosive trend along with random walk trend.

Indeed, a global consistency result can be formulated. In the asymptotic expansions the test statistic is found to be the sum of two asymptotically independent terms. The first statistic is mixed Gaussian and relates to the restrictions on the cointegration vectors. The analysis of the second term involves a central limit argument and relates to the restrictions on the adjustment vectors $\alpha_1, \alpha_\rho, \Phi_j$.

The estimators in the model M_{1DSB} are asymptotically normal. The estimator for R converges at an exponential rate $(1 + R)^T$ while the adjustment coefficients are standard $T^{1/2}$ -consistent. Theorem B.11 in the appendix gives a precise statement.

4 Empirical illustration: The US stock market 1974-2000

The proposed method is applied to the annual US stock price and dividend series tabulated by Robert J. Shiller and available at www.robertshiller.com. Here, P_t is the real S&P Composite stock price index at January at year t , and D_t denotes the associated real dividends paid during year $t - 1$. The sub-sample 1974-2000 is considered so as to focus on the recent "bubble period" that ended in 2000 and - according to Shiller (2000) - began to build up from the beginning of the 1980s, shortly after the 1974 collapse. The software OxMetrics by Doornik and Hendry (2001) is used for the analysis.

Figure 1(a, c) shows dividend growth $\Delta_1 D_t$ and prices P_t . An exponential pattern

Test	P_t	D_t	Test	system
$\chi^2_{normality} (2)$	0.6 [0.73]	1.7 [0.42]	$\chi^2_{normality} (4)$	2.0 [0.74]
$F_{ar,1-2} (2, 18)$	0.1 [0.95]	1.4 [0.26]	$F_{ar,1-2} (8, 30)$	1.4 [0.22]
$F_{arch,1-1} (1, 23)$	0.3 [0.58]	0.6 [0.44]		

Table 1: Specification tests for the unrestricted vector autoregression.

Hypothesis	Likelihood	Test statistic	p -value
$r \leq 2$	108.71		
$r \leq 1$	105.91	$LR(r \leq 1 M) = 5.6$	0.23
$r = 0$	98.21	$LR(r = 0 M) = 21.1$	0.04

Table 2: Cointegration rank determination with constant restricted to cointegration space. Critical values based on Johansen (1995, Table 15.2) and Doornik (1998).

is seen for P_t while $\Delta_1 D_t$ appears not to have exponential growth. The initial model is a vector autoregression (2.6) with two lags. The characteristic roots are 1.258, $0.675 \pm 0.354i$, and -0.295 , thus indicating an explosive root in the system.

Table 1 reports specification tests which do not provide evidence against the initial model. The validity of the autocorrelation tests in the explosive context are established in Nielsen (2006a,b). We believe the reported tests for normality and no autoregressive conditional heteroscedasticity are valid, although proofs have not been published. Quantile-Quantile plots and recursive cumulated sums of squares plots, not shown here, do also not provide evidence against the model. The validity of those plots is shown in Engler and Nielsen (2009) and Nielsen and Sohkanen (2011).

Table 2 reports cointegration rank tests. The standard Dickey-Fuller distribution of Johansen (1995, Table 15.2) is used even in the presence of an explosive root due to Theorem 3.1. The tests point to a rank of unity, although the rejection of the hypothesis of no cointegration is marginal. Imposing a unit root changes the largest root slightly to 1.223. The estimated cointegration vector is $\beta'_1 = (1, 29.296)$. As seen, the coefficient to D_t has the 'wrong' sign implying a negative expected return. However, inspection of the likelihood function reveals that it is extremely flat around the optimum; and tests of the hypotheses that β'_1 is $(1, 0)$ or $(0, 1)$ cannot be rejected at the 5% level. Thus, the data are not very informative about the value of the expected return parameter R via the cointegrating vector β_1 .

Table 3 reports the various tests associated with the bubble hypothesis. First, the test for the hypothesis H_D that dividends are non-explosive against M_1 is asymptotically χ^2 due to Theorem 3.2 and gives a p -value of 0.99 while $\hat{\rho} = 1.224$ is nearly unchanged. This confirms the impression from Figure 1 that the explosive root in the

Model	Hypothesis	Likelihood	Test statistic	d.f.	p-value
M_1	$H_1, r = 1$	105.91			
M_{1D}	H_1, H_D	105.91	$LR(H_D M_1) = 0.0002$	1	0.99
M_{1DS}	H_1, H_D, H_S	105.75	$LR(H_S M_{1D}) = 0.32$	1	0.57
			$LR(H_D, H_S M_1) = 0.32$	2	0.85
M_{1DSB}	H_1, H_D, H_S, H_B	102.20	$LR(H_B M_{1DS}) = 7.10$	3	0.07
			$LR(H_S, H_B M_{1D}) = 7.40$	4	0.12
			$LR(H_D, H_S, H_B M_1) = 7.40$	5	0.19

Table 3: Tests of the rational bubble restrictions.

system belongs to P_t and not D_t . Secondly, the test for the joint bubble hypothesis H_S, H_B against M_{1D} is asymptotically χ^2 due to Theorem 3.3 and gives a p-value of 0.12 when tested against M_{1D} . Since the latter test has 4 degrees of freedom and a relatively low p-value it is of interest to evaluate various intermediate hypotheses noting that a χ^2 -distribution theory has not been formally established.

The first of the intermediate hypotheses is the H_S hypothesis that $\beta'_1 = (1, -1/R)$ with $\rho = 1 + R$. The p-value is 0.57 when tested against M_{1D} and $\hat{\rho} = 1.263$, which implies $R = 26.1\%$, clearly not an economically reasonable estimate of the expected annual return. Testing the final hypothesis, H_B , gives a p-value of 0.07 if tested against model M_{1DS} , and a p-value of 0.19 if tested against model M_1 . The decision not to reject the hypothesis H_B is marginal against M_{1DS} albeit more convincing against M_1 . The problem arises from the constant. Imposing only $(1, 1)\alpha_1 = -1$ and $(1, 1)\alpha_\rho = -(1 + R)^2/R$ but leaving the level parameter ζ_1 unconstrained gives a likelihood of 104.65, so a test statistic of 2.20 (p-value: 0.33) against M_{1DS} . Thus the test statistic for M_{1DSB} against this intermediate hypothesis is 4.89 (p-value: 0.027).

Proceeding with the model M_{1DSB} we note that the estimate of the explosive root now becomes $\hat{\rho} = 1.156$ (standard error: 0.023) implying $\hat{R} = 15.6\%$ which is lower than before but still quite high. Here the standard error is computed from the second derivative of the profile likelihood and is valid under the normality assumption which is not rejected. The final estimated model under M_{1DSB} is

$$\begin{aligned}
M_t &= \hat{\varepsilon}_{M,t}, & \text{sdv} &= 0.357, \\
\Delta_1 D_t &= \underset{(0.0015)}{0.0032} \Delta_\rho S_{t-1} + \underset{(0.17)}{0.51} \Delta_1 D_{t-1} - \underset{(0.0017)}{0.0021} M_t + \hat{\varepsilon}_{D \cdot M,t}, & \text{sdv} &= 0.00295.
\end{aligned}$$

Figure 1(a, b) shows graphs of the fitted values of $\Delta_1 D_t$ and M_t . Panel (c) shows the actual prices P_t and the explosive trend for P_t computed using the formula $(\rho - 1)^{-1}(1, 0)C_\rho \rho^t (A_\rho + \sum_{s=1}^t \rho^{-s} \hat{\varepsilon}_s)$ given in Theorem 2.1. This explosive trend contributes with a substantial part of the movements in prices over time consistent with the bubble model. Panel (d) compares the non-explosive part of P_t , that

is the difference between P_t and the explosive trend, with the random walk trend $(1 - \rho)^{-1}(1, 0)C_1(A_1 + \sum_{s=1}^t \hat{\varepsilon}_s)$, see Theorem 2.1.

The conclusion from the analysis is that real stock prices contain an explosive component, and the formal restrictions implied by the rational bubble model cannot be rejected, although the test for the bubble hypothesis itself, H_B against M_{1DS} , is marginal. The estimate of the expected return parameter R is rather high. It is found that the hypotheses that the spread is cointegrating and that P_t has no unit root are both consistent with the data. The data are simply not able to discriminate between these two hypotheses. From an economic point of view (c.f. Shiller, 2000), and looking at Figure 1, it is not unreasonable to consider prices as being bubble-driven with no connection at all to fundamental variables like dividends.

A Proof of Granger-Johansen representation

Proof of Theorem 2.1. *Homogenous equation.* Let $\tilde{X}_t = X_t + \bar{\beta}'_1 \zeta_1 / (1 - \rho)$ so $\Delta_1 X_t = \Delta_1 \tilde{X}_t$ and $\Delta_\rho X_t = \Delta_\rho \tilde{X}_t + \bar{\beta}'_1 \zeta_1$. Since $\mu = \alpha_1 \zeta_1$ then $\mu + \alpha_1 \beta'_1 \bar{\beta}'_1 \zeta_1 = 0$. Insert this in (2.7) to see that $\tilde{X}_t$ solves the

$$\Delta_1 \Delta_\rho \tilde{X}_t = \alpha_1 \beta'_1 \Delta_\rho \tilde{X}_{t-1} + \alpha_\rho \beta'_\rho \Delta_1 \tilde{X}_{t-1} + \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho \tilde{X}_{t-j} + \varepsilon_t. \quad (\text{A.1})$$

This homogenous equation is studied in Theorem 1 of Nielsen (2010) which gives therefore gives the desired representation for X_t apart from the explicit expressions for the terms A_1, A_ρ, Y_t .

Decomposition. To find the expressions for A_1, A_ρ, Y_t decompose

$$\tilde{X}_t = \frac{y \Delta_x \tilde{X}_t - x \Delta_y \tilde{X}_t}{y - x} = \frac{y(\beta_{y\perp} \bar{\beta}'_{y\perp} + \bar{\beta}_y \beta'_y) \Delta_x \tilde{X}_t}{y - x} + \frac{x(\beta_{x\perp} \bar{\beta}'_{x\perp} + \bar{\beta}_x \beta'_x) \Delta_y \tilde{X}_t}{x - y}. \quad (\text{A.2})$$

It will be shown below that

$$\begin{aligned} x \beta_{x\perp} \bar{\beta}'_{x\perp} \Delta_y \tilde{X}_t &= C_x x^t (A_x + \sum_{s=1}^t x^{-s} \varepsilon_s) + C_x \frac{\alpha_y \beta'_y}{y - x} \Delta_x \tilde{X}_t \\ &\quad - C_x \Psi_x \bar{\beta}_x \beta'_x \Delta_y \tilde{X}_t - C_x \sum_{j=1}^{k-2} \Phi_j \sum_{h=0}^{j-1} x^{h-j} \Delta_x \Delta_y \tilde{X}_{t-h}. \end{aligned} \quad (\text{A.3})$$

Inserting (A.3) into (A.2) and keeping track of the terms gives the expressions for A_1, A_ρ, Y_t .

The identity (A.3). Note the identities

$$\begin{aligned} (y - x) \Delta_x \tilde{X}_{s-1} &= \Delta_x^2 \tilde{X}_s - \Delta_x \Delta_y \tilde{X}_s, \\ \sum_{h=0}^{j-1} x^h \Delta_x^2 \Delta_y \tilde{X}_{s-h} &= \Delta_x \Delta_y \tilde{X}_s - x^j \Delta_x \Delta_y \tilde{X}_{s-j}. \end{aligned}$$

Insert these in the homogeneous model equation and pre-multiply by $\alpha'_{x\perp}$ to get

$$\begin{aligned}\alpha'_{x\perp}\Delta_x\Delta_y\tilde{X}_s &= \alpha'_{x\perp}\varepsilon_s + (y-x)^{-1}\alpha'_{x\perp}\alpha_y\beta'_y(\Delta_x^2\tilde{X}_s - \Delta_x\Delta_y\tilde{X}_s) \\ &\quad + \alpha'_{x\perp}\sum_{j=1}^{k-2}x^{-j}\Phi_j(\Delta_x\Delta_y\tilde{X}_s - \sum_{h=0}^{j-1}x^h\Delta_x^2\Delta_y\tilde{X}_{s-h})\end{aligned}$$

Gather $\Delta_x\Delta_y\tilde{X}_s$ -terms and use $\Psi_x = I_p + (y-x)^{-1}\alpha_y\beta'_y - \sum_{j=1}^{k-2}x^{-j}\Phi_j$ to get

$$\alpha'_{x\perp}\Psi_x\Delta_x\Delta_y\tilde{X}_s = \alpha'_{x\perp}\varepsilon_s + \alpha'_{x\perp}\frac{\alpha_y\beta'_y}{y-x}\Delta_x^2\tilde{X}_s - \alpha'_{x\perp}\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}\Delta_x^2\Delta_y\tilde{X}_{s-h}.$$

Multiply with x^{-s} and sum over s and multiply with x^t to get

$$\begin{aligned}\alpha'_{x\perp}\Psi_x\left(\Delta_y\tilde{X}_t - x^t\Delta_y\tilde{X}_0\right) &= \alpha'_{x\perp}\sum_{s=1}^t x^{t-s}\varepsilon_s + \alpha'_{x\perp}\frac{\alpha_y\beta'_y}{y-x}(\Delta_x\tilde{X}_t - x^t\Delta_x\tilde{X}_0) \\ &\quad - \alpha'_{x\perp}\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}(\Delta_x\Delta_y\tilde{X}_{t-h} - x^t\Delta_x\Delta_y\tilde{X}_{-h}).\end{aligned}$$

Pre-multiply $\Delta_y X_t$ with $I_p = \bar{\beta}_x\beta'_x + \beta_{x\perp}\bar{\beta}'_{x\perp}$, recall A_x and rearrange to get

$$\begin{aligned}\alpha'_{x\perp}\Psi_x\beta_{x\perp}\bar{\beta}'_{x\perp}\Delta_y\tilde{X}_t &= \alpha'_{x\perp}x^t(A_x + \sum_{s=1}^t x^{-s}\varepsilon_s) + \alpha'_{x\perp}\frac{\alpha_y\beta'_y}{y-x}\Delta_x\tilde{X}_t \\ &\quad - \alpha'_{x\perp}\Psi_x\bar{\beta}_x\beta'_x\Delta_y\tilde{X}_t - \alpha'_{x\perp}\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}\Delta_x\Delta_y\tilde{X}_{t-h}.\end{aligned}$$

The matrix $\alpha'_{x\perp}\Psi_x\beta_{x\perp}$ is invertible by Assumption A, so pre-multiply with its inverse and then by $x\beta'_{x\perp}$ to get (A.3). ■

B Proof of asymptotic result

The abbreviations *a.s.*, P and D are used for properties holding almost surely, in probability, and in distribution respectively. For a matrix m let $m^{\otimes 2} = mm'$.

