Analysis of co-explosive processes

Bent Nielsen[1]

Department of Economics, University of Oxford

Address for correspondence: Nuffield College, Oxford OX1 1NF, UK
Email: bent.nielsen@nuf.ox.ac.uk
Web: http://www.nuff.ox.ac.uk/users/nielsen

30 March 2005

Summary: A vector autoregressive model allowing for unit roots as well as explosive characteristic roots is developed. The Granger-Johansen representation shows that this results in processes with two common features: a random walk and an explosively growing process. Co-integrating and co-explosive vectors can be found which eliminate these common factors. Likelihood ratio tests for linear restrictions on the co-explosive vectors are derived. As an empirical illustration the method is applied to data from the extreme Yugoslavian hyper-inflation of the 1990s.

Keywords: *Asymptotic normality, Co-explosiveness, Cointegration, Explosive processes, Hyper-inflation, Likelihood ratio tests, Vector autoregression.*

1 Introduction

Several empirical studies of hyper-inflation have been based on cointegration analysis. Following the work of Taylor (1991) variables like prices, money and exchange rate have often been modelled as integrated of second order, $\mathsf{I}(2)$. Such a specification is not adequate for cases where the inflation accelerates. Within the class of vector autoregressive models accelerating inflations can be captured better by allowing for an explosive characteristic root generating a common explosive trend as suggested in a recent analysis of the Yugoslavian economy from the 1980s by Juselius and Mladenović (2002). In the following a formal econometric theory for analysis of co-explosive processes is therefore developed.

The proposed statistical model is based on a vector autoregression and focuses on unit roots and a single explosive root. A Granger-Johansen representation is derived

[1] I am grateful to Søren Johansen who read an earlier version of this paper very carefully and presented it at the conference on Common Features held in Rio de Janeiro in July 2002. I have also benefitted from discussions with Martin Wagner. Computations were done using PcGive (Doornik and Hendry, 2001).

showing the structure of the common stochastic trends. It is demonstrated that there are cointegrating vectors eliminating random walk common trends and 'co-explosive' vectors eliminating an explosive common trend. While linear restrictions on the cointegrating vectors can be tested by the approach of Johansen (1996), see Nielsen (2000), it is here shown how to test linear restrictions on 'co-explosive' relations.

The method is illustrated using data from the extreme Yugoslavian hyper-inflation of the early 1990s, previously analysed by Petrović and Mladenović (2000). This gives some new insight to the nature of hyper-inflations, with a more complete empirical analysis appearing in a companion paper, Nielsen (2004).

The outline of the paper is that in §2 the model is developed and a Granger-Johansen type representation theorem is presented. In §3 hypotheses on the co-explosive vectors are formulated and the related likelihood ratio tests are derived. The empirical work is presented in §4 while §5 concludes. Mathematical proofs are given in Appendices.

The following notation is used throughout the paper: For a matrix α with full column rank let $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$ while $\alpha_\perp$ denotes the orthogonal complement so $\alpha'_\perp\alpha = 0$ and $(\alpha, \alpha_\perp)$ is invertible. The abbreviations $a.s.$, P and D are used for properties holding almost surely, in probability, and in distribution respectively.

2 The model

The starting point for the analysis is a brief review of the cointegrated vector autoregressive model of Johansen (1996). Explosiveness and in particular co-explosiveness is then presented as a restriction to that model. The Granger-Johansen representation follows.

2.1 *The cointegrated vector autoregressive model*

Suppose a p-dimensional time series, $X_{1-k}, \ldots, X_0, \ldots, X_T$ is observed. The cointegrated vector autoregressive model with k lags and cointegration rank r can then be written in terms of the equilibrium correction equation

$$\Delta_1 X_t = \alpha(\beta_1' X_{t-1} + \delta_1' t) + \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} + \mu + \varepsilon_t, \qquad \text{for } t = 1, \ldots, T, \tag{2.1}$$

conditional on the initial values $X_{1-k}, \ldots, X_0$, and where Δ_1 is the usual difference operator. The innovations ε_t are assumed independent, identically $\mathsf{N}_p(0, \Omega)$ distributed, and the parameters vary freely so $\Gamma_j, \Omega \in \mathbf{R}^{p\times p}$ where Ω is positive definite, $\alpha, \beta \in \mathbf{R}^{p\times r}$, while $\mu \in \mathbf{R}^p$ and $\delta_1 \in \mathbf{R}^r$.

The characteristic polynomial for a process of the type (2.1) is given by the determinant of

$$\left(1-z^{-1}\right) I_p - z^{-1}\alpha\beta_1' - \sum_{j=1}^{k-1} z^{-j}(1-z^{-1})\Gamma_j = \frac{z-1}{z}\left(I_p + \frac{\alpha\beta_1'}{1-z} - \sum_{j=1}^{k-1} z^{-j}\Gamma_j\right). \quad (2.2)$$

In the asymptotic analysis of the cointegration model it is usually assumed that the roots of the characteristic polynomial are either stationary, $|z| < 1$, or at one, $z = 1$. Here, the situation where there is also a single explosive root, $z > 1$, is analysed.

Most diagnostic tests for vector autoregressive models are derived under an assumption of stationarity, $|z| < 1$, or at most a random walk like behaviour, $|z| \leq 1$. It is worth noting that while such an assumption is mathematically convenient it may not be necessary. For instance, the lag length can be determined consistently regardless of the location of the characteristic roots when using likelihood methods, see Nielsen (2001a), but not when using methods based on the Yule-Walker equations. This means that correlograms cannot be used as they appear in econometric and software at the moment, see Nielsen (2003). Likewise, the cointegration rank r can be determined in the usual way as will be discussed in §3.4.

In the cointegration analysis of Johansen (1996) the cointegrating relations β_1 are central as the linear combinations eliminating the common $\mathsf{I}(1)$ trends in the process. The Granger-Johansen representation given in Theorem 2.2 below shows that this interpretation carries through in the presence of explosive roots. In the usual situation without explosiveness the cointegrating relations can be given stationary initial distributions, but this is generally not the case in the presence of explosiveness.

2.2 *A cointegration model with an explosive root*

The presence of a single explosive root, $\rho > 1$, in the estimated characteristic polynomial is an indication of an explosive common trend. By reformulating the model such a trend can be analysed in conjunction with a random walk common trend.

