*Econometrics Journal* (2012), volume **15**, pp. 226–254.
doi: 10.1111/j.1368-423X.2012.00369.x

# Testing for rational bubbles in a coexplosive vector autoregression

TOM ENGSTED[†] AND BENT NIELSEN[‡]

[†]*CREATES, Department of Economics and Business, Aarhus University, DK-8000 Aarhus C, Denmark.*
E-mail: tengsted@creates.au.dk

[‡]*Nuffield College, New Road, Oxford OX1 1NF, UK.*
E-mail: bent.nielsen@nuffield.ox.ac.uk

First version received: March 2011; final version accepted: December 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:** *Cointegration, Explosiveness and coexplosiveness, Likelihood ratio tests, Rational bubbles, Vector autoregression.*

## 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 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

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

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. Equation (2.1) can be reformulated as

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

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 (1988, 1987) 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

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

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 $\Delta_{1+R}$ 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

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

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

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

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 coexplosive 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

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

where $A_j \in \mathbb{R}^{2\times 2}$ and $\mu \in \mathbb{R}^2$. The errors, $\varepsilon_t$, are independent $\mathsf{N}_2(0, \Omega)$ -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 $\mathsf{E}$. Within this model, the efficient marked hypothesis amounts to $\mathsf{E}(M_t|\mathcal{F}_{t-1}) = 0$.

## 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

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

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

$$\mathsf{H}_1: \qquad (\Pi_1, \mu) = \alpha_1(\beta_1', \zeta_1), \qquad \Pi_\rho = \alpha_\rho \beta_\rho',$$

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

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 $\beta_1$ and $\beta_\rho$ are therefore referred to as the cointegrating vector and the coexplosive 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 $\alpha$ with full column rank $\alpha_\perp$ denotes an orthogonal complement of a matrix $\alpha$ so $\alpha_\perp' \alpha = 0$ and $(\alpha, \alpha_\perp)$ is invertible while $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$.

ASSUMPTION 2.1. *Consider the model* M *and the hypothesis* $\mathsf{H}_1$. *Suppose* (*a*) *The vectors* $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}$ *are non-zero;* (*b*) *The non-stationary characteristic roots of* $X_t$ *are at* 1 *or* $\rho > 0$; (*c*) $\det(\alpha'_{1\perp}\Psi_1\beta_{1\perp}) \neq 0$ *and* $\det(\alpha'_{\rho\perp}\Psi_\rho\beta_{\rho\perp}) \neq 0$ *where*

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

THEOREM 2.1. *Suppose Assumption 2.1 holds. Then,*

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

*where* $Y_t$ *is a stationary process. The impact matrices and initial values are*

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

*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* $\tilde{X}_t = X_t - \overline{\beta}_1\zeta_1/(1-\rho)$ *as*

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

*while the parameter* $\theta_U$ *is defined by*

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

Theorem 2.1 is proved in Appendix A. Note that the definition of $C_x$ includes a factor $x$ which is erroneously left out in Nielsen (2010).

The cointegration rank, i.e. the rank of the matrix $\Pi_1$, 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 $\Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}$; see also Nielsen (2010, section 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

$$\mathsf{H}_D: \qquad \beta_\rho = (0, 1)'.$$

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

$$\mathsf{M}_{1D}: \qquad \Delta_1\Delta_\rho X_t = \alpha_1\beta_1^{*\prime}\Delta_\rho X_{t-1}^* + \alpha_\rho\Delta_1 D_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \varepsilon_t, \tag{2.8}$$

where $\beta_1^* = (\beta_1', \zeta_1')'$ and $\Delta_\rho X_{t-1}^* = (\Delta_\rho X_{t-1}, 1)$. For a given value of $\rho$, 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, section 3.3).

### *2.5. Testing the bubble model*

The bubble model imposes two restrictions on the model $\mathsf{M}_{1D}$. 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

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

The second restriction is the martingale restriction (2.2). To formulate this hypothesis pre-multiply 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

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

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

It is convenient to reparametrise the restricted model $\mathsf{M}_{1DSB}$. To deal with the correlation structure for the errors, let $\omega$ 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

$$\mathsf{M}_{1DSB}: \qquad M_t = \varepsilon_{M,t}, \tag{2.11}$$

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

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 $\mathsf{H}_1$, $\mathsf{H}_D$, $\mathsf{H}_S$, $\mathsf{H}_B$ 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 $\varepsilon_t$ 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 3.1. *For some* $\gamma > 2$ *it holds* $\sup_t \mathsf{E}((\varepsilon_t'\varepsilon_t)^{(2+\gamma)/2}|\mathcal{F}_{t-1}) < \infty$ *a.s.*

ASSUMPTION 3.2. $\mathsf{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Omega$ *a.s. where* $\Omega$ *is positive definite.*

The purpose of Assumption 3.1 is to give a bound to the fluctuations of the paths whereas Assumption 3.2 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 $\mathsf{M}$ 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 $\varepsilon_t$ can be applied.

As a next step, the test for cointegration rank, i.e. the likelihood ratio test for $\mathsf{H}_1$ within $\mathsf{M}$, 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* $\mathsf{M}_1$ *and Assumptions 2.1, 3.1, 3.2 hold with* $\gamma > 0$ *only. Then,* $LR(\mathsf{M}_1|\mathsf{M})$ *has the same limiting distribution as in Johansen (1995, section 6), namely*

$$LR(\mathsf{M}_1|\mathsf{M}) \overset{\mathsf{d}}{\to} \operatorname{tr}\left\{\int_0^1 dB_u F_u' \left(\int_0^1 F_u F_u' du\right)^{-1} \int_0^1 F_u dB_u'\right\},$$

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

Thirdly, the likelihood ratio test statistic for $\mathsf{H}_D$ within $\mathsf{M}_1$ is asymptotically $\chi^2$. As this statistic is concerned exclusively with the explosive terms normality is required. Some discussion of the normality assumption is given in Nielsen (2010, section 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 $\tau = (I_2 - \tau_\perp\overline{\tau}_\perp')\alpha_\rho$.

THEOREM 3.2. *Suppose model* $\mathsf{M}_{1D}$ *with* $\rho \geq \varrho$ *for some* $\varrho > 1$ *and Assumptions 2.1, 3.1, 3.2 hold, and that* $\tau'\varepsilon_t$ *is independent* $\mathsf{N}(0, \tau'\Omega\tau)$*-distributed. Then,* $LR(\mathsf{M}_{1D}|\mathsf{M}_1)$ *is asymptotically* $\chi^2(1)$.

Finally, the main result of this paper concerns the likelihood ratio test for the joint hypothesis $\mathsf{H}_S$, $\mathsf{H}_B$ within the model $\mathsf{M}_{1D}$. The test statistic is asymptotically $\chi^2(2k)$. The hypotheses $\mathsf{H}_S$ and $\mathsf{H}_B$ link the cointegrating vector, the explosive root, the dynamic adjustment parameters and

**Table 1.** Specification tests for the unrestricted vector autoregression.

| 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] | | |

the intercept in a non-linear fashion and it is not immediately obvious that a $\chi^2$ result should hold, but it actually does under martingale difference assumptions.

THEOREM 3.3. *Suppose model* $\mathsf{M}_{1DSB}$ *and Assumptions 2.1, 3.1, 3.2 are satisfied. Then the test statistic* $\mathsf{LR}(\mathsf{M}_{1DSB}|\mathsf{M}_{1D})$ *is asymptotically* $\chi^2(2k)$.

The proof is given in Appendix B. It is related to the analysis in Nielsen (2010). Because the model $\mathsf{M}_{1DSB}$ 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. 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 $\mathsf{M}_{1DSB}$ 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.1 in Appendix B 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 subsample 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.

Figures 1(a) and (c) show dividend growth $\Delta_1 D_t$ and prices $P_t$. An exponential pattern 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 $\pm$ 0.354$i$ 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 ARCH 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).

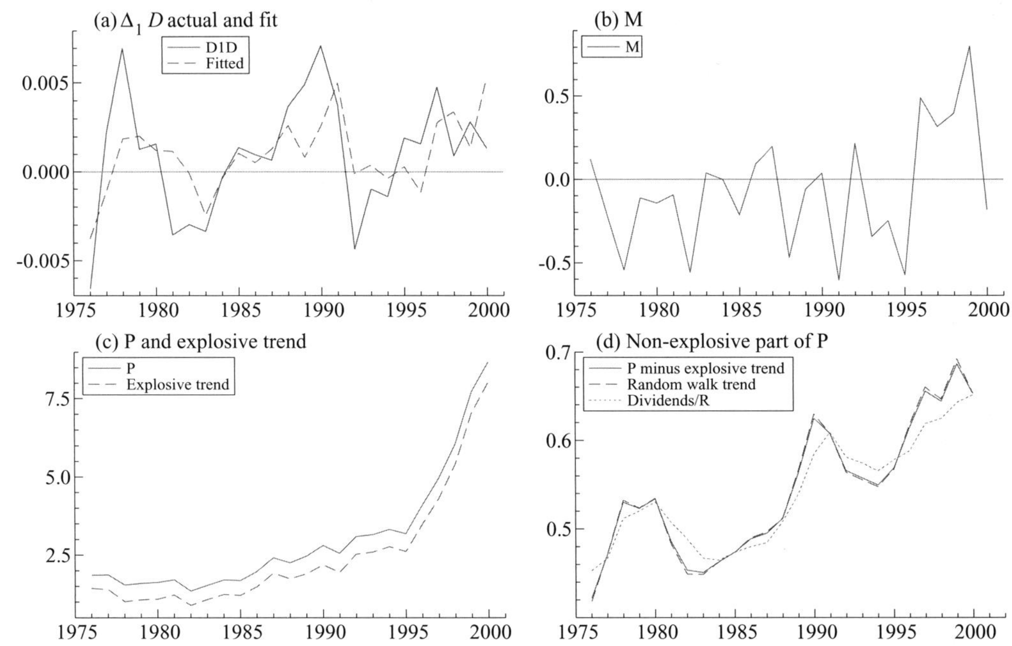


**Figure 1.** Results from $\mathsf{M}_{1DSB}$.

**Table 2.** Cointegration rank determination.

| Hypothesis | Likelihood | Test statistic | $P$ value |
|---|---|---|---|
| $r \leq 2$ | 108.71 | | |
| $r \leq 1$ | 105.91 | $LR(r \leq 1 \mid \mathsf{M}) = 5.6$ | 0.23 |
| $r = 0$ | 98.21 | $LR(r = 0 \mid \mathsf{M}) = 21.1$ | 0.04 |

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

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 $\beta_1'$ 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 $\beta_1$.