B.1 Notation and preliminary asymptotic results

B.1.1 Rotating the data vector

A feature of the vector autoregressive setup is its invariance to linear transformations. In the main discussion of the results the vector $X_t = (P_t, D_t)'$ is analysed. In the

proofs it is convenient to choose X_t in a different way. The issue is that the bubble hypothesis is that

$$M_t = P_t + D_t - (1 + R)P_{t-1} \quad (\text{B.1})$$

is a martingale difference where the contemporaneous component of M_t is $P_t + D_t$. For the proof it is convenient to choose

$$X_t = \begin{pmatrix} P_t + D_t \\ D_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} P_t \\ D_t \end{pmatrix}. \quad (\text{B.2})$$

Accordingly the error term of the model equation is denoted

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{M,t} \\ \varepsilon_{D,t} \end{pmatrix} \quad \text{with} \quad \Omega = \text{Cov}(\varepsilon_t) = \begin{pmatrix} \sigma_{MM} & \sigma_{MD} \\ \sigma_{DM} & \sigma_{DD} \end{pmatrix},$$

and the conditional error $\varepsilon_{D \cdot M,t} = \varepsilon_{M,t} - \omega \varepsilon_{D,t}$ where $\omega = \sigma_{DM} \sigma_{MM}^{-1}$. The spread is

$$S_t = \beta'_1 X_t = P_t - R^{-1} D_t = (P_t + D_t) - \mathcal{G} D_t \quad \text{with} \quad \mathcal{G} = R^{-1} + 1. \quad (\text{B.3})$$

Accordingly the cointegrating and the coexplosive vectors for the rotated system are

$$\beta_1 = \begin{pmatrix} 1 \\ -\mathcal{G} \end{pmatrix}, \quad \beta_{1\perp} = \begin{pmatrix} \mathcal{G} \\ 1 \end{pmatrix}, \quad \beta_\rho = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \beta_{\rho\perp} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \quad (\text{B.4})$$

B.1.2 The data generating process

In the probabilistic analysis the properties of the likelihood function will be analysed for each parameter $(\vartheta^\circ, \Omega_\circ)$ satisfying the restricted model $\mathbf{M}_{1DSB}$. Introduce the vector $\mathcal{S}_t^\circ = (U_t^{\circ'}, V_t^{\circ'}, W_t^{\circ'})'$ where

$$U_t^\circ = \{(\beta_1^{\circ'} \Delta_{\rho_\circ} X_t)', (\beta_\rho^{\circ'} \Delta_1 X_t)', (\Delta_1 \Delta_{\rho_\circ} X_t)', \dots, (\Delta_1 \Delta_{\rho_\circ} X_{t-k+3})'\}', \quad (\text{B.5})$$

$$V_t^\circ = \beta_{1\perp}^{\circ'} \Delta_{\rho_\circ} X_t, \quad W_t^\circ = \beta_{\rho\perp}^{\circ'} \Delta_1 X_t, \quad \mathcal{R}_t^\circ = (M_t^\circ, U_{t-1}^{\circ'})'. \quad (\text{B.6})$$

The data generating process is

$$M_t = \varepsilon_{M,t}^\circ, \quad \Delta_1 D_t = (\omega^\circ, \theta^{\circ'})' \mathcal{R}_t + \varepsilon_{D \cdot M,t}^\circ, \quad (\text{B.7})$$

where $\vartheta^\circ = (\omega^\circ, \theta^{\circ'})'$ and $\theta^{\circ'} = (\alpha_{1,D}^\circ, \alpha_{\rho,D}^\circ, \Phi_{1,D}^{\circ'}, \dots, \Phi_{k-2,D}^{\circ'})$.

B.1.3 Some further parameters

From Nielsen (2010) it is known that the analysis of the unrestricted model $\mathbf{M}_{1D}$ involves the parameters

$$\tau_\perp^\circ = \Psi_\rho^\circ \bar{\beta}_{\rho\perp}^\circ, \quad \Psi_\rho^\circ = I_p + \frac{\Pi_1^\circ}{1 - \rho_\circ} - \sum_{j=1}^{k-2} \rho_\circ^{-j} \Phi_j^\circ, \quad (\text{B.8})$$

as well as the projection matrices

$$\mathcal{P}_{\tau_{\perp}}^{\circ} = \tau_{\perp}^{\circ} (\tau_{\perp}^{\circ\prime} \Omega_{\circ}^{-1} \tau_{\perp}^{\circ})^{-1} \tau_{\perp}^{\circ\prime} \Omega_{\circ}^{-1}, \quad \mathcal{P}_{\alpha}^{\circ} = \alpha_1^{\circ} (\alpha_1^{\circ\prime} \Omega_{\circ}^{-1} \alpha_1^{\circ})^{-1} \alpha_1^{\circ\prime} \Omega_{\circ}^{-1}. \quad (\text{B.9})$$

In the restricted model $\mathbf{M}_{1DSB}$ the restrictions (2.9), (2.10) imply

$$\tau_{\perp}^{\circ} = \begin{pmatrix} \mathcal{G}^{\circ} \\ -R_{\circ}^{-1} \alpha_{1,D}^{\circ} - \sum_{j=1}^{k-2} (1 + R_{\circ})^{-j} \Phi_{j,DP}^{\circ} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\theta^{\circ\prime} \end{pmatrix} \mathcal{H}^{\circ}, \quad (\text{B.10})$$

where the coefficient $\mathcal{G}^{\circ}$ and the vector $\mathcal{H}^{\circ}$ are given by

$$\mathcal{G}^{\circ} = R_{\circ}^{-1} + 1, \quad \mathcal{H}^{\circ} = \{\mathcal{G}^{\circ}, R_{\circ}^{-1}, 0, (1 + R_{\circ})^{-1} \bar{\beta}_{\rho\perp}^{\circ}, \dots, (1 + R_{\circ})^{2-k} \bar{\beta}_{\rho\perp}^{\circ}\}'. \quad (\text{B.11})$$

B.1.4 A preliminary asymptotic result

It is convenient to state a modified version of Lemma A.1 in Nielsen (2010). That result was derived from Chan and Wei (1988) and Nielsen (2005), which in turn is based on Lai and Wei (1985); see also Mynbaev (2011, §6, §8.5). Introduce the block diagonal normalisation matrix $N_{\mathcal{S}} = \text{diag}(I_{\dim U}, N_V, N_W, 1)$ where

$$N_V = \text{diag}\{T^{-1/2} I_{p-r}, (1 - \rho)^{-1}\}, \quad N_W = T^{1/2} \rho_{\circ}^{-T}. \quad (\text{B.12})$$

Lemma B.1 *Let X_t satisfy $\mathbf{M}_{1DSB}$ and be given by (B.7). Assume A, B, C . Let ξ be a constant satisfying $\xi < \gamma/(2 + \gamma)$, recalling γ defined in Assumption B.*

Define sample variances $\widehat{\text{var}}(x_t) = T^{-1} \sum_{t=1}^T x_t x_t'$. Then stochastic matrices Σ_{WW} , Σ_{VV} , Σ_{SS} and a deterministic matrix Σ_{UU} exist so

- (i) $\widehat{\text{var}}(\varepsilon_t^{\circ}) \stackrel{a.s.}{=} \Omega_{\circ} + o(T^{-\xi}) + o(T^{\eta-1/2})$ for all $\eta > 0$.
- (ii) $\widehat{\text{var}}(U_{t-1}^{\circ}) \stackrel{a.s.}{\rightarrow} \Sigma_{UU} > 0$.
- (iii) $\hat{\Sigma}_{WW} = \rho_{\circ}^{-2T} \sum_{t=1}^T (W_{t-1}^{\circ})^2 = \rho_{\circ}^{-2T} T \widehat{\text{var}}(W_{t-1}^{\circ}) \stackrel{a.s.}{\rightarrow} \Sigma_{WW} \stackrel{a.s.}{>} 0$.
- (iv) $\hat{\Sigma}_{VV} = \widehat{\text{var}}(N_V V_{t-1}^{\circ}) \stackrel{D}{\rightarrow} \Sigma_{VV} \stackrel{a.s.}{>} 0$.
- (v) $\widehat{\text{var}}(N_S S_{t-1}^{\circ}) \stackrel{D}{\rightarrow} \Sigma_{SS} \stackrel{a.s.}{>} 0$.

Define sample correlations $\widehat{\text{corr}}(x_t, y_t) = (\sum_{t=1}^T x_t^{\otimes 2})^{-1/2} \sum_{t=1}^T x_t y_t' (\sum_{t=1}^T y_t^{\otimes 2})^{-1/2}$, so

- (vi) $\widehat{\text{corr}}(S_{t-1}^{\circ}, \varepsilon_t^{\circ}) \stackrel{a.s.}{=} o(T^{-\xi/2})$.
- (vii) $\widehat{\text{corr}}\{(U_{t-1}^{\circ\prime}, V_{t-1}^{\circ\prime}, 1)', \varepsilon_t^{\circ}\} = O_P(T^{-1/2})$.

In addition it holds jointly for some stochastic matrices $\Sigma_{V\varepsilon}, \Sigma_{U\varepsilon}, \Sigma_{VU}$ that

- (viii) $\hat{\Sigma}_{U\varepsilon} = T^{-1/2} \sum_{t=1}^T (U_{t-1}^{\circ}) \varepsilon_t^{\circ\prime} \stackrel{D}{\rightarrow} \Sigma_{U\varepsilon}$. and $\hat{\Sigma}_{U\varepsilon} \stackrel{a.s.}{=} o(T^{\eta})$ for all $\eta > 0$.
- (ix) $\hat{\Sigma}_{V\varepsilon} = T^{-1/2} \sum_{t=1}^T N_V (V_{t-1}^{\circ}) \varepsilon_t^{\circ\prime} \stackrel{D}{\rightarrow} \Sigma_{V\varepsilon}$.
- (x) $\hat{\Sigma}_{W\varepsilon} = \rho_{\circ}^{-T} \sum_{t=1}^T W_{t-1}^{\circ} \varepsilon_t^{\circ} \stackrel{a.s.}{=} o(T^{(1-\xi)/2})$.

Proof of Lemma B.1. Similar results are proved in Lemma A.1 in Nielsen (2010) for linear trend model. The difference is largely notational because regression on an intercept is avoided in the present context and $\beta_1^{*\prime} \Delta_{\rho} X_t = \beta_1' \Delta_{\rho} X_t + \zeta_1$ replaces $\beta_1^{*\prime} \Delta_{\rho} X_t = \beta_1' \Delta_{\rho} X_t + \delta_1' (1 - \rho)t$. ■

B.1.5 The test for M_1 and M_{1D} .

Proof of Theorem 3.1. Lemma B.1 only reformulates a subset of the statements of Lemma A.1 in Nielsen (2010). The remaining statements can be reformulated in a similar fashion. This then feeds into Lemma A.2 of Nielsen (2010) which in turn feeds into Lemma 11.1 of Johansen (1995). ■

Proof of Theorem 3.2. The polynomial order of the deterministic term is not crucial in the proof of Theorem 4 in Nielsen (2010). So that proof can be modified along the lines sketched in the proof of Lemma B.1 above. ■

B.2 Some asymptotic results for a given value of R

The proof will involve analysis of product sums involving $\mathcal{R}_t$ defined as

$$\mathcal{R}_t = (M_t, U'_{t-1})', \quad U_t = \{(\beta'_1 \Delta_\rho X_t)', (\beta'_\rho \Delta_1 X_t)', (\Delta_1 \Delta_\rho X_t)', \dots, (\Delta_1 \Delta_\rho X_{t-k+3})'\}'.$$

for some ρ . These product sums are expanded in terms of the estimation error for R . They will be used three times. First, in Lemma B.4 to show that $\hat{R}$ is $T^{-1/2}\rho^T$ -consistent. Secondly, in Lemma B.6 to establish an expansion of the likelihood which will lead to an improved consistency rate for $\hat{R}$. Thirdly, in Theorem B.11 to find the asymptotic distribution of the estimators in which the higher order terms of the expansion can be eliminated. It is convenient to introduce the notation

$$\begin{aligned} \begin{pmatrix} S_{\varepsilon_M \varepsilon_M}^\circ & S_{\varepsilon_M W}^\circ \\ S_{W \varepsilon_M}^\circ & S_{WW}^\circ \end{pmatrix} &= \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} M_t^\circ \\ W_{t-1}^\circ \end{pmatrix}^{\otimes 2}, \\ \begin{pmatrix} S_{DD \cdot M}^\circ & S_{\varepsilon_{D \cdot M} \mathcal{R}}^\circ & S_{\varepsilon_{D \cdot M} W}^\circ \\ S_{\mathcal{R} \varepsilon_{D \cdot M}}^\circ & S_{\mathcal{R} \mathcal{R}}^\circ & S_{\mathcal{R} W}^\circ \end{pmatrix} &= \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varepsilon_{D \cdot M, t}^\circ \\ \mathcal{R}_t^\circ \end{pmatrix} (\varepsilon_{D \cdot M, t}^\circ, \mathcal{R}_t^{\circ'}, W_{t-1}^\circ), \end{aligned}$$

the sum $S_{\Delta\Delta} = T^{-1} \sum_{t=1}^T (\Delta_1 D_t)^2$ as well as the partial product sums

$$\begin{pmatrix} S_{\mathcal{R} \mathcal{R} \cdot \Delta}^\circ & S_{\mathcal{R} W \cdot \Delta}^\circ \\ S_{W \mathcal{R} \cdot \Delta}^\circ & S_{W W \cdot \Delta}^\circ \end{pmatrix} = \begin{pmatrix} S_{\mathcal{R} \mathcal{R}}^\circ & S_{\mathcal{R} W}^\circ \\ S_{W \mathcal{R}}^\circ & S_{W W}^\circ \end{pmatrix} - \begin{pmatrix} S_{\mathcal{R} \Delta}^\circ \\ S_{W \Delta}^\circ \end{pmatrix} S_{\Delta\Delta}^{-1} (S_{\Delta \mathcal{R}}^\circ, S_{\Delta W}^\circ).$$

