Testing for rational bubbles in a co-explosive vector autoregression

TOM ENGSTED[1] & BENT NIELSEN[2]

24 June 2010

Summary: We derive the parameter restrictions that a standard equity market model implies for a bivariate vector autoregression for stock prices and dividends, and we show how to test these restrictions using likelihood ratio tests. The restrictions, which imply that stock returns are unpredictable, are derived both for a model without bubbles and for a model with a rational bubble. In both cases we show how the restrictions can be tested through standard chi-squared inference. The analysis for the no-bubble case is done within the traditional *Johansen model* for I(1) variables, while the bubble model is analysed using a co-explosive framework. The methodology is illustrated using US stock prices and dividends for the period 1872-2000.

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

1   Introduction

During the 1980s rational speculative asset bubbles were a hot topic in both theoretical and empirical asset pricing. Tirole (1982, 1985) and Diba and Grossman (1988a), among others, investigated the conditions under which bubbles could occur as an equilibrium phenomenon under rational expectations and informationally efficient capital markets, and Flood and Garber (1980), West (1987), Diba and Grossman (1988b), and Froot and Obstfeld (1991) developed econometric testing procedures for rational bubbles. During the 1990s the research agenda for bubbles was broadened to include irrational motives for the occurance of bubbles. Here, the main focus was on how investor overoptimism and herding behaviour may generate - through various amplifying feed-back mechanisms - long-lasting and irrational deviations between market prices and fundamental values (e.g. Shiller, 2000, and Shleifer, 2000), or on how rational and irrational agents together may generate long-lasting bubbles in models with limits to arbitrage (e.g. Abreu and Brunnermeier, 2003).

The increased recent focus in the literature on irrational and behavioral motives for bubbles has not diminished the professions interest in rational bubbles. Some researchers continue to consider rational bubbles more plausible than irrationally generated bubbles (e.g. Leroy, 2004), and the dramatic stock price increases during the 1990s have led to a revival of econometric analyses of rational bubbles, see e.g. Bohl (2003), Engsted (2006), and Balke and Wohar (2009). The characterizing feature of a rational bubble is that it is explosive, i.e. it generates an explosive root in the autoregressive representation for prices. Diba and Grossman (1988b) used Bhargava's (1986) von Neumann-like statistic to test the null hypothesis of a unit root against the explosive alternative. They also tested for cointegration between prices and fundamentals (dividends), arguing that in a constant discount factor present value model cointegration precludes bubbles while no cointegration would be consistent with the presence of a rational bubble. Diba and Grossman's methodology has been one of the most cited and applied methodologies in the empirical bubble literature, and the appealing feature of the approach is that it explicitly addresses the explosive nature of a rational bubble. However, the approach has a number of important limitations, as emphasized by Engsted (2006). In particular, by using Bhargava's (1986) test for explosiveness the variable under consideration needs to be at most a first-order autoregressive process, and in carrying out this test the discount factor cannot be estimated but is assumed to be known in advance.

In the present paper we suggest an econometric procedure for analyzing rational bubbles that overcomes the limitations of Diba and Grossman's procedure, and at the same time contains statistical tests of restrictions directly implied by the presence of a rational bubble. Our procedure builds on the co-explosive framework developed by Nielsen (2010) and originally designed for hyper-inflation data, see Nielsen (2008a). It is similar in spirit to the work of Campbell and Shiller (1987, 1988a) and Johansen and Swensen (1999, 2004), but with some important differences. Campbell and Shiller investigate the present value model for stock prices without bubbles. Similar to the analysis in Diba and Grossman (1988b), a central variable in Campbell and Shiller's analysis is the "spread", $S_t = P_t - D_t/R$, between prices P_t and dividends D_t normalized by a constant expected return R. They show that if prices and dividends satisfy a vector autoregression integrated of order one, then S_t will be a cointegrating relation and the short term dynamics will satisfy certain restrictions. They propose a two-step procedure in which S_t is first constructed through a single-equation cointegration regression. In a second step additional restrictions implied by the model are imposed on the short term dynamics of a vector autoregression for the "spread" S_t and the dividend growth $\Delta_1 D_t = D_t - D_{t-1}$. As shown by Johansen and Swensen (1999, 2004) then, given S_t, the latter restrictions can also be imposed on a vector autoregression for the observables P_t and D_t.

We review how the restrictions in terms of cointegration and short term dynamics

can be imposed jointly on a cointegrated vector autoregression for P_t, D_t. A key feature of these restrictions is that excess returns follow a martingale difference sequence. In this context the unknown constant expected return parameter, R, is just-identified. If the focus is on the estimation of R in the no-bubble model, only little is gained by the joint estimation approach compared to the two-step procedure. The advantages of the joint approach do, however, become apparent in the bubble case.

In the bubble situation the vector autoregression of prices P_t and dividends D_t involves both a unit root and an explosive root as alluded to by Diba and Grossman (1988b). The "spread" S_t is then a cointegrating relation in the sense of not having a unit root while being explosive, so it is not immediately clear how to estimate the "spread" S_t in a first step of a two-step approach. Instead, the joint restrictions relating to the cointegrating relation S_t, the short term dynamics, as well as non-explosiveness of dividends D_t, can be viewed as joint restrictions on the vector autoregression for P_t, D_t. These restrictions can be analysed in the context of the co-explosive model suggested by Nielsen (2010). We show how the restrictions can be tested through standard χ^2 inference. It is worth noting that in the bubble model the return parameter R is actually over-identified.

We illustrate the methodology using annual US stock prices and dividends for the period 1872-2000. We analyse both the full period and a smaller sub-period, 1974-2000. We find clear evidence of prices being explosive while dividends are non-explosive. We also find a common unit root between prices and dividends (i.e. "cointegration"). The over-identifying restrictions implied by the bubble model are rejected at standard significance levels in the full sample but not in the sub-sample. Hence, our empirical analysis provides some support for the rational bubble hypothesis.

The rest of the paper is organized as follows. In §2 we describe the standard equity market model for stock price determination and explain what a rational bubble implies for prices in this model. We derive the restrictions that the no-bubble model implies for the standard I(1) cointegrated vector autoregression, and we compare these restrictions with the restrictions derived by Johansen and Swensen (1999, 2004). Next, we derive the testable restrictions of a rational bubble on a bivariate co-explosive vector autoregression for prices and dividends. §3 contains the likelihood analyses of the models with the various tests on the cointegrating and co-explosive vectors, their associated factor loadings, and the remaining short term parameters. §4 contains the asymptotic analysis with limiting distributions of parameter estimates and test statistics. §5 reports the empirical results of using the co-explosive framework to analyse US stock prices and dividends. In §6 we briefly compare our testing procedure with earlier bubbles tests. Finally, §7 concludes while proofs are given in an appendix.

2 Stock price determination with and without bubbles

The standard stock price determination model is presented along with the well-known bubble solution. As this model does not specify the behaviour for dividends it is only a partial model so a joint model for prices and dividends is needed for empirical analysis. It is reviewed how the model without bubbles can be embedded in a cointegrated vector autoregression. This leads on to showing how the model with bubbles can be embedded in a co-explosive vector autoregression.

2.1 Standard model for stock price determination

The standard model for stock price determination is

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

where P_t and D_t are real stock prices and dividends, respectively, and R is the expected (required) one-period return on the stock which is assumed to be constant and positive, i.e. $R > 0$. The conditional expectations operator $\mathcal{E}_t$ is taken with respect to the information set at time t, which contains current and past prices and dividends, i.e. P_s, D_s for $s \leq t$, but the exact dependence is left unspecified for the moment.

The stock price determination model has some implications for the variable $S_t \equiv P_t - D_t/R$, which is called a "spread" by Campbell and Shiller (1987, 1988a). Subtracting P_t from equation (2.1), and multiplying by $1 + R$, gives the equation

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

In other words, M_t is a martingale difference. Defining the one-period stock return as $R_t = (P_t + D_t - P_{t-1})/P_{t-1}$, then (2.2) implies that $P_{t-1}^{-1}\mathcal{E}_{t-1}M_t = \mathcal{E}_{t-1}R_t - R = 0$. Thus, the economic interpretation of (2.2) is that the one-period stock return in excess of a constant mean is a martingale difference and thereby unpredictable given lagged information, i.e. the classical version of *The Efficient Markets Hypothesis*, c.f. Leroy (1989). This martingale difference can be expressed in terms of the "spread" variable by adding and subtracting RP_t, and writing price growth as $\Delta_1 P_t \equiv P_t - P_{t-1}$, to get

$$M_t = (1+R)\Delta_1 P_t - RS_t. \tag{2.3}$$

Equation (2.3) shows that if, as often assumed, $X_t = (P_t, D_t)$ is an I(1) variable then S_t must be a cointegrating relation and related to $\Delta_1 P_t$ in a specific way.

This cointegrating relationship can also be formulated as a link between dividend growth $\Delta_1 D_t$ and the spread. To see this add and subtract $\{(1+R)/R\}\Delta_1 D_t$ to get

$$M_t = \Delta_{1+R}S_t + (1 + R^{-1})\Delta_1 D_t, \tag{2.4}$$

4

where $\Delta_{1+R}S_t \equiv S_t - (1 + R)S_{t-1}$. Equation (2.4) shows that if P_t is explosive, as in a rational bubble, while D_t remains $\mathsf{I}(1)$, then S_t will be explosive but without a unit root. The explosiveness is eliminated by the operator Δ_{1+R}.

As indicated, the stock price model is a *partial model*. In order to make interesting statements and to make inference, some further assumptions to the stochastic behaviour of X_t are needed. In effect the partial model is completed by making an assumption to the behaviour of dividends. At first it is useful to discuss the present value formulation of the model, which is common in the literature.

2.2   Present value formulation

The stock price determination model is often stated by writing prices as a present value of expected future dividends. Specifically, the model (2.1) has solution

$$P_t = \sum_{s=1}^{\infty} \frac{\mathcal{E}_t D_{t+s}}{(1 + R)^s} + bB_t, \tag{2.5}$$

for $b \in \mathbb{R}$ and where B_t obeys (c.f. Diba and Grossman, 1988a,b)

$$B_t = \frac{1}{1 + R}\mathcal{E}_t B_{t+1}. \tag{2.6}$$

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 (2.5) and (2.6) solve (2.1) insert these on the right hand side of (2.1). The solution (2.5) requires that the sum $\sum_{s=1}^{\infty}(1 + R)^{-s}\mathcal{E}_t D_{t+s}$ is well-defined.

The rational bubble B_t defined by (2.6) also satisfies

$$\Delta_{1+R}B_{t+1} \equiv B_{t+1} - (1 + R)B_t = \xi_{t+1},$$

where the rational expectations error $\xi_{t+1} = B_{t+1} - \mathcal{E}_t B_{t+1}$ satisfies $\mathcal{E}_t\xi_{t+1} = 0$. Since $1 + R > 1$, the bubble is explosive so unless $b = 0$ it induces explosiveness in P_t.

The present value formulation implies an interesting behaviour for the "spread" variable $S_t = P_t - D_t/R$. Due to the identity $R^{-1} = \sum_{s=1}^{\infty}(1 + R)^{-s}$ then

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

Noting that $D_{t+s} - D_t = \sum_{r=1}^{s}\Delta_1 D_{t+r}$ with $\Delta_1 D_{t+r} \equiv D_{t+r} - D_{t+r-1}$, and that $(1 + R)^r \sum_{s=r}^{\infty}(1 + R)^{-s} = (1 + R)/R$, this in turn implies

$$S_t = P_t - \frac{1}{R}D_t = \frac{1 + R}{R} \sum_{r=1}^{\infty} \frac{\mathcal{E}_t(\Delta_1 D_{t+r})}{(1 + R)^r} + bB_t. \tag{2.7}$$

This equation shows that S_t enherits its stochastic behaviour from $\Delta_1 D_t$ and B_t.

In the following we will make some additional assumptions to the stochastic behaviour of X_t and study their implications. In a first step X_t is assumed vector autoregressive in which case the model (2.1) implies certain linear restrictions on the vector autoregressive coefficients. The solutions with and without bubbles emerge if some additional assumptions are imposed. The first case reviews Campbell and Shiller (1987, 1988a) and Johansen and Swensen (1999, 2004) where the bubble is absent, $b = 0$. Here X_t is assumed to have a unit root so $\Delta_1 D_t$, S_t and, hence, $\Delta_1 P_t$ are stationary and linked as in (2.3). In the second case a bubble is present so $b \neq 0$. Here X_t is assumed to have a unit root and an explosive root so $\Delta_1 D_t$ and $\Delta_{1+R} S_t$ are stationary and linked through (2.4).

2.3 A vector autoregressive framework

The model (2.1) has some interesting implications if $X_t = (P_t, D_t)$ is assumed vector autoregressive. This assumption is common in the literature, see e.g. Campbell and Shiller (1988a) and Johansen and Swensen (1999, 2004), and it is testable. For the moment no assumptions are made to the location of the characteristic roots. A vector autoregression of order k has the format

$$X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu + \varepsilon_t, \tag{2.8}$$

where $A_j \in \mathbb{R}^{2 \times 2}$ and $\mu \in \mathbb{R}^2$. The errors, ε_t, are independent $\mathsf{N}_2(0, \Omega)$-distributed, or more generally a martingale difference sequence.