Table 3 reports the various tests associated with the bubble hypothesis. First, the test for the hypothesis $\mathsf{H}_D$ that dividends are non-explosive against $\mathsf{M}_1$ is asymptotically $\chi^2$ 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 system belongs to $P_t$ and not $D_t$. Secondly, the test for the joint bubble hypothesis $\mathsf{H}_S, \mathsf{H}_B$ against $\mathsf{M}_{1D}$ is asymptotically $\chi^2$ due

**Table 3.** Tests of the rational bubble restrictions.

| Model | Hypothesis | Likelihood | Test statistic | d.f. | $P$ value |
|---|---|---|---|---|---|
| $\mathsf{M}_1$ | $\mathsf{H}_1, r = 1$ | 105.91 | | | |
| $\mathsf{M}_{1D}$ | $\mathsf{H}_1, \mathsf{H}_D$ | 105.91 | $LR(\mathsf{M}_{1D}\vert\mathsf{M}_1) = 0.0002$ | 1 | 0.99 |
| $\mathsf{M}_{1DS}$ | $\mathsf{H}_1, \mathsf{H}_D, \mathsf{H}_S$ | 105.75 | $LR(\mathsf{M}_{1DS}\vert\mathsf{M}_{1D}) = 0.32$ | 1 | 0.57 |
| | | | $LR(\mathsf{M}_{1DS}\vert\mathsf{M}_1) = 0.32$ | 2 | 0.85 |
| $\mathsf{M}_{1DSB}$ | $\mathsf{H}_1, \mathsf{H}_D, \mathsf{H}_S, \mathsf{H}_B$ | 102.20 | $LR(\mathsf{M}_{1DSB}\vert\mathsf{M}_{1DS}) = 7.10$ | 3 | 0.07 |
| | | | $LR(\mathsf{M}_{1DSB}\vert\mathsf{M}_{1D}) = 7.40$ | 4 | 0.12 |
| | | | $LR(\mathsf{M}_{1DSB}\vert\mathsf{M}_1) = 7.40$ | 5 | 0.19 |

to Theorem 3.3 and gives a $P$ value of 0.12 when tested against $\mathsf{M}_{1D}$. Since the latter test has four degrees of freedom and a relatively low $P$ value it is of interest to evaluate various intermediate hypotheses noting that a $\chi^2$-distribution theory has not been formally established.

The first of the intermediate hypotheses is the $\mathsf{H}_S$ hypothesis that $\beta_1' = (1, -1/R)$ with $\rho = 1 + R$. The $P$ value is 0.57 when tested against $\mathsf{M}_{1D}$ 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, $\mathsf{H}_B$, gives a $P$ value of 0.07 if tested against model $\mathsf{M}_{1DS}$, and a $P$ value of 0.19 if tested against model $\mathsf{M}_1$. The decision not to reject the hypothesis $\mathsf{H}_B$ is marginal against $\mathsf{M}_{1DS}$ albeit more convincing against $\mathsf{M}_1$. 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 $\zeta_1$ unconstrained gives a likelihood of 104.65, so a test statistic of 2.20 ($P$ value: 0.33) against $\mathsf{M}_{1DS}$. Thus the test statistic for $\mathsf{M}_{1DSB}$ against this intermediate hypothesis is 4.89 ($P$ value: 0.027).

Proceeding with the model $\mathsf{M}_{1DSB}$, 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 $\mathsf{M}_{1DSB}$ is

$$M_t = \hat{\varepsilon}_{M,t}, \quad \mathsf{sdv} = 0.354,$$

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

Figures 1(a) and (b) show graphs of the fitted values of $\Delta_1 D_t$ and $M_t$. Figure 1(c) shows the actual prices $P_t$ and the fitted explosive trend $(\hat{\rho} - 1)^{-1}(1, 0)\hat{C}_\rho \hat{\rho}^t(\hat{A}_\rho + \sum_{s=1}^t \hat{\rho}^{-s}\hat{\varepsilon}_s)$ given by Theorem 2.1. This explosive trend contributes with a substantial part of the movements in prices over time consistent with the bubble model. Figure 1(d) compares the non-explosive part of $P_t$, i.e. the difference between $P_t$ and the explosive trend, with the dividends scaled by $R$ and 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, $\mathsf{H}_B$ against $\mathsf{M}_{1DS}$, 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.

## ACKNOWLEDGMENTS

Comments from the referees are gratefully acknowledged. The first author was supported by CREATES, which is financed by the Danish National Research Foundation. The second author was supported by ESRC grant RES-000-27-0179 and the Institute for New Economic Thinking at the Oxford Martin School.

## REFERENCES

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics 30*, 676–87.

Brown, B. M. and G. K. Eagleson (1971). Martingale convergence to infinitely divisible laws with finite variance. *Transactions of the American Mathematical Society 162*, 449–53.

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

Campbell, J. Y. and R. J. Shiller (1988). Interpreting cointegrated models. *Journal of Economic Dynamics and Control 12*, 505–22.

Chan, N. H. and C. Z. Wei (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*, 1533–75.

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

Doornik, J. A. (1998). Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys 12*, 573–93.

Doornik, J. A. and D. F. Hendry (2001). *Empirical Econometric Modelling Using PcGive 10, Volumes 1 and 2*. London: Timberlake Consultants Press.

Engler, E. and B. Nielsen (2009). The empirical process of autoregressive residuals. *Econometrics Journal 12*, 367–81.

Engsted, T. (2006). Explosive bubbles in the cointegrated VAR model. *Finance Research Letters 3*, 154–62.

Flood, R. P. and P. M. Garber (1980). Market fundamentals versus price-level bubbles: the first tests. *Journal of Political Economy 88*, 745–70.

Froot, K. A. and M. Obstfeld (1991). Intrinsic bubbles: the case of stock prices. *American Economic Review 81*, 1189–214.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Johansen, S. and A. R. Swensen (1999). Testing exact rational expectations in cointegrated vector autoregressive models. *Journal of Econometrics 93*, 73–91.

Johansen, S. and A. R. Swensen (2004). More on testing exact rational expectations in cointegrated vector autoregressive models: restricted constant and linear term. *Econometrics Journal 7*, 389–97.

Johansen, S. and A. R. Swensen (2011). On a graphical technique for evaluating some rational expectations models. *Journal of Time Series Econometrics 3*, Issue 1, Article 9.

Lai, T. L. and C. Z. Wei (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P. R. Krishnaiah (Ed.), *Multivariate Analysis, Volume VI*, 375–93. Amsterdam: Elsevier.

Leroy, S. F. (1989). Efficient capital markets and martingales. *Journal of Economic Literature 27*, 1583–621.

Mirsky, L. (1961). *An Introduction to Linear Algebra.* Oxford: Clarendon Press.

Mynbaev, K. T. (2011). *Short-Memory Linear Processes and Econometric Applications.* Hoboken, NJ: Wiley.

Nielsen, B. (2001). The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica 69*, 211–19.

Nielsen, B. (2005). Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econometric Theory 21*, 534–61.

Nielsen, B. (2006a). Correlograms for non-stationary autoregressions. *Journal of the Royal Statistical Society*, Series B *68*, 707–20.

Nielsen, B. (2006b). Order determination in general vector autoregressions. In H.-C. Ho, C.-K. Ing and T. L. Lai (Eds.), *Time Series and Related Topics: In Memory of Ching-Zong Wei, IMS Lecture Notes and Monograph Series, Volume 52*, 93–112. Beachwood, OH: Institute of Mathematical Statistics.

Nielsen, B. (2010). Analysis of coexplosive processes. *Econometric Theory 26*, 882–915.

Nielsen, B. and J. S. Sohkanen (2011). Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends. *Econometric Theory 27*, 913–27.

Shiller, R. J. (2000). *Irrational Exuberance.* Princeton, NJ: Princeton University Press.

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*, 553–80.

## APPENDIX A: PROOF OF GRANGER–JOHANSEN REPRESENTATION

**Proof of Theorem 2.1: Homogeneous equation.** Let $\tilde{X}_t = X_t + \overline{\beta}_1'\zeta_1/(1-\rho)$ so $\Delta_1 X_t = \Delta_1 \tilde{X}_t$ and $\Delta_\rho X_t = \Delta_\rho \tilde{X}_t + \overline{\beta}_1'\zeta_1$. Since $\mu = \alpha_1\zeta_1$ then $\mu + \alpha_1\beta_1'\overline{\beta}_1'\zeta_1 = 0$. Insert this in (2.7) to see that $\tilde{X}_t$ solves

$$\Delta_1\Delta_\rho\tilde{X}_t = \alpha_1\beta_1'\Delta_\rho\tilde{X}_{t-1} + \alpha_\rho\beta_\rho'\Delta_1\tilde{X}_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho\tilde{X}_{t-j} + \varepsilon_t. \tag{A.1}$$

This homogeneous equation is studied in Nielsen (2010, Theorem 1) which therefore gives the desired representation for $X_t$ apart from the explicit expressions for the terms $A_1, A_\rho, Y_t$. □

**Decomposition.** To find the expressions for $A_1, A_\rho, Y_t$ decompose

$$\tilde{X}_t = \frac{y\Delta_x\tilde{X}_t - x\Delta_y\tilde{X}_t}{y-x} = \frac{y(\beta_{y\perp}\overline{\beta}_{y\perp}' + \overline{\beta}_y\beta_y')\Delta_x\tilde{X}_t}{y-x} + \frac{x(\beta_{x\perp}\overline{\beta}_{x\perp}' + \overline{\beta}_x\beta_x')\Delta_y\tilde{X}_t}{x-y}. \tag{A.2}$$

It will be shown below that

$$x\beta_{x\perp}\overline{\beta}_{x\perp}'\Delta_y\tilde{X}_t = C_x x^t\left(A_x + \sum_{s=1}^{t}x^{-s}\varepsilon_s\right) + C_x\frac{\alpha_y\beta_y'}{y-x}\Delta_x\tilde{X}_t$$
$$-C_x\Psi_x\overline{\beta}_x\beta_x'\Delta_y\tilde{X}_t - C_x\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}\Delta_x\Delta_y\tilde{X}_{t-h}. \tag{A.3}$$