Just as the cointegration model (2.1) is formulated in terms of first differences $\Delta_1 X_t$ the filter $\Delta_\rho X_t = X_t - \rho X_{t-1}$ will be needed where ρ is a freely varying parameter. Applying the identities

$$\begin{aligned} (1-\rho) X_{t-1} &= \Delta_\rho X_{t-1} - \rho\Delta_1 X_{t-1}, & (2.3) \\ \rho^j \Delta_1 X_{t-j-1} &= \Delta_1 X_{t-1} - \sum_{l=1}^{j} \rho^{l-1}\Delta_1\Delta_\rho X_{t-l}, & (2.4) \end{aligned}$$

to the model equation (2.1), defining $\Delta_\rho X_{t-1}^* = \{\Delta_\rho X_{t-1}', (1-\rho)\, t\}'$, and assuming the lag length is at least two, $k \geq 2$, then equation (2.1) can be rewritten as

$$\mathsf{M}: \qquad \Delta_1\Delta_\rho X_t = \alpha_1\beta_1^{*\prime}\Delta_\rho X_{t-1}^* + \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j\Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t. \quad (2.5)$$

The new parameters $\alpha_1, \alpha_\rho, \beta_\rho, \Phi_j$ relate to the original parameters by

$$\alpha_1 = \frac{\alpha}{1-\rho}, \quad \alpha_\rho \beta_\rho' = -\frac{1}{\rho}\left(I_p + \alpha_1\beta_1' - \sum_{j=1}^{k-1} \rho^{-j}\Gamma_j\right), \quad \Phi_j = \sum_{l=j+1}^{k-1} \rho^{j-l}\Gamma_l, \qquad (2.6)$$

while $\beta_1^* = (\beta_1', \delta_1')'$. It follows from the expression for the characteristic polynomial in (2.2) that the matrix $\alpha_\rho\beta_\rho'$ has rank $p-1$ precisely when ρ is a single characteristic root. The new parameter space is therefore given in terms of freely varying parameters satisfying $\alpha_1, \beta_1 \in \mathbf{R}^{p\times r}$, $\alpha_\rho, \beta_\rho \in \mathbf{R}^{p\times(p-1)}$, $\mu \in \mathbf{R}^p$, $\delta_1' \in \mathbf{R}^r$, $\Phi_j, \Omega \in \mathbf{R}^{p\times p}$ so Ω is positive definite, and $\rho \in \mathbf{R}$.

With the maximum likelihood estimators from the model formulated by (2.1) at hand it is easy to estimate the parameters of (2.5). The explosive root ρ is estimated by the explosive root of the sample characteristic polynomial for (2.1). The other parameters are estimated using (2.6). Since $\hat{\alpha}_\rho\hat{\beta}_\rho'$ has reduced rank of $(p-1)$ by construction, α_ρ and β_ρ can be estimated by the associated left- and right-eigenvectors.

In order to interpret the process a Granger-Johansen representation is needed. This will be formulated in terms of restrictions to the parameters resembling those of Johansen (1996, Theorem 4.2).

Assumption 2.1 (A) *The matrices* α_1, $\beta_1, \alpha_\rho, \beta_\rho$ *have full column rank.*
(B) *The non-stationary characteristic roots of* X_t *are at* 1 *or at* ρ *where* $\rho > 1$.
(C) $\det(\alpha_{1\perp}'\Psi_1\beta_{1\perp}) \neq 0$ *and* $\det(\alpha_{\rho\perp}'\Psi_\rho\beta_{\rho\perp}) \neq 0$ *where*

$$\Psi_1 = I_p + \frac{\alpha_\rho\beta_\rho'}{\rho-1} - \sum_{j=1}^{k-2}\Phi_j, \qquad \Psi_\rho = I_p + \frac{\alpha_1\beta_1'}{1-\rho} - \sum_{j=1}^{k-2}\rho^{-j}\Phi_j.$$

(D) *The process* X_t *has one explosive characteristic root and* $p-r$ *unit roots so* $\operatorname{rank}(\alpha_\rho\beta_\rho') = p-1$ *and* $\operatorname{rank}(\alpha_1\beta_1') = r$.

The representation theorem can now be formulated. This shows that processes satifying (2.5) have two common features in form of a random walk component and an explosive component. The parameters β_1 and β_ρ have interpretation as co-integrating and co-explosive relationships in that $\beta_1'X_{t-1}$ has no random walk component while $\beta_\rho'X_{t-1}$ has no explosive trend. Both stochastic trends will be removed by the $r-1$ linear relations given by $\operatorname{span}(\beta_1)\cap\operatorname{span}(\beta_\rho)$. The proofs follow in §A.

Theorem 2.2 *Consider a process of the form* (2.5) *satisfying Assumption 2.1. Then*

$$U_t = \{(\Delta_1X_t)'\beta_\rho, (\Delta_\rho X_t^*)'\beta_1^*, (\Delta_1\Delta_\rho X_t)', \ldots, (\Delta_1\Delta_\rho X_{t-k+3})'\}'$$

can be given a stationary initial distribution ensuring the representation

$$X_t \overset{D}{=} \frac{1}{1-\rho} C_1 \sum_{s=1}^{t} \varepsilon_s + \frac{1}{\rho - 1} C_\rho \sum_{s=1}^{t} \rho^{t-s} \varepsilon_s + Y_t + \tau_c + \tau_l t + \tau_x \rho^t,$$

where $C_x = \beta_{x\perp} \left(\alpha'_{x\perp} \Psi_x \beta_{x\perp} \right)^{-1} \alpha'_{x\perp}$ *and* Y_t *is a stationary process. In particular,* $\beta'_c X_{t-1}$ *can be given a stationary initial distribution for any* $\beta_c \in \mathrm{span}(\beta_1) \cap \mathrm{span}(\beta_\rho)$*. The linear slope coefficient can be expressed as*

$$\tau_l = C_1 \mu / (1 - \rho) + (C_1 \Psi_1 - I_p) \overline{\beta}_1 \delta'_1,$$

and in particular $\beta'_1 \tau_l + \delta'_1 = 0$*. The coefficients for the exponential term and the constant level depend on the initial values in such a way that* $\beta'_\rho \tau_x = 0$ *and*

$$\beta'_1 \tau_c = \overline{\alpha}'_1 (C_1 \Psi_1 - I_p) \mu / (1 - \rho) + \overline{\alpha}'_1 (\Psi_1 C_1 \Psi_1 - \Psi_1) \overline{\beta}_1 \delta'_1 + \delta'_1 \rho / (1 - \rho).$$

Finally, it holds that the process $\tilde{X}_t = X_t - \tau_c - \tau_l t$ *satisfies the equation*

$$\Delta_1 \Delta_\rho \tilde{X}_t = \alpha_1 \beta'_1 \Delta_\rho \tilde{X}_{t-1} + \alpha_\rho \beta'_\rho \Delta_1 \tilde{X}_{t-1} + \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho \tilde{X}_{t-j} + \varepsilon_t. \tag{2.7}$$

In this representation the first component $\sum_{s=1}^{t} \varepsilon_s$ is a random walk which is extensively studied in the econometric literature. The second component is ρ^t times the sum $\sum_{s=1}^{t} \rho^{-s} \varepsilon_s$. That sum converges almost surely for increasing t according to the Marcinkiewicz-Zygmund theorem. The convergence holds even if ε_t is a martingale difference sequence as proved by Lai and Wei (1983).

3 Testing a simple hypothesis on the co-explosive parameters

In this section it is discussed how to test linear restrictions on the co-explosive vectors β_ρ. First, the hypothesis and the estimation in the resulting restricted model are discussed. Then the maximum likelihood estimators are argued to be consistent. This result is rather robust in that it is valid also for martingale difference innovations. Thirdly, the asymptotic distributions of $\hat{\rho}$ and of the likelihood ratio test statistic for testing the hypothesis on β_ρ follow. These distribution are normal and χ^2 respectively, but those results rely on the normality of the innovations. Finally, tests for cointegration rank and linear restrictions on β_1^* are discussed.

3.1 *Hypothesis and estimation*

The hypothesis of interest is that the co-explosive vectors are known:

$$\mathsf{H}: \qquad \beta_\rho = \beta_\rho^\circ. \tag{3.1}$$

This hypothesis can equivalently be formulated as a simple hypothesis on the orthogonal complement $\beta_{\rho\perp}$, which is a p-vector. In the empirical illustration reported in §4 and in the empirical analysis of Nielsen (2004) the hypotheses of interest are simple homogeneity restrictions.