To see the implications of (2.1) for the vector autoregression (2.8), it is convenient to write (2.8) in companion form

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \boldsymbol{\mu} + \mathbf{e}_t$$

where

$$\mathbf{A} = \left(\begin{array}{c|c} A_1, \ldots A_{k-1} & A_k \\ \hline I_{2k-2} & 0 \end{array} \right), \quad \mathbf{X}_{t-1} = \left(\begin{array}{c} X_{t-1} \\ \vdots \\ X_{t-k} \end{array} \right), \quad \boldsymbol{\mu} = \left(\begin{array}{c} \mu \\ 0 \end{array} \right), \quad \mathbf{e}_t = \left(\begin{array}{c} \varepsilon_t \\ 0 \end{array} \right).$$

Equation (2.2) implies that the companion vector $\mathbf{X}_t$ satisfies

$$\nu' \mathcal{E}_{t-1} \mathbf{X}_t = 0 \qquad \text{where} \qquad \nu' = \{1, 1, -(1+R), 0, \ldots, 0\}.$$

As the innovations satisfy $\mathcal{E}_{t-1} \mathbf{e}_t = 0$, it must hold that $\nu' \mathbf{A} = 0$, which is equivalent to the constraints, for $j = 2, \ldots, k$,

$$(1,1)A_1 = (1+R, 0), \quad (1,1)A_j = 0, \quad (1,1)\mu = 0. \tag{2.9}$$

Thus, (2.1) together with the vector autoregressive assumption implies the constraints (2.9). Some further assumptions to the locations of the characteristic roots of the vector autoregression are needed for two reasons: first, to be able to distinguish between the cases with and without bubbles, and, secondly, to be able to conduct reliable inference about the constraints (2.9).

2.4   The case without a rational bubble

We now analyse the standard case without a rational bubble. Campbell and Shiller (1987, 1988a) discuss this case under the additional assumption that the vector autoregression has one unit root in such a way that the spread, $S_t = P_t - D_t/R$, and dividend growth, $\Delta_1 D_t$, are stationary. Campbell and Shiller estimate the structural parameter R super-consistently in a first-step cointegration regression, and this parameter is subsequently treated as known in a vector autoregressive analysis for $\Delta_1 D_t$ are S_t. They derive the restrictions that (2.1) implies for this particular vector autoregression. Johansen and Swensen (1999, 2004) instead work with a vector autoregression for $X_t = (P_t, D_t)'$ and show how the restrictions implied by (2.1) can be tested in an I(1) framework. In their theoretical analysis they assume that the cointegrating vector S_t is known; however, in their empirical illustrations they discuss the case of unknown cointegration vector. We discuss the latter case in further detail.

It is convenient to reparametrise the vector autoregression in error correction form:

$$\Delta_1 X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} + \mu + \varepsilon_t, \tag{2.10}$$

where $\Pi, \Gamma_j \in \mathbb{R}^{2\times 2}$ and $\mu \in \mathbb{R}^2$. This is equivalent to the vector autoregression (2.8) with $\Pi = \sum_{j=1}^{k} A_j - I_2$ and $\Gamma_j = -\sum_{\ell=j+1}^{k} A_\ell$, see Johansen (1995, §4.1). The restrictions to the A_j-coefficients in (2.9) are then equivalent to, for $j = 1, \ldots, k-1$,

$$(1,1)\Pi = R(1, -1/R), \quad (1,1)\Gamma_j = 0, \quad (1,1)\mu = 0. \tag{2.11}$$

These are the restrictions discussed in the theory part of Johansen and Swensen (1999, 2004) under the assumption that $X_t \sim I(1)$ and reduced rank of Π.

With the assumption that X_t is I(1) with one unit root and the remaining roots stationary, Π has a rank of unity so $\Pi = \alpha\beta'$ where $\beta \in \mathbb{R}^{2\times 1}$ is the cointegrating vector and $\alpha \in \mathbb{R}^{2\times 1}$ is the associated vector of factor loadings. The restriction $(1,1)\Pi = R(1, -1/R)$ in (2.11) then implies that

$$\beta = (1, -1/R)'. \tag{2.12}$$

The last argument also shows that $(1,1)\alpha = R$. It is interesting to note that these arguments do not exploit the multiplicity of the unit root so they would also apply

7

to I(2) situations and explosive situations as will be seen in §2.5. Note also that the restriction in (2.12) identifies the parameter R as pointed out in connection with the illustration of Johansen and Swensen (2004).

Another representation of the process is the Granger-Johansen representation: With the assumption that X_t is I(1) with one unit root, and the remaining roots stationary, the process can be interpreted through its Granger-Johansen representation, see Johansen (1995, Theorem 4.1). This shows that $\beta' X_t$ and $\Delta_1 X_t$ can be given a stationary distribution while

$$X_t = C \sum_{s=1}^{t} \varepsilon_s + Y_t + \tau_c + \tau_l t,$$

where $C = \beta_\perp \left(\alpha'_\perp \Psi \beta_\perp \right)^{-1} \alpha'_\perp$ for $\Psi = I_2 - \sum_{j=1}^{k-1} \Gamma$ while Y_t is a stationary process. The deterministic components are $\tau_l = C\mu$ while τ_c depends on parameters and the initial observations. The linear trend is avoided by assuming $\mu = \alpha\delta$ for some $\delta \in \mathbb{R}$ so $\tau_l = C\mu = 0$. The restriction $(1,1)\mu = 0$ in (2.11) along with the finding that $(1,1)\alpha = R$ then implies that $\delta = 0$.

We now have two sets of restrictions: those in (2.11) arising from the vector autoregressive assumption, and the restrictions to β and μ arising from the unit root assumption. These restrictions imply

$$(1,1)\alpha = R, \qquad \beta = (1, -1/R)', \qquad (1,1)\Gamma_j = 0, \qquad (1,1)\mu = 0, \qquad (2.13)$$

in which R appears twice, or, equivalently,

$$(1,1)\Pi(0,1)' = -1, \qquad \beta = (1, -1/R)', \qquad (1,1)\Gamma_j = 0, \qquad (1,1)\mu = 0. \quad (2.14)$$

With these restrictions the model (2.10) can be rewritten in terms of an equation for the martingale difference M_t in (2.2) and an equation for dividend growth. These equations are obtained by taking the linear combinations $(1,1)\Delta_1 X_t$ and $(0,1)\Delta_1 X_t$ of equation (2.10). They are given by

$$M_t = \varepsilon_{M,t}, \qquad (2.15)$$

$$\Delta_1 D_t = \alpha_D S_{t-1} + \sum_{j=1}^{k-1} \Gamma_{j,D} \Delta_1 X_{t-j} + \mu_D + \varepsilon_{D,t}, \qquad (2.16)$$

where $(\alpha_D, \Gamma_{j,D}, \mu_D) = (0,1)(\alpha, \Gamma_j, \mu)$ while $\varepsilon_{M,t} = (1,1)\varepsilon_t$ and $\varepsilon_{D,t} = (0,1)\varepsilon_t$. These equations show that in the context of a vector autoregression for X_t, which is I(1), then the stock price model implies that M_t is a martingale difference along with an equation for dividend growth in terms of the lagged "spread" and lagged growth of dividends and prices. In other words, the equation for dividends show how the partial model (2.1) is completed in the context of an I(1) vector autoregression. Johansen and Swensen (2004) discuss equations for $(1 + R)^{-1} M_t$ and $\Delta_1(P_t - D_t)$ which will give an equivalent representation of the model.

8

2.4.1 Weak exogeneity

We close this section with a brief discussion of weak exogeneity. A feature of equation (2.16) is that there is feedback from the lagged "spread", S_{t-1}, to dividend growth, $\Delta_1 D_t$. For a first order model, $k = 1$, the only feedback is through the coefficient α_D which has to be positive under the $\mathsf{I}(1)$ assumption. In other words, dividends cannot be weakly exogenous for the "spread". In higher-order models there are two sources of feedback: directly from S_{t-1} with coefficient α_D as well as through lagged price growth with coefficients $\Gamma_{j,D}(1,0)'$. In this situation the $\mathsf{I}(1)$ assumption is consistent with the coefficient α_D being positive or negative or indeed zero.

For a first order model, $k = 1$, the argument is as follows. The Granger-Johansen representation (Johansen 1995, Theorem 4.2) gives a condition for stationarity of $S_t = \beta' X_t$ which is that $|1 + \beta'\alpha| < 1$. Due to the restriction $(1,1)\alpha = R$ we can write $\alpha = (\lambda, R - \lambda)'$ for some parameter λ. Since $\beta = (1, -1/R)$ in (2.13) then $1 + \beta'\alpha = 1 + \lambda - (1 - \lambda/R)$. This is bounded by one in absolute value when $|\lambda| < R/(1 + R)$. In particular, the coefficient $\alpha_D = R - \lambda$ must be in the positive range from $R^2/(1 + R) > 0$ to $R + R/(1 + R)$. In particular, the possibility that $\alpha_D = 0$, implying that dividends form a pure random walk, is then ruled out.

For a second order model, $k = 2$, it is complicated to analyse the characteristic polynomial in general. An example, that obeys (2.2) and where $\alpha_D = 0$, is

$$\Delta_1 \begin{pmatrix} P_t \\ D_t \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix} S_{t-1} + \begin{pmatrix} -2R \\ 2R \end{pmatrix} \Delta_1 P_{t-1} + \varepsilon_t,$$

for which the characteristic polynomial is $(1-z)\{1-(1-R)z\}$ giving a unit root and a stationary root of $1/(1 - R) > 1$. In this example the apparent explosive reaction in prices to S_{t-1}, and the lack of reaction in dividends to S_{t-1}, are compensated for by the lagged price growth. This example is akin to the example in Johansen (1995, Exercise 4.3). It shows that for $k > 1$, one should be careful in interpreting the signs and magnitudes of the individual factor loadings in α. All one can say is that under the stock price model (2.1), the restriction $(1,1)\alpha = R$ has to hold, see (2.13).

In summary, in the context of the cointegrated vector autoregression (2.10), the stock price model (2.1) without a bubble implies the parameter restrictions in (2.13). These restrictions imply that M_t, as defined in (2.2), is a martingale difference meaning that stock returns are unpredictable. In §3 we explain how to test these restrictions within a likelihood framework.

2.5 The case with a rational bubble

The case with a rational bubble can be analysed much the same way as the case without a bubble. In this case the additional assumption is made that the vector autoregression has one unit root, and one explosive root $\rho > 1$.

Before exploiting the additional assumptions to the characteristic roots it is convenient to reparametrise the vector autoregression in error correction form aimed at co-explosive behaviour, c.f. Nielsen (2010):

$$\Delta_1\Delta_\rho X_t = \Pi_1\Delta_\rho X_{t-1} + \Pi_\rho\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t, \qquad (2.17)$$

where $\Delta_\rho X_t \equiv X_t - \rho X_{t-1}$, and where $\Pi_1, \Pi_\rho, \Phi_j \in \mathbb{R}^{2\times 2}$ and $\rho \in \mathbb{R}$. This is equivalent to the vector autoregression (2.8) with

$$\Pi_1 = \frac{\Pi}{1-\rho}, \quad \Pi_\rho = -\rho\left(I_p + \Pi_1 - \sum_{j=1}^{k-1}\rho^{-j}\Gamma_j\right), \quad \Phi_j = \sum_{l=j+1}^{k-1}\rho^{j-l}\Gamma_l, \qquad (2.18)$$

see Nielsen (2010, §2.3). The restrictions to the Π, Γ_j, and μ coefficients in (2.11) are then equivalent to, for $j = 1, \ldots, k-2$,

$$(1,1)\Pi_1 = \frac{R}{1-\rho}(1, -1/R), \qquad (1,1)\Phi_j = 0, \qquad (1,1)\mu = 0, \qquad (2.19)$$

$$(1,1)\Pi_\rho = \frac{-\rho}{1-\rho}(R, -1) - \rho(1,1) = \frac{-\rho}{1-\rho}(R+1-\rho, -\rho). \qquad (2.20)$$

The additional assumptions that X_t has one unit root and one explosive root and that the remaining roots are stationary are accommodated by reduced rank restrictions so $\Pi_1 = \alpha_1\beta_1'$ and $\Pi_\rho = \alpha_\rho\beta_\rho'$ where $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}^2$. The process can be interpreted through its Granger-Johansen representation. Such a representation was given in Nielsen (2010, Theorem 1) and a more detailed version is given as Theorem A.1 in the Appendix. 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 while

$$X_t = \frac{1}{1-\rho}C_1\left(A_1 + \sum_{s=1}^{t}\varepsilon_s\right) + \frac{1}{\rho-1}C_\rho\rho^t\left(A_\rho + \sum_{s=1}^{t}\rho^{-s}\varepsilon_s\right) + Y_t + \tau.$$

Here Y_t is some stationary process, A_1, A_ρ depend on parameters and initial values. The impact matrices satisfy $\beta_1'C_1 = 0$ so $\beta_1'X_t$ has no random walk component while $\beta_\rho'C_\rho = 0$ so $\beta_\rho'X_t$ has no explosive component. Thus, β_1 are cointegrating vectors and β_ρ are co-explosive vectors. The explosive common trend $\rho^t(A_\rho + \sum_{s=1}^{t}\rho^{-s}\varepsilon_s)$ then represents the bubble B_t in (2.6). Detailed definitions of these coefficients are given in Theorem A.1 in the appendix.