Lemma B.2 Assume A , B , C . Define $\mathcal{D}_R = (R - R_\circ)$ and $\mathcal{I}_R = (R^{-1} - R_\circ^{-1})$. Let $\mathfrak{o}_{pol} = \mathfrak{o}(T^{-k})$ for some finite k not depending on R . Then

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ \mathcal{R}_t' &\stackrel{a.s.}{=} S_{\varepsilon_{D \cdot M} \mathcal{R}}^\circ - \mathcal{D}_R S_{\varepsilon_{D \cdot M} W}^\circ \mathcal{H}^{\circ'} + (\mathcal{D}_R + \mathcal{I}_R) \mathfrak{o}_{pol}, \\ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t^\circ \mathcal{R}_t' &\stackrel{a.s.}{=} S_{\mathcal{R} \mathcal{R}}^\circ - \mathcal{D}_R S_{\mathcal{R} W}^\circ \mathcal{H}^{\circ'} + (\mathcal{D}_R + \mathcal{I}_R) \mathfrak{o}_{pol}, \\ \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' &\stackrel{a.s.}{=} S_{\mathcal{R} \mathcal{R}}^\circ - \mathcal{D}_R (S_{\mathcal{R} W}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W \mathcal{R}}^\circ) + \mathcal{D}_R^2 \mathcal{H}^\circ S_{W W}^\circ \mathcal{H}^{\circ'} \\ &\quad + (\mathcal{D}_R + \mathcal{I}_R) (1 + \rho_\circ^T \mathcal{D}_R + \mathcal{D}_R + \mathcal{I}_R) \mathfrak{o}_{pol}. \end{aligned}$$

In particular it holds

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ M_t' &\stackrel{a.s.}{=} S_{\varepsilon_{D \cdot M} \varepsilon_M}^\circ - \mathcal{D}_R(\mathcal{G}^\circ S_{\varepsilon_{D \cdot M} W}^\circ + o_{pol}), \\ \frac{1}{T} \sum_{t=1}^T M_t^2 &\stackrel{a.s.}{=} S_{\varepsilon_M \varepsilon_M}^\circ - 2\mathcal{D}_R(\mathcal{G}^\circ S_{\varepsilon_M W}^\circ + o_{pol}) + \mathcal{D}_R^2\{\mathcal{G}^{\circ 2} S_{WW}^\circ + o(1)\} \end{aligned} \quad (\text{B.13})$$

Proof of Lemma B.2. *Identities.* Recall the definition of X in (B.2) and of the cointegrating and the coexplosive vectors in (B.4). Then it holds

$$\begin{pmatrix} \beta_1' \Delta_{1+R} \\ \beta_\rho' \Delta_1 \\ \beta_{1\perp}' \Delta_{1+R} \\ \beta_{\rho\perp}' \Delta_1 \end{pmatrix} X_{t-1} = \begin{pmatrix} 1 & -\frac{1}{R} & -(1+R) & \frac{1+R}{R} \\ 0 & 1 & 0 & -1 \\ 1 & R & -(1+R) & -R(1+R) \\ 1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} P_{t-1} \\ D_{t-1} \\ P_{t-2} \\ D_{t-2} \end{pmatrix}$$

which has the solution

$$\begin{pmatrix} P_{t-1} \\ D_{t-1} \\ P_{t-2} \\ D_{t-2} \end{pmatrix} = \begin{pmatrix} -R & 0 & -1 & 1+R \\ 1 & 1+R & -R & 0 \\ -R & 0 & -1 & 1 \\ 1 & 1 & -R & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{1+R^2} \beta_1' \Delta_{1+R} \\ \frac{1}{R} \beta_\rho' \Delta_1 \\ \frac{1}{R(1+R^2)} \beta_{1\perp}' \Delta_{1+R} \\ \frac{1}{R} \beta_{\rho\perp}' \Delta_1 \end{pmatrix} X_{t-1}. \quad (\text{B.14})$$

It also holds, see Nielsen (2010, equations A.17, A.18), that

$$\begin{aligned} \Delta_1 \Delta_{1+R} X_{t-j} &= \Delta_1 \Delta_{1+R_o} X_{t-j} \\ &+ (R_o - R)(1 + R_o)^{-j} \{ \Delta_1 X_{t-1} - \sum_{l=1}^j (1 + R_o)^{l-1} \Delta_1 \Delta_{1+R_o} X_{t-l} \}. \end{aligned} \quad (\text{B.15})$$

Expansion of $\mathcal{R}$. It is to be derived that $\mathcal{R}_t = (M_t, U_{t-1}')'$, see (B.6), satisfies

$$\mathcal{R}_t = \mathcal{R}_t^\circ - \mathcal{D}_R \mathcal{H}^\circ W_{t-1}^\circ - (\mathcal{D}_R \mathcal{H}_2 + \mathcal{I}_R e_2 \mathcal{H}_3')(U_{t-1}^\circ, V_{t-1}^\circ)',$$

for some matrices $\mathcal{H}_2 \in \mathbb{R}^{(2k-2) \times (2k-1)}$ and $\mathcal{H}_3 \in \mathbb{R}^{2k-1}$ not depending on R and where $e_2 = (0, 1, 0_{1 \times 2(k-2)})'$.

The expression for M_t . Since $M_t = P_t + D_t - (1+R)P_{t-1}$ by (B.1) then $M_t = M_t^\circ - \mathcal{D}_R P_{t-1}$. Due to (B.14) then P_{t-1} is the sum of $(R_o^{-1} + 1)W_{t-1}^\circ$ and some linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$.

The first coordinate of U_{t-1} is $\beta_1' \Delta_{1+R} X_{t-1}$. Using (2.3) write

$$\beta_1' \Delta_{1+R} X_{t-1} = \Delta_{1+R} S_{t-1} = M_{t-1} - (R^{-1} + 1) \Delta_1 D_{t-1}.$$

Writing $M_{t-1} = M_{t-1}^\circ - \mathcal{D}_R P_{t-2}$ and adding and subtracting $R_\circ^{-1} \Delta_1 D_{t-1}$ shows

$$\beta'_1 \Delta_{1+R} X_{t-1} = \beta_1^{\circ'} \Delta_{1+R_\circ} X_{t-1} - \mathcal{D}_R P_{t-2} - \mathcal{I}_R \Delta_1 D_{t-1}.$$

Due to (B.14) then P_{t-2} is the sum of $R_\circ^{-1} W_{t-1}^\circ$ and some linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$, while $\Delta_1 D_{t-1}$ is some other linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$.

The second coordinate of U_{t-1} is $\beta'_\rho \Delta_1 X_{t-1} = \Delta_1 D_{t-1}$ and does not depend on R .

The remaining coordinates of U_{t-1} are of the type $\Delta_1 \Delta_{1+R} X_{t-j}$. These are rewritten using (B.15). Thus, pre-multiplying $\Delta_1 X_{t-1}$ by $I_p = \bar{\beta}_{\rho\perp}^\circ \beta_{\rho\perp}^{\circ'} + \bar{\beta}_\rho^\circ \beta_\rho^{\circ'}$ it is seen that $\Delta_1 \Delta_{1+R} X_{t-j}$ is the sum of $\Delta_1 \Delta_{1+R_\circ} X_{t-j} - \mathcal{D}_R (1+R_\circ)^{-j} \bar{\beta}_{\rho\perp}^\circ \beta_{\rho\perp}^{\circ'} \Delta_1 X_{t-1}$ and some linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$.

Product sums. The first component of interest is

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ \mathcal{R}'_t &\stackrel{a.s.}{=} \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ \mathcal{R}_t^{\circ'} - \mathcal{D}_R \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ W_{t-1}^\circ \mathcal{H}' \\ &\quad - \frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ) (\mathcal{D}_R \mathcal{H}'_2 + \mathcal{I}_R \mathcal{H}_3 e'_2). \end{aligned}$$

The processes $U_{t-1}^\circ, V_{t-1}^\circ$ are of polynomial order, see Nielsen (2005, Theorem 5.1), so $T^{-1} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ)' = o_{pol} \text{ a.s.}$ Note that the first coordinate $\frac{1}{T} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ M_t$ does not have an $\mathcal{I}_R$ component.

By a similar argument then, with $S_1 = T^{-1} \sum_{t=1}^T \mathcal{R}_t^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ)' (\mathcal{D}_R \mathcal{H}'_2 + \mathcal{I}_R \mathcal{H}_3 e'_2)$,

$$\frac{1}{T} \sum_{t=1}^T \mathcal{R}_t^\circ \mathcal{R}'_t = \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t^\circ \mathcal{R}_t^{\circ'} - \mathcal{D}_R \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t^\circ W_{t-1}^\circ \mathcal{H}^{\circ'} - S_1,$$

The processes $\mathcal{R}_t^\circ, U_{t-1}^\circ, V_{t-1}^\circ$ are of polynomial order, see Nielsen (2005, Theorem 5.1) so $\sum_{t=1}^T \mathcal{R}_t^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ)' = o_{pol} \text{ a.s.}$

By a similar argument then, with S_1 as above and

$$S_2 = \mathcal{D}_R \mathcal{H}^\circ \frac{1}{T} \sum_{t=1}^T W_{t-1}^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ) (\mathcal{D}_R \mathcal{H}'_2 + \mathcal{I}_R \mathcal{H}_3 e'_2),$$

$$S_3 = (\mathcal{D}_R \mathcal{H}_2 + \mathcal{I}_R e_2 \mathcal{H}'_3) \frac{1}{T} \sum_{t=1}^T (U_{t-1}^{\circ'}, V_{t-1}^\circ) (U_{t-1}^{\circ'}, V_{t-1}^\circ) (\mathcal{D}_R \mathcal{H}'_2 + \mathcal{I}_R \mathcal{H}_3 e'_2),$$

it holds that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}'_t &= \frac{1}{T} \sum_{t=1}^T \mathcal{R}_t^\circ \mathcal{R}_t^{\circ'} - \mathcal{D}_R \frac{1}{T} \sum_{t=1}^T (\mathcal{R}_t^\circ W_{t-1}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ W_{t-1}^\circ \mathcal{R}_t^{\circ'})' \\ &\quad + \mathcal{D}_R^2 \frac{1}{T} \sum_{t=1}^T \mathcal{H}^\circ (W_{t-1}^\circ)^2 \mathcal{H}^{\circ'} - S_1 - S'_1 + S_2 + S'_2 + S_3. \end{aligned}$$

As argued above $\sum_{t=1}^T \mathcal{R}_t^\circ (U_{t-1}^{\circ'}, V_{t-1}^\circ)'$ and $\sum_{t=1}^T (U_{t-1}^{\circ'}, V_{t-1}^\circ) (U_{t-1}^{\circ'}, V_{t-1}^\circ)$ are of polynomial order while $T^{-1/2} \sum_{t=1}^T \mathcal{R}_t^\circ W_{t-1}^\circ = o(\rho_\circ^T) \text{ a.s.}$ by Lemma B.1(i, v, vi). ■

B.3 Consistency under H_B

The expression for the martingale difference M_t is quadratic in the unknown parameters, so global consistency can be proved under H_B in contrast to the co-explosive analysis in Nielsen (2010). The starting point is the profile log likelihood for the parameter R . Let $\hat{\sigma}_M^2(R)$ and $\hat{\sigma}_{D.M}^2(R)$ denote the residual variances of the regression equations (2.11), (2.12). Then the profile log likelihood is

$$\ell(R) = -\frac{T}{2} \log\{\hat{\sigma}_M^2(R)\hat{\sigma}_{D.M}^2(R)\}. \quad (\text{B.16})$$

The residual variances at R_o satisfy the following results.

Lemma B.3 *Assume A, B, C. Then*

- (i) $\hat{\sigma}_M^2(R_o) = T^{-1} \sum_{t=1}^T M_t^2 = S_{\varepsilon_M \varepsilon_M}^\circ \xrightarrow{a.s.} \sigma_{MM}^\circ$,
- (ii) $\hat{\sigma}_{D.M}^2(R_o) \xrightarrow{a.s.} \sigma_{DD.M}^\circ$.
- (iii) $2\ell(R_o) = -T \log \det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD.M}^{-1} \hat{\Sigma}_{\varepsilon_{D.M}U} \Sigma_{UU}^{-1} \hat{\Sigma}_{U\varepsilon_{D.M}} + o(1)$.

Proof of Lemma B.3. (i) Use $M_t^\circ = \varepsilon_{M,t}^\circ$.

(ii) Note $\hat{\sigma}_{D.M}^2(R_o) = T^{-1} \sum_{t=1}^T (\varepsilon_{D.M,t}^\circ | \mathcal{R}_t^\circ)^2$. Since $\mathcal{R}_t = (\varepsilon_{M,t}^\circ, U_{t-1}^\circ)'$, see (B.6), and $\varepsilon_{D.M,t}^\circ, \varepsilon_{M,t}^\circ, U_{t-1}^\circ$ are asymptotically uncorrelated then Lemma B.1(i) implies

$$\hat{\sigma}_{D.M}^2(R_o) \stackrel{a.s.}{=} S_{\varepsilon_{D.M} \varepsilon_{D.M}}^\circ - \{S_{\varepsilon_{D.M} \varepsilon_M}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M \varepsilon_{D.M}}^\circ + S_{\varepsilon_{D.M} U}^\circ S_{UU}^{\circ-1} S_{U \varepsilon_{D.M}}^\circ\} \{1 + o(T^{-1/4})\}.$$

Lemma B.1(ii,vii) shows $S_{\varepsilon_{D.M} \varepsilon_M}^\circ, S_{\varepsilon_{D.M} U}^\circ = o(T^{-3/8})$ while $S_{\varepsilon_M \varepsilon_M}^{\circ-1}, S_{UU}^{\circ-1}$ converge so

$$\hat{\sigma}_{D.M}^2(R_o) \stackrel{a.s.}{=} S_{\varepsilon_{D.M} \varepsilon_{D.M}}^\circ - S_{\varepsilon_{D.M} \varepsilon_M}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M \varepsilon_{D.M}}^\circ - S_{\varepsilon_{D.M} U}^\circ S_{UU}^{\circ-1} S_{U \varepsilon_{D.M}}^\circ + o(T^{-1}), \quad (\text{B.17})$$

and in particular $\hat{\sigma}_{D.M}^2(R_o) \rightarrow \sigma_{DD.M}^\circ$ as desired.