Under the hypothesis H the likelihood function has to be maximised by numerical methods. For a given value of ρ the remaining parameters, θ say, can be estimated by reduced rank regression as discussed by Johansen (1996, §6) giving a profile estimator $\hat{\theta}(\rho)$. The profile likelihood function for ρ is then $L(\rho) = \max_\theta L(\rho, \theta) = L\{\rho, \hat{\theta}(\rho)\}$ and can be maximised by grid search.

In principle there is an identification problem. As an example, a bivariate first order autoregression may result in the fitted model $X_t = \hat{A}X_{t-1} + \hat{\varepsilon}_t$ where the estimator $\hat{A}$ has eigenvalues $\hat{\rho}_1 = 1/2$ and $\hat{\rho}_2 = 2$. The model can then be reformulated as $\Delta_{\hat{\rho}_j} X_t = \Pi_j X_{t-1} + \hat{\varepsilon}_t$ where $\Pi_j = \hat{A} - \hat{\rho}_j I_2$. For both choices of j the matrix Π_j has reduced rank and the likelihood achieves the same maximum $\max_{\rho_j, \Pi_j} L\left(\rho_j, \Pi_j\right) = \max_A L\left(A\right)$. The obvious estimator for the explosive root is of course $\hat{\rho}_2 = 2$. In the more complicated model (2.5) the grid search for $\hat{\rho}$ can be done for $\rho > 1$. While this estimator is not necessarily unique this results in a consistent estimator. If none or more than one characteristic root are explosive the specification of the model should be investigated. The same applies if the likelihood function has a local maximum at $\rho = 1$ since this happens with vanishing probability in increasing samples if the model is correct.

3.2 *Consistency in the restricted model*

In the following consistency is stated for the maximum likelihood estimators in the restricted model H. The result is proved in §B,C and is based on a study of the level curves of the likelihood function showing that the likelihood function achieves its maximum in a neighbourhood of the true parameter in large samples.

The asymptotic theory is derived using results of Lai and Wei (1985) and later adaptations by Nielsen (2005). While the likelihood is based on the Gaussian likelihood derived from the model in §2 the assumptions to the innovations can be relaxed. 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 shows that this process converges almost surely. In its original formulation the innovations are required to be independent and identically distributed, but following Lai and Wei (1983) 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 3.1 *For some $\gamma > 0$ it holds $\sup_t \mathsf{E}\{(\varepsilon_t'\varepsilon_t)^{1+\gamma}|\mathcal{F}_{t-1}\} < \infty$ a.s.*

On the other hand it is necessary that the random walk common trend $\sum_{s=1}^t \varepsilon_s$ converges in distribution. Although many flexible Central Limit Theorems are available the assumption of Chan and Wei (1988) to the conditional variance of ε_t is adopted since Assumption 3.1 already bounds the conditional moments.

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

The consistency result for the maximum likelihood estimators under the hypothesis H can now be formulated. The estimator $\hat{\rho}$ is found by searching over a closed set so $\rho > 1$ as described in §3.1, and consistency is established for that situation. It is likely that consistency actually holds for $\rho \geq 1$, as it can also be shown that the likelihood function does not achieve it is at $\rho = 1$.

Theorem 3.3 *Consider the restricted model given by H, and where $\rho \geq \rho_{\min}$ for some $\rho_{\min} > 1$. Suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Identify $\hat{\beta}_1^*$ by*

$$\hat{\beta}_1^{*\prime}\overline{\beta}_1^* = I_r, \tag{3.2}$$

and define the block-diagonal normalisation matrices

$$N_V = \operatorname{diag}(T^{-1/2}I_{p-r}, T^{-1}) \qquad \text{and} \qquad N_W = T^{1/2}\rho^{-T}. \tag{3.3}$$

Then it holds

(*i*) $(\hat{\Omega}, \hat{\mu}) = (\Omega, \mu) + \mathrm{o}_\mathsf{P}(1)$,
(*ii*) $\hat{\rho} = \rho_\circ + \mathrm{o}_\mathsf{P}(N_W)$,
(*iii*) $(\hat{\alpha}_\rho, \hat{\alpha}_1, \hat{\Phi}_1, \ldots, \hat{\Phi}_{k-2}) = (\alpha_\rho, \alpha_1, \Phi_1, \ldots, \Phi_{k-2}) + \mathrm{o}_\mathsf{P}(1)$,
(*iv*) $\hat{\beta}_1^{*\prime}\overline{\beta}_{1\perp}^* N_V^{-1} \xrightarrow{\mathsf{P}} 0$.

The choice of identification of $\hat{\beta}_1^*$ in (3.2) follows the approach of Johansen (1996, §13.2). Identification could alternatively be obtained through a matrix c so that for instance $c'\hat{\beta}_1^* = I_r$. Consistency under that type of identification also follows from Theorem 3.3 by an argument as that of Johansen (1996, §13.2).

3.3 *Asymptotic distribution of estimators and test statistic*

In the following the asymptotic distribution of the likelihood ratio tests statistic for H is described. In contrast to the consistency results described above the results depends on the exact distribution of the innovations corresponding to the results for the univariate first order autoregression of Anderson (1959). Anderson's result will be discussed briefly and then results for the test statistic are given.

Anderson (1959) considered the maximum likelihood estimator for the autoregressive coefficient in a first order autoregression

$$w_t = \rho w_{t-1} + \varepsilon_t,$$

when $|\rho| > 1$. From the Marcinkiewicz-Zygmund result discussed above it follows that

$$\rho^{-t} w_t = w_0 + \rho^{-1}\varepsilon_1 + \rho^{-2}\varepsilon_2 + \cdots + \rho^{-t}\varepsilon_t \overset{a.s.}{\to} \mathcal{W}, \tag{3.4}$$

for some positive, continuous random variable $\mathcal{W}$. Anderson (1959) then proved that the normalised least squares estimator satisfies

$$\frac{\sum_{t=1}^{T} w_{t-1}\varepsilon_t}{\sqrt{\sum_{t=1}^{T} w_{t-1}^2}} = \frac{\mathcal{W}\sum_{t=1}^{T} \rho^{t-T}\varepsilon_t}{\sqrt{\mathcal{W}^2 \sum_{t=1}^{T} \rho^{2(t-T)}}} + \mathrm{o}_{\mathsf{P}}(1), \tag{3.5}$$

where the $\mathcal{W}$'s cancel out. This result holds almost surely and quite generally under the Assumption 3.1, 3.2. The last expression is a weighted average of the innovations with most of the weight on the most recent observations. It is exactly normally distributed when the innovations are normal and is seen to converge in distribution as long as the innovations are independent and identically distributed. A similar result can be established for the likelihood ratio test statistic for the hypothesis H.

To formulate the asymptotic result, introduce a parameter $\tau_\perp = \Psi_\rho \beta_{\rho\perp}$. This is a p-vector which is non-zero due to the Assumption 2.1(C), and its orthogonal complement can be chosen as $\tau = (I_p - \tau_\perp \overline{\tau}'_\perp)\alpha_\rho$, see Johansen (1996, Exercise 3.7).