With the above assumptions, three additional restrictions emerge and the restrictions (2.19)-(2.20) simplify. In order to match the explosive common trend and the bubble, it must hold that $\rho = 1 + R$, see (2.6). The additional restrictions can be

deduced in two ways. First, $\Delta_\rho X_t$ and $\Delta_\rho S_t$ must be non-explosive while $\Delta_1 D_t$ has no unit root. Equation (2.4) then implies that $\Delta_\rho S_t$ and $\Delta_1 D_t$ must both be stationary. Thus, S_t is the cointegrating relation and D_t is the co-explosive relation so that

$$\beta_1 = (1, -1/R)', \qquad \beta_\rho = (0, 1)'. \tag{2.21}$$

Secondly, following Nielsen (2010), the assumption to the roots implies that Π_1, Π_ρ both have rank of unity so $\Pi_1 = \alpha_1 \beta_1'$ and $\Pi_\rho = \alpha_\rho \beta_\rho'$ where $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}^{2\times 1}$. Since the explosive root is $\rho = 1 + R$, the restrictions (2.19)-(2.20) simplify as

$$(1,1)\Pi_1 = (-1, 1/R), \quad (1,1)\Pi_\rho = \{0, -(1+R)^2/R\}, \quad (1,1)\Phi_j = 0, \quad (1,1)\mu = 0.$$

This in turn implies the expressions for β_1 and β_ρ in (2.21). The latter argument also shows that $(1,1)\alpha_1 = -1$ and $(1,1)\alpha_\rho = -(1+R)^2/R$.

We now have two sets of restrictions: those in (2.19)-(2.20) arising from the vector autoregressive assumption as well as the restriction to ρ, β_1 and β_ρ arising from the assumptions to the characteristic roots. These restrictions together imply

$$\begin{aligned}
(1,1)\alpha_1 &= -1, & \beta_1 = (1, -1/R)', & \quad \rho = 1 + R, & (1,1)\mu = 0, \\
(1,1)\alpha_\rho &= -(1+R)^2/R, & \beta_\rho = (0,1)', & \quad (1,1)\Phi_j = 0.
\end{aligned} \tag{2.22}$$

With these restrictions the model (2.17) can be rewritten in terms of an equation for the martingale difference M_t in (2.2) and an equation for dividend growth. These equations arise by taking the linear combinations $(1,1)\Delta_1\Delta_\rho X_t$ and $(0,1)\Delta_1\Delta_\rho X_t$ of equation (2.17). They are given by

$$M_t = \varepsilon_{M,t}, \tag{2.23}$$

$$\begin{aligned}
\Delta_1\Delta_\rho D_t = {} & \alpha_{1,D}\Delta_\rho S_{t-1} + \alpha_{\rho,D}\Delta_1 D_{t-1} \\
& + \sum_{j=1}^{k-2} \Phi_{j,D}\Delta_1\Delta_\rho X_{t-j} + \mu_D + \varepsilon_{D,t},
\end{aligned} \tag{2.24}$$

where $(\alpha_{1,D}, \alpha_{\rho,D}, \Phi_{j,D}, \mu_D) = (0,1)(\alpha_1, \alpha_\rho, \Phi_j, \mu)$ while $\varepsilon_{M,t} = (1,1)\varepsilon_t$ and $\varepsilon_{D,t} = (0,1)\varepsilon_t$. These equations show that in the context of a vector autoregression for X_t, which is I(1) with an explosive root, the stock price model implies that M_t is a martingale difference along with an equation for dividends in terms of the lagged "spread" and lagged filtered growth of dividends and prices. As before, the equation for dividends shows how the partial model (2.1) is completed in the context of an explosive vector autoregression.

There are no simple restrictions to the feedback coefficients $\alpha_{1,D}$ and $\alpha_{\rho,D}$ in the dividend equation (2.24). This is because the simplest model has two lags, $k = 2$,

11

for which the dynamics is rather complicated as we saw for the non-bubble model in §2.4.1. As an illustration, parametrise the model (2.17) with restrictions (2.22) as

$$\Delta_1 \Delta_\rho \begin{pmatrix} P_t \\ D_t \end{pmatrix} = \begin{pmatrix} \lambda \\ -1-\lambda \end{pmatrix} \Delta_\rho S_{t-1} + \begin{pmatrix} \gamma \\ -(1+R)^2/R-\gamma \end{pmatrix} \Delta_1 D_{t-1} + \varepsilon_t.$$

The characteristic polynomial is $(1-z)\{1-(1+R)z\}\{1+(\gamma-\lambda-\lambda/R)z\}$, which has one unit root, one explosive root at $1/(1+R)$, and one stationary root if $-1 < \gamma - \lambda - \lambda/R < 1$. It links $\alpha_{1,D}$ and $\alpha_{\rho,D}$ in a complicated way. For instance, it is possible that $\alpha_{1,D} = 0$, that is $\lambda = -1$, as long as $-1 < \gamma + 1 + 1/R < 1$, that is $|\alpha_{\rho,D} + 1 + R| < 1$.

In summary, in the context of the co-explosive vector autoregression (2.17), the stock price model (2.1) with a bubble implies the parameter restrictions in (2.22). Just as in the case with no bubble, these restrictions imply that M_t, as defined in (2.2), is a martingale difference meaning that stock returns are unpredictable. In the next section we explain how to test these restrictions within a likelihood framework.

3   Likelihood analysis of the models

In the previous section the stock price model was analysed in the context of a vector autoregression for $X_t = (P_t, D_t)'$. In particular, it was shown which parameter restrictions are implied by a rational bubble. In the following it is shown how these parameter restrictions can be tested through a likelihood analysis.

3.1   Unrestricted vector autoregression

The data consists of a bivariate time series $X_{1-k}, \ldots, X_0, X_1, \ldots, X_T$ where $X_t = (P_t, D_t)'$. The unrestricted vector autoregression of (2.8) is of the form

$$\mathsf{M}: \qquad X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu + \varepsilon_t, \qquad (3.1)$$

conditional on the initial observations $X_{1-k}, \ldots, X_0$. The parameters satisfy $A_j \in \mathbb{R}^{2\times 2}$, $\mu \in \mathbb{R}^2$ and ε_t are independently $\mathsf{N}_2(0, \Omega)$-distributed. While the normality assumption is important for defining the likelihood it can be replaced with a martingale difference assumptions for most purposes. Indeed, for consistency of the estimators and all but one of the suggested tests, martingale difference assumptions suffice.

From the least squares estimates of the dynamic parameters A_j the characteristic roots can be computed. A first indication of explosiveness would be if one of these roots appear in the explosive region.

12

To secure that this model is statistically well-specified, the usual diagnostic tests are carried out as far as they are valid in the presence of both a unit root and an explosive root. Nielsen (2006a,b, 2008b) has shown that the usual procedures for lag-length determinations are valid. This includes the information criteria approaches of Schwarz (1978) and Hannan and Quinn (1979) and the auxillary regression approaches of Godfrey (1978), Breusch (1978) and Pagan (1984). Engler and Nielsen (2009) show that QQ-plots for comparing the empirical distribution of the residuals with a normal distribution are valid. The properties of tests for autoregressive conditional heteroskedasticity by Engle (1982) and the recursive Chow-type tests of Doornik and Hendry (2001) are currently explored by one of us.

3.2   Testing the cointegration rank

The next step in the analysis is to establish the cointegration rank. With that in mind the model is reparametrised in error correction form, as in (2.10), as

$$\mathsf{M}: \qquad \Delta_1 X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} + \mu + \varepsilon_t. \qquad (3.2)$$

The cointegration rank hypothesis of interest is that

$$\mathsf{H}_1: \qquad r = \mathrm{rank}(\Pi, \mu) = 1 \quad \text{or} \quad (\Pi, \mu) = \alpha(\beta' + \delta') \quad \text{for} \quad \alpha, \beta \in \mathbb{R}^{2 \times 1}, \delta \in \mathbb{R}.$$

Under this reduced rank hypothesis, we denote the model M_1. The constant is restricted to the cointegrating space partly because it is appropriate in the empirical illustration and partly because it simplifies the determination of the rank, see Nielsen and Rahbek (2000). It would be possible to extend the analysis to the case of an unrestricted constant or even a restricted linear trend.

In practice, the rank is estimated through a sequential testing procedure as shown by Johansen (1995, §13), see also Nielsen and Rahbek (2000). The rank is estimated to be unity if the hypothesis

$$\mathsf{H}_{1,0}: \qquad \Pi = 0, \qquad \mu = 0,$$

is rejected and the hypothesis

$$\mathsf{H}_{1,1}: \qquad \mathrm{rank}(\Pi, \mu) \le 1,$$

cannot be rejected. By testing the hypotheses in this sequence the asymptotic properties of the rank estimator can be controlled.

Results concerning the likelihood ratio tests for these hypotheses against M are available for both the non-bubble and the bubble case. In the non-bubble case it

can be assumed that the process is I(1), that is, it has unit roots of multiplicity one along stationary roots. This situation is analysed by Johansen (1995, Theorem 6.2) and the test statistics are found to have Dickey-Fuller type distributions. For the bubble case where the process also has one explosive root, results are available for the univariate case in Nielsen (2001) and for the multivariate case in Nielsen (2010, Theorem 4). Strictly speaking, the latter multivariate result covers the situation with a linear trend component, but the result could presumably be extended to the specification of deterministic components used here.

The further analysis now depends on the presence of an explosive root and hence a bubble. For the non-bubble case the analysis is given in §3.3, whereas for the bubble case the analysis is given in §3.4.

3.3 Testing the model without a bubble

The theory model without a bubble implies the restricted model given by (2.15)-(2.16). The errors $\varepsilon_{M,t}$ and $\varepsilon_{D,t}$ may be correlated, but by assuming normality and letting ω denote the population regression coefficient of $\varepsilon_{D,t} = (0,1)\varepsilon_t$ on $\varepsilon_{M,t} = (1,1)\varepsilon_t$, these equations can be rewritten as

$$
\begin{aligned}
M_t &= \varepsilon_{M,t}, \\
\Delta_1 D_t &= \alpha_D S_{t-1} + \sum_{j=1}^{k-1} \Gamma_{j,D} \Delta_1 X_{t-j} + \mu_D + \omega M_t + \varepsilon_{D\cdot M,t},
\end{aligned}
$$

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. For a known R, the profile likelihood is maximised by estimating these two unrelated regressions. The likelihood is then maximised by maximising over R.

The likelihood ratio test statistic for the restrictions (2.14) within the model M_1 would presumably by χ^2 with $1+2(k-1)+1 = 2k$ degrees of freedom. In order to prove this formally one would have to combine techniques from two sets of results. The first result concerns a test for the restrictions on Π in the model M_1. Boswijk and Doornik (2004) show that this is χ^2. The second result concerns a test for the restrictions on the Γ_j-parameters. Johansen (1995, Theorem 13.5) shows that the unrestricted estimators $\hat{\Gamma}_j$ in model M_1 are asymptotically normal which would presumably imply that the test on the Γ_j-parameters is also χ^2. Johansen and Swensen (2004, Remark 4) discuss the joint test of all the restrictions, albeit with a known value of R.

3.4 Testing the model with a bubble

We now derive the tests of the restrictions implied by the bubble hypothesis, i.e. (2.22). First, we reparametrise the vector autoregression in co-explosive form and,

based on that, we then set up null hypotheses on the cointegrating and co-explosive vectors, their factor loadings, and the remaining short term parameters. In the end this leads to a four-step procedure for testing for bubbles.

3.4.1 Testing that dividends are non-explosive

Suppose the cointegrating rank has been determined to be unity by the approach in §3.2 and that estimated parameters imply one explosive root, $\hat{\rho}$, say. The vector autoregressive model can then be rewritten in co-explosive form as discussed in §2.5:

$$\mathsf{M}_1: \qquad \Delta_1\Delta_\rho X_t = \alpha_1\beta_1'\Delta_\rho X_{t-1} + \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t. \quad (3.3)$$

Estimates for the parameters can be computed from those of a standard cointegrated vector autoregression using the identities (2.18). By construction the estimate $\hat{\alpha}_\rho\hat{\beta}_\rho'$ will have reduced rank of unity, so the estimates $\hat{\alpha}_\rho, \hat{\beta}_\rho \in \mathbb{R}^2$ can be found as the left and right eigenvectors, respectively.

A central assumption in §2.5 is that dividends are difference stationary without any explosive features. This results in a simple hypothesis on the co-explosive vector

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

Under the hypothesis H_D, the model M_1 reduces to

$$\mathsf{M}_{1D}: \qquad \Delta_1\Delta_\rho X_t = \alpha_1\beta_1'\Delta_\rho X_{t-1} + \alpha_\rho\Delta_1 D_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t. \quad (3.4)$$

The likelihood of the model M_{1D} is analysed through a profile argument. For a given ρ, the likelihood is maximised by a reduced rank regression of $\Delta_1\Delta_\rho X_t$ on $\Delta_\rho X_{t-1}$ correcting for $\Delta_1 D_{t-1}$, lags $\Delta_1\Delta_\rho X_{t-j}$, and a constant. This in turn can be maximised by a grid search over ρ.