(iii) Apply the expansion $\log(1+h) = h + O(h^2)$ to (B.17) keeping the first two terms as the main term and noting $S_{\varepsilon_{D.M} \varepsilon_M}^\circ, S_{\varepsilon_{D.M} U}^\circ = o(T^{-3/8})$ to get

$$\begin{aligned} -T \log\{\hat{\sigma}_{D.M}^2(R_o)\} &\stackrel{a.s.}{=} -T \log(S_{\varepsilon_{D.M} \varepsilon_{D.M}}^\circ - S_{\varepsilon_{D.M} \varepsilon_M}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M \varepsilon_{D.M}}^\circ) \\ &\quad + T S_{\varepsilon_{D.M} \varepsilon_{D.M}}^{\circ-1} S_{\varepsilon_{D.M} U}^\circ S_{UU}^{\circ-1} S_{U \varepsilon_{D.M}}^\circ + o(1). \end{aligned}$$

Insert this and $-T \log\{\hat{\sigma}_M^2(R_o)\} = -T \log S_{\varepsilon_M \varepsilon_M}^\circ$ into (B.16) to get

$$2\ell(R_o) = -T \log(S_{\varepsilon_M \varepsilon_M}^\circ S_{\varepsilon_{D.M} \varepsilon_{D.M}}^\circ - S_{\varepsilon_{D.M} \varepsilon_M}^{\circ 2}) + T S_{\varepsilon_{D.M} \varepsilon_{D.M}}^{\circ-1} S_{\varepsilon_{D.M} U}^\circ S_{UU}^{\circ-1} S_{U \varepsilon_{D.M}}^\circ + o(1).$$

Due to the identity

$$\det(S_{\varepsilon\varepsilon}^\circ) = \det\left\{\frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varepsilon_{M,t}^\circ \\ \varepsilon_{D,t}^\circ \end{pmatrix} \right\}^{\otimes 2} = \det\left\{\frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varepsilon_{D.M,t}^\circ \\ \varepsilon_{D,t}^\circ \end{pmatrix} \right\}^{\otimes 2}$$

the first term of $2\ell(R_o)$ is $-T \log \det(S_{\varepsilon\varepsilon}^\circ)$. For the second term note $S_{UU}^\circ \rightarrow \Sigma_{UU}^\circ$ and $S_{\varepsilon_{D.M} \varepsilon_{D.M}}^\circ \rightarrow \sigma_{DD.M}^\circ$ while $T^{1/2} S_{\varepsilon_{D.M} U}^\circ = \hat{\Sigma}_{\varepsilon_{D.M} U}$. ■

Lemma B.4 Consider the maximum likelihood estimators under $\mathbf{H}_B$. Assume A , B , C . Then $\hat{R} - R = o(T^{1/2}\rho^{-T})$ a.s.

Proof of Lemma B.4. Let R_o denote the true value of R .

Likelihood value at R_o . Lemma B.3 shows $\hat{\sigma}_M^2(R_o) = T^{-1} \sum_{t=1}^T \varepsilon_{M,t}^{\circ 2}$ and $\hat{\sigma}_{D.M}^2(R_o) = T^{-1} \sum_{t=1}^T (\varepsilon_{D.M,t}^{\circ} | \mathcal{R}_t^{\circ})^2$. Lemma B.1(i,vi) then implies

$$\hat{\sigma}_M^2(R_o) \xrightarrow{a.s.} \sigma_M^{\circ 2}, \quad \hat{\sigma}_{D.M}^2(R_o) \xrightarrow{a.s.} \sigma_{D.M}^{\circ 2}.$$

Likelihood value outside neighbourhood of R_o . For any $\delta > 0$ consider an R so $|N_W^{-1}(R - R_o)| > \delta$. It is to be shown that

$$\liminf_{T \rightarrow \infty} \hat{\sigma}_M^2(R_o) \stackrel{a.s.}{=} \sigma_M^{\circ 2} + \delta^2 \kappa_P, \quad \liminf_{T \rightarrow \infty} \hat{\sigma}_{D.M}^2(R_o) \stackrel{a.s.}{\geq} \sigma_{D.M}^{\circ 2},$$

where $\kappa_P = \mathcal{G}^{\circ 2} \lim_{T \rightarrow \infty} \rho_o^{-2T} \sum_{t=1}^T (W_{t-1}^{\circ})^2 > 0$ a.s. for $\rho_o = 1 + R_o$.

For the first result note that by (B.13) then

$$\hat{\sigma}_M^2(R) \stackrel{a.s.}{=} S_{MM}^{\circ} + 2(R_o - R)\mathcal{G}^{\circ} S_{MW}^{\circ} + (R_o - R)^2 \mathcal{G}^{\circ 2} S_{WW}^{\circ} + O(T^{-1}). \quad (\text{B.18})$$

Due to Lemma B.1(i,iii,vi) then

$$\hat{\sigma}_M^2(R) \stackrel{a.s.}{=} \sigma_M^{\circ 2} + T^{-1} \rho_o^{2T} (R_o - R)^2 \kappa_P + 2T^{-1/2} \rho_o^T (R_o - R) o(1) + o(1).$$

Thus, for any R outside a neighbourhood of R_o this has the stated limes inferior.

For the second result write the residual variance as

$$\hat{\sigma}_{D.M}^2(R) = \frac{1}{T} \sum_{t=1}^T (\Delta_1 D_t | \mathcal{R}_t)^2.$$

As $M_t = M_t^{\circ} + (R_o - R)P_{t-1}$ the regressor $\mathcal{R}_t$ is a linear combination of $M_t^{\circ}, \mathcal{S}_{t-1}^{\circ}$. Moreover, $\Delta_1 D_t = \theta_o' \mathcal{R}_t^{\circ} + \varepsilon_{D.M,t}^{\circ}$ for some θ_o while $\mathcal{R}_t^{\circ}$ is a linear combination of $M_t^{\circ}, \mathcal{S}_{t-1}^{\circ}$ and $M_t^{\circ} = \varepsilon_{M,t}^{\circ}$. Thus

$$\hat{\sigma}_{D.M}^2(R) \geq \frac{1}{T} \sum_{t=1}^T (\Delta_1 D_t | M_t^{\circ}, \mathcal{S}_{t-1}^{\circ})^2 = \frac{1}{T} \sum_{t=1}^T (\varepsilon_{D.M,t}^{\circ} | \varepsilon_{M,t}^{\circ}, \mathcal{S}_{t-1}^{\circ})^2.$$

Since the sample correlations of $\varepsilon_{D.M,t}^{\circ}$, $\varepsilon_{M,t}^{\circ}$ and $\mathcal{S}_{t-1}^{\circ}$ vanish asymptotically then $\hat{\sigma}_{D.M}^2(R)$ has the stated limes inferior.

Continuity of likelihood function. The profile log likelihood $\ell(R)$ is continuous and will, asymptotically, attain its minimum in a compact interval $|N_W^{-1}(R - R_o)| \leq \delta$ as it is large outside the interval. This shows the desired consistency. ■

B.4 Expanding likelihood under H_B

The profile likelihood for R is analysed. The first Lemma expands log determinants.

Lemma B.5 $\log \det(I + h) = \text{tr}(h) - \frac{1}{2}\text{tr}(h^2) + \frac{1}{3}\text{tr}(h^3) + O(\|h\|^4)$.

Proof of Lemma B.5. Any matrix h can be decomposed as $h = ABA^{-1}$ where B is a triangular, possibly complex matrix with diagonal elements λ_j , see Mirsky (1961, p. 266, 307). Thus, $I + h = A(I + B)A^{-1}$ and $\det(I + h) = \det(I + B) = \prod_{j=1}^{\dim h} (1 + \lambda_j)$. By the expansion $\log(1 + x) = x - x^2/2 + x^3/6 + O(x^4)$ it holds

$$\log \det(I + J) = \sum_{j=1}^{\dim h} \log(1 + \lambda_j) = \sum_{j=1}^{\dim h} \left\{ \lambda_j - \frac{1}{2}\lambda_j^2 + \frac{1}{3}\lambda_j^3 + O(\lambda_j^4) \right\}.$$

Noting that $\text{tr}(B^k) = \sum_{j=1}^{\dim h} \lambda_j^k$ and $\text{tr}(A^k) = \text{tr}(B^k)$ the desired result follows. ■

The next step is to write the profile likelihood in terms quadratic functions in R .

Lemma B.6 Assume A, B, C . Under H_B the profile likelihood has expansion

$$2\{\ell(R) - \ell(R_\circ)\} \stackrel{a.s.}{=} 2\{\tilde{\ell}(R) - \tilde{\ell}(R_\circ)\} + o(1)$$

for $|R - R_\circ| \leq cT^{1/2}\rho_\circ^{-T}$ for any $c > 0$. Here $\tilde{\ell}(R) = \tilde{\ell}_M(R) + \tilde{\ell}_{\mathcal{R} \cdot \Delta}(R) - \tilde{\ell}_{\mathcal{R}}(R)$ with

$$\begin{aligned} \tilde{\ell}_M(R) &= -\frac{T}{2} \log(S_{\varepsilon_M \varepsilon_M}^\circ - 2\mathcal{D}_R \mathcal{G}^\circ S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2 \mathcal{G}^{\circ 2} S_{WW}^\circ), \\ \tilde{\ell}_{\mathcal{R} \cdot \Delta}(R) &= -\frac{T}{2} \log \det\{S_{\mathcal{R} \cdot \Delta}^\circ - \mathcal{D}_R(S_{\mathcal{R} W \cdot \Delta}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W \mathcal{R} \cdot \Delta}^\circ) + \mathcal{D}_R^2 \mathcal{H}^\circ S_{WW \cdot \Delta}^\circ \mathcal{H}^{\circ'}\}, \\ \tilde{\ell}_{\mathcal{R}}(R) &= -\frac{T}{2} \log \det\{S_{\mathcal{R} \mathcal{R}}^\circ - \mathcal{D}_R(S_{\mathcal{R} W}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W \mathcal{R}}^\circ) + \mathcal{D}_R^2 \mathcal{H}^\circ S_{WW}^\circ \mathcal{H}^{\circ'}\}. \end{aligned}$$

Proof of Lemma B.6. *Profile likelihood.* This is given by

$$2\ell(R) = -T \log\{\hat{\sigma}_M^2(R)\} - T \log\{\hat{\sigma}_{D \cdot M}^2(R)\}. \quad (\text{B.19})$$

It will be shown that this is quadratic in R up to an approximation.

Component involving $\hat{\sigma}_M^2(R)$. Since $\hat{\sigma}_M^2(R) = T^{-1} \sum_{t=1}^T M_t^2$ consider the expansion (B.13). Since $\mathcal{D}_R = o(T^{1/2}\rho_\circ^{-T})$ then

$$\hat{\sigma}_M^2(R) \stackrel{a.s.}{=} S_{\varepsilon_M \varepsilon_M}^\circ - 2\mathcal{D}_R \mathcal{G}^\circ S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2 \mathcal{G}^{\circ 2} S_{WW}^\circ + o(T^{-1}).$$

The expansion $\log(1 + h) = O(h)$ shows

$$\log\{\hat{\sigma}_M^2(R)\} \stackrel{a.s.}{=} \log(S_{\varepsilon_M \varepsilon_M}^\circ - 2\mathcal{D}_R \mathcal{G}^\circ S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2 \mathcal{G}^{\circ 2} S_{WW}^\circ) + o(T^{-1}). \quad (\text{B.20})$$

Component involving $\hat{\sigma}_{D.M}^2(R)$. First, by the rule for taking determinants of partitioned matrices

$$\begin{aligned} \log\{\hat{\sigma}_{D.M}^2(R)\} &= \log \det\{T^{-1} \sum_{t=1}^T (\mathcal{R}_t | \Delta_1 D_t)^{\otimes 2}\} \\ &\quad - \log \det\{T^{-1} \sum_{t=1}^T (\mathcal{R}_t)^{\otimes 2}\} + \log\{T^{-1} \sum_{t=1}^T (\Delta_1 D_t)^2\}. \end{aligned} \quad (\text{B.21})$$

The last term does not depend on R .

For the second term of (B.21) apply Lemma B.2 to get

$$\frac{1}{T} \sum_{t=1}^T \mathcal{R}_t \mathcal{R}_t' \stackrel{a.s.}{=} S_{\mathcal{R}\mathcal{R}}^\circ - \mathcal{D}_R(S_{\mathcal{R}W}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W\mathcal{R}}^\circ) + \mathcal{D}_R^2 \mathcal{H}^\circ S_{WW}^\circ \mathcal{H}^{\circ'} + o(T^{-1}).$$

Applying the log determinant expansion in Lemma B.5 it follows that

$$\log \det\{T^{-1} \sum_{t=1}^T (\mathcal{R}_t)^{\otimes 2}\} \stackrel{a.s.}{=} \tilde{\ell}_R(R) + o(T^{-1}). \quad (\text{B.22})$$

Apply a similar argument for the first term of (B.21) to get

$$\log \det\{T^{-1} \sum_{t=1}^T (\mathcal{R}_t | \Delta_1 D_t)^{\otimes 2}\} \stackrel{a.s.}{=} \tilde{\ell}_{R,\Delta}(R) + o(T^{-1}). \quad (\text{B.23})$$

Profile likelihood expansion. Insert (B.22) and (B.23) into the expression for $\log(\hat{\sigma}_{D.M}^2)$ in (B.21) and in turn insert this and the expression (B.20) for $\log(\hat{\sigma}_M^2)$ into the profile likelihood (B.19) to get

$$2\ell(R) \stackrel{a.s.}{=} 2\tilde{\ell}(R) - T \log(S_{\Delta\Delta}) + o(1).$$

Finally note that $\ell_R(R_\circ) = \tilde{\ell}_R(R_\circ) - T \log(S_{\Delta\Delta})$. ■

The derivatives of the approximation $\tilde{\ell}$ to the profile likelihood are considered.