Inserting (A.3) into (A.2) and keeping track of the terms gives the expressions for $A_1, A_\rho, Y_t$. □

**The identity (A.3).** Note the identities

$$(y-x)\,\Delta_x \tilde{X}_{s-1} = \Delta_x^2 \tilde{X}_s - \Delta_x \Delta_y \tilde{X}_s,$$

$$\sum_{h=0}^{j-1} x^h \Delta_x^2 \Delta_y \tilde{X}_{s-h} = \Delta_x \Delta_y \tilde{X}_s - x^j \Delta_x \Delta_y \tilde{X}_{s-j}.$$

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

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

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 $I_p = \overline{\beta}_x\beta'_x + \beta_{x\perp}\overline{\beta}'_{x\perp}$, recall $A_x$ and rearrange to get

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

The matrix $\alpha'_{x\perp}\Psi_x\beta_{x\perp}$ is invertible by Assumption 2.1, so pre-multiply with its inverse and then by $x\beta'_{x\perp}$ to get (A.3). □

## APPENDIX B: PROOF OF ASYMPTOTIC RESULTS

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

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

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

$$X_t = \begin{pmatrix} P_t + D_t \\ D_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} P_t \\ D_t \end{pmatrix}. \tag{B.2}$$

Accordingly, the error term of the model equation is denoted

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

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

$$S_t = \beta_1' X_t = P_t - R^{-1} D_t = (P_t + D_t) - \mathcal{G} D_t \qquad \text{with} \qquad \mathcal{G} = R^{-1} + 1. \tag{B.3}$$

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

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

*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 $\mathsf{M}_{1DSB}$. Introduce the vector $\mathcal{S}_t^\circ = (U_t^{\circ\prime}, V_t^{\circ\prime}, W_t^{\circ\prime})'$, where

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

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

The data generating process is

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

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

*B.1.3. Some further parameters.* From Nielsen (2010), it is known that the analysis of the unrestricted model $\mathsf{M}_{1D}$ involves the parameters

$$\tau_\perp^\circ = \Psi_\rho^\circ \overline{\beta}_{\rho\perp}^\circ, \qquad \Psi_\rho^\circ = I_p + \frac{\Pi_1^\circ}{1-\rho_\circ} - \sum_{j=1}^{k-2} \rho_\circ^{-j}\Phi_j^\circ, \tag{B.8}$$

as well as the projection matrices

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

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

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

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

$$\mathcal{G}^\circ = R_\circ^{-1} + 1, \qquad \mathcal{H}^\circ = \{\mathcal{G}^\circ, R_\circ^{-1}, 0, (1+R_\circ)^{-1}\overline{\beta}^\circ_{\rho\perp}, \ldots, (1+R_\circ)^{2-k}\overline{\beta}^\circ_{\rho\perp}\}'. \tag{B.11}$$

*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, sections 6 and 8.5). Introduce the block diagonal normalisation matrix $N_S = \mathrm{diag}(I_{\dim U}, N_V, N_W, 1)$ where

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

LEMMA B.1. *Let $X_t$ satisfy $\mathsf{M}_{1DSB}$ and be given by (B.7). Assume Assumptions 2.1, 3.1, 3.2. Let $\xi$ be a constant satisfying $\xi < \gamma/(2+\gamma)$, recalling $\gamma$ defined in Assumption 3.1.*

*Define sample variances $\widehat{\mathsf{var}}(x_t) = T^{-1}\sum_{t=1}^T x_t x_t'$. Then stochastic matrices $\Sigma_{WW}$, $\Sigma_{VV}$, $\Sigma_{SS}$ and a deterministic matrix $\Sigma_{UU}$ exist so (a) $\widehat{\mathsf{var}}(\varepsilon_t^\circ) \overset{a.s.}{=} \Omega_\circ + \mathrm{o}(T^{-\xi}) + \mathrm{o}(T^{\eta-1/2})$ for all $\eta > 0$, (b) $\widehat{\mathsf{var}}(U_{t-1}^\circ) \overset{a.s.}{\to} \Sigma_{UU} > 0$, (c) $\hat\Sigma_{WW} = \rho_\circ^{-2T}\sum_{t=1}^T (W_{t-1}^\circ)^2 = \rho_\circ^{-2T} T\widehat{\mathsf{var}}(W_{t-1}^\circ) \overset{a.s.}{\to} \Sigma_{WW} \overset{a.s.}{>} 0$, (d) $\hat\Sigma_{VV} = \widehat{\mathsf{var}}(N_V V_{t-1}^\circ) \overset{\mathsf{d}}{\to} \Sigma_{VV} \overset{a.s.}{>} 0$, (e) $\widehat{\mathsf{var}}(N_S S_{t-1}^\circ) \overset{\mathsf{d}}{\to} \Sigma_{SS} \overset{a.s.}{>} 0$.*

*Define sample correlations $\widehat{\mathsf{corr}}(x_t, y_t) = (\sum_{t=1}^T x_t^{\otimes 2})^{-1/2}\sum_{t=1}^T x_t y_t'(\sum_{t=1}^T y_t^{\otimes 2})^{-1/2}$, so (f) $\widehat{\mathsf{corr}}(S_{t-1}^\circ, \varepsilon_t^\circ) \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2})$, (g) $\widehat{\mathsf{corr}}\{(U_{t-1}^{\circ\prime}, V_{t-1}^{\circ\prime}, 1)', \varepsilon_t^\circ\} = \mathrm{O}_p(T^{-1/2})$.*

*In addition, it holds jointly for some stochastic matrices $\Sigma_{V\varepsilon}, \Sigma_{U\varepsilon}, \Sigma_{VU}$ that (h) $\hat\Sigma_{U\epsilon} = T^{-1/2}\sum_{t=1}^T (U_{t-1}^\circ)\varepsilon_t^{\circ\prime} \overset{\mathsf{d}}{\to} \Sigma_{U\varepsilon}$. and $\hat\Sigma_{U\epsilon} \overset{a.s.}{=} \mathrm{o}(T^\eta)$ for all $\eta > 0$, (i) $\hat\Sigma_{V\epsilon} = T^{-1/2}\sum_{t=1}^T N_V(V_{t-1}^\circ)\varepsilon_t^{\circ\prime} \overset{\mathsf{d}}{\to} \Sigma_{V\varepsilon}$, (j) $\hat\Sigma_{W\epsilon} = \rho_\circ^{-T}\sum_{t=1}^T W_{t-1}^\circ \epsilon_t^\circ \overset{a.s.}{=} \mathrm{o}(T^{(1-\xi)/2})$.*

**Proof:** Similar results are proved in Nielsen (2010, Lemma A.1) for the 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 $\mathsf{M}_1$ and $\mathsf{M}_{1D}$*

**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 Nielsen (2010, Lemma A.1) which in turn feeds into Johansen (1995, Lemma 11.1). □

**Proof of Theorem 3.2:** The polynomial order of the deterministic term is not crucial in the proof of Nielsen (2010, Theorem 4). 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

$$\mathcal{R}_t = (M_t, U_{t-1}')', \quad U_t = \{(\beta_1'\Delta_\rho X_t)', (\beta_\rho'\Delta_1 X_t)', (\Delta_1\Delta_\rho X_t)', \ldots, (\Delta_1\Delta_\rho X_{t-k+3})'\}'$$

for some $\rho$. 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.1 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

$$\begin{pmatrix} S^{\circ}_{\varepsilon_M \varepsilon_M} & S^{\circ}_{\varepsilon_M W} \\ S^{\circ}_{W \varepsilon_M} & S^{\circ}_{WW} \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} M^{\circ}_t \\ W^{\circ}_{t-1} \end{pmatrix}^{\otimes 2},$$

$$\begin{pmatrix} S^{\circ}_{DD\cdot M} & S^{\circ}_{\varepsilon_{D\cdot M} \mathcal{R}} & S^{\circ}_{\varepsilon_{D\cdot M} W} \\ S^{\circ}_{\mathcal{R} \varepsilon_{D\cdot M}} & S^{\circ}_{\mathcal{R}\mathcal{R}} & S^{\circ}_{\mathcal{R}W} \end{pmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \varepsilon^{\circ}_{D\cdot M,t} \\ \mathcal{R}^{\circ}_t \end{pmatrix} \left( \varepsilon^{\circ}_{D\cdot M,t}, \mathcal{R}^{\circ\prime}_t, W^{\circ}_{t-1} \right),$$

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

$$\begin{pmatrix} S^{\circ}_{\mathcal{R}\mathcal{R}\cdot\Delta} & S^{\circ}_{\mathcal{R}W\cdot\Delta} \\ S^{\circ}_{W\mathcal{R}\cdot\Delta} & S^{\circ}_{WW\cdot\Delta} \end{pmatrix} = \begin{pmatrix} S^{\circ}_{\mathcal{R}\mathcal{R}} & S^{\circ}_{\mathcal{R}W} \\ S^{\circ}_{W\mathcal{R}} & S^{\circ}_{WW} \end{pmatrix} - \begin{pmatrix} S^{\circ}_{\mathcal{R}\Delta} \\ S^{\circ}_{W\Delta} \end{pmatrix} S^{-1}_{\Delta\Delta} \left( S^{\circ}_{\Delta\mathcal{R}}, S^{\circ}_{\Delta W} \right).$$

LEMMA B.2. *Assume Assumptions 2.1, 3.1, 3.2. Define* $\mathcal{D}_R = (R - R_\circ)$ *and* $\mathcal{I}_R = (R^{-1} - R_\circ^{-1})$. *Let* $\mathrm{o}_{pol} = \mathrm{o}(T^{-k})$ *for some finite k not depending on R. Then*