The likelihood ratio test statistic for testing H_D in M_1 is asymptotically $\chi^2(p-1)$ as shown by Nielsen (2010, Corollary 1). In our case the dimension of X_t is $p = 2$. This result requires that a certain linear combination of the innovations are normally distributed. The linear combination in question is $\tau'\varepsilon_t$ where τ is the orthogonal complement of $\Psi_\rho\beta_{\rho\perp}$.

3.4.2 Testing the link between the spread and the explosive root

Conditional on H_D above, the restriction between the cointegrating vector and the explosive root can be formulated as a "spread" restriction:

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

Under this restriction, the model reduces to

$$\mathsf{M}_{1DS}: \qquad \Delta_1\Delta_\rho X_t = \alpha_1\Delta_\rho S_{t-1} + \alpha_\rho\Delta_1 D_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t, \qquad (3.5)$$

where $S_t = P_t - \frac{1}{R}D_t$ is denoted the "spread". Again, the likelihood is maximized over ρ. The likelihood ratio test statistic for testing H_S in M_{1D} is asymptotically $\chi^2(1)$. We do not show this directly. Rather, in §4 we show that the test statistic for the joint hypothesis $\mathsf{H}_S, \mathsf{H}_B$ in M_{1D} is a sum of two χ^2 statistics of which one corresponds to the test for H_S in M_{1D}.

3.4.3 Testing the martingale difference restrictions

The final set of restrictions in (2.22) implied by the theory model with a bubble is:

$$\mathsf{H}_B: \quad (1,1)\alpha_1 = -1, \quad (1,1)\alpha_\rho = -(1+R)^2/R, \quad (1,1)\Phi_j = 0, \quad (1,1)\mu = 0.$$

In combination with the restricted constant, $\mu = \alpha_1\delta'$, the above constraints imply that $\mu = 0$. Under these restrictions, the model reduces to the least squares regressions given by (2.23)-(2.24). As in section 3.3 the errors may be correlated, so by assuming normality and letting ω denote the population regression coefficient of $\varepsilon_{D,t} = (0,1)\varepsilon_t$ on $\varepsilon_{M,t} = (1,1)\varepsilon_t$, these equations can be rewritten as

$$\mathsf{M}_{1DSB}: \qquad M_t = \varepsilon_{M,t}, \qquad\qquad\qquad\qquad\qquad (3.6)$$
$$\Delta_1 D_t = \alpha_{1,D}\Delta_\rho S_{t-1} + \alpha_{\rho,D}\Delta_1 D_{t-1}$$
$$+ \sum_{j=1}^{k-2}\Phi_{j,D}\Delta_1\Delta_\rho X_{t-j} + \mu_D + \omega M_t + \varepsilon_{D\cdot M,t}, \qquad (3.7)$$

where the errors $\varepsilon_{M,t}$ and $\varepsilon_{D\cdot M,t} = \varepsilon_{D,t} - \omega\varepsilon_{M,t}$ are uncorrelated. We denote this model M_{1DSB}. The likelihood implied by these equations is maximised through a profile argument using that for a known R the regressions (3.6) and (3.7) are unrelated. The likelihood is then maximised by maximising over R.

The likelihood ratio test statistic for the restricted model M_{1DSB}, i.e. (3.6)-(3.7), within the model M_{1D}, is shown to be asymptotically $\chi^2(2k)$ in §4.

3.5 Summary of procedure for testing the model with a bubble

To summarize, our suggested procedure for testing for a rational bubble consists of the following four steps:

Step 1: Fit an unrestricted VAR model for $X_t = (P_t, D_t)'$ and check that it is econometrically well-specified using misspecification tests. Compute the characteristic

roots from the matrix lag polynomial and see if the largest root is larger than 1, i.e. $\rho > 1$. Test that the cointegration rank is $r = 1$ using Johansen's test. The restricted model is denoted M_1.

Step 2: Given a cointegration rank of $r = 1$ and that $\rho > 1$, test the hypothesis H_D: $\beta_\rho = (0, 1)'$, meaning that D_t is non-explosive. The restricted model is denoted M_{1D}. The test statistic for H_D in M_1 is asymptotically $\chi^2(1)$.

Step 3: Test the hypothesis H_S: $\beta_1 = (1, -R^{-1})'$ with $R = \rho - 1$. This restriction relates the expected return in the theory model linearly to the explosive root in the econometric model. The restricted model is denoted M_{1DS}. The test statistic for H_S in M_{1D} is asymptotically $\chi^2(1)$.

Step 4: Test the final restrictions implied by the bubble model, H_B: $(1, 1)\alpha_1 = -1$, $(1, 1)\alpha_\rho = -(1 + R)^2/R$, $(1, 1)\mu = 0$, and $(1, 1)\Phi_j = 0$. Together with the restriction in H_S, these restrictions imply a martingale difference sequence for stock returns. The restricted model is denoted M_{1DSB}. The test statistic for H_B is asymptotically $\chi^2(2k - 1)$ if tested against M_{1DS} and $\chi^2(2k)$ if tested against M_{1D}.

4   Asymptotic analysis

Under the bubble hypothesis the conditions for the Granger-Johansen representation (Nielsen 2010, Assumption 1) reduce as follows.

Assumption A *Under the bubble hypothesis the parameters satisfy* (2.22) *with* $R > 0$. *In addition, the parameters satisfy*
(i) The nonstationary characteristic roots of X_t *are at 1 or* $1 + R$.
(ii) $\det(\alpha'_{1\perp} \Psi_1 \beta_{1\perp}) \neq 0$ *and* $\det(\alpha'_{\rho\perp} \Psi_\rho \beta_{\rho\perp}) \neq 0$ *where*

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

The basic building blocks for the asymptotic theory are given by Lai and Wei (1985) and later adaptations by Nielsen (2005) to the case with deterministic terms. While the likelihood funtion is based on the assumption of independent, normal innovations the assumptions to the innovations can be relaxed in the asymptotic theory. The relaxed assumptions have to be formulated with two types of results in mind. On the one hand it is needed that the main component of the explosive common trend, $\sum_{s=1}^{t} \rho^{-s}\varepsilon_s$, converges. The Marcinkiewicz-Zygmund Theorem, see Lai and Wei (1983), shows that this process converges to a random vector almost surely. It suffices to assume that $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence for some filtration $\mathcal{F}_t$ satisfying the following condition.

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

On the other hand it is necessary that the normalised random walk $T^{-1/2} \sum_{s=1}^{t} \varepsilon_s$ converges to a Brownian motion. The assumption of Chan and Wei (1988) to the conditional variance of ε_t is adopted since Assumption B already bounds the conditional moments.

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

Under these assumptions it can be shown that the likelihood ratio test statistic for the explosive rational bubble model M_{1DSB} against M_{1D} is asymptotically χ^2. The test statistic can be decomposed as

$$\mathsf{LR}(\mathsf{H}_B, \mathsf{H}_S | \mathsf{M}_{1D}) = \mathsf{LR}(\mathsf{H}_A | \mathsf{M}_{1D}) + \mathsf{LR}(\mathsf{H}_B | \mathsf{M}_{1DA}),$$

where H_A is the hypothesis related to the extended cointegration vector $(\beta_1', \delta) = (1, -R^{-1}, 0)$. Accordingly the first statistic relates to the restrictions on the cointegration vectors while the second statistic relates to the restrictions on the adjustment vectors $\alpha_1, \alpha_\rho, \Phi_j$. In the asymptotic analysis it is found that $\mathsf{LR}(\mathsf{H}_B, \mathsf{H}_S | \mathsf{M}_{1D})$ is the sum of two components: one involving a mixed Gaussian term and one involving a central limit theorem. This corresponds to the results for the I(1)-model, see Johansen (1995, §13), but the proof is somewhat more complicated.

Theorem 4.1 *Suppose model* M_{1DSB} *is satisfied. Assume A, B, C. Then the test statistic* $\mathsf{LR}(\mathsf{H}_B, \mathsf{H}_S | \mathsf{M}_{1D})$ *is asymptotically* $\chi^2(2k)$.

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

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

5   Empirical illustration: The US stock market

We apply the methods 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 $t - 1$. These data have been used in numerous previous studies, including many of the earlier empirical bubble studies mentioned in the introduction (and in section 6 below).
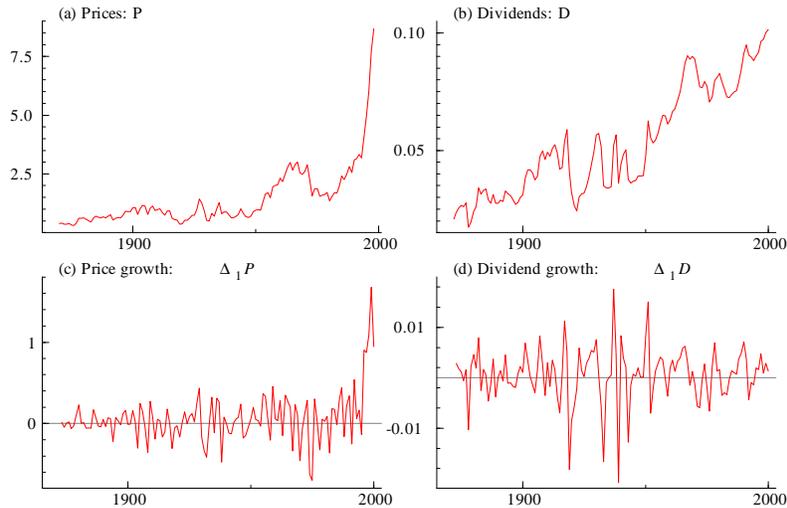
Figure 1: Data for prices, P_t, and dividends, D_t, and their differences. Note the exponential behaviour of P_t and $\Delta_1 P_t$ and the lack of exponential behaviour for dividends.

Figure 1 shows prices and dividends in levels and differences over the full sample period. An exponential pattern is seen both for levels and differences of P_t while D_t appears not to have exponential growth. The dramatic price increases that took place during the 1990s, and which have led many to argue that a bubble was driving the market in that period, are clearly seen in the figure. Shiller (2000), however, argues that not just the 1990s, but also some earlier periods (e.g. the 1920s and 1960s) were driven by speculative bubbles. We will therefore analyse both the full sample period 1872-2000 as well as a more recent sub-sample, 1974-2000. The full sample analysis serves mainly as an illustration as the empirical model is misspecified for the full sample, presumably due to the multiple bubbles. The sub-sample analysis is more substantial as it has one dominant bubble and the empirical model is well-specified.

The software OxMetrics by Doornik and Hendry (2001) is used for the analysis.

5.1 Full sample period, 1872-2000

Here the full sample is analysed using the procedure summarised in §3.5. The initial cointegration analysis is similar to that reported by Engsted (2006).

The first step is to fit a bivariate unrestricted vector-autoregression for $X_t = (P_t, D_t)'$. The Hannan-Quinn criterion picks a two-lag model. Table 1 shows specification

Test	P_t	D_t	Test	system
$\chi^2_{normality}(2)$	14.9 [<0.001]	24.7 [<0.001]	$\chi^2_{normality}(4)$	39.5 [<0.001]
$F_{ar,1-2}(2,120)$	1.6 [0.21]	2.7 [0.07]	$F_{ar,1-2}(8,234)$	1.7 [0.10]
$F_{arch,1-1}(1,120)$	16.7 [<0.001]	1.2 [0.28]		

Table 1: Specification tests for the unrestricted vector autoregression with 2 lags. US 1871 - 2000. All tests are asymptotically valid regardless of the location of the parameters. For the normality tests see Engler and Nielsen (2009), for autoregressive tests see Nielsen (2006a,b), for ARCH test see Doornik and Hendry (2001). p-values in square brackets.

Hypothesis	Likelihood	Test statistic	p-value
$r \leq 2$	494.19		
$r \leq 1$	492.25	$LR(r \leq 1|M) = 3.9$	0.44
$r = 0$	480.80	$LR(r = 0|M) = 26.8$	0.004

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

tests for this model. While the specification appears to capture the autocorrelation in the data, it fails with respect to normality and autocorrelation in the squared residuals. Various recursive tests not shown here also indicate that the model parameters may not be stable over time. In the subsequent section 5.2 we analyse a sub-samble without this specification problem. Nonetheless, for comparison with existing results in the literature we do not attempt to obtain a better specified model for the full sample, and continue with the model as it is.

The four characteristic roots of the unrestricted vector autoregression, M, are 1.183, 0.9404 and $0.1103 \pm 0.1904i$. The largest root is clearly above one, and the second largest root is below - but close to - one. The unrestricted maximum likelihood value is 494.19, see Table 2.