Lemma B.7 *Assume A, B, C . Under $\mathbf{H}_B$ then*

$$\begin{aligned} \tilde{\ell}'(R_\circ) &= \rho_\circ^T \{\mathcal{G}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} \hat{\Sigma}_{\varepsilon_M W} + \mathcal{H}^{\circ'} (S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W \cdot \Delta} - S_{\mathcal{R}\mathcal{R}}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W})\}, \\ \tilde{\ell}''(R_\circ) &\stackrel{a.s.}{=} -\rho_\circ^{2T} \Sigma_{WW}^\circ \{\mathcal{G}^{\circ 2} S_{\varepsilon_M \varepsilon_M}^{\circ-1} + \mathcal{H}^{\circ'} (S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ-1} - S_{\mathcal{R}\mathcal{R}}^{\circ-1}) \mathcal{H}^\circ\} \{1 + o(1)\}, \end{aligned}$$

where $\hat{\Sigma}_{\mathcal{R}W} = \sum_{t=1}^T \mathcal{R}_t^\circ W_{t-1}^\circ$. It holds that

$$\tilde{\ell}'(R_\circ) \stackrel{a.s.}{=} o(T^{1/4} \rho_\circ^T), \quad \{\tilde{\ell}''(R_\circ)\}^{-1} = O(\rho_\circ^{-2T}), \quad \tilde{\ell}'''(R_\circ) = o(T^{-3/4} \rho_\circ^{3T}).$$

Proof of Lemma B.7. *Term $\tilde{\ell}_M(R)$.* This satisfies

$$-(2/T)\tilde{\ell}_M(R) = \log(S_{\varepsilon_M \varepsilon_M}^\circ) + \log(1 + h).$$

where $h = -2\mathcal{D}_R \mathcal{G}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2 \mathcal{G}^{\circ 2} S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{WW}^\circ$. Apply the log expansion in Lemma B.5. Rearrange to get an expansion in $\mathcal{D}_R$ which is

$$\begin{aligned} \log(1 + h) &= -2\mathcal{D}_R \mathcal{G}^\circ S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2 \mathcal{G}^{\circ 2} (S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{WW}^\circ - 2S_{\varepsilon_M \varepsilon_M}^{\circ-2} S_{\varepsilon_M W}^{\circ 2}) \\ &\quad + \mathcal{D}_R^3 \mathcal{G}^{\circ 3} (2S_{\varepsilon_M \varepsilon_M}^{\circ-2} S_{\varepsilon_M W}^\circ S_{WW}^\circ - \frac{8}{3} S_{\varepsilon_M \varepsilon_M}^{\circ-3} S_{\varepsilon_M W}^{\circ 3}) + O(\mathcal{D}_R^4) \end{aligned}$$

Hence the coefficient to $\mathcal{D}_R$ gives the first derivative $\tilde{\ell}'_M(R_o) = T\mathcal{G} S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{\varepsilon_M W}^\circ$. Replacing $T S_{\varepsilon_M W}^\circ = T^{1/2}(\rho_o^{-T} \sum_{t=1}^T \varepsilon_{M,t}^\circ W_{t-1}^\circ)(T^{-1/2} \rho_o^T) = T^{1/2} \hat{\Sigma}_{\varepsilon_M W} N_W^{-1}$ gives

$$\tilde{\ell}'_M(R_o) = T^{1/2} \mathcal{G} S_{\varepsilon_M \varepsilon_M}^{\circ-1} \hat{\Sigma}_{\varepsilon_M W} N_W^{-1}.$$

Likewise the second and third derivatives are

$$\begin{aligned} \tilde{\ell}''_M(R_o) &= (2!)(-T/2) \mathcal{G}^{\circ 2} (S_{\varepsilon_M \varepsilon_M}^{\circ-1} S_{WW}^\circ - 2S_{\varepsilon_M \varepsilon_M}^{\circ-2} S_{\varepsilon_M W}^{\circ 2}) \\ \tilde{\ell}'''_M(R_o) &= (3!)(-T/2) \mathcal{G}^{\circ 3} (2S_{\varepsilon_M \varepsilon_M}^{\circ-2} S_{\varepsilon_M W}^\circ S_{WW}^\circ - \frac{8}{3} S_{\varepsilon_M \varepsilon_M}^{\circ-3} S_{\varepsilon_M W}^{\circ 3}) \end{aligned}$$

Noting that $S_{\varepsilon_M \varepsilon_M}^{\circ-1}, S_{WW} N_W^2 = \hat{\Sigma}_{WW}$ are convergent while $S_{\varepsilon_M W}^\circ N_W = o(T^{-1/4})$ then

$$\tilde{\ell}''_M(R_o) = -T \mathcal{G}^{\circ 2} S_{\varepsilon_M \varepsilon_M}^{\circ-1} \Sigma_{WW} N_W^{-2} \{1 + o(1)\}, \quad \tilde{\ell}'''_M(R_o) \stackrel{a.s.}{=} o(T^{-3/4} \rho_o^{3T}).$$

Term $\tilde{\ell}_{\mathcal{R}}(R)$. This satisfies

$$-(2/T)\tilde{\ell}_{\mathcal{R}}(R) = \log \det(S_{\mathcal{R}\mathcal{R}}) + \log \det(I_{\dim \mathcal{R}} + h)$$

where $h = -\mathcal{D}_R S_{\mathcal{R}\mathcal{R}}^{\circ-1} (S_{\mathcal{R}W}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W\mathcal{R}}^\circ) + \mathcal{D}_R^2 S_{\mathcal{R}\mathcal{R}}^{\circ-1} \mathcal{H}^\circ S_{WW}^\circ \mathcal{H}^{\circ'}$. Apply the log expansion in Lemma B.5. Rearrange to get an expansion in $\mathcal{D}_R$ which is

$$\begin{aligned} \log \det(I + h) &= -2\mathcal{D}_R \text{tr}\{S_{\mathcal{R}\mathcal{R}}^{\circ-1} S_{\mathcal{R}W}^\circ \mathcal{H}^{\circ'}\} + \mathcal{D}_R^2 \{\text{tr}(S_{\mathcal{R}\mathcal{R}}^{\circ-1} \mathcal{H}^\circ S_{WW}^\circ \mathcal{H}^{\circ'}) - \text{tr}(\mathcal{B}^2)/2\} \\ &\quad + \mathcal{D}_R^3 (\mathcal{B} S_{\mathcal{R}\mathcal{R}}^{\circ-1} \mathcal{H}^\circ S_{WW}^\circ \mathcal{H}^{\circ'}/2 - \mathcal{B}^3/3) + O(\mathcal{D}_R^4), \end{aligned}$$

where $\mathcal{B} = S_{\mathcal{R}\mathcal{R}}^{\circ-1} (S_{\mathcal{R}W}^\circ \mathcal{H}^{\circ'} + \mathcal{H}^\circ S_{W\mathcal{R}}^\circ)$. By considerations as above it is seen that $\mathcal{B} = o(T^{-3/4} \rho_o^T)$ and the derivatives satisfy

$$\begin{aligned} \tilde{\ell}'_{\mathcal{R}}(R_o) &= T^{1/2} \text{tr}\{\mathcal{H}^{\circ'} S_{\mathcal{R}\mathcal{R}}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W}\} N_W^{-1}, \quad \tilde{\ell}'''_{\mathcal{R}}(R_o) \stackrel{a.s.}{=} o(T^{-3/4} \rho_o^{3T}), \\ \tilde{\ell}''_{\mathcal{R}}(R_o) &\stackrel{a.s.}{=} -T \text{tr}(\mathcal{H}^{\circ'} S_{\mathcal{R}\mathcal{R}}^{\circ-1} \mathcal{H}^\circ) \Sigma_{WW} N_W^{-2} \{1 + o(1)\}. \end{aligned}$$

Term $\tilde{\ell}_{\mathcal{R} \cdot \Delta}(R)$. Same derivation as for $\tilde{\ell}_{\mathcal{R}}(R)$ replacing $S_{\mathcal{R}\mathcal{R}}, S_{\mathcal{R}W}$ and S_{WW} by $S_{\mathcal{R}\mathcal{R} \cdot \Delta}, S_{\mathcal{R}W \cdot \Delta}$ and $S_{WW \cdot \Delta} = S_{WW} \{1 + o(1)\}$. ■

The expressions for the $\tilde{\ell}'$ and $\tilde{\ell}''$ are simplified using the parameter $\tau_\perp^\circ$ from (B.10).

Lemma B.8 Assume A, B, C . Under H_B then

$$\begin{aligned}\tilde{\ell}'(R_o) &\stackrel{a.s.}{=} \rho_o^T \{\tau_\perp^o \Omega_o^{-1} \hat{\Sigma}_{\varepsilon W} + o(T^{-1/4})\} \\ -\tilde{\ell}''(R_o) &\stackrel{a.s.}{=} \rho_o^{2T} \tau_\perp^o \Omega_o^{-1} \tau_\perp^o \Sigma_{WW}^o \{1 + o(1)\}.\end{aligned}$$

Proof of Lemma B.8. *Product moment matrices.* Recall from (B.7) that $\Delta_1 D_t = (\omega^\circ, \theta^\circ) \mathcal{R}_t^\circ + \varepsilon_{D \cdot M, t}^\circ$ and note $\omega^\circ = \sigma_{MM}^{\circ-1} \sigma_{DM}^\circ$. It holds, for all $\eta > 0$, see Lemma B.1(i,vii),

$$S_{\mathcal{R}\mathcal{R}}^\circ \xrightarrow{a.s.} \begin{pmatrix} \sigma_{MM}^\circ & 0 \\ 0 & \Sigma_{UU}^\circ \end{pmatrix}, \quad (B.24)$$

$$S_{\mathcal{R}\Delta}^\circ \stackrel{a.s.}{=} \begin{pmatrix} \sigma_{MM}^\circ \omega^\circ \\ \Sigma_{UU}^\circ \theta^\circ \end{pmatrix} = \begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ \theta^\circ \end{pmatrix} + o(T^{\eta-1/2}). \quad (B.25)$$

Since $\Delta_1 D_t$ also satisfies $\Delta_1 D_t = \theta^{\circ'} \mathcal{R}_t^\circ + \varepsilon_{D, t}^\circ$ then

$$S_{\Delta\Delta}^\circ \stackrel{a.s.}{=} \sigma_{DD}^\circ + \theta^{\circ'} \Sigma_{UU}^\circ \theta^\circ + o(T^{\eta-1/2}). \quad (B.26)$$

Moreover, exploiting $\Delta_1 D_t = (\omega^\circ, \theta^\circ) \mathcal{R}_t^\circ + \varepsilon_{D \cdot M, t}^\circ$ and $\varepsilon_{D, t}^\circ = \varepsilon_{D \cdot M, t}^\circ + \omega^\circ \varepsilon_{M, t}^\circ$ it holds

$$TS_{\mathcal{R}W} = \sum_{t=1}^T \begin{pmatrix} \varepsilon_{M, t}^\circ \\ U_{t-1}^\circ \end{pmatrix} W_{t-1}^\circ, \quad TS_{\Delta W} = \hat{\Sigma}_{\Delta W} = (1, \theta') \sum_{t=1}^T \begin{pmatrix} \varepsilon_{D, t}^\circ \\ U_{t-1}^\circ \end{pmatrix} W_{t-1}^\circ. \quad (B.27)$$

Information. Combine the expressions (B.24), (B.25), (B.26) to get

$$S_{\mathcal{R}\mathcal{R} \cdot \Delta}^\circ = S_{\mathcal{R}\mathcal{R}}^\circ - S_{\mathcal{R}\Delta}^\circ S_{\Delta\Delta}^{-1} S_{\Delta\mathcal{R}}^\circ \xrightarrow{a.s.} \begin{pmatrix} \sigma_{MM}^\circ & 0 \\ 0 & \Sigma_{UU}^\circ \end{pmatrix} - \begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ \theta^\circ \end{pmatrix}^{\otimes 2} \frac{1}{\sigma_{DD}^\circ + \theta^{\circ'} \Sigma_{UU}^\circ \theta^\circ}.$$

The partitioned inversion formula $A_{11 \cdot 2}^{-1} = A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22 \cdot 1}^{-1} A_{21} A_{11}^{-1}$ shows, noting that $\omega^\circ = \sigma_{MM}^{\circ-1} \sigma_{DM}^\circ$,

$$\begin{aligned}& S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ-1} - S_{\mathcal{R}\mathcal{R}}^{\circ-1} \\ & \stackrel{a.s.}{=} \frac{\begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UY}^{\circ-1} \end{pmatrix} \begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ \theta^\circ \end{pmatrix} (\sigma_{DM}^\circ, \theta^{\circ'} \Sigma_{UU}^\circ) \begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UU}^{\circ-1} \end{pmatrix}}{\sigma_{DD}^\circ + \theta^{\circ'} \Sigma_{UU}^\circ \theta^\circ - (\sigma_{DM}^\circ, \theta^{\circ'} \Sigma_{UU}^\circ) \begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UU}^{\circ-1} \end{pmatrix} \begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ \theta^\circ \end{pmatrix}} + o(T^{-1/4}) \\ & = \frac{1}{\sigma_{DD \cdot M}^\circ} \begin{pmatrix} \omega^\circ \\ \theta^\circ \end{pmatrix}^{\otimes 2} + o(T^{-1/4}). \quad (B.28)\end{aligned}$$

Further, note that $S_{\varepsilon_M \varepsilon_M}^\circ \xrightarrow{a.s.} \sigma_{MM}^\circ$ while the definition of $\tau_\perp^\circ$ in (B.10) implies

$$(\omega^\circ, \theta^{\circ'}) \mathcal{H} = (\omega^\circ, -1) \tau_\perp^\circ, \quad \mathcal{G}^\circ = (1, 0) \tau_\perp^\circ. \quad (B.29)$$

Combining these expressions shows

$$\mathcal{G}^{\circ 2} S_{\varepsilon_M \varepsilon_M}^{\circ -1} + \mathcal{H}'(S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{-1} - S_{\mathcal{R}\mathcal{R}}^{-1}) \mathcal{H} \xrightarrow{a.s.} \tau_{\perp}^{\circ'} \{ \sigma_{MM}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes 2} + \sigma_{DD \cdot M}^{-1} \begin{pmatrix} \omega^{\circ} \\ -1 \end{pmatrix}^{\otimes 2} \} \tau_{\perp}^{\circ}.$$

Finally, the desired expression follows since by partitioned inversion

$$\Omega_{\circ}^{-1} = \sigma_{MM}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes 2} + \sigma_{DD \cdot M}^{-1} \begin{pmatrix} \omega^{\circ} \\ -1 \end{pmatrix}^{\otimes 2}. \quad (\text{B.30})$$