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon^{\circ}_{D\cdot M,t} \mathcal{R}'_t \overset{a.s.}{=} S^{\circ}_{\varepsilon_{D\cdot M} \mathcal{R}} - \mathcal{D}_R S^{\circ}_{\varepsilon_{D\cdot M} W} \mathcal{H}^{\circ\prime} + (\mathcal{D}_R + \mathcal{I}_R) \mathrm{o}_{pol},$$

$$\frac{1}{T} \sum_{t=1}^{T} \mathcal{R}^{\circ}_t \mathcal{R}'_t \overset{a.s.}{=} S^{\circ}_{\mathcal{R}\mathcal{R}} - \mathcal{D}_R S^{\circ}_{\mathcal{R}W} \mathcal{H}^{\circ\prime} + (\mathcal{D}_R + \mathcal{I}_R) \mathrm{o}_{pol},$$

$$\frac{1}{T} \sum_{t=1}^{T} \mathcal{R}_t \mathcal{R}'_t \overset{a.s.}{=} S^{\circ}_{\mathcal{R}\mathcal{R}} - \mathcal{D}_R (S^{\circ}_{\mathcal{R}W} \mathcal{H}^{\circ\prime} + \mathcal{H}^{\circ} S^{\circ}_{W\mathcal{R}}) + \mathcal{D}_R^2 \mathcal{H}^{\circ} S^{\circ}_{WW} \mathcal{H}^{\circ\prime}$$
$$+ (\mathcal{D}_R + \mathcal{I}_R)(1 + \rho_\circ^T \mathcal{D}_R + \mathcal{D}_R + \mathcal{I}_R) \mathrm{o}_{pol}.$$

*In particular,*

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon^{\circ}_{D\cdot M,t} M'_t \overset{a.s.}{=} S^{\circ}_{\varepsilon_{D\cdot M} \varepsilon_M} - \mathcal{D}_R (\mathcal{G}^{\circ} S^{\circ}_{\varepsilon_{D\cdot M} W} + \mathrm{o}_{pol}),$$

$$\frac{1}{T} \sum_{t=1}^{T} M_t^2 \overset{a.s.}{=} S^{\circ}_{\varepsilon_M \varepsilon_M} - 2\mathcal{D}_R (\mathcal{G}^{\circ} S^{\circ}_{\varepsilon_M W} + \mathrm{o}_{pol}) + \mathcal{D}_R^2 \{\mathcal{G}^{\circ 2} S^{\circ}_{WW} + \mathrm{o}(1)\}. \tag{B.13}$$

**Proof:** **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

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

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

$$\Delta_1 \Delta_{1+R} X_{t-j} = \Delta_1 \Delta_{1+R_\circ} X_{t-j} + (R_\circ - R)(1+R_\circ)^{-j} \left\{ \Delta_1 X_{t-1} - \sum_{l=1}^{j} (1+R_\circ)^{l-1} \Delta_1 \Delta_{1+R_\circ} X_{t-l} \right\}. \tag{B.15}$$

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

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

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 $(R_\circ^{-1} + 1)W_{t-1}^\circ$ 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

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

Due to (B.14) then $P_{t-2}$ is the sum of $R_\circ^{-1} W_{t-1}^\circ$ and a 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 $I_p = \overline{\beta}_{\rho\perp}^\circ \beta_{\rho\perp}^{\circ\prime} + \overline{\beta}_\rho^\circ \beta_\rho^{\circ\prime}$, it is seen that $\Delta_1 \Delta_{1+R} X_{t-j}$ is the sum of $\Delta_1 \Delta_{1+R_\circ} X_{t-j} - \mathcal{D}_R (1+R_\circ)^{-j} \overline{\beta}_{\rho\perp}^\circ \beta_{\rho\perp}^{\circ\prime} \Delta_1 X_{t-1}$ and some linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$.

**Product sums.** The first component of interest is

$$\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{D\cdot M,t}^\circ \mathcal{R}_t' \overset{a.s}{=} \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{D\cdot M,t}^\circ \mathcal{R}_t^{\circ\prime} - \mathcal{D}_R \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{D\cdot M,t}^\circ W_{t-1}^\circ \mathcal{H}' - \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{D\cdot M,t}^\circ (U_{t-1}^{\circ\prime}, V_{t-1}^\circ)(\mathcal{D}_R \mathcal{H}_2' + \mathcal{I}_R \mathcal{H}_3 e_2').$$

The processes $U_{t-1}^\circ, V_{t-1}^\circ$ are of polynomial order, see Nielsen (2005, Theorem 5.1), so $T^{-1}\sum_{t=1}^{T} \varepsilon_{D\cdot M,t}^\circ (U_{t-1}^{\circ\prime}, V_{t-1}^\circ)' = \mathrm{o}_{pol}$ $a.s$. 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 $S_1 = T^{-1}\sum_{t=1}^{T} \mathcal{R}_t^\circ (U_{t-1}^{\circ\prime}, V_{t-1}^\circ)'(\mathcal{D}_R \mathcal{H}_2' + \mathcal{I}_R \mathcal{H}_3 e_2')$,

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

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 $\sum_{t=1}^T \mathcal{R}_t^\circ(U_{t-1}^{\circ\prime}, V_{t-1}^\circ)' = \mathrm{o}_{pol}\ a.s.$

By a similar argument then, with $S_1$ as above and

$$S_2 = \mathcal{D}_R\mathcal{H}^\circ \frac{1}{T}\sum_{t=1}^T W_{t-1}^\circ(U_{t-1}^{\circ\prime}, V_{t-1}^\circ)(\mathcal{D}_R\mathcal{H}_2' + \mathcal{I}_R\mathcal{H}_3 e_2'),$$

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

it holds that

$$\frac{1}{T}\sum_{t=1}^T \mathcal{R}_t\mathcal{R}_t' = \frac{1}{T}\sum_{t=1}^T \mathcal{R}_t^\circ\mathcal{R}_t^{\circ\prime} - \mathcal{D}_R\frac{1}{T}\sum_{t=1}^T (\mathcal{R}_t^\circ W_{t-1}^\circ\mathcal{H}^{\circ\prime} + \mathcal{H}^\circ W_{t-1}^\circ\mathcal{R}_t^\circ)'$$
$$+ \mathcal{D}_R^2\frac{1}{T}\sum_{t=1}^T \mathcal{H}^\circ(W_{t-1}^\circ)^2\mathcal{H}^{\circ\prime} - S_1 - S_1' + S_2 + S_2' + S_3.$$

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

### *B.3. Consistency under* $\mathsf{H}_B$

The expression for the martingale difference $M_t$ is quadratic in the unknown parameters, so global consistency can be proved under $\mathsf{H}_B$ in contrast to the coexplosive analysis in Nielsen (2010). The starting point is the profile log likelihood for the parameter $R$. Let $\hat\sigma_M^2(R)$ and $\hat\sigma_{D\cdot M}^2(R)$ denote the residual variances of the regression equations (2.11) and (2.12). Then the profile log likelihood is

$$\ell(R) = -\frac{T}{2}\log\left\{\hat\sigma_M^2(R)\hat\sigma_{D\cdot M}^2(R)\right\}. \tag{B.16}$$

The residual variances at $R_\circ$ satisfy the following results.

LEMMA B.3. *Assume Assumptions 2.1, 3.1, 3.2. Then (a)* $\hat\sigma_M^2(R_\circ) = T^{-1}\sum_{t=1}^T M_t^{\circ 2} = S_{\varepsilon_M\varepsilon_M}^\circ \overset{a.s.}{\to} \sigma_{MM}^\circ$, *(b)* $\hat\sigma_{D\cdot M}^2(R_\circ) \overset{a.s.}{\to} \sigma_{DD\cdot M}^\circ$, *(c)* $2\ell(R_\circ) = -T\log\det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD\cdot M}^{-1}\hat\Sigma_{\varepsilon_{D\cdot M}U}\Sigma_{UU}^{\circ-1}\hat\Sigma_{U\varepsilon_{D\cdot M}} + \mathrm{o}(1)$.

**Proof:** (a) Use $M_t^\circ = \varepsilon_{M,t}^\circ$.

(b) Note $\hat\sigma_{D\cdot M}^2(R_\circ) = T^{-1}\sum_{t=1}^T(\varepsilon_{D\cdot M,t}^\circ|\mathcal{R}_t^\circ)^2$. Since $\mathcal{R}_t = (\varepsilon_{M,t}^\circ, U_{t-1}^{\circ\prime})'$, see (B.6), and $\varepsilon_{D\cdot M,t}^\circ, \varepsilon_{M,t}^\circ, U_{t-1}^\circ$ are asymptotically uncorrelated then Lemma B.1(a) implies

$$\hat\sigma_{D\cdot M}^2(R_\circ) \overset{a.s.}{=} S_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}}^\circ$$
$$- \left\{S_{\varepsilon_{D\cdot M}\varepsilon_M}^\circ S_{\varepsilon_M\varepsilon_M}^{\circ-1}S_{\varepsilon_M\varepsilon_{D\cdot M}}^\circ + S_{\varepsilon_{D\cdot M}U}^\circ S_{UU}^{\circ-1}S_{U\varepsilon_{D\cdot M}}^\circ\right\}\{1 + \mathrm{o}(T^{-1/4})\}.$$

Lemma B.1(b) (g) show $S_{\varepsilon_{D\cdot M}\varepsilon_M}^\circ, S_{\varepsilon_{D\cdot M}U}^\circ = \mathrm{o}(T^{-3/8})$ while $S_{\varepsilon_M\varepsilon_M}^{\circ-1}, S_{UU}^{\circ-1}$ converge so

$$\hat\sigma_{D\cdot M}^2(R_\circ) \overset{a.s.}{=} S_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}}^\circ - S_{\varepsilon_{D\cdot M}\varepsilon_M}^\circ S_{\varepsilon_M\varepsilon_M}^{\circ-1}S_{\varepsilon_M\varepsilon_{D\cdot M}}^\circ - S_{\varepsilon_{D\cdot M}U}^\circ S_{UU}^{\circ-1}S_{U\varepsilon_{D\cdot M}}^\circ + \mathrm{o}(T^{-1}), \tag{B.17}$$

and in particular $\hat\sigma_{D\cdot M}^2(R_\circ) \to \sigma_{DD\cdot M}^\circ$ as desired.