The cointegration rank can be determined through Johansen's rank test with a constant restricted to the cointegration space. This test applies even with an explosive root. The results are reported in Table 2. The restricted model M_1 with no cointegration, $r = 0$, is firmly rejected, whereas the hypothesis of at most one cointegrating relation, $r \leq 1$ cannot be rejected, with a maximum likelihood value of 492.25. Thus a single unit root appears present as required in the theory model, meaning that prices and dividends are 'cointegrated' in the sense that they share an I(1) trend. Imposing a unit root as in model M_1 changes only slightly the largest characteristic root, to $\hat{\rho} = 1.190$. The estimated cointegrating vector is $\hat{\beta}_1 = (1, -22.2)$. This implies

Model	Hypothesis	Likelihood	Test statistic	d.f.	p-value
M_1	$H_1, r = 1$	492.25			
M_{1D}	H_1, H_D	491.93	$LR(H_D|M_1) = 0.64$	1	0.42
M_{1DS}	H_1, H_D, H_S	491.12	$LR(H_S|M_{1D}) = 1.62$	1	0.20
			$LR(H_S|M_1) = 2.26$	2	0.32
M_{1DSB}	H_1, H_D, H_S, H_B	485.23	$LR(H_B|M_{1DS}) = 11.80$	3	0.008
			$LR(H_B|M_{1D}) = 13.40$	4	0.009
			$LR(H_B|M_1) = 14.04$	5	0.015

Table 3: Tests of the rational bubble restrictions. See section 3.4 for a description of the various models and hypotheses.

an expected annual real return of $\hat{R} = 1/22.2 = 4.5\%$. This value is quite low and $1 + \hat{R} = 1.045$ deviates somewhat from the explosive root estimate. Below we test whether they are statistically different.

The next step is to test the hypothesis H_D, i.e. that the co-explosive vector takes the form $\beta'_\rho = (0, 1)$ implying that dividends are non-explosive. The model here is M_{1D} in (3.4) which for a given ρ is estimated by reduced rank regression, c.f. section 3.4.1. The maximum likelihood estimate of ρ is then obtained by a grid search over ρ with $\hat{\rho} = 1.192$. The estimated cointegrating vector is $\hat{\beta}_1 = (1, -20.7)$ with an implied expected annual real return of $\hat{R} = 4.9\%$. The maximum likelihood value is 491.93, and the $\chi^2(1)$ test statistic for the hypothesis H_D is 0.64 with a p-value of 0.42, see Table 3. Thus, we cannot reject that dividends are non-explosive.

The third step is to test the hypothesis H_S which relates R in the cointegrating vector to the explosive root ρ, i.e. $\beta_1 = (1, -1/R)'$ and $\rho = 1 + R$, see section 3.4.2. The model here is M_{1DS} as in (3.5). The explosive root is estimated by $\hat{\rho} = 1.138$ implying an expected annual real stock return of $R = 13.8\%$. The likelihood value is 491.12. The test statistic for the hypothesis H_S is 1.62 with a p-value of 0.20 if tested against M_{1D} and 2.26 with a p-value of 0.32 in the $\chi^2(2)$ distribution if tested against M_1. Thus, despite the estimated difference between R and $\rho - 1$ seen above, they are not statistically different.

The estimated system (3.5) under M_{1DS} with $\hat{R} = 0.138$ is (with standard errors in parentheses):

$$\Delta_1\Delta_\rho P_t = \underset{(0.088)}{-0.856}(\Delta_\rho S_{t-1} + \underset{(0.024)}{0.083}) - \underset{(4.4)}{14.0}\Delta_1 D_{t-1},$$

$$\Delta_1\Delta_\rho D_t = \underset{(0.0016)}{-0.0086}(\Delta_\rho S_{t-1} + \underset{(0.024)}{0.083}) - \underset{(0.080)}{0.926}\Delta_1 D_{t-1},$$

$$\hat{\sigma}_{PP} = 0.261, \quad \hat{\sigma}_{DD} = 0.00475, \quad corr = -0.0064.$$

The last line reports the estimated residual standard deviations for the two equations along with the residual correlation. As seen, the estimated adjustment coefficients are

$\widehat{\alpha}_1 = (\widehat{\alpha}_{1,P}, \ \widehat{\alpha}_{1,D})' = (-0.856, \ -0.0086)'$ and $\widehat{\alpha}_\rho = (\widehat{\alpha}_{\rho,P}, \ \widehat{\alpha}_{\rho,D})' = (-14.0, \ -0.926)'$, all highly statistically significant. As also seen, these unrestricted estimates are not wildly at odds with the restrictions in (2.22), that is, $(1,1)\alpha_1 = -1$ and $(1,1)\alpha_\rho = -(1+R)^2/R = -9.38$ (for $R = 0.138$), whereas the constraint $(1,1)\mu = 0$ appears invalid. In the final step we explicitly test these restrictions.

The last step is to test the bubble hypothesis H_B. In the restricted model M_{1DSB} the explosive root is estimated by $\hat{\rho} = 1.102$ implying an expected annual real stock return of $R = 10.2\%$. The likelihood value is 485.23. The test statistics for the hypothesis H_B are 11.80 (p-value: 0.008) in M_{1DS} and 14.04 (p-value: 0.015) in M_1. Thus the stock price model with a rational bubble is not supported by the data.

The problem in rejecting the bubble hypothesis lies primarily with the restriction to the constant and to some extent with the restriction to α_1. We tried to impose the restrictions $(1,1)\alpha_\rho = -(1+R)^2/R$, $(1,1)\alpha_1 = -1$, and $(1,1)\mu = 0$ individually on the model M_{1DS}. This gave test statistics of 1.50 (p-value: 0.22), 2.88 (p-value: 0.09) and 4.34 (p-value: 0.04), respectively, in a $\chi^2(1)$ distribution. Imposing both $(1,1)\alpha_1 = -1$ and $(1,1)\alpha_\rho = -(1+R)^2/R$, but leaving the restricted constant unconstrained, gave a test statistic of 5.08 (p-value: 0.08).
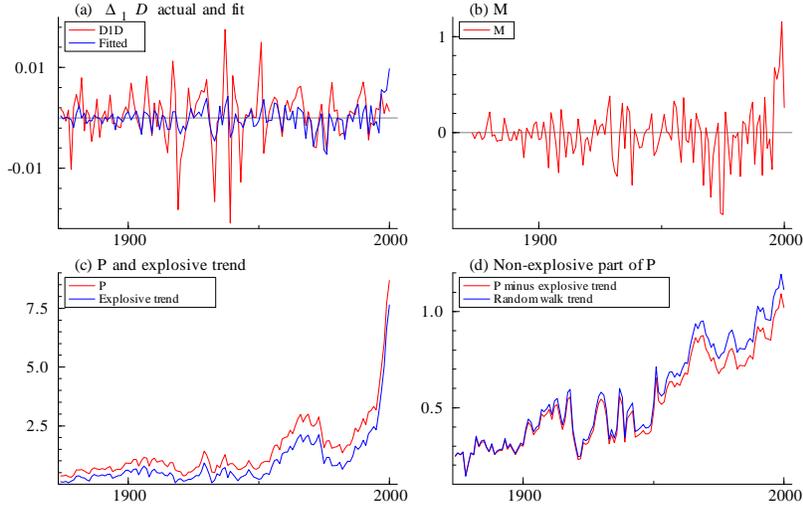

Figure 2: Results from M_{1DSB} applied to full sample. (a) shows $\Delta_1 D_t$ and the fit - note that fit drifts towards the end. (b) shows M_t - note drifting mean after 1960. (c) shows P_t and the estimated explosive trend. (d) shows P_t minus explosive trend along with random walk trend.

It is interesting to consider graphical properties of the rejected bubble model

M_{1DSB}. The equations (3.6), (3.7) for the martingale difference M_t and dividend growth $\Delta_1 D_t$ were fitted. The drifting tendencies in panels (a) and (b) of Figure 2 contribute to the rejection of H_B. Panel (c) shows the actual prices P_t and the explosive trend for P_t computed using the formula $(\rho-1)^{-1}(1,0)C_\rho\rho^t(A_\rho+\sum_{s=1}^{t}\rho^{-s}\hat{\varepsilon}_s)$ given in Theorem A.1. This explosive trend is the estimated bubble component in the model. We see that the model implies that the bubble contributes with a substantial part of the movements in prices over time. Panel (d) compares the non-explosive part of P_t, that is the difference between P_t and the explosive trend, with the random walk trend $(1-\rho)^{-1}(1,0)C_1(A_1+\sum_{s=1}^{t}\hat{\varepsilon}_s)$, see Theorem A.1.

The conclusion from the above results is that US stock prices are explosive while dividends are non-explosive, and the two variables have a common stochastic I(1) trend. The over-identifying restrictions implied by a standard stock price model with a constant discount rate and a rational bubble are however not supported at standard significance levels, implying that excess returns do not behave as a martingale difference. However, since the specification tests in Table 1 show evidence of misspecification, we are reluctant to interpret these results as providing strong evidence against the standard stock price model. We consider the results indicative of the presence of a bubble.

5.2   Sub-sample, 1974-2000

In this section we will analyse a smaller sub-sample that contains data for 1976-2000, using the 1974 and 1975 observations as initial observations in the estimation of the vector autoregression. By letting the sample start in the 1970s we are also able to focus on the recent "bubble period" that ended in 2000 and - according to Shiller (2000) - began to build up already from the beginning of the 1980s. The proposed method is designed to work well for such a single bubble period.

Tables 4 and 5 report, respectively, specification tests for an unrestricted model with two lags and tests for the cointegration rank. There is no evidence of misspecification so the unrestricted vector autoregression seems to be econometrically well-specified. The characteristic roots are 1.258, $0.675\pm0.354i$, and -0.295, thus clearly indicating an explosive root in the system. The rank tests point to a cointegration rank of unity, although the rejection of the hypothesis of no cointegration is marginal. Imposing a unit root changes only slightly the largest root, to 1.223. In the following we will assume the presence of a unit root.

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 either $(1, 0)$ or $(0, 1)$ are contained in the cointegration space cannot be rejected at the 5% level of significance, which implies that we can

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

Table 4: Specification tests for the unrestricted vector autoregression with 2 lags. US 1973 - 2000. All tests are asymptotically valid regardless of the location of the parameters. For the normality tests see Engler and Nielsen (2007), for autoregressive tests see Nielsen (2006a,b), for ARCH test see Doornik and Hendry (2007). p-values in square brackets.

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

Table 5: Cointegration rank determination with constant restricted to cointegration space. US 1973 - 2000. Critical values based on Johansen (1995, Table 15.2) and Doornik (1998).

treat either P_t or D_t as having no unit root. Thus, unfortunately the data are not very informative about the value of the expected return parameter R.

Table 6 reports the various tests associated with the bubble hypothesis. First, the H_D hypothesis that dividends are non-explosive gives a p-value of 0.99 while $\hat{\rho} = 1.224$ is nearly unchanged. This is as expected from Figure 1 which clearly shows that the explosive root in the system belongs to P_t and not D_t. Second, the H_S hypothesis that $\beta'_1 = (1, -1/R)$ with $\rho = 1 + R$, gives a p-value of 0.57 when tested against M_{1D} and $\widehat{\rho} = 1.263$, which implies $R = 26.1\%$, clearly not an economically reasonable estimate of the expected annual return.

Testing the final hypothesis, H_B, gives a p-value of 0.07 if tested against model M_{1DS}, and a p-value of 0.19 if tested against model M_1. The decision not to reject the hypothesis H_B is marginal against M_{1DS} albeit more convincing against M_1, so all in all the sub-sample is more supportive of the standard stock market model than the full sample. As for the full sample the difficulties arise from the constant. Imposing only $(1,1)\alpha_1 = -1$ and $(1,1)\alpha_\rho = -(1 + R)^2/R$ but leaving the restricted constant unconstrained gives a likelihood of 104.68, so a test statistic of 2.20 (p-value: 0.33) against M_{1DS}. Thus the test statistic for M_{1DSB} against this intermediate hypothesis is 4.89 (p-value: 0.027).

Proceeding with the model M_{1DSB} we note that the estimate of the explosive root

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

Table 6: Tests of the rational bubble restrictions. US 1973 - 2000. See section 3.4 for a description of the various models and hypotheses.

now becomes $\widehat{\rho} = 1.156$ (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 estimated model under M_{1DSB} can be presented in two ways. First, the system representation is

$$\Delta_1\Delta_\rho P_t = \underset{(0.0014)}{-1.0032}\Delta_\rho S_{t-1} - \underset{(0.16)}{7.92}\Delta_1 D_{t-1}$$

$$\Delta_1\Delta_\rho D_t = \underset{(*)}{0.0032}\Delta_\rho S_{t-1} - \underset{(*)}{0.65}\Delta_1 D_{t-1}$$

$$\hat{\sigma}_{PP} = 0.362, \qquad \hat{\sigma}_{DD} = 0.00292, \qquad corr = -0.262,$$

while the martingale representation, as in (3.6) and (3.7), is

$$M_t = \hat{\varepsilon}_{M,t}$$

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

$$\hat{\sigma}_{MM} = 0.126, \qquad \hat{\sigma}_{DD\cdot M} = 0.00295, \qquad corr = 0.$$

Figure 3 shows a graph of M_t along with graphs of actual and fitted values of $\Delta_1 D_t$ as well as a break down of the explosive trend and the random walk trend as for Figure 2. Note the improved fit for the sub-sample as compared with the full sample.