Score. Combine (B.25), (B.26), (B.27) to get

$$\begin{aligned} \hat{\Sigma}_{\mathcal{R}W \cdot \Delta} &= \hat{\Sigma}_{\mathcal{R}W} - S_{\mathcal{R}\Delta}^{\circ} S_{\Delta\Delta}^{-1} \hat{\Sigma}_{\Delta W} \\ &\stackrel{a.s.}{=} \hat{\Sigma}_{\mathcal{R}W} - \frac{1 + o(T^{\eta-1/2})}{\sigma_{DD}^{\circ} + \theta^{\circ'} \Sigma_{UU}^{\circ} \theta^{\circ}} \begin{pmatrix} \sigma_{DM}^{\circ} \\ \Sigma_{UU}^{\circ} \theta^{\circ} \end{pmatrix} (1, \theta^{\circ'}) \hat{\Sigma}_{\Delta W}. \end{aligned}$$

In a similar way write

$$\begin{aligned} S_{\mathcal{R}\mathcal{R}}^{\circ -1} \hat{\Sigma}_{\mathcal{R}W} &= S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ -1} S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ} S_{\mathcal{R}\mathcal{R}}^{\circ -1} \hat{\Sigma}_{\mathcal{R}W} \\ &\stackrel{a.s.}{=} S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ -1} \{ I_2 - \begin{pmatrix} \sigma_{DM}^{\circ} \\ \Sigma_{UU}^{\circ} \theta^{\circ} \end{pmatrix} \frac{1 + o(T^{\eta-1/2})}{\sigma_{DD}^{\circ} + \theta^{\circ'} \Sigma_{UU}^{\circ} \theta^{\circ}} \} (\omega^{\circ}, \theta^{\circ'}) \hat{\Sigma}_{\mathcal{R}W}. \end{aligned}$$

These expression combine as

$$\begin{aligned} &S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ -1} \hat{\Sigma}_{\mathcal{R}W \cdot \Delta} - S_{\mathcal{R}\mathcal{R}}^{\circ -1} \hat{\Sigma}_{\mathcal{R}W} \\ &\stackrel{a.s.}{=} S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ -1} \frac{\{1 + o(T^{\eta-1/2})\}}{\sigma_{DD}^{\circ} + \theta^{\circ'} \Sigma_{UU}^{\circ} \theta^{\circ}} \begin{pmatrix} \sigma_{DM}^{\circ} \\ \Sigma_{UU}^{\circ} \theta^{\circ} \end{pmatrix} \{ (\omega^{\circ}, \theta^{\circ'}) \hat{\Sigma}_{\mathcal{R}W} - (1, \theta^{\circ'}) \hat{\Sigma}_{\Delta W} \}. \end{aligned} \quad (\text{B.31})$$

Noting that, see (B.27),

$$\hat{\Sigma}_{\mathcal{R}W} = \begin{pmatrix} (1, 0) \hat{\Sigma}_{\varepsilon W} \\ \hat{\Sigma}_{UW} \end{pmatrix}, \quad \hat{\Sigma}_{\Delta W} = \begin{pmatrix} (0, 1) \hat{\Sigma}_{\varepsilon W} \\ \hat{\Sigma}_{UW} \end{pmatrix},$$

it is seen that $(\omega^{\circ}, \theta^{\circ'}) \hat{\Sigma}_{\mathcal{R}W} - (1, \theta^{\circ'}) \hat{\Sigma}_{\Delta W} = (\omega^{\circ}, -1) \hat{\Sigma}_{\varepsilon W}$. The expression for $S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{-1}$ in (B.28) implies

$$\begin{aligned} &S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ -1} \begin{pmatrix} \sigma_{DM}^{\circ} \\ \Sigma_{UU}^{\circ} \theta^{\circ} \end{pmatrix} \\ &\stackrel{a.s.}{=} \left\{ \begin{pmatrix} \sigma_{MM}^{\circ -1} & 0 \\ 0 & \Sigma_{UU}^{\circ -1} \end{pmatrix} + \frac{1}{\sigma_{DD \cdot M}^{\circ}} \begin{pmatrix} \omega^{\circ} \\ \theta^{\circ} \end{pmatrix}^{\otimes 2} \right\} \begin{pmatrix} \sigma_{DM}^{\circ} \\ \Sigma_{UU}^{\circ} \theta^{\circ} \end{pmatrix} + o(T^{\eta-1/2}) \\ &= \begin{pmatrix} \omega^{\circ} \\ \theta^{\circ} \end{pmatrix} \frac{\sigma_{DD}^{\circ} + \theta^{\circ'} \Sigma_{UU}^{\circ} \theta^{\circ}}{\sigma_{DD \cdot M}^{\circ}} + o(T^{\eta-1/2}), \end{aligned}$$

where $(\omega^\circ, \theta^{\circ'})\mathcal{H}^\circ = (\omega^\circ, -1)\tau_\perp^\circ$ by (B.10). Inserting these results in (B.31) shows

$$\mathcal{H}^{\circ'}(S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W \cdot \Delta} - S_{\mathcal{R}\mathcal{R}}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W}) \stackrel{a.s.}{=} \tau_\perp^{\circ'} \begin{pmatrix} \omega^\circ \\ -1 \end{pmatrix}^{\otimes 2} \sigma_{DD \cdot M}^{\circ-1} \hat{\Sigma}_{\varepsilon W} \{1 + o(T^{\eta-1/2})\}.$$

Further, note that $S_{\varepsilon_M \varepsilon_M}^\circ \stackrel{a.s.}{=} \sigma_{MM}^\circ + T^{-1/4}$ and $\hat{\Sigma}_{\varepsilon_M W} = (1, 0) \hat{\Sigma}_{\varepsilon W}$ along with the identities (B.29) to see

$$\mathcal{G}S_{MM}^{-1} \hat{\Sigma}_{\varepsilon_M W} \stackrel{a.s.}{=} \sigma_{MM}^{-1} \tau_\perp^{\circ'} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes 2} \hat{\Sigma}_{\varepsilon W} \{1 + o(T^{\eta-1/2})\}.$$

Combining the two last expressions and noting $\hat{\Sigma}_{\varepsilon W} = o(T^{1/4-\eta})$ for some $\eta > 0$ shows

$$\begin{aligned} \mathcal{G}S_{MM}^{-1} \hat{\Sigma}_{\varepsilon_M W} + \mathcal{H}'(S_{\mathcal{R}\mathcal{R} \cdot \Delta}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W \cdot \Delta} - S_{\mathcal{R}\mathcal{R}}^{\circ-1} \hat{\Sigma}_{\mathcal{R}W}) \\ \stackrel{a.s.}{=} \tau_\perp^{\circ'} \{ \sigma_{MM}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes 2} + \sigma_{DD \cdot M}^{-1} \begin{pmatrix} \omega^\circ \\ -1 \end{pmatrix}^{\otimes 2} \} \hat{\Sigma}_{\varepsilon W} + o(T^{-1/4}). \end{aligned}$$

Finally, the desired result follows by the partitioned inversion formula (B.30). ■

B.5 Improving the rate of consistency

Lemma B.9 *Consider the maximum likelihood estimators in model $\mathbf{M}_{1DSB}$. Assume A, B, C . Then $\hat{R} - R = o(T^{1/4} \rho^{-T})$ a.s.*

Proof of Lemma B.9. Lemma B.4 shows that $\hat{R} - R_\circ = o(T^{1/2} \rho^{-T})$. Thus it suffices to analyse the profile likelihood $\ell(R)$ in a neighbourhood of $R_\circ$. Lemma B.6 shows that the profile likelihood $\ell(R)$ is maximised by maximising $\tilde{\ell}(R)$ up to an error of order $o(1)$ uniformly over intervals $|R - R_\circ| \leq cT^{1/2} \rho_\circ^{-T}$ for any $c > 0$. Thus, consider the approximate score equation

$$0 = \tilde{\ell}'(R) = \tilde{\ell}'(R_\circ) + \tilde{\ell}''(R_\circ)(R - R_\circ) + \frac{1}{2} \tilde{\ell}'''(R_\circ)(R - R_\circ)^2$$

for some R_* so $|R_* - R_\circ| \leq |R - R_\circ|$. Thus it holds

$$R - R_\circ = \frac{\tilde{\ell}'(R_\circ) + 2^{-1} \tilde{\ell}'''(R_\circ)(R_* - R_\circ)^2}{-\tilde{\ell}''(R_\circ)}. \quad (\text{B.32})$$

Insert the results of Lemma B.7 to get the desired result. ■

Lemma B.10 Consider the maximum likelihood estimators in model M_{1DSB} . Assume A, B, C . Then

- (i) $\hat{R} - R_o \stackrel{a.s.}{=} \{-\tilde{\ell}''(R_o)\}^{-1}\tilde{\ell}'(R_o) + o(T^{-1/4}\rho_o^{-T})$.
- (ii) $\rho_o^T(\hat{R} - R_o)\tau_\perp^\circ \stackrel{a.s.}{=} \mathcal{P}_{\tau_\perp}^\circ \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1} \{1 + o(1)\} + o(1)$.

Proof of Lemma B.10. (i) Lemma B.9 shows that $\hat{R} - R_o = o(T^{1/4}\rho_o^{-T})$. Insert this and the results of Lemma B.7 into (B.32) to get

$$\hat{R} - R_o = \{-\tilde{\ell}''(R_o)\}^{-1}\tilde{\ell}'(R_o) + \{-2\tilde{\ell}''(R_o)\}\tilde{\ell}'''(R_o)(R_* - R_o)^2$$

where $|R_* - R_o| \leq |\hat{R} - R_o|$. The second term is $o(T^{-3/4}\rho_o^{3T})(T^{1/4}\rho_o^{-T})^2\rho_o^{-2T} = o(T^{-1/4}\rho_o^{-T})$ so (i) follows.

(ii) Insert the expressions in Lemma B.8 into (i) so that

$$\hat{R} - R_o \stackrel{a.s.}{=} \frac{\rho_o^T \{\tau_\perp^{\circ'} \Omega_o^{-1} \hat{\Sigma}_{\varepsilon W} + o(T^{-1/4})\}}{\rho_o^{2T} \tau_\perp^{\circ'} \Omega_o^{-1} \tau_\perp^\circ \Sigma_{WW}^\circ} \{1 + o(1)\} + o(\rho_o^{-T}).$$

Rearrange to get the desired result. ■

B.6 Asymptotic distribution of estimators

Theorem B.11 Consider the maximum likelihood estimators in model M_{1DSB} . Assume A, B, C . Then

- (i) $\{(\sum_{t=1}^T M_t^2)^{1/2}(\hat{\omega} - \omega^\circ), (\sum_{t=1}^T U_t^2)^{1/2}(\hat{\theta} - \theta^\circ)\} \xrightarrow{D} N(0, \sigma_{DD \cdot M}^2 I_{2k-1})$.
- (ii) $\hat{\sigma}_{MM} \rightarrow \sigma_{MM}, \hat{\sigma}_{DD \cdot M} \rightarrow \sigma_{DD \cdot M}$ a.s.
- (iii) Let $H = (\tau_\perp^{\circ'} \Omega_o^{-1} \tau_\perp^\circ)^{-1/2} \tau_\perp^{\circ'} \Omega_o^{-1} \{\sum_{t=1}^T \rho^{2(t-T)}\}^{-1/2} \sum_{t=1}^T \rho^{t-T} \varepsilon_t^\circ$. Then it holds $\{-\ell''(\hat{R})\}^{1/2}(\hat{R} - R_o) = H + o(1)$ a.s.
- (iv) If $(\tau_\perp^{\circ'} \Omega_o^{-1} \tau_\perp^\circ)^{-1/2} \tau_\perp^{\circ'} \Omega_o^{-1} \varepsilon_t^\circ$ are independent $N(0, 1)$ then H is $N(0, 1)$.

Proof of Theorem B.11. (i) Since $\hat{\rho} - \rho_o = o(T^{1/4}\rho_o^{-T})$ by Lemma B.9 and since $\Delta_1 D_t = (\omega^\circ, \theta^{\circ'}) \mathcal{R}_t^\circ + \varepsilon_{D \cdot M, t}^\circ$ then Lemma B.2 implies that

$$\begin{aligned} (\hat{\omega} - \omega^\circ, \hat{\theta}' - \theta^{\circ'}) \left(\sum_{t=1}^T \mathcal{R}_t^{\otimes 2} \right)^{1/2} &= \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ \mathcal{R}_t' \left(\sum_{t=1}^T \mathcal{R}_t^{\otimes 2} \right)^{-1/2} \\ &\stackrel{a.s.}{=} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^\circ \mathcal{R}_t^{\circ'} \left(\sum_{t=1}^T \mathcal{R}_t^{\circ \otimes 2} \right)^{-1/2} + o(1), \end{aligned}$$

which is asymptotic normal. Similarly $\sum_{t=1}^T \mathcal{R}_t^{\otimes 2} = \sum_{t=1}^T \mathcal{R}_t^{\circ \otimes 2} \{1 + o(1)\}$ where $\mathcal{R}_t^\circ$ has asymptotically uncorrelated components $\varepsilon_{M, t}^\circ, U_{t-1}^\circ$.

(ii) First, consider $\hat{\sigma}_{MM} = T^{-1} \sum_{t=1}^T M_t^2$. Since $\hat{\rho} - \rho_o = o(T^{1/4}\rho_o^{-T})$ by Lemma B.9 then Lemma B.2 implies $\hat{\sigma}_{MM} = T^{-1} \sum_{t=1}^T M_t^2 + o(1)$ which has the desired limit.

Secondly, consider $\hat{\sigma}_{DD \cdot M} = T^{-1} \sum_{t=1}^T (\Delta_1 D_t | \mathcal{R}_t)$. Noting $\Delta_1 D_t = (\omega^\circ, \theta^\circ) \mathcal{R}_t^\circ + \varepsilon_{D \cdot M, t}^\circ$ then in the same way Lemma B.2 implies $\hat{\sigma}_{DD \cdot M} = T^{-1} \sum_{t=1}^T \varepsilon_{D \cdot M, t}^{\circ 2} + o(1)$ which has the desired limit.

(iii) Combine Lemmas B.6, B.8, B.10(ii) to see that

$$\{-\ell''(\hat{R})\}^{1/2}(\hat{R} - R_\circ) \stackrel{a.s.}{=} (\tau_\perp^{\circ'} \Omega_\circ^{-1} \tau_\perp^\circ)^{-1/2} \tau_\perp^{\circ'} \Omega_\circ^{-1} \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1/2} + o(1).$$

then $\rho_\circ^T(\hat{R} - R_\circ) = (\tau_\perp^{\circ'} \Omega_\circ^{-1} \tau_\perp^\circ)^{-1} \tau_\perp^{\circ'} \Omega_\circ^{-1} \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1} \{1 + o(1)\} + o(1)$ *a.s.* By an argument as in Anderson (1959), see also Nielsen (2010, Theorem 4) then $H = \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1/2} = \{\sum_{t=1}^T \rho_\circ^{2(t-T)}\}^{-1/2} \sum_{t=1}^T \rho_\circ^{t-T} \varepsilon_t^\circ + o(1)$ giving the desired result.