Theorem 3.4 *Suppose the parameters satisfy Assumption 2.1 and the innovations satisfy Assumptions 3.1, 3.2 with $\gamma > 1$. Then, the log likelihood ratio test statistic for H in M satisfies*

$$LR(\mathsf{H}|\mathsf{M}) = H'\tau(\tau'\Omega\tau)^{-1}\tau' H \{1 + \mathrm{o}_{\mathsf{P}}(1)\} + \mathrm{o}_{\mathsf{P}}(1),$$

where

$$H = \frac{\sum_{t=1}^{T} \varepsilon_t(\beta'_{\rho\perp}\Delta_1 X_{t-1})}{\{\sum_{t=1}^{T}(\beta'_{\rho\perp}\Delta_1 X_{t-1})^2\}^{1/2}} \overset{a.s.}{=} \frac{\sum_{t=1}^{T} \rho^{t-T}\varepsilon_t}{\{\sum_{t=1}^{T} \rho^{2(t-T)}\}^{1/2}} + \mathrm{o}\,(1)\,.$$

If the innovations are independent, identically normal distributed then $LR \overset{\mathsf{D}}{\to} \chi^2(p-1)$.

The asymptotic normality in Theorem 3.4 relies directly on the normality assumption by using that a weighted average of independent normal distributed variables is normal. In practice the best we can do is to check if the normality assumption looks reasonable and then use the above test. It may not hold exactly so infences should only be made cautiously when test statistics are close to the chosen critical value. Having said that, the remainder term in Theorem 3.4 disappears under quite general martingale difference assumptions, while the leading term is exactly normal under the normality assumptions. This type of result may therefore not be that different from usual regression models which are used in finite samples so the robustness induced by the Central Limit Theorem can be of limited use.

3.4 Determination of the cointegration rank

In order to determine the cointegrating rank, r, in the model M consider the more general unrestricted vector autoregression where the parameter matrix $(\Pi, \Pi_l) = \alpha\beta_1^{*\prime}$ is unrestricted. It is then of interest to consider hypotheses of the type

$$\mathsf{H}(r): \qquad \text{rank } (\Pi, \Pi_l) \leq r.$$

The cointegration rank is estimated to be r if indeed the hypothesis $\mathsf{H}(r)$ is accepted while $\mathsf{H}(r-1)$ is rejected. These hypotheses can be tested using Johansen's (1996) likelihood procedure. In his original derivation of the asymptotic theory it was assumed that (i) the number of unit roots is $p-r$ and (ii) the remaining characteristic roots are stationary. The latter assumption is actually not necessary, so the procedure can be used with the usual asymptotic distributions even in the presence of explosive roots. While the asymptotic derivations in the Appendices focus on the results that will be reviewed in §3 below, those results immediately lead to the following special case of the rank test results. General results are given by Nielsen (2000, 2001).

Theorem 3.5 *Suppose the model* M *and Assumption 2.1 are satisfied, while the innovations satisfy the Assumptions 3.1, 3.2. Then, the log likelihood ratio test statistic for cointegration rank has the usual asymptotic distribution as given in Johansen* (*1995, Theorem 6.1*).

3.5 *Linear restrictions on cointegrating vectors*

Johansen (1996, §7) discusses a variety of linear restrictions on the cointegrating vectors β_1^*. The most general hypothesis for which the likelihood function can be analysed analytically is the hypothesis that some cointegrating relations are known and the others satisfy some linear restrictions. This hypothesis can be formulated as a restriction on the space spanned by β_1^*,

$$\mathsf{H}_1: \qquad \text{span}\,(\beta_1^*) = \text{span}\,(B_1, B_{1\perp}D_1\varphi_1)\,, \tag{3.6}$$

for some known matrices B_1, D_1 of dimension $\{(p+1)\times b_1\}$ and $\{(p+1-b_1)\times d_1\}$ and a parameter φ_1 of dimension $(d_1\times r)$, all with full column rank.

Hypotheses of this type can be tested within the model M given in (2.5) using χ^2 inference. This is proved by Johansen (1996, §13) for a variety of hypotheses assuming that (i) the number of unit roots is $p-r$, and (ii) the remaining characteristic roots are stationary. The results also hold when the stationarity assumption (ii) is relaxed as shown by Nielsen (2000) for the hypothesis $\beta = B\varphi$.

It may also be of interest to test the hypothesis H on the co-explosive vectors given in (3.1) within the model M restricted by H_1. This can be done using χ^2 inference as outlined in Theorem 3.4. The proof is similar to that presented in Appendix C albeit notationally more burdensome due to the more complicated structure of β_1^*.

4 Empirical illustration: The extreme Yugoslavian hyper-inflation

The presented model is now applied to data from the extreme Yugoslavian hyper-inflation of the 1990s. A more complete empirical analysis can be found in the companion paper, Nielsen (2004).

The institutional background for the extreme Yugoslavian hyper-inflation is described in Petrović and Vujošević (1996) and Petrović, Bogetić and Vujošević (1999). In short, the former federal republic of Yugoslavia was falling apart in 1991, the war started and United Nations embargo was introduced in May 1992. This situation led to decreased output and fiscal revenue while transfers to the Serbian population in Croatia and Bosnia-Herzegovina as well as military expenditure added to the fiscal problems. The monthly inflation rose above 50% in February 1992 and accelerated further, a price freeze was attempted in the end of August 1993 and the inflation finally ended on 24 January 1994 with a currency reform after prices had risen by a factor of 1.6×10^{21} over 24 months.

The main economic theory for hyper-inflation is due to Cagan (1956). Taylor (1991) reformulated the model in a cointegration setup in terms of the equations

$$\begin{aligned} m_t - p_t &= -\alpha\Delta p^e_{t+1} + \zeta_t, &(4.1)\\ \Delta p^e_{t+1} &= \Delta p_{t+1} - \epsilon_{t+1}, &(4.2)\end{aligned}$$

where $m_t - p_t$ is the log of real money, Δp_t is the monthly growth in log prices, and Δp^e_{t+1} measures the expected inflation in period $t+1$, while ζ_t, ϵ_{t+1} are stationary error terms. Dividing by $-\alpha$, and then subtracting Δp_t on both sides leads to

$$\Delta^2 p_{t+1} = -\alpha^{-1}(m_t - p_t + \alpha\Delta p_t) + \left(\epsilon_{t+1} - \alpha^{-1}\zeta_t\right). \qquad (4.3)$$

Assuming that m_t and p_t are both $\mathsf{I}(2)$ variables it can be tested whether the real money $m_t - p_t$ is $\mathsf{I}(1)$ and in turn whether $m_t - p_t$ and Δp_t cointegrate to $\mathsf{I}(0)$. In

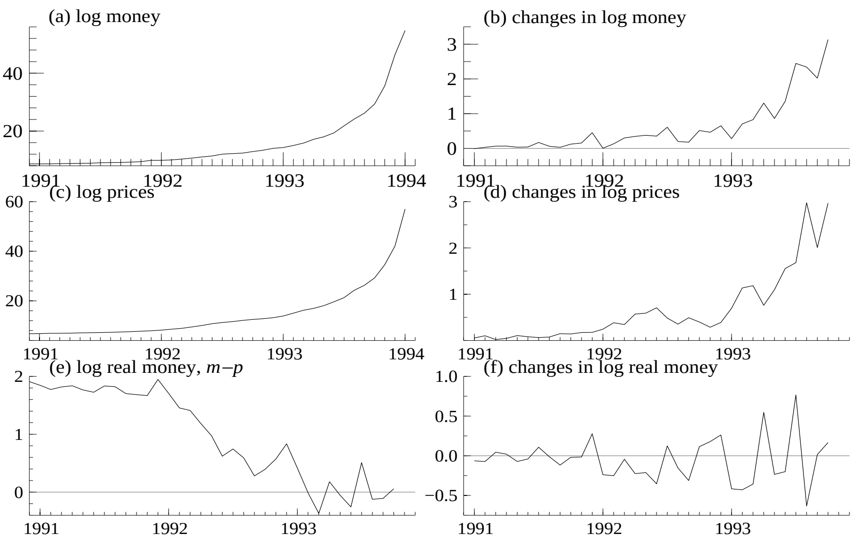

Figure 1: Data in levels for full period until 1994:1. Data in differences (using Δ_1-operator) for shorter period until 1993:10.

this integrated framework the semi-elasticity α in the money demand schedule (4.1) therefore shows up as the coefficient to Δp_t in a cointegrating relation.