The conclusion from this sub-sample analysis, 1974-2000, is that real stock prices clearly contain an explosive component, and the formal restrictions implied by the rational bubble model cannot be rejected statistically at conventional significance levels, although the test for the bubble hypothesis itself, H_B against M_{1DS}, is marginal. The analysis does provide a rather high value of the expected return parameter R. It is also found that two quite different models for the relation between prices and dividends are both consistent with the data: Either P_t does share the unit root in D_t (i.e. 'cointegration' between P_t and D_t), or it does not. The data are simply not
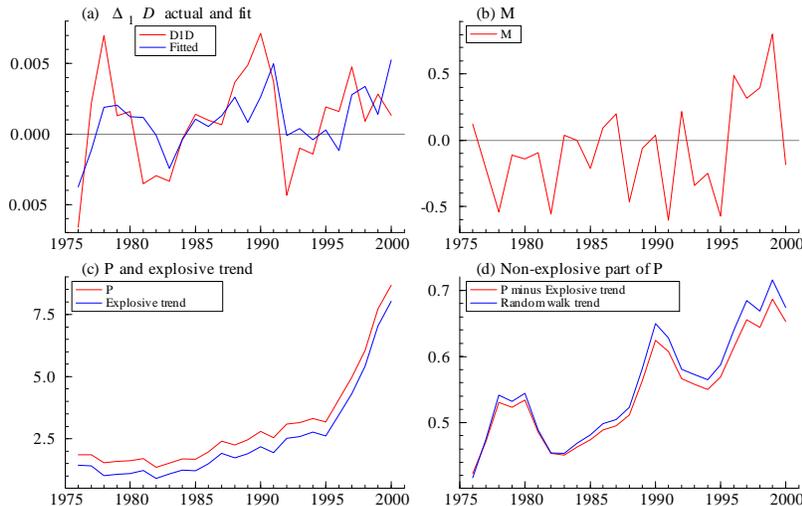
Figure 3: Results from M_{1DSB} applied to sub-sample. (a) shows $\Delta_1 D_t$ and the fit. (b) shows M_t. (c) shows P_t and the estimated explosive trend. (d) shows P_t minus explosive trend along with random walk trend.

able to discriminate between these two hypotheses. From an economic point of view (c.f. Shiller, 2000), and looking at figure 1, it is not totally unreasonable to consider prices during the 1980s and 1990s as being bubble-driven with no connection at all to fundamental variables like dividends.

6 Comparing our test with earlier bubble tests

In this section we briefly discuss our bubble testing procedure in relation to other bubble tests proposed in the literature. Earlier in the paper we discussed Diba and Grossman's (1988b) approach and, as we saw, in contrast to their analysis our approach makes it possible to estimate both the explosive root and the cointegrating vector between the I(1) components of P_t and D_t, and it allows testing the restriction that a rational bubble imposes on the relation between the the explosive root and the cointegrating vector.

An important assumption in both Diba and Grossman and our analysis is that the discount factor, i.e. expected return, is constant. Craine (1993) modified Diba and Grossman's approach by showing that with a time-varying (but stationary) discount factor, the price-dividend *ratio*, P_t/D_t, should be stationary under no bubbles. The appealing feature of Craine's approach is that no unknown cointegrating vector has

to be estimated. It suffices to just test the stationarity of P_t/D_t. On the other hand, no other testable restrictions follow from this approach. When we fit a first order autoregression to P_t/D_t on our US 1872-2000 data, the autoregressive root is above one (1.0465) and the null hypothesis of a unit root against the one-sided explosive alternative is strongly rejected at the right tail of the Dickey-Fuller distribution (ADF-value: 1.04), consistent with the results in §5.

A potential weakness of cointegration based tests for bubbles is that they will have difficulties discovering periodically collapsing bubbles. Evans (1991) showed that in a finite sample unit root and cointegration tests will typically not identify the explosive component of periodically collapsing bubbles. Our co-explosive framework may not perform well in that situation. Indeed, the IT bubble is so dominant that explosiveness is picked up in a full sample ending in the years 1996 to 2000 but not if ending in 1995, thus supporting Evans' (1991) conjecture; see also discussion of Engsted (2006). When looking exclusively at the sub-sample from 1974 onward with a single bubble the explosiveness is, however, picked up. For instance for the sample ending in 1990 there is an explosive root of 1.10.

In this paper we have confined attention to stochastic explosive bubbles driven by extraneous sources. We have not considered the kind of purely deterministic explosive bubbles analysed by Flood and Garber (1980), or the 'intrinsic' explosive bubble analysed by Froot and Obstfeld (1991) where the bubble is a non-linear deterministic function of fundamentals. For the kind of stochastic bubbles we analyse, West (1987) devised an often cited Hausman-type specification test, which compares two sets of estimates of the underlying asset pricing model. The first set of estimates is consistent both with and without a bubble, whereas the second set is only consistent in the absence of a bubble. The specification test then tests for equality of the two sets of estimates. Thus, the null hypothesis is no bubble, while the presence of a bubble should lead to rejection of the null hypothesis. One troublesome aspect of West's procedure (noted by West himself in West (1985), i.e. the working paper version of West (1987)), is that the test is not consistent: under the alternative hypothesis that bubbles are present, the probability that the test will reject the null of no bubbles will not go to unity asymptotically. The problem is directly a result of the explosiveness of prices under the alternative. The testing procedure we develop in the present paper does not face such problems because our null hypothesis explicitly involves bubbles.

Finally, in some recent literature (e.g. Cochrane, 2008; Balke and Wohar, 2009), bubbles are discussed within the log-linearized present value model of Campbell and Shiller (1988b). However, that model is based on a Taylor series expansion around the unconditional mean log dividend-price ratio which needs to be stationary in order for the log-linearization to be valid. Thus, that model in fact rules out bubbles from the outset.

7   Conclusions

In this paper we have developed a procedure for testing a constant discount rate equity market model within a vector autoregressive framework. The special feature of our approach is that it allows for both unit roots in prices and dividends and an explosive root due to a rational speculative bubble. The analysis has been conducted within the co-explosive vector autoregressive framework of Nielsen (2010), which is an extension of the standard *Johansen model* for I(1) variables to the case with one explosive root in the system. We have developed likelihood ratio tests of various restrictions implied by the bubble model; these restrictions can be tested using standard χ^2-inference.

In an empirical illustration using US stock market data that includes the "bubble period" in the 1990s, we find stong evidence of an explosive root in stock prices, while dividends do not contain explosive elements, thus lending support to the argument that US stock prices in part have been driven by a speculative bubble. For the full sample 1872-2000 the bubble restrictions are, however, rejected at standard significance levels - and this is when ignoring that the empirical model is mis-specified. For the sub-sample analysis for the period 1974-2000 the empirical model appears well-specified and the bubble restrictions cannot be rejected.

References

Abreu, D. and Brunnermeier, M.K. (2003) Bubbles and crashes. *Econometrica* 71, 173-204.

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.

Balke, N.S. and Wohar, M.E. (2009) Market fundamentals versus rational bubbles in stock prices: A Bayesian perspective. *Journal of Applied Econometrics* 24, 35-75.

Bhargava, A. (1986) On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 53, 369-384.

Bohl, M.T. (2003) Periodically collapsing bubbles in the US stock market? *International Review of Economics and Finance* 12, 385-397.

Boswijk, H.P. and Doornik, J.A. (2004) Identifying, estimating and testing restricted cointegrated systems: An overview. *Statistica Neerlandica* 58, 440-465.

Breusch, T.S. (1978) Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17, 334-355.

Brown, B.M. and Eagleson, G.K. (1971). Martingale convergence to infinitely divisible laws with finite variance. *Transactions of the American Mathematical Society* 162, 449-453.

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

Campbell, J.Y. and Shiller, R.J. (1988a) Interpreting cointegrated models. *Journal of Economic Dynamics and Control* 12, 505-522.

Campbell, J.Y. and Shiller, R.J. (1988b) The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 5, 243-280.

Chan, N.H. and Wei, C.Z. (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

Cochrane, J.H. (2008) The dog that did not bark: A defense of return predictability. *Review of Financial Studies* 21, 1533-1575.

Craine, R. (1993) Rational bubbles: A test. *Journal of Economic Dynamics and Control* 17, 829-846.

Diba, B.T. and Grossman, H.I. (1988a) The theory of rational bubbles in stock prices. *Economic Journal* 98, 746-757.

Diba, B.T. and Grossman, H.I. (1988b) Explosive rational bubbles in stock prices. *American Economic Review* 78, 520-530.

Doornik, J. (1998) Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573–593.

Doornik, J.A. and Hendry, D.F. (2001) *Empirical econometric modelling using Pc-Give 10, vol. 1 and 2.* London: Timberlake Consultants Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R.F., Hendry, D.F. and Richard, J.-F. (1983) Exogeneity. *Econometrica* 51, 277-304.

Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometric Journal* 12, 367-381.

Engsted, T. (2006) Explosive bubbles in the cointegrated VAR model. *Finance Research Letters* 3, 154-162.

Evans, G.W. (1991) Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review* 81, 922-930.

Flood, R.P. and Garber, P.M. (1980) Market fundamentals versus price-level bubbles: The first tests. *Journal of Political Economy* 88, 745-770.

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

Godfrey, L.G. (1978) Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46, 1293-1301.

Hannan, E.J. and Quinn, B.G. (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society* B41, 190-195.

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

Johansen, S. (1995) *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

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

Johansen, S. and Swensen, A.R. (2004) More on testing exact rational expectations in cointegrated vector autoregressive models: Restricted constant and linear term. *Econometrics Journal* 7, 389-397.

Lai, T.L. and C.Z. Wei (1983) A note on martingale difference sequences satisfy-ing the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics, Academia Sinica* 11, 1-13.

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 VI*, pp. 375-393. Amsterdam: Elsevier Science Publishers.

Leroy, S.F. (1989) Efficient capital markets and martingales. *Journal of Economic Literature* 27, 1583-1621.

Leroy, S.F. (2004) Rational exuberance. *Journal of Economic Literature* 42, 783-804.

Nielsen, B. and Rahbek, A. (2000) Similarity issues in cointegration models. *Oxford Bulletin of Economics and Statistics* 62, 5-22.

Nielsen, B. (2001) The Asymptotic Distribution of Unit Root Tests of Unstable Autoregressive Processes. *Econometrica* 69, 211-219.

Nielsen, B. (2005) Strong consistency results for least squares statistics in general vector autoregressive models. *Econometric Theory* 21, 534-561.

Nielsen, B. (2006a) Correlograms for non-stationary autoregressions. To appear in *Journal of the Royal Statistical Society* B68, 707-720.

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

Nielsen, B. (2008a) On the Explosive Nature of Hyper-Inflation Data. *Economics: The Open-Access, Open-Assessment E-Journal* 2, 2008-21.

Nielsen, B. (2008b) Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination. Discussion paper, Nuffield College.

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

Pagan, A.R. (1984) Model evaluation by variable addition. In D.F. Hendry and K.F. Wallis (eds.) *Econometrics and Quantitative Economics.* Oxford: Basil Blackwell.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

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

Shleifer, A. (2000) *Inefficient Markets. An Introduction Behavioral Finance.* Oxford University Press, Oxford.

Tirole, J. (1982) On the possibility of speculation under rational expectations. *Econometrica* 50, 1163-1181.

Tirole, J. (1985) Asset bubbles and overlapping generations. *Econometrica* 53, 1499-1528.

West, K.D. (1985) A specification test for speculative bubbles. *Working Paper*, Princeton University.

West, K.D. (1987) A specification test for speculative bubbles. *Quarterly Journal of Economics* 102, 553-580.

A Appendix: Granger-Johansen representation

In this appendix a detailed representation theorem is derived for the co-explosive equation

$$\Delta_1\Delta_\rho X_t = \alpha_1\beta_1'\Delta_\rho X_{t-1} + \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \alpha_1\delta' + \varepsilon_t \qquad \text{(A.1)}$$

in which $\alpha_1, \beta_1, \alpha_\rho, \beta_\rho \in \mathbb{R}^2$ and $\delta' \in \mathbb{R}$.

The following notation is used throughout the appendices: For a matrix α with full column rank let $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$ while $\alpha_\perp$ denotes a basis to the orthogonal complement of the span of α so $\alpha_\perp'\alpha = 0$ and $(\alpha, \alpha_\perp)$ is invertible.

Theorem A.1 *Suppose Assumption A holds. Then it holds*

$$X_t = \frac{1}{1-\rho}C_1(A_1 + \sum_{s=1}^{t}\varepsilon_s) + \frac{1}{\rho-1}C_\rho\rho^t(A_\rho + \sum_{s=1}^{t}\rho^{-s}\varepsilon_s) + Y_t - \frac{\overline{\beta}_1\delta'}{1-\rho}.$$

where $Y_t = \theta_U U_t$ and

$$
\begin{aligned}
\Psi_x &= I_p + \frac{\alpha_y\beta_y'}{y-x} - \sum_{j=1}^{k-2}x^{-j}\Phi_j \\
C_x &= x\beta_{x\perp}(\alpha_{x\perp}'\Psi_x\beta_{x\perp})^{-1}\alpha_{x\perp}', \\
A_x &= \Psi_x\Delta_y X_0 - \frac{\alpha_y\beta_y'}{y-x}\Delta_x X_0 + \sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}\Delta_x\Delta_y X_{-h}, \\
\theta_U &= (G_{1,\rho}, G_{\rho,1}, H_{1,\rho,1} + H_{\rho,1,1}, \ldots, H_{1,\rho,k-2} + H_{\rho,1,k-2}), \\
U_{t-1} &= \{(\beta_1'\Delta_\rho X_{t-1} + \delta')', (\beta_\rho'\Delta_1 X_{t-1})', \Delta_1\Delta_\rho X_{t-1}, \ldots, \Delta_1\Delta_\rho X_{t-k+2}\}, \\
G_{x,y} &= -\frac{C_y\alpha_x}{(y-x)^2} - \frac{C_x\Psi_x\overline{\beta}_x}{x-y} + \frac{x\overline{\beta}_x}{x-y}, \qquad H_{x,y,n} = \frac{C_x x^{n-1}}{x-y}\sum_{j=n}^{k-2}\Phi_j x^{-j}.
\end{aligned}
$$

The process U_t satisfies a first order vector autoregression and can be given a stationary initial distribution.