(iv) Under the normality assumption then H is a linear combination of normals, so normal itself. ■

B.7 Likelihood in restricted model

Lemma B.12 *Consider the maximum likelihood estimators in model M_{1DSB} . Assume A, B, C . Then*

$$2\ell(\hat{R}) \stackrel{a.s.}{=} -T \log \det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD \cdot M}^{-1} \hat{\Sigma}_{\varepsilon D \cdot M} U \Sigma_{UU}^{\circ-1} \hat{\Sigma}_{U \varepsilon D \cdot M} + \text{tr}(\Omega_\circ^{-1} \mathcal{P}_{\tau_\perp}^\circ \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1} \hat{\Sigma}_{W\varepsilon}) + o(1).$$

Proof of Lemma B.12. The profile log likelihood is given in Lemma B.6 as

$$2\{\ell(\hat{R}) - \ell(R_\circ)\} = 2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_\circ)\} + o(1). \quad (\text{B.33})$$

Expanding $\tilde{\ell}(\hat{R})$ around $R_\circ$ then gives

$$2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_\circ)\} = 2\tilde{\ell}'(R_\circ)(\hat{R} - R_\circ) + \tilde{\ell}''(R_\circ)(\hat{R} - R_\circ)^2 + \frac{1}{3}\tilde{\ell}'''(R_\circ)(\hat{R} - R_\circ)^3.$$

where $|R_* - R_\circ| \leq |\hat{R} - R_\circ|$. Insert the expression for $\hat{R} - R_\circ$ from Lemma B.10(i) and use the bound $\hat{R} - R_\circ = o(T^{1/4} \rho_\circ^{-T})$ from Lemma B.4 to get

$$2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_\circ)\} \stackrel{a.s.}{=} -\{\tilde{\ell}''(R_\circ)\}^{-1} \{\tilde{\ell}'(R_\circ)\}^2 + o\{T^{-1/4} \rho_\circ^{-T} \tilde{\ell}'(R_\circ) + \rho_\circ^{-2T} \tilde{\ell}''(R_\circ) + T^{3/4} \rho_\circ^{-3T} \tilde{\ell}'''(R_\circ)\}.$$

Insert the bounds and the expressions for the derivatives established in Lemmas B.7, B.8 to see

$$2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_\circ)\} \stackrel{a.s.}{=} -\frac{\rho_\circ^{2T} \{\tau_\perp^{\circ'} \Omega_\circ^{-1} \hat{\Sigma}_{\varepsilon W} + o(T^{-1/4})\}^2}{\rho_\circ^{2T} \tau_\perp^{\circ'} \Omega_\circ^{-1} \tau_\perp^\circ \Sigma_{WW}^\circ} + o(1).$$

Noting that $\hat{\Sigma}_{\varepsilon W} = o(T^{1/4})$ this reduces to

$$2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_o)\} \stackrel{a.s.}{=} -(\tau_{\perp}^{\circ'} \Omega_o^{-1} \tau_{\perp}^{\circ})^{-1} \tau_{\perp}^{\circ'} \Omega_o^{-1} \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1} \Sigma_{W\varepsilon}^{\circ} \Omega_o^{-1} \tau_{\perp}^{\circ} + o(1).$$

Taking trace and rearranging shows

$$2\{\tilde{\ell}(\hat{R}) - \tilde{\ell}(R_o)\} \stackrel{a.s.}{=} \text{tr}(\Omega_o^{-1} \mathcal{P}_{\tau_{\perp}}^{\circ} \hat{\Sigma}_{\varepsilon W} \Sigma_{WW}^{\circ-1} \Sigma_{W\varepsilon}^{\circ}) + o(1)$$

Insert this in (B.33) and use the expression for $2\ell(R_o)$ in Lemma B.3. ■

B.8 Likelihood in unrestricted model

The unrestricted model M_{1D} is now analysed. The analysis of Nielsen (2010) needs to be elaborated. In particular, improved convergence results are needed for the estimator of the adjustment parameters $\alpha = (\alpha_1, \alpha_{\rho}, \Phi_1, \dots, \Phi_{k-2})$ and for Ω . Given the estimators $\hat{\rho}_D, \hat{\beta}_1^{*D}$ the maximum likelihood estimator for α is found by regression of $\Delta_1 \Delta_{\hat{\rho}_D} X_t$ on

$$U_{t-1}^D = (\hat{\beta}_1^{*D'} \Delta_{\hat{\rho}_D} X_{t-1}^*, \beta_{\rho}^{\circ'} \Delta_1 X_{t-1}, \Delta_1 \Delta_{\hat{\rho}_D} X_{t-1}, \dots, \Delta_1 \Delta_{\hat{\rho}_D} X_{t-k+2})'$$

For the results of this subsection the data generating process is

$$\Delta_1 \Delta_{\rho_o} X_t = \alpha_1^{\circ} \beta_1^{*o'} \Delta_{\rho_o} X_{t-1}^* + \alpha_{\rho}^{\circ} \beta_{\rho}^{\circ'} \Delta_1 X_{t-1} + \sum_{j=1}^{k-2} \Phi_j^{\circ} \Delta_1 \Delta_{\rho_o} X_{t-j} + \varepsilon_t^{\circ}.$$

This of course encompasses the data generating process (B.7) under M_{1DSB} .

Lemma B.13 *Suppose M_{1D} holds with $\rho_o \geq \varrho$ for some $\varrho > 1$. Assume A, B, C . Recall the definitions of $\tau_{\perp}^{\circ}, \mathcal{P}_{\tau_{\perp}}^{\circ}, \mathcal{P}_{\alpha}^{\circ}$ in (B.8), (B.9). Then*

- (i) $\hat{\alpha} - \alpha = T^{-1/2} \hat{\Sigma}_{\varepsilon U} \Sigma_{UU}^{-1} + o_P(T^{-\xi}) = O_P(T^{-1/2})$.
- (ii) $\alpha_1^{\circ} (\hat{\beta}_1^{*D} - \beta_1^{*o})' \hat{\beta}_{1\perp}^{*D} N_V^{-1} = T^{-1/2} \mathcal{P}_{\alpha}^{\circ} \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} + o_P(T^{-1/2}) = O_P(T^{-1/2})$.
- (iii) $\rho_o^{-T} \hat{\tau}^{D'} \sum_{t=1}^T \hat{\varepsilon}_t^D W_{t-1}^{\circ} = \rho_o^{-T} \tau^{\circ'} \sum_{t=1}^T \varepsilon_t^{\circ} W_{t-1}^{\circ} + o_P(1)$.
- (iv) $\hat{\Omega}_D - \Omega_o = T^{-1} \sum_{t=1}^T \{(\varepsilon_t^{\circ})^{\otimes 2} - \Omega_o\} + o_P(T^{-1/2}) = O_P(T^{-1/2})$.
- (v) $\tau_{\perp}^{\circ} (\hat{\rho}_D - \rho_o) = \mathcal{P}_{\tau_{\perp}} \sum_{t=1}^T \varepsilon_t^{\circ} W_{t-1}^{\circ} \{\sum_{t=1}^T (W_{t-1}^{\circ})^2\}^{-1} \{1 + O_P(T^{-1/2})\} + o_P(T^{-1/4} \rho_o^{-T})$.

Proof of Lemma B.13. *Product moments.* Let V_{t-1}^D denote V_{t-1} computed at $\hat{\rho}_D$. Combine (A.12) and the first display on p.911 of Nielsen (2010) to get

$$\begin{aligned} & \begin{pmatrix} \hat{S}_{UU} & \hat{S}_{UV} & \hat{S}_{U\varepsilon} & \hat{S}_{UW} \\ * & \hat{S}_{VV} & \hat{S}_{V\varepsilon} & \hat{S}_{VW} \\ * & * & \hat{S}_{\varepsilon\varepsilon} & \hat{S}_{\varepsilon W} \\ * & * & * & \hat{S}_{WW} \end{pmatrix} = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} U_{t-1}^D \\ N_V V_{t-1}^D \\ \varepsilon_t^\circ \\ N_W W_{t-1}^\circ \end{pmatrix}^{\otimes 2} \\ & = \begin{pmatrix} \Sigma_{UU} + \text{op}(1) & \text{op}(T^{-\xi/2}) & T^{-1/2}\{\hat{\Sigma}_{U\varepsilon} + \text{op}(1)\} & \text{op}(T^{-\xi/2}) \\ * & \hat{\Sigma}_{VV} + \text{op}(1) & T^{-1/2}\{\hat{\Sigma}_{V\varepsilon} + \text{op}(1)\} & \text{op}(T^{-\xi/2}) \\ * & * & \Omega_\circ + \text{op}(1) & T^{-1/2}\hat{\Sigma}_{\varepsilon W} \\ * & * & * & \Sigma_{WW} + \text{op}(1) \end{pmatrix} \end{aligned} \quad (\text{B.34})$$

where $\hat{\Sigma}_{W\varepsilon} = \rho_\circ^{-T} \sum_{t=1}^T W_{t-1}^\circ \varepsilon_t^\circ$, $\hat{\Sigma}_{U\varepsilon} = T^{-1/2} \sum_{t=1}^T U_{t-1}^\circ \varepsilon_t^\circ$, $\hat{\Sigma}_{V\varepsilon} = T^{-1/2} N_V \sum_{t=1}^T V_{t-1}^\circ \varepsilon_t^\circ$, and $\hat{\Sigma}_{VV} = T^{-1} \sum_{t=1}^T (N_V V_{t-1}^\circ)^{\otimes 2}$.

(i) An equation shown in the proof of Lemma A.7 of Nielsen (2010) gives

$$\Delta_1 \Delta_{\hat{\rho}} X_t = \varepsilon_t^\circ + (\boldsymbol{\alpha} + \hat{\delta}_U) U_{t-1}^D + \hat{\delta}_V N_V V_{t-1}^D + \hat{\delta}_W W_{t-1}^\circ, \quad (\text{B.35})$$

where $\hat{\delta}_U = (\hat{\rho}_D - \rho_\circ) \{ \alpha_1^\circ / (1 - \hat{\rho}_D), -\check{\Psi}_{\hat{\rho}}^D \bar{\beta}_1, \hat{\rho}_D^0 \sum_{j=1}^{k-2} \Phi_j^\circ, \dots, \hat{\rho}_D^{k-3} \sum_{j=k-2}^{k-2} \Phi_j^\circ \}$ and $\hat{\delta}_V = \alpha_1^\circ \beta_1^{*\circ} \bar{\beta}_{1\perp}^{*D} N_V^{-1} (1 - \rho_\circ) / (1 - \hat{\rho}_D)$ and $\hat{\delta}_W = -(\hat{\rho}_D - \rho_\circ) \check{\tau}_\perp^D$. It follows that

$$\begin{aligned} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} &= \sum_{t=1}^T \varepsilon_t^\circ (U_{t-1}^D)' \{ \sum_{t=1}^T (U_{t-1}^D)^{\otimes 2} \}^{-1} \\ &\quad + \hat{\delta}_U + \sum_{t=1}^T (\hat{\delta}_V N_V V_{t-1}^D + \hat{\delta}_W W_{t-1}^\circ) (U_{t-1}^D)' \{ \sum_{t=1}^T (U_{t-1}^D)^{\otimes 2} \}^{-1}. \end{aligned}$$

Lemma A.11 of Nielsen (2010) shows that $T^{1/2} \rho_\circ^T (\hat{\rho}_D - \rho_\circ)$ and $\beta_1^{*\circ} \bar{\beta}_{1\perp}^{*D} N_V^{-1}$ are $\text{op}(T^{-\xi/2})$. This implies that $\hat{\delta}_U, \hat{\delta}_W = \text{op}(\rho_\circ^{-T} T^{(1-\xi)/2})$ and $\hat{\delta}_V = \text{op}(T^{-\xi/2})$. From (B.34) it follows that

$$\begin{aligned} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} &= \{ T^{-1/2} \Sigma_{\varepsilon U} + \text{op}(T^{-1/2}) + \text{op}(T^{-\xi}) \} \{ \Sigma_{UU}^{-1} + \text{op}(T^{-\xi/2}) \} \\ &\quad + \text{op}(\rho_\circ^{-T} T^{(1-\xi)/2}) + \text{op}(T^{-\xi/2}) \text{op}(T^{-\xi/2}) + \text{op}(\rho_\circ^{-T} T^{(1-\xi)/2}) \text{op}(T^{(-1-\xi)/2} \rho_\circ^T). \end{aligned}$$

By Assumption B then $\xi > 1/2$ and the desired result follows.

(ii, iii) Statement of Lemma A.12(ii, iii) of Nielsen (2010).

(iv) The variance estimator is $\hat{\Omega}_D = T^{-1} \sum_{t=1}^T (\Delta_1 \Delta_{\hat{\rho}} X_t | U_{t-1}^D)^{\otimes 2}$. Due to (B.35) then

$$\hat{\Omega}_D = T^{-1} \sum_{t=1}^T (\varepsilon_t^\circ + \hat{\delta}_V N_V V_{t-1}^D + \hat{\delta}_W W_{t-1}^\circ | U_{t-1}^D)^{\otimes 2}.$$

From (ii, iii) it follows that $\hat{\delta}_V = -T^{-1/2}\mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} + o_P(T^{-\xi}) = o_P(T^{-1/2})$. Inserting this and using (B.34) it follows that $\hat{\Omega}_D = S_{\varepsilon\varepsilon}^\circ + T^{-1}G$ where

$$\begin{aligned} G &= -\Sigma_{\varepsilon U} \Sigma_{UU}^{-1} \Sigma_{U\varepsilon} + o_P(1) + \hat{\delta}_W o_P(\rho_o^T T^{-1-\xi/2}) \\ &\quad + \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} \mathcal{P}_\alpha^{\circ'} - \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} - \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} \mathcal{P}_\alpha^{\circ'} \\ &\quad + \hat{\delta}_W \hat{\Sigma}_{WW} \hat{\delta}_W' \{1 + o_P(T^{-1/2})\} + \rho_o^{-T} (\hat{\delta}_W \hat{\Sigma}_{W\varepsilon} + \hat{\Sigma}_{\varepsilon W} \hat{\delta}_W'). \end{aligned}$$

Since $\hat{\delta}_W = o_P(\rho_o^{-T} T^{(1-\xi)/2})$, $\hat{\Sigma}_{WW} = O(1)$ and $S_{W\varepsilon} = o_P(T^{(1-\xi)/2})$ and $\xi > 1/2$ then

$$\begin{aligned} G &= -\Sigma_{\varepsilon U} \Sigma_{UU}^{-1} \Sigma_{U\varepsilon} + \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} \mathcal{P}_\alpha^{\circ'} - \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} - \hat{\Sigma}_{\varepsilon V} \hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{V\varepsilon} \mathcal{P}_\alpha^{\circ'} \\ &\quad + \hat{\delta}_W \hat{\Sigma}_{WW} \hat{\delta}_W' + \rho_o^{-T} (\hat{\delta}_W \hat{\Sigma}_{W\varepsilon} + \hat{\Sigma}_{\varepsilon W} \hat{\delta}_W') + o_P(1), \end{aligned} \quad (\text{B.36})$$

and $G = o_P(T^{1/2})$. It follows that $\hat{\Omega}_D = S_{\varepsilon\varepsilon}^\circ + o_P(T^{-1/2})$. Since $S_{\varepsilon\varepsilon}^\circ = \Omega_o + o_P(T^{-1/2})$ the desired result follows.