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

$$-T\log\left\{\hat{\sigma}^2_{D\cdot M}(R_\circ)\right\} \overset{a.s.}{=} -T\log\left(S^\circ_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}} - S^\circ_{\varepsilon_{D\cdot M}\varepsilon_M} S^{\circ-1}_{\varepsilon_M\varepsilon_M} S^\circ_{\varepsilon_M\varepsilon_{D\cdot M}}\right) + T S^{\circ-1}_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}} S^\circ_{\varepsilon_{D\cdot M}U} S^{\circ-1}_{UU} S^\circ_{U\varepsilon_{D\cdot M}} + \mathrm{o}(1).$$

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

$$2\ell(R_\circ) = -T\log\left(S^\circ_{\varepsilon_M\varepsilon_M} S^\circ_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}} - S^{\circ 2}_{\varepsilon_{D\cdot M}\varepsilon_M}\right) + T S^{\circ-1}_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}} S^\circ_{\varepsilon_{D\cdot M}U} S^{\circ-1}_{UU} S^\circ_{U\varepsilon_{D\cdot M}} + \mathrm{o}(1).$$

Due to the identity

$$\det(S^\circ_{\varepsilon\varepsilon}) = \det\left\{\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \varepsilon^\circ_{M,t} \\ \varepsilon^\circ_{D,t} \end{pmatrix}^{\otimes 2}\right\} = \det\left\{\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \varepsilon^\circ_{D\cdot M,t} \\ \varepsilon^\circ_{D,t} \end{pmatrix}^{\otimes 2}\right\},$$

the first term of $2\ell(R_\circ)$ is $-T\log\det(S^\circ_{\varepsilon\varepsilon})$. For the second term, note $S^\circ_{UU} \to \Sigma^\circ_{UU}$ and $S^\circ_{\varepsilon_{D\cdot M}\varepsilon_{D\cdot M}} \to \sigma_{DD\cdot M}$ while $T^{1/2} S^\circ_{\varepsilon_{D\cdot M}U} = \hat{\Sigma}_{\varepsilon_{D\cdot M}U}$. □

LEMMA B.4. *Consider the maximum likelihood estimators under* $\mathsf{H}_B$. *Assume Assumptions 2.1, 3.1, 3.2. Then* $\hat{R} - R = \mathrm{o}(T^{1/2}\rho^{-T})$ *a.s.*

**Proof:** Let $R_\circ$ denote the true value of $R$.

*Likelihood value at* $R_\circ$. Lemma B.3 shows $\hat{\sigma}^2_M(R_\circ) = T^{-1}\sum_{t=1}^T \varepsilon^{\circ 2}_{M,t}$ and $\hat{\sigma}^2_{D\cdot M}(R_\circ) = T^{-1}\sum_{t=1}^T (\varepsilon^\circ_{D\cdot M,t}|\mathcal{R}^\circ_t)^2$. Lemma B.1(a), (f) then imply

$$\hat{\sigma}^2_M(R_\circ) \overset{a.s.}{\to} \sigma^{\circ 2}_M, \qquad \hat{\sigma}^2_{D\cdot M}(R_\circ) \overset{a.s.}{\to} \sigma^{\circ 2}_{D\cdot M}.$$

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

$$\liminf_{T\to\infty} \hat{\sigma}^2_M(R_\circ) \overset{a.s.}{=} \sigma^{\circ 2}_M + \delta^2\kappa_P, \qquad \liminf_{T\to\infty} \hat{\sigma}^2_{D\cdot M}(R_\circ) \overset{a.s.}{\geq} \sigma^2_{D\cdot M},$$

where $\kappa_P = \mathcal{G}^{\circ 2}\lim_{T\to\infty}\rho_\circ^{-2T}\sum_{t=1}^T (W^\circ_{t-1})^2 > 0$ *a.s.* for $\rho_\circ = 1 + R_\circ$.

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

$$\hat{\sigma}^2_M(R) \overset{a.s.}{=} S^\circ_{MM} + 2(R_\circ - R)\mathcal{G}^\circ S^\circ_{MW} + (R_\circ - R)^2\mathcal{G}^{\circ 2} S^\circ_{WW} + \mathrm{O}(T^{-1}). \tag{B.18}$$

Due to Lemma B.1(a), (c) and (f) then

$$\hat{\sigma}^2_M(R) \overset{a.s.}{=} \sigma^{\circ 2}_M + T^{-1}\rho_\circ^{2T}(R_\circ - R)^2\kappa_P + 2T^{-1/2}\rho_\circ^T(R_\circ - R)\mathrm{o}(1) + \mathrm{o}(1).$$

Thus, for any $R$ outside a neighbourhood of $R_\circ$ this has the stated limes inferior.

For the second result, write the residual variance as

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

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

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

The sample correlations of $\varepsilon_{D\cdot M,t}^\circ$, $\varepsilon_{M,t}^\circ$ and $\mathcal{S}_{t-1}^\circ$ vanish asymptotically, which implies that $\hat{\sigma}_{D\cdot M}^2(R)$ 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 $|N_W^{-1}(R - R_\circ)| \leq \delta$ as it is large outside the interval. This shows the desired consistency. □

### *B.4 Expanding likelihood under* $\mathsf{H}_B$

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

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

**Proof:** Any matrix $h$ can be decomposed as $h = ABA^{-1}$ where $B$ is a triangular, possibly complex matrix with diagonal elements $\lambda_j$, 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 $\log(1 + x) = x - x^2/2 + x^3/6 + \mathrm{O}(x^4)$ it holds

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

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

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

LEMMA B.6. *Let Assumptions 2.1, 3.1, 3.2 hold. Under* $\mathsf{H}_B$ *the profile likelihood has expansion*

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

*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*

$$\tilde{\ell}_M(R) = -\frac{T}{2}\log\left(S_{\varepsilon_M\varepsilon_M}^\circ - 2\mathcal{D}_R\mathcal{G}^\circ S_{\varepsilon_M W}^\circ + \mathcal{D}_R^2\mathcal{G}^{\circ 2}S_{WW}^\circ\right),$$

$$\tilde{\ell}_{\mathcal{R}\cdot\Delta}(R) = -\frac{T}{2}\log\det\left\{S_{\mathcal{R}\mathcal{R}\cdot\Delta}^\circ - \mathcal{D}_R(S_{\mathcal{R}W\cdot\Delta}^\circ\mathcal{H}^{\circ\prime} + \mathcal{H}^\circ S_{W\mathcal{R}\cdot\Delta}^\circ) + \mathcal{D}_R^2\mathcal{H}^\circ S_{WW\cdot\Delta}^\circ\mathcal{H}^{\circ\prime}\right\},$$

$$\tilde{\ell}_{\mathcal{R}}(R) = -\frac{T}{2}\log\det\left\{S_{\mathcal{R}\mathcal{R}}^\circ - \mathcal{D}_R(S_{\mathcal{R}W}^\circ\mathcal{H}^{\circ\prime} + \mathcal{H}^\circ S_{W\mathcal{R}}^\circ) + \mathcal{D}_R^2\mathcal{H}^\circ S_{WW}^\circ\mathcal{H}^{\circ\prime}\right\}.$$

**Proof:** **Profile likelihood.** This is given by

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

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

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

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

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

$$\log\{\hat{\sigma}^2_M(R)\} \overset{a.s.}{=} \log(S^\circ_{\varepsilon_M\varepsilon_M} - 2\mathcal{D}_R\mathcal{G}^\circ S^\circ_{\varepsilon_M W} + \mathcal{D}_R^2\mathcal{G}^{\circ 2}S^\circ_{WW}) + \mathrm{o}(T^{-1}). \tag{B.20}$$

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

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

The last term does not depend on $R$.

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

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

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

$$\log\det\left\{T^{-1}\sum_{t=1}^T(\mathcal{R}_t)^{\otimes 2}\right\} \overset{a.s.}{=} \tilde{\ell}_R(R) + \mathrm{o}(T^{-1}). \tag{B.22}$$

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

$$\log\det\left\{T^{-1}\sum_{t=1}^T(\mathcal{R}_t|\Delta_1 D_t)^{\otimes 2}\right\} \overset{a.s.}{=} \tilde{\ell}_{R\cdot\Delta}(R) + \mathrm{o}(T^{-1}). \tag{B.23}$$

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

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

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. *Let Assumptions 2.1, 3.1, 3.2 hold. Under* $\mathsf{H}_B$ *then*

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

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

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

**Proof:** **Term** $\tilde{\ell}_M(R)$. This satisfies

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

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

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

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

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

Likewise, the second and third derivatives are

$$\begin{aligned}\tilde{\ell}''_M(R_\circ) &= (2!)(-T/2)\mathcal{G}^{\circ 2}\Big(S^{\circ -1}_{\varepsilon_M\varepsilon_M}S_{WW} - 2S^{\circ -2}_{\varepsilon_M\varepsilon_M}S^{\circ 2}_{\varepsilon_M W}\Big)\\ \tilde{\ell}'''_M(R_\circ) &= (3!)(-T/2)\mathcal{G}^{\circ 3}\Big(2S^{\circ -2}_{\varepsilon_M\varepsilon_M}S^{\circ}_{\varepsilon_M W}S_{WW} - \tfrac{8}{3}S^{\circ -3}_{\varepsilon_M\varepsilon_M}S^{\circ 3}_{\varepsilon_M W}\Big).\end{aligned}$$

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

$$\tilde{\ell}''_M(R_\circ) = -T\mathcal{G}^{\circ 2}S^{\circ -1}_{\varepsilon_M\varepsilon_M}\Sigma^{\circ}_{WW}N_W^{-2}\{1+\mathrm{o}(1)\}, \qquad \tilde{\ell}'''_M(R_\circ) \overset{a.s.}{=} \mathrm{o}(T^{-3/4}\rho_\circ^{3T}).$$

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

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

where $\mathcal{B} = S^{\circ -1}_{\mathcal{R}\mathcal{R}}(S^{\circ}_{\mathcal{R}W}\mathcal{H}^{\circ\prime} + \mathcal{H}^{\circ}S^{\circ}_{W\mathcal{R}})$. By considerations as above, it is seen that $\mathcal{B} = \mathrm{o}(T^{-3/4}\rho_\circ^{T})$ and the derivatives satisfy

$$\begin{aligned}\tilde{\ell}'_{\mathcal{R}}(R_\circ) &= T^{1/2}\mathrm{tr}\{\mathcal{H}^{\circ\prime}S^{\circ -1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W}\}N_W^{-1}, \qquad \tilde{\ell}'''_M(R_\circ) \overset{a.s.}{=} \mathrm{o}(T^{-3/4}\rho_\circ^{3T}),\\ \tilde{\ell}''_M(R_\circ) &\overset{a.s.}{=} -T\,\mathrm{tr}(\mathcal{H}^{\circ\prime}S^{\circ -1}_{\mathcal{R}\mathcal{R}}\mathcal{H}^{\circ})\Sigma^{\circ}_{WW}N_W^{-2}\{1+\mathrm{o}(1)\}.\end{aligned}$$