Remark A.2 *A similar result can be derived for the model with a linear trend*

$$\Delta_1\Delta_\rho X_t = \alpha_1(\beta_1', \delta')\begin{pmatrix}\Delta_\rho X_{t-1} \\ (1-\rho)t\end{pmatrix} + \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t,$$

see Nielsen (2010). Define coefficients τ_c, τ_l by

$$
\begin{aligned}
\tau_l &= C_1\mu/(1-\rho) + (C_1\Psi_1 - I_p)\overline{\beta}_1\delta_1', \\
\beta_1'\tau_c &= \overline{\alpha}_1'(\Psi_1 C_1 - I_p)\mu/(1-\rho) + \overline{\alpha}_1'\Psi_1(C_1\Psi_1 - I_p)\overline{\beta}_1\delta_1' + \delta_1'\rho/(1-\rho).
\end{aligned}
$$

Then $\tilde{X}_t = X_t - \overline{\beta}_1\beta_1'\tau_c - \tau_l t$ solves the homogeneous solution $(A.2)$. As a consequence

$$X_t = \frac{1}{1-\rho}C_1(A_1 + \sum_{s=1}^{t}\varepsilon_s) + \frac{1}{\rho-1}C_\rho\rho^t(A_\rho + \sum_{s=1}^{t}\rho^{-s}\varepsilon_s) + Y_t + \overline{\beta}_1\beta_1'\tau_c + \tau_l t.$$

Proof of Theorem A.1. *Homogeneous equation.* Let $\tilde{X}_t = X_t + \overline{\beta}_1\delta'/(1-\rho)$. Insert $\Delta_1 X_t = \Delta_1\tilde{X}_t$ and $\beta_1'\Delta_\rho X_{t-1} = \beta_1'\Delta_\rho\tilde{X}_t - \delta'$ in $(A.1)$ to see that $\tilde{X}_t$ solves

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

It suffices to find the representation for this homogeneous equation. For the general case subtract $\overline{\beta}_1\delta'/(1-\rho)$ from the representation.

Some identities. Note first that

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

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

Representation for $\beta_{x\perp}'\Delta_y\tilde{X}_t$. It is first argued that

$$x\beta_{x\perp}\overline{\beta}_{x\perp}'\Delta_y\tilde{X}_t = C_x x^t(A_x + \sum_{s=1}^{t}x^{-s}\varepsilon_s) + C_x\frac{\alpha_y\beta_y'}{y-x}\Delta_x\tilde{X}_t$$

$$-C_x\Psi_x\overline{\beta}_x\beta_x'\Delta_y\tilde{X}_t - C_x\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}\Delta_x\Delta_y\tilde{X}_{t-h}. \qquad (A.3)$$

To get that expression pre-multiply the model equation by $\alpha_{x\perp}'$ and insert the above identities to get

$$\alpha_{x\perp}'\Delta_x\Delta_y\tilde{X}_s = \alpha_{x\perp}'\varepsilon_s + (y-x)^{-1}\alpha_{x\perp}'\alpha_y\beta_y'(\Delta_x^2\tilde{X}_s - \Delta_x\Delta_y\tilde{X}_s)$$

$$+\alpha_{x\perp}'\sum_{j=1}^{k-2}x^{-j}\Phi_j(\Delta_x\Delta_y\tilde{X}_s - \sum_{h=0}^{j-1}x^h\Delta_x^2\Delta_y\tilde{X}_{s-h})$$

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

$$\alpha_{x\perp}'\Psi_x\left(\Delta_y\tilde{X}_t - x^t\Delta_y\tilde{X}_0\right) = \alpha_{x\perp}'\sum_{s=1}^{t}x^{t-s}\varepsilon_s + \alpha_{x\perp}'\frac{\alpha_y\beta_y'}{y-x}(\Delta_x\tilde{X}_t - x^t\Delta_x\tilde{X}_0)$$

$$-\alpha_{x\perp}'\sum_{j=1}^{k-2}\Phi_j\sum_{h=0}^{j-1}x^{h-j}(\Delta_x\Delta_y\tilde{X}_{t-h} - x^t\Delta_x\Delta_y\tilde{X}_{-h}).$$

34

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

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

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

Representation. Note the identity

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

Thus, $\tilde{X}_t$ has the desired representation in which

$$
\begin{aligned}
Y_t &= \frac{C_1}{1 - \rho} \left\{ \frac{\alpha_\rho \beta_\rho'}{\rho - 1} \Delta_1 X_t - \Psi_1 \overline{\beta}_1 \beta_1' \Delta_\rho X_t - \sum_{h=0}^{k-3} \left(\sum_{j=h+1}^{k-2} \Phi_j \right) \Delta_1 \Delta_\rho X_{t-h} \right\} \\
&\quad + \frac{C_\rho}{\rho - 1} \left\{ \frac{\alpha_1 \beta_1'}{1 - \rho} \Delta_\rho X_t - \Psi_\rho \overline{\beta}_\rho \beta_\rho' \Delta_1 X_t - \sum_{h=0}^{k-3} \rho^h \left(\sum_{j=h+1}^{k-2} \Phi_j \rho^{-j} \right) \Delta_1 \Delta_\rho X_{t-h} \right\} \\
&\quad + \frac{1}{1 - \rho} \overline{\beta}_1 \beta_1' \Delta_\rho X_t + \frac{\rho}{\rho - 1} \overline{\beta}_\rho \beta_\rho' \Delta_1 X_t.
\end{aligned}
$$

Rearrange the terms of Y_t to get the desired expression for Y_t.

The process U_t. The model equation implies a first order vector autoregression for U_t. Assuming A this can be given a stationary initial distribution as stated in Nielsen (2010, Theorem 1). $\blacksquare$

B   Appendix: Proof of asymptotic results

In this appendix the asymptotic results are proved.

The abbreviations $a.s.$, P and D are used for properties holding almost surely, in probability, and in distribution respectively.

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

$$
\begin{aligned}
U_t^\circ &= \{(\beta_1^{\circ\prime}\Delta_{\rho_\circ} X_t)', (\beta_\rho^{\circ\prime}\Delta_1 X_t)', (\Delta_1\Delta_{\rho_\circ} X_t)', \ldots, (\Delta_1\Delta_{\rho_\circ} X_{t-k+3})'\}', & \text{(B.5)} \\
V_t^\circ &= \beta_{1\perp}^{\circ\prime}\Delta_{\rho_\circ} X_t, \qquad W_t^\circ = \beta_{\rho\perp}^{\circ\prime}\Delta_1 X_t, \qquad \mathcal{R}_t^\circ = (M_t^\circ, U_{t-1}^{\circ\prime})'. & \text{(B.6)}
\end{aligned}
$$

The data generating process is

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

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

B.1.3   Some further parameters

From Nielsen (2010) it is known that the analysis of the unrestricted model M_{1D} involves the parameters

$$\tau^\circ_\perp = \Psi^\circ_\rho \overline{\beta}^\circ_{\rho\perp}, \qquad \Psi^\circ_\rho = I_p + \frac{\Pi^\circ_1}{1 - \rho_\circ} - \sum_{j=1}^{k-2} \rho_\circ^{-j} \Phi^\circ_j, \tag{B.8}$$

as well as the projection matrices

$$\mathcal{P}^\circ_{\tau_\perp} = \tau^\circ_\perp (\tau^{\circ\prime}_\perp \Omega^{-1}_\circ \tau^\circ_\perp)^{-1} \tau^{\circ\prime}_\perp \Omega^{-1}_\circ, \qquad \mathcal{P}^\circ_\alpha = \alpha^\circ_1 (\alpha^{\circ\prime}_1 \Omega^{-1}_\circ \alpha^\circ_1)^{-1} \alpha^{\circ\prime}_1 \Omega^{-1}_\circ. \tag{B.9}$$

In the restricted model M_{1DSB} the restrictions (2.22) imply

$$\tau^\circ_\perp = \begin{pmatrix} \mathcal{G}^\circ \\ -R^{-1}_\circ \alpha^\circ_{1,D} - \sum_{j=1}^{k-2} (1 + R_\circ)^{-j} \Phi^\circ_{j,DP} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\theta^{\circ\prime} \end{pmatrix} \mathcal{H}^\circ, \tag{B.10}$$

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

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

B.1.4   A preliminary asymptotic result

It is convenient to recall the following asymptotic result stated as Lemma A.1 in Nielsen (2010), but derived from Chan and Wei (1988) and Nielsen (2005). Introduce the block diagonal normalisation matrix $N_\mathcal{S} = \mathrm{diag}(I_{\dim U}, N_V, N_W, 1)$ where

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

Lemma B.1 *Let X_t satisfy M_{1DSB} and be given by $(B.7)$. Assume A, B, C. Let ξ be a constant satisfying $\xi < \gamma/(2 + \gamma)$, recalling γ defined in Assumption B.*
Define sample variances $\widehat{\mathsf{var}}(x_t) = T^{-1} \sum_{t=1}^T x_t x_t'$. Then stochastic matrices Σ_{WW}, Σ_{VV}, Σ_{SS} and a deterministic matrix Σ_{UU} exist so

(i) $\widehat{\mathsf{var}}(\varepsilon^\circ_t) \overset{a.s.}{=} \Omega_\circ + \mathrm{o}(T^{-\xi}) + \mathrm{o}(T^{\eta-1/2})$ *for all $\eta > 0$.*
(ii) $\widehat{\mathsf{var}}(U^\circ_{t-1}) \overset{a.s.}{\to} \Sigma_{UU} > 0.$
(iii) $\hat{\Sigma}_{WW} = \rho^{-2T}_\circ \sum_{t=1}^T (W^\circ_{t-1})^2 = \rho^{-2T}_\circ T \widehat{\mathsf{var}}(W^\circ_{t-1}) \overset{a.s.}{\to} \Sigma_{WW} \overset{a.s.}{>} 0.$

(iv) $\hat{\Sigma}_{VV} = \widehat{\mathsf{var}}(N_V V^{\circ}_{t-1}) \overset{\mathsf{D}}{\to} \Sigma_{VV} \overset{a.s.}{>} 0.$

(v) $\widehat{\mathsf{var}}\left(N_S S^{\circ}_{t-1}\right) \overset{\mathsf{D}}{\to} \Sigma_{SS} \overset{a.s.}{>} 0.$

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

(vi) $\widehat{\mathsf{corr}}(S^{\circ}_{t-1}, \varepsilon^{\circ}_t) \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2}).$
(vii) $\widehat{\mathsf{corr}}\{(U^{\circ\prime}_{t-1}, V^{\circ\prime}_{t-1}, 1)', \varepsilon^{\circ}_t\} = \mathrm{O_P}(T^{-1/2}).$

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

(vii) $\hat{\Sigma}_{U\epsilon} = T^{-1/2} \sum_{t=1}^{T} (U^{\circ}_{t-1}) \varepsilon^{\circ\prime}_t \overset{\mathsf{D}}{\to} \Sigma_{U\varepsilon}.$ and $\hat{\Sigma}_{U\epsilon} \overset{a.s.}{=} \mathrm{o}(T^{\eta})$ for all $\eta > 0.$
$(viii)$ $\hat{\Sigma}_{V\epsilon} = T^{-1/2} \sum_{t=1}^{T} N_V (V^{\circ}_{t-1}) \varepsilon^{\circ\prime}_t \overset{\mathsf{D}}{\to} \Sigma_{V\varepsilon}.$
(ix) $\hat{\Sigma}_{W\epsilon} = \rho_{\circ}^{-T} \sum_{t=1}^{T} W^{\circ}_{t-1} \epsilon^{\circ}_t \overset{a.s.}{=} \mathrm{o}(T^{(1-\xi)/2}).$

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 ρ. Some expansions of these product sums are needed. 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

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

Lemma B.2 *Assume A, B, C. 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*

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

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

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

$$+ (\mathcal{D}_R + \mathcal{I}_R)(1 + \rho_{\circ}^T \mathcal{D}_R + \mathcal{D}_R + \mathcal{I}_R) \mathrm{o}_{pol}.$$

38

In particular it holds

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{D\cdot M,t}^{\circ}M_t' \stackrel{a.s.}{=} S_{\varepsilon_{D\cdot M}\varepsilon_M}^{\circ} - \mathcal{D}_R(\mathcal{G}^{\circ}S_{\varepsilon_{D\cdot M}W}^{\circ} + o_{pol}),$$

$$\frac{1}{T}\sum_{t=1}^{T}M_t^2 \stackrel{a.s.}{=} S_{\varepsilon_M\varepsilon_M}^{\circ} - 2\mathcal{D}_R(\mathcal{G}^{\circ}S_{\varepsilon_M W}^{\circ} + o_{pol}) + \mathcal{D}_R^2\{\mathcal{G}^{\circ 2}S_{WW}^{\circ} + o(1)\}. \quad \text{(B.13)}$$

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

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

which has the solution

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

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

$$\Delta_1\Delta_{1+R}X_{t-j} = \Delta_1\Delta_{1+R_\circ}X_{t-j} \quad \text{(B.15)}$$

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

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

$$\mathcal{R}_t = \mathcal{R}_t^{\circ} - \mathcal{D}_R\mathcal{H}^{\circ}W_{t-1}^{\circ} - (\mathcal{D}_R\mathcal{H}_2 + \mathcal{I}_Re_2\mathcal{H}_3')(U_{t-1}^{\circ\prime}, V_{t-1}^{\circ})',$$

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 $e_2 = (0, 1, 0_{1\times2(k-2)})'$.