Figure $1(a, c)$ show two time series of monthly data relating to the period 1990:12 to 1994:1. The variables are narrow money measured as M1, m_t, and a price index, p_t, both reported on a logarithmic scale. The sources for the data are documented in Petrović and Mladenović (2000). They consider the prices for 1993:12 and 1994:1 to be unreliable, so following precedence in the empirical hyper-inflation literature only the data until 1993:10 rather than the full sample are analysed. Figure $1(b, d)$ show first differences of the series. Both in levels and in differences the series show an exponentially growth over time and hence an increasing growth in prices. Real money measured as $m_t - p_t$ is shown in levels and in differences in Figure $1(e, f)$. It appears that real money, $m_t - p_t$, has a random walk behaviour while the depreciation rate, $\Delta_1 p_t$, is explosive, indicating that the cointegration analysis proposed by Taylor cannot be applied directly in this case. Such exlosive behaviour was also observed for the earlier Yugoslavian hyper-inflation of the 1980s by Juselius and Mladenović (2002). Using the developed statistical model this can now be analysed formally.

A vector autoregression with a constant trend and three lags is fitted to the data up to 1993:10. On the one hand this gives a model that has admittedly few degrees of freedom in that each equation has 7 mean parameters which are fitted using $T = 32$ observations. On the other hand a lot of information should be available in these

Test	m	p	Test	(m,p)
$\chi^2_{normality}(2)$	2.3 [0.32]	2.4 [0.31]	$\chi^2_{normality}(4)$	7.1 [0.13]
$F_{AR(3)}(3,21)$	1.2 [0.34]	1.6 [0.22]	$F_{AR(3)}(12,34)$	0.7 [0.73]
$F_{Hetero}(14,9)$	1.8 [0.19]	3.9 [0.02]	$F_{Hetero}(42,21)$	2.0 [0.05]
$F_{ARCH(3)}(3,21)$	1.0 [0.42]	0.5 [0.70]		

Table 1: Mis-specification tests for the vector autoregressive model for m, p. Single-equation as well as system tests are reported. p-values are given in brackets.

Rank	Log likelihood	Test against full rank
0	-0.41	29.6 [0.01]
1	10.77	7.2 [0.33]
2	14.39	

Table 2: Test for cointegration rank. p-values are given in brackets.

explosively growing time series. Formal mis-specification tests are reported in Table 1. Interpreting these in the usual way indicates that the model is well-specified with respect to normality and autocorrelation. There is possibly some unmodelled heterogeneity. In doing so it is assumed that the usual asymptotic theory is valid although this has only been proved for the test for autocorrelation in the residuals, see Nielsen (2001a), whereas it is unclear whether for instance the test for heterogeneity is valid. Some of the test statistics are reported in an F-form as advocated by Doornik and Hendry (2001) in an attempt to deal with finite sample issues for these tests even though it has not yet been argued whether this represents an improvement. This model has a single explosive characteristic root of $\hat{\rho} = 1.216$.

A cointegration analysis can be carried out in the usual way as described in §3.4. The results are given in Table 2, pointing towards a rank of one. This explosive characteristic root is now estimated by $\hat{\rho} = 1.205$. The model M given in (2.5) with $r = 1$ therefore appears to give a reasonable description of the data.

The hypothesis that β_ρ is known and given by the homogeneity condition

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

can now be tested. This would imply that real money, $m_t - p_t$, is a co-explosive relation with random walk behaviour, but no explosive behaviour. Since β_ρ is completely specified the model can be estimated by regression for each value of ρ. This in turn results in a profile likelihood in ρ which can then be maximised by a grid search. Searching in the region $\rho > 1$ there is a unique maximum to the likelihood function of 10.74 at $\hat{\rho} = 1.200$. The test statistic for H against M is 0.06 which is small compared to the $\chi^2(1)$ distribution arising from Theorem 3.4.

Returning to the cointegrating vector β_1^* it can be tested that the trend is absent, so $\delta_1 = 0$. This reduces the likelihood further to 10.64 for $\hat{\rho} = 1.204$. The test statistic for this hypothesis within the model H is therefore 0.20 which is also small compared to a $\chi^2(1)$ distribution. The estimated cointegrating vector is $\hat{\beta} = (1, -0.59)$, indicating that $m_t - 0.59p_t$ is explosive, but has no random walk or linear trend behaviour.

In summary, the above analysis shows that the two variables m_t, p_t each has an explosive common trend and a random walk trend. The series co-explode so $m_t - p_t$ is an $\mathsf{I}(1)$ process, while the differenced series $\Delta_1 m_t, \Delta_1 p_t$ are explosive, but without a random walk component. This indicates that linking for instance $m_t - p_t$ with Δp_t as in the model of Taylor (1991) may not give a balanced regression in this situation.

A suggestion for getting around the issue of the unbalanced regression is given in Nielsen (2004). The idea is to measure inflation as the cost of holding money, $c_t = \Delta_1 P_t / P_t = 1 - \exp(-\Delta_1 p_t)$ and a corresponding measure for the depreciation rate, d_t say. It turns out that a system of $m_t - s_t, c_t, d_t$ can be analysed by standard $\mathsf{I}(1)$ cointegration analysis right through to the end of the hyper-inflation. In that framework Cagan's money demand schedule, and, moreover, Cagan's notion of 'optimal' inflation tax can be discussed.

5 Conclusion

Hyperinflation data have traditionally been analysed using $\mathsf{I}(2)$ models, so real money, $m_t - p_t$, and price growth Δp_t become $\mathsf{I}(1)$ and Cagan's money demand schedule can be found as a cointegrating relation of these. At least for the extreme Yugoslavian hyper-inflation these assumptions are unrealistic as the series involved appear to grow explosively. The presented statistical model gives a new tool for analysing such data.

When applying the presented statistical model to data it is found that $m_t - p_t$ can very well be $\mathsf{I}(1)$ whereas Δp_t is explosive. This leads to the somewhat negative conclusion that in this case these variables cannot be linked as usually done in the empirical hyper-inflation literature. A solution to this empirical problem is proposed in the companion empirical analysis in Nielsen (2004).

A Proof of representation theorem

Proof of Theorem 2.2. When μ, δ_1 are both zero the result follows as in Johansen and Schaumburg (1998, Theorem 4). That result is formulated for processes without explosive roots, but the proof actually applies without that restriction. The only difference is that their use of complex conjugates has to be replaced by inverses, which of course would amount to the same if z were on the complex unit circle.