**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 $S_{WW\cdot\Delta} = S_{WW}\{1+\mathrm{o}(1)\}$. □

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

LEMMA B.8. *Let Assumptions 2.1, 3.1, 3.2 hold. Under* $\mathsf{H}_B$ *then*

$$\begin{aligned}\tilde{\ell}'(R_\circ) &\overset{a.s.}{=} \rho_\circ^{T}\{\tau^{\circ\prime}_{\perp}\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon W} + \mathrm{o}(T^{-1/4})\}\\ -\tilde{\ell}''(R_\circ) &\overset{a.s.}{=} \rho_\circ^{2T}\tau^{\circ\prime}_{\perp}\Omega_\circ^{-1}\tau^{\circ}_{\perp}\Sigma^{\circ}_{WW}\{1+\mathrm{o}(1)\}.\end{aligned}$$

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

$$S_{\mathcal{R}\mathcal{R}}^\circ \overset{a.s.}{\to} \begin{pmatrix} \sigma_{MM}^\circ & 0 \\ 0 & \Sigma_{UU}^\circ \end{pmatrix}, \tag{B.24}$$

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

Since $\Delta_1 D_t$ also satisfies $\Delta_1 D_t = \theta^{\circ\prime}\mathcal{R}_t^\circ + \varepsilon_{D,t}^\circ$ then

$$S_{\Delta\Delta} \overset{a.s.}{=} \sigma_{DD}^\circ + \theta^{\circ\prime}\Sigma_{UU}^\circ\theta^\circ + \mathrm{o}(T^{\eta-1/2}). \tag{B.26}$$

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

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

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

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

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 $\omega^\circ = \sigma_{MM}^{\circ-1}\sigma_{DM}^\circ$,

$$\begin{aligned} S_{\mathcal{R}\mathcal{R}\cdot\Delta}^{\circ-1} - S_{\mathcal{R}\mathcal{R}}^{\circ-1} &\overset{a.s.}{=} \frac{\begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UY}^{\circ-1} \end{pmatrix}\begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ\theta^\circ \end{pmatrix}\left(\sigma_{DM}^\circ, \theta^{\circ\prime}\Sigma_{UU}^\circ\right)\begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UU}^{\circ-1} \end{pmatrix}}{\sigma_{DD}^\circ + \theta^{\circ\prime}\Sigma_{UU}^\circ\theta^\circ - \left(\sigma_{DM}^\circ, \theta^{\circ\prime}\Sigma_{UU}^\circ\right)\begin{pmatrix} \sigma_{MM}^{\circ-1} & 0 \\ 0 & \Sigma_{UU}^{\circ-1} \end{pmatrix}\begin{pmatrix} \sigma_{DM}^\circ \\ \Sigma_{UU}^\circ\theta^\circ \end{pmatrix}} + \mathrm{o}(T^{-1/4}) \\ &= \frac{1}{\sigma_{DD\cdot M}^\circ}\begin{pmatrix} \omega^\circ \\ \theta^\circ \end{pmatrix}^{\otimes 2} + \mathrm{o}(T^{-1/4}). \end{aligned} \tag{B.28}$$

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

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

Combining these expressions shows

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

Finally, the desired expression follows since by partitioned inversion

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

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

$$\hat{\Sigma}_{\mathcal{R}W\cdot\Delta} = \hat{\Sigma}_{\mathcal{R}W} - S^{\circ}_{\mathcal{R}\Delta}S^{-1}_{\Delta\Delta}\hat{\Sigma}_{\Delta W}$$
$$\overset{a.s.}{=} \hat{\Sigma}_{\mathcal{R}W} - \frac{1+\mathrm{o}(T^{\eta-1/2})}{\sigma^{\circ}_{DD}+\theta^{\circ\prime}\Sigma^{\circ}_{UU}\theta^{\circ}}\begin{pmatrix}\sigma^{\circ}_{DM}\\ \Sigma^{\circ}_{UU}\theta^{\circ}\end{pmatrix}(1,\theta^{\circ\prime})\hat{\Sigma}_{\Delta W}.$$

In a similar way, write

$$S^{\circ-1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W} = S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}S^{\circ}_{\mathcal{R}\mathcal{R}\cdot\Delta}S^{\circ-1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W}$$
$$\overset{a.s.}{=} S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}\left\{I_2 - \begin{pmatrix}\sigma^{\circ}_{DM}\\ \Sigma^{\circ}_{UU}\theta^{\circ}\end{pmatrix}\frac{1+\mathrm{o}(T^{\eta-1/2})}{\sigma^{\circ}_{DD}+\theta^{\circ\prime}\Sigma^{\circ}_{UU}\theta^{\circ}}\right\}(\omega^{\circ},\theta^{\circ\prime})\hat{\Sigma}_{\mathcal{R}W}.$$

These expressions combine as

$$S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}\hat{\Sigma}_{\mathcal{R}W\cdot\Delta} - S^{\circ-1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W}$$
$$\overset{a.s.}{=} S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}\frac{\{1+\mathrm{o}(T^{\eta-1/2})\}}{\sigma^{\circ}_{DD}+\theta^{\circ\prime}\Sigma^{\circ}_{UU}\theta^{\circ}}\begin{pmatrix}\sigma^{\circ}_{DM}\\ \Sigma^{\circ}_{UU}\theta^{\circ}\end{pmatrix}\{(\omega^{\circ},\theta^{\circ\prime})\hat{\Sigma}_{\mathcal{R}W} - (1,\theta^{\circ\prime})\hat{\Sigma}_{\Delta W}\}. \tag{B.31}$$

Noting that, see (B.27),

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

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

$$S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}(\sigma^{\circ}_{DM}\Sigma^{\circ}_{UU}\theta^{\circ}) \overset{a.s.}{=} \left\{\begin{pmatrix}\sigma^{\circ-1}_{MM} & 0\\ 0 & \Sigma^{\circ-1}_{UU}\end{pmatrix} + \frac{1}{\sigma^{\circ}_{DD\cdot M}}\begin{pmatrix}\omega^{\circ}\\ \theta^{\circ}\end{pmatrix}^{\otimes 2}\right\}\begin{pmatrix}\sigma^{\circ}_{DM}\\ \Sigma^{\circ}_{UU}\theta^{\circ}\end{pmatrix} + \mathrm{o}(T^{\eta-1/2})$$
$$= \begin{pmatrix}\omega^{\circ}\\ \theta^{\circ}\end{pmatrix}\frac{\sigma^{\circ}_{DD}+\theta^{\circ\prime}\Sigma^{\circ}_{UU}\theta^{\circ}}{\sigma^{\circ}_{DD\cdot M}} + \mathrm{o}(T^{\eta-1/2}),$$

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

$$\mathcal{H}^{\circ\prime}\left(S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}\hat{\Sigma}_{\mathcal{R}W\cdot\Delta} - S^{\circ-1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W}\right) \overset{a.s.}{=} \tau^{\circ\prime}_{\perp}\begin{pmatrix}\omega^{\circ}\\ -1\end{pmatrix}^{\otimes 2}\sigma^{\circ-1}_{DD\cdot M}\hat{\Sigma}_{\varepsilon W}\{1+\mathrm{o}(T^{\eta-1/2})\}.$$

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

$$\mathcal{G}S^{-1}_{MM}\hat{\Sigma}_{\varepsilon_M W} \overset{a.s.}{=} \sigma^{-1}_{MM}\tau^{\circ\prime}_{\perp}\begin{pmatrix}1\\ 0\end{pmatrix}^{\otimes 2}\hat{\Sigma}_{\varepsilon W}\{1+\mathrm{o}(T^{\eta-1/2})\}.$$

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

$$\mathcal{G}S^{-1}_{MM}\hat{\Sigma}_{\varepsilon_M W} + \mathcal{H}'\left(S^{\circ-1}_{\mathcal{R}\mathcal{R}\cdot\Delta}\hat{\Sigma}_{\mathcal{R}W\cdot\Delta} - S^{\circ-1}_{\mathcal{R}\mathcal{R}}\hat{\Sigma}_{\mathcal{R}W}\right)$$
$$\overset{a.s.}{=} \tau^{\circ\prime}_{\perp}\left\{\sigma^{-1}_{MM}\begin{pmatrix}1\\ 0\end{pmatrix}^{\otimes 2} + \sigma^{-1}_{DD\cdot M}\begin{pmatrix}\omega^{\circ}\\ -1\end{pmatrix}^{\otimes 2}\right\}\hat{\Sigma}_{\varepsilon W} + \mathrm{o}(T^{-1/4}).$$

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 $\mathsf{M}_{1DSB}$. Let Assumptions 2.1, 3.1, 3.2 hold. Then $\hat{R} - R = \mathrm{o}(T^{1/4}\rho^{-T})$ a.s.*

**Proof:** Lemma B.4 shows that $\hat{R} - R_\circ = \mathrm{o}(T^{1/2}\rho^{-T})$. 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

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

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

LEMMA B.10. *Consider the maximum likelihood estimators in model $\mathsf{M}_{1DSB}$. Let Assumptions 2.1, 3.1, 3.2 hold. Then (a) $\hat{R} - R_\circ \overset{a.s.}{=} \{-\tilde{\ell}''(R_\circ)\}^{-1}\tilde{\ell}'(R_\circ) + \mathrm{o}(T^{-1/4}\rho_\circ^{-T})$, (b) $\rho_\circ^T(\hat{R} - R_\circ)\tau_\perp^\circ \overset{a.s.}{=} \mathcal{P}_{\tau_\perp}^\circ \hat{\Sigma}_{\varepsilon W}\Sigma_{WW}^{\circ -1}\{1 + \mathrm{o}(1)\} + \mathrm{o}(1)$.*

**Proof:** (a) Lemma B.9 shows that $\hat{R} - R_\circ = \mathrm{o}(T^{1/4}\rho^{-T})$. Insert this and the results of Lemma B.7 into (B.32) to get

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

where $|R_* - R_\circ| \leq |\hat{R} - R_\circ|$. The second term is $\mathrm{o}(T^{-3/4}\rho_\circ^{3T})(T^{1/4}\rho_\circ^{-T})^2\rho_\circ^{-2T}$, which reduces to $\mathrm{o}(T^{-1/4}\rho_\circ^{-T})$ so $(i)$ follows.