The expression for M_t. Since $M_t = P_t + D_t - (1+R)P_{t-1}$ by (B.1) then $M_t = M_t^{\circ} - \mathcal{D}_RP_{t-1}$. 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 (2.4) 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 some linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$, while $\Delta_1 D_{t-1}$ is some other linear combination of $U_{t-1}^\circ, V_{t-1}^\circ$.

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

The remaining coordinates of U_{t-1} are of the type $\Delta_1\Delta_{1+R}X_{t-j}$. These are rewritten using (B.15). Thus, pre-multiplying $\Delta_1 X_{t-1}$ by $I_p = \overline{\beta}_{\rho\perp}^\circ\beta_{\rho\perp}^{\circ\prime} + \overline{\beta}_\rho^\circ\beta_\rho^{\circ\prime}$ 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 $T^{-1}\sum_{t=1}^T \varepsilon_{D\cdot M,t}^\circ (U_{t-1}^{\circ\prime}, V_{t-1}^\circ)' = o_{pol}$ 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 $\sum_{t=1}^T \mathcal{R}_t^\circ(U_{t-1}^{\circ\prime}, V_{t-1}^\circ)' = o_{pol}$ 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 $T^{-1/2}\sum_{t=1}^T \mathcal{R}_t^\circ W_{t-1}^\circ = o(\rho_\circ^T)$ a.s. by Lemma B.1(*i,v,vi*). ∎

B.3 Consistency under H_B

The expression for the martingale difference M_t is quadratic in the unknown parameters, so global consistency can be proved under H_B in contrast to the co-explosive analysis in Nielsen (2010). The starting point is the profile log likelihood for the parameter R. Let $\hat{\sigma}_M^2(R)$ and $\hat{\sigma}_{D \cdot M}^2(R)$ denote the residual variances of the regression equations (3.6), (3.7). Then the profile log likelihood is

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

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

Lemma B.3 *Assume A, B, C. Then*
(i) $\hat{\sigma}_M^2(R_\circ) = T^{-1} \sum_{t=1}^T M_t^{\circ 2} = S_{\varepsilon_M \varepsilon_M}^\circ \overset{a.s.}{\to} \sigma_{MM}^\circ,$
(ii) $\hat{\sigma}_{D \cdot M}^2(R_\circ) \overset{a.s.}{\to} \sigma_{DD \cdot M}^\circ.$
(iii) $2\ell(R_\circ) = -T \log \det(S_{\varepsilon\varepsilon}^\circ) + \sigma_{DD \cdot M}^{-1} \hat{\Sigma}_{\varepsilon_{D \cdot M} U} \Sigma_{UU}^{\circ-1} \hat{\Sigma}_{U\varepsilon_{D \cdot M}} + o(1).$

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

(ii) 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(i) implies

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

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

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

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

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

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

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

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

Due to the identity

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

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

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

Proof of Lemma B.4. Let $R_\circ$ denote the true value of R.

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

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

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

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

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

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

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

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

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

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

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

As $M_t = M_t^\circ + (R_\circ - R)P_{t-1}$ 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

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

Since the sample correlations of $\varepsilon_{D\cdot M,t}^\circ$, $\varepsilon_{M,t}^\circ$ and $\mathcal{S}_{t-1}^\circ$ vanish asymptotically then $\hat{\sigma}_{D\cdot M}^2(R)$ 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 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 of Lemma B.5. The matrix h can be decomposed as $h = AJA^{-1}$ where J is a triangular, Jordan matrix with diagonal elements λ_j. Thus, $I + h = A(I + J)A^{-1}$ and $\det(I + h) = \det(I + J) = \prod_{j=1}^{\dim h}(1 + \lambda_j)$. By the expansion $\log(1 + x) = x - x^2/2 + x^3/6 + \mathrm{O}(x^4)$ it holds

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Component involving $\hat{\sigma}^2_{D\cdot M}(R)$. First, use partitioned inversion to get

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

The last term does not depend on R.

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

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

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

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

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

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

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 *Assume A, B, C. Under* H_B *then*

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

where $\hat{\Sigma}_{\mathcal{RW}} = \sum_{t=1}^{T}\mathcal{R}^\circ_t W^\circ_{t-1}$. *It holds that*

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

Proof of Lemma B.7. *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

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

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 derivates are

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

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

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

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

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

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

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

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 *Assume A, B, C. Under* H_B *then*

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

Proof of Lemma B.8. *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(*i,vii*),

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

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

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

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

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

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

The partitioned inversion formula $A_{11\cdot2}^{-1} = A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22\cdot1}^{-1} A_{21} A_{11}^{-1}$ shows, noting that $\omega^\circ = \sigma_{MM}^{\circ-1} \sigma_{DM}^\circ$,

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

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

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

46

Combining these expressions shows

$$\mathcal{G}^{\circ 2}S^{\circ-1}_{\varepsilon_M \varepsilon_M} + \mathcal{H}'(S^{-1}_{\mathcal{RR}\cdot\Delta} - S^{-1}_{\mathcal{RR}})\mathcal{H} \overset{a.s.}{\to} \tau^{\circ\prime}_{\perp}\{\sigma^{-1}_{MM}\begin{pmatrix} 1 \\ 0 \end{pmatrix}^{\otimes 2} + \sigma^{-1}_{DD\cdot M}\begin{pmatrix} \omega^{\circ} \\ -1 \end{pmatrix}^{\otimes 2}\}\tau^{\circ}_{\perp}.$$

Finally, the desired expression follows since by partitioned inversion

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

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

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

In a similar way write

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

These expression combine as

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

Noting that, see (B.27),

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

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

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

47

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

$$\mathcal{H}^{\circ\prime}(S_{\mathcal{RR}\cdot\Delta}^{\circ-1}\hat{\Sigma}_{\mathcal{RW}\cdot\Delta} - S_{\mathcal{RR}}^{\circ-1}\hat{\Sigma}_{\mathcal{RW}}) \overset{a.s.}{=} \tau_\perp^{\circ\prime}\begin{pmatrix}\omega^\circ\\-1\end{pmatrix}^{\otimes 2}\sigma_{DD\cdot M}^{\circ-1}\hat{\Sigma}_{\varepsilon W}\{1 + \mathrm{o}(T^{\eta-1/2})\}.$$

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

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

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

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

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* M_{1DSB}. *Assume A, B, C. Then* $\hat{R} - R = \mathrm{o}(T^{1/4}\rho^{-T})$ *a.s.*

Proof of Lemma B.9. 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 $\mathrm{o}(1)$ 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* M_{1DSB}. *Assume A, B, C. Then*

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

(ii) $\rho_\circ^T(\hat{R} - R_\circ)\tau_\perp^\circ \overset{a.s.}{=} \mathcal{P}_{\tau_\perp}^\circ \hat{\Sigma}_{\varepsilon W}\Sigma_{WW}^{\circ-1}\{1 + o(1)\} + o(1)$.

Proof of Lemma B.10. *(i)* Lemma B.9 shows that $\hat{R} - R_\circ = o(T^{1/4}\rho^{-T})$. 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 $o(T^{-3/4}\rho_\circ^{3T})(T^{1/4}\rho_\circ^{-T})^2\rho_\circ^{-2T} = o(T^{-1/4}\rho_\circ^{-T})$ so *(i)* follows.

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

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

Rearrange to get the desired result. ∎

B.6 Asymptotic distribution of estimators

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

(i) $\{(\sum_{t=1}^T M_t^2)^{1/2}(\hat{\omega} - \omega^\circ), (\sum_{t=1}^T U_t^2)^{1/2}(\hat{\theta} - \theta^\circ)\} \overset{\mathsf{D}}{\to} \mathsf{N}(0, \sigma_{DD\cdot M}^2 I_{2k-1})$.

(ii) $\hat{\sigma}_{MM} \to \sigma_{MM}, \hat{\sigma}_{DD\cdot M} \to \sigma_{DD\cdot M}$ *a.s.*

(iii) *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 it holds*
$\{-\ell''(\hat{R})\}^{1/2}(\hat{R} - R_\circ) = H + o(1)$ *a.s.*

(iv) *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 of Theorem B.11. *(i)* Since $\hat{\rho} - \rho_\circ = o(T^{1/4}\rho_\circ^{-T})$ 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

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

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

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

49

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

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

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

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

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

B.7 Likelihood in restricted model

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

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

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

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

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

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

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

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

50

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 M_{1D} is now analysed. An expression for the likelihood value was given in Nielsen (2010) but this is not directly applicable here and needs to be elaborated. Lemma A.12 of the same paper gives an improved consistency rate for some of the parameters of M_{1D}. That result is now extended for all parameters. For results in the unrestricted model the data generating process is

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

where $\beta_1^* = (\beta_1', \delta')'$ and $\Delta_{\rho_\circ}X_{t-1}^* = (\Delta_{\rho_\circ}X_{t-1}, 1)$. This of course encompasses the data generating process (B.7) under M_{1DSB}.

Lemma B.13 *Suppose M_{1D} holds with $\rho_\circ \geq \varrho$ for some $\varrho > 1$. Assume A, B, C. Recall the definitions of $\tau_\perp^\circ, \mathcal{P}_{\tau_\perp}^\circ, \mathcal{P}_\alpha^\circ$ in (B.8), (B.9). Then*
(i) $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = T^{-1/2}\hat{\Sigma}_{\varepsilon U}\Sigma_{UU}^{-1} + \mathrm{o}_\mathsf{P}(T^{-\xi}) = \mathrm{O}_\mathsf{P}(T^{-1/2})$.
(ii) $\alpha_1^\circ(\hat{\beta}_1^{*D} - \beta_1^{*\circ})'\overline{\beta}_{1\perp}^{*D}N_V^{-1} = T^{-1/2}\mathcal{P}_\alpha^\circ\hat{\Sigma}_{\varepsilon V}\hat{\Sigma}_{VV}^{-1} + \mathrm{o}_\mathsf{P}(T^{-1/2}) = \mathrm{O}_\mathsf{P}(T^{-1/2})$.
(iii) $\rho_\circ^{-T}\check{\tau}^{D\prime}\sum_{t=1}^T\hat{\varepsilon}_t^D W_{t-1}^\circ = \rho_\circ^{-T}\tau^{\circ\prime}\sum_{t=1}^T\varepsilon_t^\circ W_{t-1}^\circ + \mathrm{o}_\mathsf{P}(1)$.
(iv) $\hat{\Omega}_D - \Omega_\circ = T^{-1}\sum_{t=1}^T\{(\varepsilon_t^\circ)^{\otimes 2} - \Omega_\circ\} + \mathrm{o}_\mathsf{P}(T^{-1/2}) = \mathrm{O}_\mathsf{P}(T^{-1/2})$.
(v) $\tau_\perp^\circ(\hat{\rho}_H - \rho_\circ) = \mathcal{P}_{\tau_\perp}\sum_{t=1}^T\varepsilon_t^\circ W_{t-1}^\circ\{\sum_{t=1}^T(W_{t-1}^\circ)^2\}^{-1}\{1 + \mathrm{O}_\mathsf{P}(T^{-1/2})\} + \mathrm{o}_\mathsf{P}(T^{-1/4}\rho_\circ^{-T})$.

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

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

51

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

(i) The estimator for $\hat{\boldsymbol{\alpha}}$ is then

$$\hat{\boldsymbol{\alpha}} = \sum_{t=1}^{T}(\Delta_1\Delta_{\hat{\rho}}X_t)(U_{t-1}^D)'\{\sum_{t=1}^{T}(U_{t-1}^D)^{\otimes 2}\}^{-1}.$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Exploit the $T^{-1/4}\rho_\circ^T$ consistency of $\hat{\rho}_H$ as well as (B.35) to get the desired result. ∎

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

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

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

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

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

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

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

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

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

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

B.9 Proof of main theorem

Proof of Theorem 4.1. It holds that

$$\mathsf{LR}(\mathsf{H}_B, \mathsf{H}_S | \mathsf{M}_{1D}) = 2\hat{\ell}_{1D} - 2\hat{\ell}_{1DSB}$$

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

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

This reduces to

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

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

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

so the test statistic satisfies

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

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

The term $\mathcal{P}_\alpha^\circ \hat{\Sigma}_{\varepsilon V}$ is the stochastic integral of $B_{1,t,T} = N_V V_{t-1}^\circ$ with respect to $c = T^{-1/2} \sum_{s=1}^t \alpha_1^{\circ\prime} \Omega_\circ^{-1} \varepsilon_s$. 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 asymptocally independent processes, so by a mixed Gaussian argument, see Johansen (1995, §13.1), the last term is χ^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 then follows that LR is asymptotically χ^2 with $(2k - 2) + 2 = 2k$ degrees of freedom. ∎