For the general result replace ε_t by $\varepsilon_t+\mu+\alpha_1\delta_1'(1-\rho)t$. To see that the deterministic term is of the desired form note that $C_1\sum_{s=1}^t\{\mu+\alpha_1\delta_1'(1-\rho)s\}$ gives a linear trend while $C_\rho\sum_{s=1}^t\rho^{t-s}\{\mu+\alpha_1\delta_1'(1-\rho)s\}$ gives a linear trend and an exponential trend due to the formulas $\sum_{s=1}^t\rho^{t-s}(1-\rho)=(1-\rho^t)$ and $\sum_{s=1}^t s\rho^{t-s}(1-\rho)=t-\rho(1-\rho^t)/(1-\rho)$.

Let $D_t=\tau_c+\tau_l t+\tau_x\rho^t$ denote the deterministic trend of X_t. Replace X_t by $\tilde{X}_t+D_t$ in the model equation (2.5) giving

$$\begin{aligned}\Delta_1\Delta_\rho\tilde{X}_t+(1-\rho)\tau_l &= \alpha_1\beta_1'\left\{\Delta_\rho\tilde{X}_{t-1}+(1-\rho)\tau_c+\rho\tau_l+(1-\rho)\tau_l t\right\}\\ &\quad+(1-\rho)\alpha_1\delta_1't+\alpha_\rho\beta_\rho'\left\{\Delta_1\tilde{X}_{t-1}+\tau_l-(1-\rho)\rho^{t-1}\tau_x\right\}\\ &\quad+\sum_{j=1}^{k-2}\Phi_j\left\{\Delta_1\Delta_\rho\tilde{X}_{t-j}+(1-\rho)\tau_l\right\}+\mu+\varepsilon_t.\end{aligned}$$

Collecting the coefficient to $\rho^{t-1}, t, 1$ then results in the equations

$$\beta_\rho'\tau_x=0,\qquad \beta_1'\tau_l+\delta_1'=0,\qquad (1-\rho)\Psi_1\tau_l=\alpha_1\beta_1'\{(1-\rho)\tau_c+\rho\tau_l\}+\mu.$$

The first two equations are easily solved for $\beta_\rho'\tau_x$ and $\beta_1'\tau_l$. To get $\beta_{1\perp}'\tau_l$ and thereby τ_l pre-multiply the third equation with $\alpha_{1\perp}'$ and post-multiply Ψ_1 with $I_p=\overline{\beta}_1\beta_1'+\beta_{1\perp}\overline{\beta}_{1\perp}'$. Finally, to get $\beta_1'\tau_c$ pre-multiply the third equation with $\overline{\alpha}_1'$. ∎

B Analysis of the likelihood function

In this appendix, the likelihood function is analysed by analytic means. The likelihood itself is discussed in §B.1. As it is of a regression-type its maximum can be expressed in terms of the maximum likelihood estimator for the variance parameter Ω. Hence, expressions for the estimators of Ω in the unrestricted model M and under the hypothesis H are derived, which can facilitate subsequent asymptotic analysis.

Some additional notation is used. For a matrix α define $\alpha^{\otimes 2}=\alpha\alpha'$. The notation $(Y_t|Z_t)$ denotes the residual of the least squares regression of Y_t on Z_t and the symbol $S_{XY\cdot Z}$ is used for the partial sample covariance of X_t and Y_t given Z_t, that is $T^{-1}\sum_{t=1}^T X_t(Y_t|Z_t)$.

B.1 The likelihood function

The likelihood function for the unrestricted model M is written down and a profile likelihood for the parameters of the conditional expectation is derived.

Two equivalent expressions for the model equation were given in (2.1) and (2.5):

$$\begin{aligned} \varepsilon_t &= \Delta_1 X_t - \alpha_1 \beta_1^{*\prime} X_{t-1}^* - \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} - \mu, && \text{(B.1)} \\ &= \Delta_1 \Delta_\rho X_t - \alpha_\rho \beta_\rho' \Delta_1 X_{t-1} - \alpha_1 \beta_1^{*\prime} \Delta_\rho X_{t-1}^* - \sum_{j=1}^{k-2} \Phi_j \Delta_1 \Delta_\rho X_{t-j} - \mu. && \text{(B.2)} \end{aligned}$$

Due to the normality of the innovations the log likelihood function is

$$-2 \log L(\vartheta, \Omega) = T \log(\det \Omega) + \operatorname{tr}\left(\Omega^{-1} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t'\right),$$

where ϑ represents the parameters $\Pi, \Pi_l, \Gamma_1, \ldots, \Gamma_{k-1}, \mu$ in the conditional expectations equation (B.1). For each value of ϑ the likelihood function has a unique maximum for Ω given by $\hat{\Omega}(\vartheta) = T^{-1} \sum_{t=1}^{T} \varepsilon_t \varepsilon_t'$. The profile likelihood for ϑ is therefore

$$-2 \log L(\vartheta) = -2 \max_\Omega \log L(\vartheta, \Omega) = T \log \det\{\hat{\Omega}(\vartheta)\}. \qquad \text{(B.3)}$$

In the probabilistic analysis the properties of the likelihood function will be analysed for each parameter $(\vartheta^\circ, \Omega_\circ)$ satisfying the hypothesis H. Using the model equation (2.5) restricted by H the innovations of the model can then be written as

$$\varepsilon_t^\circ = \Delta_1 \Delta_{\rho_\circ} X_t - \alpha_\rho^\circ \beta_\rho^{\circ\prime} \Delta_1 X_{t-1} - \alpha_1^\circ \beta_1^{\circ *\prime} \Delta_{\rho_\circ} X_{t-1}^* - \sum_{j=1}^{k-2} \Phi_j^\circ \Delta_1 \Delta_{\rho_\circ} X_{t-j} - \mu^\circ. \qquad \text{(B.4)}$$

In particular it holds that $\hat{\Omega}(\vartheta^\circ) = T^{-1} \sum_{t=1}^{T} \varepsilon_t^\circ \varepsilon_t^{\circ\prime}$.

B.2 The unrestricted model: Variance estimator

In the following an expression for the variance estimator in the unrestricted model M is derived. Some notation is needed. Introduce the vectors

$$\begin{aligned} U_{\rho,t-1}^\circ &= \left\{ \left(\beta_\rho^{\circ\prime} \Delta_1 X_{t-1}\right)', \left(\Delta_1 \Delta_{\rho_\circ} X_{t-1}\right)', \ldots, \left(\Delta_1 \Delta_{\rho_\circ} X_{t-k+2}\right)' \right\}', \\ W_{t-1}^\circ &= \beta_{\rho\perp}^{\circ\prime} \Delta_1 X_{t-1}, && \text{(B.5)} \end{aligned}$$

and define the residuals

$$R_{0,t} = \left(\Delta_1 \Delta_{\rho_\circ} X_t \,\middle|\, U_{\rho,t-1}^\circ, W_{t-1}^\circ, 1\right), \qquad R_{1,t} = \left(\Delta_{\rho_\circ} X_{t-1} \,\middle|\, U_{\rho,t-1}^\circ, W_{t-1}^\circ, 1\right).$$

Due to the innovation equation (B.4) these are linked by

$$R_{\varepsilon,t} = \left(\varepsilon_t^{\circ} \left| U_{\rho,t-1}^{\circ}, W_{t-1}^{\circ}, 1\right.\right) = R_{0,t} - \alpha_1^{\circ} \beta_1^{*\circ\prime} R_{1,t}. \tag{B.6}$$

The following result then holds.