(b) Insert the expressions in Lemma B.8 into (a) so that

$$\hat{R} - R_\circ \overset{a.s.}{=} \frac{\rho_\circ^T\{\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon W} + \mathrm{o}(T^{-1/4})\}}{\rho_\circ^{2T}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ\Sigma_{WW}^\circ}\{1 + \mathrm{o}(1)\} + \mathrm{o}(\rho_\circ^{-T}).$$

Rearrange to get the desired result. □

### B.6. Asymptotic distribution of estimators

THEOREM B.1. *Consider the maximum likelihood estimators in model $\mathsf{M}_{1DSB}$. Let Assumptions 2.1, 3.1, 3.2 hold. Then (a) $\{(\sum_{t=1}^T M_t^2)^{1/2}(\hat{\omega} - \omega^\circ), (\sum_{t=1}^T U_t^2)^{1/2}(\hat{\theta} - \theta^\circ)\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, \sigma_{DD\cdot M}^2 I_{2k-1})$, (b) $\hat{\sigma}_{MM} \to \sigma_{MM}$, $\hat{\sigma}_{DD\cdot M} \to \sigma_{DD\cdot M}$ a.s., (c) let $H = (\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ)^{-1/2}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\{\sum_{t=1}^T \rho^{2(t-T)}\}^{-1/2}\sum_{t=1}^T \rho^{t-T}\varepsilon_t^\circ$, then $\{-\ell''(\hat{R})\}^{1/2}(\hat{R} - R_\circ) = H + \mathrm{o}(1)$ a.s., (d) if $(\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ)^{-1/2}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\varepsilon_t^\circ$ are independent $\mathsf{N}(0, 1)$ then $H$ is $\mathsf{N}(0, 1)$.*

**Proof**: (a) Since $\hat\rho - \rho_\circ = \mathrm{o}(T^{1/4}\rho_\circ^{-T})$ by Lemma B.9 and since $\Delta_1 D_t = (\omega^\circ, \theta^{\circ\prime})\mathcal{R}_t^\circ + \varepsilon_{D\cdot M,t}^\circ$ then Lemma B.2 implies that

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

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

(b) First, consider $\hat\sigma_{MM} = T^{-1}\sum_{t=1}^T M_t^2$. Since $\hat\rho - \rho_\circ = \mathrm{o}(T^{1/4}\rho_\circ^{-T})$ by Lemma B.9 then Lemma B.2 implies $\hat\sigma_{MM} = T^{-1}\sum_{t=1}^T M_t^{\circ 2} + \mathrm{o}(1)$ 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 $\Delta_1 D_t = (\omega^\circ, \theta^{\circ\prime})\mathcal{R}_t^\circ + \varepsilon_{D\cdot M,t}^\circ$ then in the same way Lemma B.2 implies $\hat\sigma_{DD\cdot M} = T^{-1}\sum_{t=1}^T \varepsilon_{D\cdot M,t}^{\circ 2} + \mathrm{o}(1)$ which has the desired limit.

(c) Combine Lemmas B.6, B.8, B.10(b) to see that

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

then $\rho_\circ^T(\hat R - R_\circ) = (\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ)^{-1}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\hat\Sigma_{\varepsilon W}\Sigma_{WW}^{\circ -1}\{1+\mathrm{o}(1)\} + \mathrm{o}(1)$ *a.s.* By an argument in Anderson (1959), see also Nielsen (2010, Theorem 4), $H = \hat\Sigma_{\varepsilon W}\Sigma_{WW}^{\circ -1/2} = \{\sum_{t=1}^T \rho^{2(t-T)}\}^{-1/2}\sum_{t=1}^T \rho^{t-T}\varepsilon_t^\circ + \mathrm{o}(1)$ giving the desired result.

(d) Under the normality assumption $H$ is a linear combination of normals, so normal itself. □

### *B.7. Likelihood in restricted model*

LEMMA B.11. *Consider the maximum likelihood estimators in model $\mathsf{M}_{1DSB}$. Let Assumptions 2.1, 3.1, 3.2 hold. Then* $2\ell(\hat R) \overset{a.s.}{=} -T\log\det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD\cdot M}^{-1}\hat\Sigma_{\varepsilon_{D\cdot M}U}\Sigma_{UU}^{\circ -1}\hat\Sigma_{U\varepsilon_{D\cdot M}} + \mathrm{tr}(\Omega_\circ^{-1}\mathcal{P}_{\tau_\perp^\circ}\hat\Sigma_{\varepsilon W}\Sigma_{WW}^{\circ -1}\hat\Sigma_{W\varepsilon}) + \mathrm{o}(1)$.

**Proof:** The profile log likelihood is given in Lemma B.6 as

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

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

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

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

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

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

$$2\{\tilde\ell(\hat R) - \tilde\ell(R_\circ)\} \overset{a.s.}{=} -\frac{\rho_\circ^{2T}\{\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\hat\Sigma_{\varepsilon W} + \mathrm{o}(T^{-1/4})\}^2}{\rho_\circ^{2T}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ\Sigma_{WW}^\circ} + \mathrm{o}(1).$$

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

$$2\{\tilde\ell(\hat R) - \tilde\ell(R_\circ)\} \overset{a.s.}{=} -(\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ)^{-1}\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\hat\Sigma_{\varepsilon W}\Sigma_{WW}^{\circ -1}\Sigma_{W\varepsilon}^\circ\Omega_\circ^{-1}\tau_\perp^\circ + \mathrm{o}(1).$$

Taking trace and rearranging shows

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

Insert this in (B.33) and use the expression for $2\ell(R_\circ)$ in Lemma B.3. □

## *B.8. Likelihood in unrestricted model*

The unrestricted model $\mathsf{M}_{1D}$ 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 $\alpha=(\alpha_1,\alpha_\rho,\Phi_1,\ldots,\Phi_{k-2})$ and for $\Omega$. Given the estimators $\hat{\rho}_D$, $\hat{\beta}_1^{*D}$ the maximum likelihood estimator for $\alpha$ is found by regression of $\Delta_1\Delta_{\hat{\rho}_D}X_t$ on

$$U_{t-1}^D=(\hat{\beta}_1^{*D\prime}\Delta_{\hat{\rho}_D}X_{t-1}^*,\beta_\rho^{\circ\prime}\Delta_1X_{t-1},\Delta_1\Delta_{\hat{\rho}_D}X_{t-1},\ldots,\Delta_1\Delta_{\hat{\rho}_D}X_{t-k+2})'.$$

For the results of this subsection, the data generating process is

$$\Delta_1\Delta_{\rho_\circ}X_t=\alpha_1^\circ\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^*+\alpha_\rho^\circ\beta_\rho^{\circ\prime}\Delta_1X_{t-1}+\sum_{j=1}^{k-2}\Phi_j^\circ\Delta_1\Delta_{\rho_\circ}X_{t-j}+\varepsilon_t^\circ.$$

This of course encompasses the data generating process (B.7) under $\mathsf{M}_{1DSB}$.

LEMMA B.12. *Suppose $\mathsf{M}_{1D}$ holds with $\rho_\circ\geq\varrho$ for some $\varrho>1$. Let Assumptions 2.1, 3.1, 3.2 hold. Recall the definitions of $\tau_\perp^\circ,\mathcal{P}_{\tau_\perp}^\circ,\mathcal{P}_\alpha^\circ$ in (B.8), (B.9). Then (a) $\hat{\alpha}-\alpha=T^{-1/2}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}+\mathrm{o}_p(T^{-\xi})=\mathrm{O}_p(T^{-1/2})$, (b) $\alpha_1^\circ(\hat{\beta}_1^{*D}-\beta_1^{*\circ})'\overline{\hat{\beta}}_{1\perp}^{*D}N_V^{-1}=T^{-1/2}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}+\mathrm{o}_p(T^{-1/2})=\mathrm{O}_p(T^{-1/2})$, (c) $\rho_\circ^{-T}\check{\tau}^{D\prime}\sum_{t=1}^T\hat{\varepsilon}_t^DW_{t-1}^\circ=\rho_\circ^{-T}\tau^{\circ\prime}\sum_{t=1}^T\varepsilon_t^\circ W_{t-1}^\circ+\mathrm{o}_p(1)$, (d) $\hat{\Omega}_D-\Omega_\circ=T^{-1}\sum_{t=1}^T\{(\varepsilon_t^\circ)^{\otimes2}-\Omega_\circ\}+\mathrm{o}_p(T^{-1/2})=\mathrm{O}_p(T^{-1/2})$. (e) $\tau_\perp^\circ(\hat{\rho}_D-\rho_\circ)=\mathcal{P}_{\tau_\perp}\sum_{t=1}^T\varepsilon_t^\circ W_{t-1}^\circ\{\sum_{t=1}^T(W_{t-1}^\circ)^2\}^{-1}\{1+\mathrm{O}_p(T^{-1/2})\}+\mathrm{o}_p(T^{-1/4}\rho_\circ^{-T})$.*

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

$$\begin{pmatrix}\hat{S}_{UU}&\hat{S}_{UV}&\hat{S}_{U\varepsilon}&\hat{S}_{UW}\\ *&\hat{S}_{VV}&\hat{S}_{V\varepsilon}&\hat{S}_{VW}\\ *&*&\hat{S}_{\varepsilon\varepsilon}&\hat{S}_{\varepsilon W}\\ *&*&*&\hat{S}_{WW}\end{pmatrix}=\frac{1}{T}\sum_{t=1}^T\begin{pmatrix}U_{t-1}^D\\ N_VV_{t-1}^D\\ \varepsilon_t^\circ\\ N_WW_{t-1}^\circ\end{pmatrix}^{\otimes2}$$

$$=\begin{pmatrix}\Sigma_{UU}+\mathrm{o}_p(1)&\mathrm{o}_p(T^{-\xi/2})&T^{-1/2}\{\hat{\Sigma}_{U\varepsilon}+\mathrm{o}_p(1)\}&\mathrm{o}_p(T^{-\xi/2})\\ *&\hat{\Sigma}_{VV}+\mathrm{o}_p(1)&T^{-1/2}\{\hat{\Sigma}_{V\varepsilon}+\mathrm{o}_p(1)\}&\mathrm{o}_p(T^{-\xi/2})\\ *&*&\Omega_\circ+\mathrm{o}_p(1)&T^{-1/2}\hat{\Sigma}_{\varepsilon W}\\ *&*&*&\Sigma_{WW}+\mathrm{o}_p(1)\end{pmatrix},\qquad(B.34)$$

where the main terms are given by $\hat{\Sigma}_{W\varepsilon}=\rho_\circ^{-T}\sum_{t=1}^TW_{t-1}^\circ\varepsilon_t^\circ$, $\hat{\Sigma}_{U\varepsilon}=T^{-1/2}\sum_{t=1}^TU_{t-1}^\circ\varepsilon_t^\circ$, $\hat{\Sigma}_{V\varepsilon}=T^{-1/2}N_V\sum_{t=1}^TV_{t-1}^\circ\varepsilon_t^\circ$ and $\hat{\Sigma}_{VV}=T^{-1}\sum_{t=1}^T(N_VV_{t-1}^\circ)^{\otimes2}$.