(v) Lemma A.9 of Nielsen (2010) shows

$$\begin{aligned} \hat{\tau}_\perp^{D'} \hat{\Omega}_D^{-1} \rho_o^T (\hat{\rho}_D - \rho_o) (\check{\tau}_\perp^D \hat{S}_{WW \cdot U} - \hat{\mathcal{P}}_\alpha \check{\tau}_\perp^D \hat{S}_{WV \cdot U} \hat{S}_{VV \cdot U}^{-1} \hat{S}_{VW \cdot U}) \\ = T^{1/2} \hat{\tau}_\perp^D \hat{\Omega}_D^{-1} (\hat{S}_{\varepsilon W \cdot U} - \hat{\mathcal{P}}_\alpha \hat{S}_{\varepsilon V \cdot U} \hat{S}_{VV \cdot U}^{-1} \hat{S}_{VW \cdot U}) \end{aligned}$$

where $\hat{\mathcal{P}}_\alpha = \check{\alpha}_1 (\hat{\alpha}_1^{D'} \hat{\Omega}_D^{-1} \check{\alpha}_1)^{-1} \hat{\alpha}_1^{D'} \hat{\Omega}_D^{-1}$. Exploit the $T^{1/2}$ -order of the $\rho, \alpha, \tau, \Omega$ estimators by (i, iii) as well as premultiplying the equation by $\tau_\perp^\circ (\tau_\perp^{\circ'} \Omega_o^{-1} \tau_\perp^\circ)^{-1}$ and postmultiplying by $\hat{S}_{WW}^{-1}$ to get

$$\begin{aligned} \tau_\perp^\circ \rho_o^T (\hat{\rho}_D - \rho_o) &= \{\rho_o^T (\hat{\rho}_D - \rho_o) \mathcal{P}_{\tau_\perp} \mathcal{P}_\alpha^\circ \tau_\perp^\circ \hat{S}_{WV \cdot U} \hat{S}_{VV \cdot U}^{-1} \hat{S}_{VW \cdot U} \hat{S}_{WW \cdot U}^{-1} \\ &\quad + \mathcal{P}_{\tau_\perp} T^{1/2} (\hat{S}_{\varepsilon W \cdot U} - \mathcal{P}_\alpha^\circ \hat{S}_{\varepsilon V \cdot U} \hat{S}_{VV \cdot U}^{-1} \hat{S}_{VW \cdot U}) \hat{S}_{WW \cdot U}^{-1}\} \{1 + o_P(T^{-1/2})\}. \end{aligned}$$

Exploit the $T^{-1/4} \rho_o^T$ consistency of $\hat{\rho}_D$ as well as (B.35) to get the desired result. ■

An expansion is needed for the variance estimator $\hat{\Omega}_D$ in the unrestricted model M_{1D} . Theorem 3 of Nielsen (2010) shows that the estimator $\hat{\Omega}_D$, called $\hat{\Omega}_H$ in that paper, is consistent.

Lemma B.14 *Suppose M_{1D} holds with $\rho_o \geq \varrho$ for some $\varrho > 1$. Assume A, B, C . Then*

$$\begin{aligned} 2\hat{\ell}_{1D} &= -T \log \det S_{\varepsilon^\circ \varepsilon^\circ} + \text{tr}(\Omega_o^{-1} \hat{\Sigma}_{\varepsilon U} \Sigma_{UU}^{-1} \hat{\Sigma}_{U\varepsilon}) \\ &\quad + \text{tr}(\Omega_o^{-1} \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V} \Sigma_{VV}^{-1} \hat{\Sigma}_{V\varepsilon}) + \text{tr}(\Omega_o^{-1} \mathcal{P}_{\tau_\perp}^\circ \hat{\Sigma}_{\varepsilon W} \hat{\Sigma}_{WW}^{-1} \hat{\Sigma}_{W\varepsilon}) + o_P(1). \end{aligned}$$

Proof of Lemma B.14. Combine (B.36) where $\hat{\delta}_W = -(\hat{\rho}_D - \rho_o)\check{\tau}_\perp^D$ with Lemma B.13(v) to get $\hat{\Omega}_D = S_{\varepsilon^\circ\varepsilon^\circ} + T^{-1}G$ where

$$\begin{aligned} G = & -\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon} + \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ'} - \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon} - \hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ'} \\ & + \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\mathcal{P}_{\tau_\perp}^{\circ'} - \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon} - \hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\mathcal{P}_{\tau_\perp}^{\circ'} + o_P(1). \end{aligned}$$

Use the log determinant expansion in Lemma B.5 to get

$$2\hat{\ell}_{1D} = -T \log \det \hat{\Omega}_D = -T \log \det S_{\varepsilon^\circ\varepsilon^\circ} - \text{tr}(\hat{\Omega}_D^{-1}G).$$

Since $\hat{\Omega}_D$ is consistent and $\mathcal{P}_\alpha^{\circ'}\Omega_o^{-1}\mathcal{P}_\alpha^\circ = \Omega_o^{-1}\mathcal{P}_\alpha^\circ$ and $\mathcal{P}_\alpha^{\circ'}\Omega_o^{-1}\mathcal{P}_\alpha^\circ = \Omega_o^{-1}\mathcal{P}_\alpha^\circ$ and using the symmetry of the trace then

$$-\text{tr}(\hat{\Omega}_D^{-1}G) = \text{tr}\{\Omega_o^{-1}(\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon} + \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon} + \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon})\} + o_P(1).$$

Insert this in the expression for $2\hat{\ell}_D$ to get the desired result. ■

B.9 Proof of main theorem

Proof of Theorem 3.3. It holds that

$$\text{LR}(\mathbf{M}_{1DSB}|\mathbf{M}_{1D}) = 2\hat{\ell}_{1D} - 2\hat{\ell}_{1DSB}$$

Inserting results from Lemmas B.12, B.14 gives

$$\begin{aligned} \text{LR} = & \{-T \log \det S_{\varepsilon^\circ\varepsilon^\circ} + \text{tr}(\Omega_o^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}) + \text{tr}(\Omega_o^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}) \\ & + \text{tr}(\Omega_o^{-1}\mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon})\} - \{-T \log \det(S_{\varepsilon\varepsilon}^\circ) \\ & + \sigma_{DD \cdot M}^{-1}\hat{\Sigma}_{\varepsilon D \cdot M U}\Sigma_{UU}^{\circ-1}\hat{\Sigma}_{U\varepsilon D \cdot M} + \text{tr}(\Omega_o^{-1}\mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\Sigma_{WW}^{\circ-1}\hat{\Sigma}_{W\varepsilon})\} + o_P(1). \end{aligned}$$

This reduces to

$$\begin{aligned} \text{LR} = & \text{tr}(\Omega_o^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}) - \sigma_{DD \cdot M}^{\circ-1}\hat{\Sigma}_{\varepsilon D \cdot M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon D \cdot M} \\ & + \text{tr}(\Omega_o^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}) + o_P(1). \end{aligned}$$

In the first term partitioned inversion of Ω_o^{-1} gives

$$\text{tr}(\Omega_o^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}) = \sigma_{MM}^{\circ-1}\hat{\Sigma}_{\varepsilon M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon M} + \sigma_{DD \cdot M}^{\circ-1}\hat{\Sigma}_{\varepsilon D \cdot M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon D \cdot M}$$

so the test statistic satisfies

$$\text{LR} = \sigma_{MM}^{\circ-1}\hat{\Sigma}_{\varepsilon M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon M} + \text{tr}(\Omega_o^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}) + o_P(1).$$

Since $\varepsilon_{M,t}, U_{t-1}$ are mutually independent a martingal central limit theorem, see Brown and Eagleson (1971), gives that the first term is asymptotically χ^2 with $\dim U = 2k - 2$ degrees of freedom.

The term $\mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V}$ is the stochastic integral of $B_{1,t,T} = N_V V_{t-1}^\circ$ with respect to $c = T^{-1/2} \sum_{s=1}^t \alpha_1^{\circ'} \Omega_\circ^{-1} \varepsilon_s$. The process $B_{1,t,T}$ is a function of $T^{-1/2} \sum_{s=1}^t \alpha_{1\perp}^{\circ'} \varepsilon_s$. Thus, $B_{1,t,T}$ and $\hat{\Sigma}_{\varepsilon V}$ converge to asymptotically independent processes, so by a mixed Gaussian argument, see Johansen (1995, §13.1), the last term is χ^2 with $\dim(\alpha_1^{\circ'} \Omega_\circ^{-1} \varepsilon_t) \dim(V) = 2$ degrees of freedom.

It is left to argue that the last term is asymptotically independent of the previous two. The last one is based on the processes $B_{1,t,T}, \hat{\Sigma}_{\varepsilon M U}$ which are asymptotically independent of $\hat{\Sigma}_{\varepsilon M \varepsilon D \cdot M}, \hat{\Sigma}_{\varepsilon M U}$, see Chan and Wei (1988, Theorem 2.2). Since $\hat{\Sigma}_{\varepsilon V}, \hat{\Sigma}_{V V}$ are functionals of $B_{1,t,T}, \hat{\Sigma}_{\varepsilon M U}$ then $\hat{\Sigma}_{\varepsilon V}, \hat{\Sigma}_{V V}$ are asymptotically independent of $\hat{\Sigma}_{\varepsilon M \varepsilon D \cdot M}, \hat{\Sigma}_{\varepsilon M U}$.

It then follows that LR is asymptotically χ^2 with $(2k - 2) + 2 = 2k$ degrees of freedom. ■

References

- Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.
- Brown, B.M. and Eagleson, G.K. (1971). Martingale convergence to infinitely divisible laws with finite variance. *Transactions of the American Mathematical Society* 162, 449-453.
- Campbell, J.Y. and Shiller, R.J. (1987) Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062-1088.
- Campbell, J.Y. and Shiller, R.J. (1988) Interpreting cointegrated models. *Journal of Economic Dynamics and Control* 12, 505-522.
- Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.
- Cochrane, J.H. (2008) The dog that did not bark: A defense of return predictability. *Review of Financial Studies* 21, 1533-1575.
- Diba, B.T. and Grossman, H.I. (1988) Explosive rational bubbles in stock prices. *American Economic Review* 78, 520-530.
- Doornik, J. (1998) Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573-593.

- Doornik, J.A. and Hendry, D.F. (2001) *Empirical econometric modelling using Pc-Give 10, vol. 1 and 2*. London: Timberlake Consultants Press.
- Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometric Journal* 12, 367-381.
- Engsted, T. (2006) Explosive bubbles in the cointegrated VAR model. *Finance Research Letters* 3, 154-162.
- Flood, R.P. and Garber, P.M. (1980) Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy* 88, 745-770.
- Froot, K.A. and Obstfeld, M. (1991) Intrinsic bubbles: the case of stock prices. *American Economic Review* 81, 1189-1214.
- Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.
- Johansen, S. (1995) *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Johansen, S. and Swensen, A.R. (1999) Testing exact rational expectations in cointegrated vector autoregressive models. *Journal of Econometrics* 93, 73-91.
- Johansen, S. and Swensen, A.R. (2004) More on testing exact rational expectations in cointegrated vector autoregressive models: Restricted constant and linear term. *Econometrics Journal* 7, 389-397.
- Johansen, S. and Swensen, A.R. (2011) On a graphical technique for evaluating some rational expectations models. *Journal of Time Series Econometrics* 3 (1), Article 9.
- Lai, T.L. and Wei, C.Z. (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah (ed.), *Multivariate Analysis VI*, pp. 375-393. Amsterdam: Elsevier Science Publishers.
- Leroy, S.F. (1989) Efficient capital markets and martingales. *Journal of Economic Literature* 27, 1583-1621.
- Mirsky, L. (1961) *An Introduction to Linear Algebra*. Oxford: Clarendon Press.
- Mynbaev, K.T. (2011) *Short-Memory Linear Processes and Econometric Applications*. Hoboken NJ: John Wiley & Sons.
- Nielsen, B. (2001) The Asymptotic Distribution of Unit Root Tests of Unstable Autoregressive Processes. *Econometrica* 69, 211-219.
- Nielsen, B. (2005) Strong consistency results for least squares statistics in general vector autoregressive models. *Econometric Theory* 21, 534-561.

- Nielsen, B. (2006a) Correlograms for non-stationary autoregressions. To appear in *Journal of the Royal Statistical Society* B68, 707-720.
- Nielsen, B. (2006b) Order determination in general vector autoregressions. In Ho, H.-C., Ing, C.-K., and Lai, T.L. (eds): *Time Series and Related Topics: In Memory of Ching-Zong Wei*. IMS Lecture Notes and Monograph Series, 52, 93-112.
- Nielsen, B. (2010) Analysis of co-explosive processes. *Econometric Theory* 26, 882–915.
- Nielsen, B. and Sohkanen, J.S. (2011) Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends. *Econometric Theory* 27, 913–927.
- Shiller, R.J. (2000) *Irrational Exuberance*. Princeton University Press. Princeton NJ.
- West, K.D. (1985) A specification test for speculative bubbles. *Working Paper*, Princeton University.
- West, K.D. (1987) A specification test for speculative bubbles. *Quarterly Journal of Economics* 102, 553-580.