Lemma B.1 *Consider the unrestricted model* M *and a process satisfying* $(B.4)$ *and hence the restricted model* H. *Then the maximum likelihood estimator for* Ω *satisfies*

$$T\hat{\Omega}_{\mathsf{M}} = \sum_{t=1}^{T} \left\{ R_{\varepsilon,t} + \left(\alpha_1^{\circ}\beta_1^{*\circ\prime} - \hat{\alpha}_1^{\mathsf{M}}\hat{\beta}_1^{*\mathsf{M}\prime}\right) R_{1,t} \right\}^{\otimes 2},$$

where $\hat{\alpha}_1^{\mathsf{M}}, \hat{\beta}_1^{*\mathsf{M}}$ *are maximum likelihood estimators for* α_1, β_1^* *in the model* M. *In particular,* $\hat{\alpha}_1^{\mathsf{M}}, \hat{\beta}_1^{*\mathsf{M}}$ *are obtained by reduced rank regression of* $R_{0,t}$ *on* $R_{1,t}$.

Proof of Lemma B.1. The model equation (B.1) is a reduced rank regression equation. The profile likelihood for $(\alpha_1, \beta_1^*) = \{\alpha/(1-\rho_\circ), \beta^*\}$ is therefore found by eliminating $\Gamma_1, \ldots, \Gamma_{k-1}, \mu$ by partial regression,

$$\hat{\Omega}_{\mathsf{M}}(\alpha, \beta^*) = \min_{\Gamma_1,\ldots,\Gamma_{k-1},\mu} \hat{\Omega}(\vartheta) = T^{-1}\sum_{t=1}^{T}\left(\tilde{R}_{0,t} - \alpha_1\beta_1^{*\prime}\tilde{R}_{1,t}\right)^{\otimes 2},$$

where, following the notation of Johansen (1996, §6), $\tilde{R}_{0,t}$ and $\tilde{R}_{1,t}$ are the residuals

$$\begin{aligned} \tilde{R}_{0,t} &= \left(\Delta_1 X_t \left| \Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}, 1\right.\right), \\ \tilde{R}_{1,t} &= \left\{(1-\rho_\circ) X_{t-1} \left| \Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}, 1\right.\right\}. \end{aligned}$$

Using the identity (2.3) it is seen that $\tilde{R}_{0,t} = R_{0,t}$ and $\tilde{R}_{1,t} = R_{1,t}$.

The profile likelihood for α_1, β_1 is maximised by reduced rank regression giving unique maximum likelihood estimators $\hat{\alpha}_1^{\mathsf{M}}, \hat{\beta}_1^{\mathsf{M}*}$ and

$$\hat{\Omega}_{\mathsf{M}} = T^{-1}\sum_{t=1}^{T}(R_{\hat{\varepsilon},t})^{\otimes 2} \qquad \text{where} \qquad R_{\hat{\varepsilon},t} = R_{0,t} - \hat{\alpha}_1^{\mathsf{M}}\hat{\beta}_1^{*\mathsf{M}\prime}R_{1,t}.$$

Subtracting and adding $R_{\varepsilon,t}$ as defined in (B.6) then gives the desired result. ∎

B.3 The restricted model: Variance estimator

In the following an expression for the variance estimator is derived for the restricted model H where $\beta_\rho = \beta_\rho^{\circ}$.

Recall the parameter $\tau_\perp = \Psi_\rho \overline{\beta}^\circ_{\rho\perp}$ with orthogonal complement $\tau = (I_p - \tau_\perp \overline{\tau}'_\perp)\alpha_\rho$. Two random versions of $\tau_\perp$ as well as a random version of α_1 are needed:

$$\hat{\tau}_\perp = \hat{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp} \qquad \breve{\tau}_\perp = \breve{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp}, \qquad \breve{\alpha}_1 = \frac{1-\rho_\circ}{1-\hat{\rho}}\alpha_1^\circ \tag{B.7}$$

defined in terms of the quantities

$$\hat{\Psi}_\rho = I + \frac{\hat{\Pi}_1}{1-\hat{\rho}} - \sum_{j=1}^{k-2} \hat{\rho}^{-j}\hat{\Phi}_j, \qquad \breve{\Psi}_\rho = I + \frac{\Pi_1^\circ}{1-\rho} - \sum_{j=1}^{k-2} \rho^{-j}\Phi_j^\circ. \tag{B.8}$$

Moreover, introduce the vector

$$U_{\hat{\rho},t-1} = \left\{ \left(\beta_\rho^{\circ\prime}\Delta_1 X_{t-1}\right)', (\Delta_1\Delta_{\hat{\rho}}X_{t-1})', \ldots, (\Delta_1\Delta_{\hat{\rho}}X_{t-k+2})' \right\}',$$

and define the residuals

$$R^{\mathsf{H}}_{0,t} = (\Delta_1\Delta_{\hat{\rho}}X_t \,|U_{\hat{\rho},t-1}, 1), \qquad R^{\mathsf{H}}_{1,t} = \left(\Delta_{\hat{\rho}}X^*_{t-1}\,|U_{\hat{\rho},t-1}, 1\right).$$

A result like Lemma B.1 for the restricted model H can now be stated.

Lemma B.2 *Consider the restricted model* H *and a process satisfying* $(B.4)$. *Then*

$$\begin{aligned} T\hat{\Omega}_{\mathsf{H}} \;&=\; \sum_{t=1}^T \left\{ R_{\varepsilon,t} + \left(\alpha_1^\circ\beta_1^{*\circ\prime} - \hat{\alpha}_1^{\mathsf{H}}\hat{\beta}_1^{*\mathsf{H}\prime}\right) R_{1,t}\right\}^{\otimes 2} \\ &\quad + \left[\hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}}(\breve{\tau}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}})^{-1}\breve{\tau}^{\mathsf{H}\prime}\sum_{t=1}^T \hat{\varepsilon}_t^{\mathsf{H}} W^\circ_{t-1}\left\{\sum_{t=1}^T \left(W^\circ_{t-1}\right)^2\right\}^{-1/2}\right]^{\otimes 2}, \end{aligned}$$

where $\hat{\alpha}_1^{\mathsf{H}}, \hat{\beta}_1^{*\mathsf{H}}, \hat{\tau}^{\mathsf{H}}$ *and* $\hat{\varepsilon}_t^{\mathsf{H}}$ *are maximum likelihood estimators and residuals. In particular,* $\hat{\alpha}_1^{\mathsf{H}}, \hat{\beta}_1^{*\mathsf{H}}$ *are obtained by reduced rank regression of* $R^{\mathsf{H}}_{0,t}$ *on* $R^{\mathsf{H}}_{1,t}$.