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

$$\Delta_1\Delta_{\hat{\rho}}X_t=\varepsilon_t^\circ+(\alpha+\hat{\delta}_U)U_{t-1}^D+\hat{\delta}_VN_VV_{t-1}^D+\hat{\delta}_WW_{t-1}^\circ,\qquad(B.35)$$

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

$$\hat{\alpha} - \alpha = \sum_{t=1}^T \varepsilon_t^\circ (U_{t-1}^D)' \left\{ \sum_{t=1}^T \left(U_{t-1}^D\right)^{\otimes 2} \right\}^{-1} + \hat{\delta}_U + \sum_{t=1}^T \left(\hat{\delta}_V N_V V_{t-1}^D + \hat{\delta}_W W_{t-1}^\circ\right)\left(U_{t-1}^D\right)' \left\{ \sum_{t=1}^T \left(U_{t-1}^D\right)^{\otimes 2} \right\}^{-1}.$$

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

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

By Assumption 3.1 then $\xi > 1/2$ and the desired result follows.

(b) (c) Statement of Nielsen (2010, Lemma A.12(b)(c)).

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

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

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

$$\begin{aligned} G = {} & -\Sigma_{\varepsilon U}\Sigma_{UU}^{-1}\Sigma_{U\varepsilon} + \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ\prime} - \mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon} - \hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ\prime} \\ & + \hat{\delta}_W \hat{\Sigma}_{WW}\hat{\delta}_W' + \rho_\circ^{-T}(\hat{\delta}_W \hat{\Sigma}_{W\varepsilon} + \hat{\Sigma}_{\varepsilon W}\hat{\delta}_W') + \mathrm{o}_p(1), \end{aligned} \tag{B.36}$$

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

(e) Nielsen (2010, Lemma A.9) shows

$$\begin{aligned} & \hat{\tau}_\perp^{D\prime}\hat{\Omega}_D^{-1}\rho_\circ^T(\hat{\rho}_D - \rho_\circ)(\check{\tau}_\perp^D \hat{S}_{WW\cdot U} - \hat{\mathcal{P}}_\alpha \check{\tau}_\perp^D \hat{S}_{WV\cdot U}\hat{S}_{VV\cdot U}^{-1}\hat{S}_{VW\cdot U}) \\ & \quad = T^{1/2}\hat{\tau}_\perp^D \hat{\Omega}_D^{-1}(\hat{S}_{\varepsilon W\cdot U} - \hat{\mathcal{P}}_\alpha \hat{S}_{\varepsilon V\cdot U}\hat{S}_{VV\cdot U}^{-1}\hat{S}_{VW\cdot U}), \end{aligned}$$

where $\hat{\mathcal{P}}_\alpha = \check{\alpha}_1(\hat{\alpha}_1^{D\prime}\hat{\Omega}_D^{-1}\check{\alpha}_1)^{-1}\hat{\alpha}_1^{D\prime}\hat{\Omega}_D^{-1}$. Exploit the $T^{1/2}$-order of the $\rho, \alpha, \tau, \Omega$ estimators by (a) (c) as well as pre-multiplying the equation by $\tau_\perp^\circ(\tau_\perp^{\circ\prime}\Omega_\circ^{-1}\tau_\perp^\circ)^{-1}$ and post-multiplying by $\hat{S}_{WW}^{-1}$ to get

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

Exploit the $T^{-1/4}\rho_\circ^T$ 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 $\mathsf{M}_{1D}$. Nielsen (2010, Theorem 3) shows that the estimator $\hat{\Omega}_D$, called $\hat{\Omega}_{\mathsf{H}}$ in that paper, is consistent.

LEMMA B.13. *Suppose $\mathsf{M}_{1D}$ holds with $\rho_\circ \geq \varrho$ for some $\varrho > 1$. Let Assumptions 2.1, 3.1, 3.2 hold. Then*

$$2\hat{\ell}_{1D} = -T \log\det S_{\varepsilon^\circ\varepsilon^\circ} + \mathrm{tr}(\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}) + \mathrm{tr}(\Omega_\circ^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}) + \mathrm{tr}(\Omega_\circ^{-1}\mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}) + \mathrm{o}_{\mathsf{P}}(1).$$

**Proof:** Combine (B.36) where $\hat{\delta}_W = -(\hat{\rho}_D - \rho_\circ)\check{\tau}_\perp^D$ with Lemma B.12(e) to get $\hat{\Omega}_D = S_{\varepsilon^\circ\varepsilon^\circ} + T^{-1}G$ where

$$G = -\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon} + \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ\prime} - \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon} - \hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\mathcal{P}_\alpha^{\circ\prime} + \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\mathcal{P}_{\tau_\perp}^{\circ\prime} - \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon} - \hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\mathcal{P}_{\tau_\perp}^{\circ\prime} + \mathrm{o}_{\mathsf{P}}(1).$$

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

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

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

$$-\mathrm{tr}\left(\hat{\Omega}_D^{-1}G\right) = \mathrm{tr}\left\{\Omega_\circ^{-1}(\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon} + \mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon} + \mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\right\} + \mathrm{o}_{\mathsf{P}}(1).$$

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

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

Inserting results from Lemmas B.11, B.13 gives

$$\mathsf{LR} = \Big\{-T\log\det S_{\varepsilon^\circ\varepsilon^\circ} + \mathrm{tr}\left(\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}\right) + \mathrm{tr}\left(\Omega_\circ^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\right) + \mathrm{tr}\left(\Omega_\circ^{-1}\mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\hat{\Sigma}_{WW}^{-1}\hat{\Sigma}_{W\varepsilon}\right)\Big\} - \Big\{-T\log\det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD\cdot M}^{-1}\hat{\Sigma}_{\varepsilon_{D\cdot M}U}\Sigma_{UU}^{\circ-1}\hat{\Sigma}_{U\varepsilon_{D\cdot M}} + \mathrm{tr}\left(\Omega_\circ^{-1}\mathcal{P}_{\tau_\perp}^\circ\hat{\Sigma}_{\varepsilon W}\Sigma_{WW}^{\circ-1}\hat{\Sigma}_{W\varepsilon}\right)\Big\} + \mathrm{o}_{\mathsf{P}}(1).$$

This reduces to

$$\mathsf{LR} = \mathrm{tr}\left(\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}\right) - \sigma_{DD\cdot M}^{\circ-1}\hat{\Sigma}_{\varepsilon_{D\cdot M}U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon_{D\cdot M}} + \mathrm{tr}\left(\Omega_\circ^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\right) + \mathrm{o}_{\mathsf{P}}(1).$$

In the first term, partitioned inversion of $\Omega_\circ^{-1}$ gives

$$\mathrm{tr}\left(\Omega_\circ^{-1}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon}\right) = \sigma_{MM}^{\circ-1}\hat{\Sigma}_{\varepsilon_M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon_M} + \sigma_{DD\cdot M}^{\circ-1}\hat{\Sigma}_{\varepsilon_{D\cdot M}U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon_{D\cdot M}},$$

so the test statistic satisfies

$$\mathsf{LR} = \sigma_{MM}^{\circ-1}\hat{\Sigma}_{\varepsilon_M U}\Sigma_{UU}^{-1}\hat{\Sigma}_{U\varepsilon_M} + \mathrm{tr}\left(\Omega_\circ^{-1}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\Sigma_{VV}^{-1}\hat{\Sigma}_{V\varepsilon}\right) + \mathrm{o}_{\mathsf{P}}(1).$$

Since $\varepsilon_{M,t}$, $U_{t-1}$ are mutually independent, a martingale central limit theorem, see Brown and Eagleson (1971), gives that the first term is asymptotically $\chi^2$ 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 $c = T^{-1/2} \sum_{s=1}^t \alpha_1^{\circ\prime} \Omega_\circ^{-1} \varepsilon_s$. The process $B_{1,t,T}$ is a function of $T^{-1/2} \sum_{s=1}^t \alpha_{1\perp}^{\circ\prime} \varepsilon_s$. Thus, $B_{1,t,T}$ and $B_{1,t,T}$ converge to asymptotically independent processes, so by a mixed Gaussian argument, see Johansen (1995, section 13.1), the last term is $\chi^2$ with $\dim(\alpha_1^{\circ\prime} \Omega_\circ^{-1} \varepsilon_t) \dim(V) = 2$ 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 $B_{1,t,T}$, $B_{1,t,T}$ 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 $\hat{\Sigma}_{\varepsilon V}$, $\hat{\Sigma}_{VV}$ are functionals of $B_{1,t,T}$, $B_{1,t,T}$ then $\hat{\Sigma}_{\varepsilon V}$, $\hat{\Sigma}_{VV}$ are asymptotically independent of $\hat{\Sigma}_{\varepsilon_M \varepsilon_{D \cdot M}}$, $\hat{\Sigma}_{\varepsilon_M U}$.

It follows that LR is asymptotically $\chi^2$ with $(2k - 2) + 2 = 2k$ degrees of freedom. □

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Table S1:** Profile likelihood under $\mathrm{M}_{1D}$
**Table S2:** Profile likelihood under $\mathrm{M}_{1DS}$
**Table S3:** Profile likelihood under $\mathrm{M}_{1DSB}$
**Table S4:** Profile likelihood under $\mathrm{M}_{1DSB}$
**Table S5:** Profile likelihood under $\mathrm{M}_{1DSB-}$

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.