Proof of Lemma B.2. The profile likelihood for $\rho, \alpha_1, \beta_1^*$ is given in terms of

$$\hat{\Omega}_{\mathsf{H}}(\rho,\alpha,\beta^*) = \min_{\alpha_1,\Phi_1,\ldots,\Phi_{k-2},\mu} \hat{\Omega}\,(\vartheta) = T^{-1}\sum_{t=1}^T \left(R^{\mathsf{H},\rho}_{0,t} - \alpha_1\beta_1^{*\prime}R^{\mathsf{H},\rho}_{1,t}\right)^{\otimes 2},$$

where the profile residuals depend on ρ and are given by

$$R^{\mathsf{H},\rho}_{0,t} = (\Delta_1\Delta_\rho X_t\,|U_{\rho,t-1}, 1), \qquad R^{\mathsf{H},\rho}_{1,t} = \left(\Delta_\rho X^*_{t-1}\,|U_{\rho,t-1},1\right).$$

Once again, profile estimates for α_1, β_1^* given ρ are found by reduced rank regression of $R^{\mathsf{H},\rho}_{0,t}$ on $R^{\mathsf{H},\rho}_{1,t}$ giving estimators $\hat{\alpha}_1^{\mathsf{H},\rho}, \hat{\beta}_1^{*\mathsf{H},\rho}$ and

$$\hat{\Omega}_{\mathsf{H}}(\rho) = T^{-1}\sum_{t=1}^T \left(\hat{\varepsilon}_t^{\mathsf{H},\rho}\right)^{\otimes 2} \qquad \text{where} \qquad \hat{\varepsilon}_t^{\mathsf{H},\rho} = R^{\mathsf{H},\rho}_{0,t} - \hat{\alpha}_1^{\mathsf{H},\rho}\hat{\beta}_1^{*\mathsf{H},\rho\prime}R^{\mathsf{H},\rho}_{1,t}.$$

Minimising the determinant of $\hat{\Omega}_{\mathsf{H}}(\rho)$ with respect to ρ gives maximum likelihood estimators $\hat{\rho}, \hat{\alpha}_1^{\mathsf{H}}, \hat{\beta}_1^{*\mathsf{H}}$ and $\hat{\Omega}_{\mathsf{H}} = T^{-1} \sum_{t=1}^{T} \left(\hat{\varepsilon}_t^{\mathsf{H}}\right)^{\otimes 2}$ where $\hat{\varepsilon}_t^{\mathsf{H}}$ can be written

$$\hat{\varepsilon}_t^{\mathsf{H}} = R_{0,t}^{\mathsf{H}} - \hat{\alpha}_1^{\mathsf{H}} \hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}} = \left(R_{0,t}^{\mathsf{H}} \left| \hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}} \right. \right) = \left(\Delta_1 \Delta_{\hat{\rho}} X_t \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right).$$

The residuals of the restricted and unrestricted model are related through partial regression. Since $U_{\hat{\rho},t}, W_{t-1}^{\circ}$ and $U_{\rho,t-1}^{\circ}, W_{t-1}^{\circ}$ span the same space it holds

$$R_{0,t} = \left(R_{0,t}^{\mathsf{H}} \left| W_{t-1}^{\circ} \right. \right), \qquad R_{1,t} = \left(R_{1,t}^{\mathsf{H}} \left| W_{t-1}^{\circ} \right. \right).$$

Using partial regression on W_{t-1}° the variance estimator therefore satisfies

$$T\hat{\Omega}_{\mathsf{H}} = \sum_{t=1}^{T} \left(\hat{\varepsilon}_t^{\mathsf{H}} \left| W_{t-1}^{\circ} \right. \right)^{\otimes 2} + \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ} \left\{ \sum_{t=1}^{T} \left(W_{t-1}^{\circ} \right)^2 \right\}^{-1} \sum_{t=1}^{T} W_{t-1}^{\circ} \hat{\varepsilon}_t^{\mathsf{H}\prime}. \tag{B.9}$$

The first term in (B.9) resembles the expression in Lemma B.1 in that

$$\sum_{t=1}^{T} \left(\hat{\varepsilon}_t^{\mathsf{H}} \left| W_{t-1}^{\circ} \right. \right)^{\otimes 2} = \sum_{t=1}^{T} \left\{ R_{\varepsilon,t} + \left(\alpha_1^{\circ} \beta_1^{*\circ\prime} - \hat{\alpha}_1^{\mathsf{H}} \hat{\beta}_1^{*\mathsf{H}\prime} \right) R_{1,t} \right\}^{\otimes 2}.$$

For the second term in (B.9), it is first argued that

$$\begin{aligned} &\left\{ \Delta_1 X_{t-1} - \hat{\Pi}_1^* \begin{pmatrix} X_{t-2} \\ t \end{pmatrix} - \sum_{j=1}^{k-2} \hat{\Gamma}_j \Delta_1 X_{t-j-1} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right\} \\ &\quad = \hat{\tau}_{\perp} \left(W_{t-1}^{\circ} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right). \end{aligned} \tag{B.10}$$

The three components in the regressand are rewritten one by one. First, noting that $I_p = \overline{\beta}_{\rho\perp}^{\circ} \beta_{\rho\perp}^{\circ\prime} + \overline{\beta}_{\rho}^{\circ} \beta_{\rho}^{\circ\prime}$ it holds

$$\left(\Delta_1 X_{t-1} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right) = \overline{\beta}_{\rho\perp}^{\circ} \left(W_{t-1}^{\circ} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right).$$

Secondly, by the identity

$$(1 - \hat{\rho}) \begin{pmatrix} X_{t-2} \\ t \end{pmatrix} = \Delta_{\hat{\rho}} X_{t-1}^* - \begin{pmatrix} \Delta_1 X_{t-1} \\ 0 \end{pmatrix}$$

it holds

$$\left\{ \hat{\Pi}_1^* \begin{pmatrix} X_{t-2} \\ t \end{pmatrix} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right\} = \hat{\Pi}_1 \overline{\beta}_{\rho\perp}^{\circ} \left(W_{t-1}^{\circ} \left| \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1 \right. \right).$$

Thirdly, use (2.4) for a similar substitution in the third term.

Returning to the second term in in (B.9) consider the likelihood equation for ρ. Following the approach of Johansen (1996, p.182) this is

$$0 = \operatorname{tr}\left[\hat{\Omega}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}\left\{\Delta_1 X_{t-1} - \hat{\Pi}_1^*\begin{pmatrix} X_{t-2} \\ t \end{pmatrix} - \sum_{j=1}^{k-2}\hat{\Gamma}_j\Delta_1 X_{t-j-1}\right\}'\right].$$

Due to the identity (B.10) this reduces to

$$0 = \operatorname{tr}\left\{\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}\left(\hat{\tau}_\perp W_{t-1}^\circ\right)'\right\} = \hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^\circ, \tag{B.11}$$

and so, by the identity $I_p = \Omega\hat{\tau}(\breve{\tau}'\Omega\hat{\tau})^{-1}\breve{\tau}' + \breve{\tau}_\perp(\hat{\tau}_\perp'\Omega^{-1}\breve{\tau}_\perp)^{-1}\hat{\tau}_\perp'\Omega^{-1}$ it holds, as desired,

$$\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^\circ = \hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}}(\breve{\tau}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}})^{-1}\breve{\tau}^{\mathsf{H}\prime}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^\circ.$$

■

B.4 The restricted model: The residuals

The residuals are more difficult to handle in the restricted model H than in the unrestricted model M. In the following an expression akin to (B.6) is therefore derived. Sub-sequently this is used for deriving an analytic expression for $\hat{\rho}$.

Lemma B.3 *Recall $\breve{\tau}_\perp$, $\breve{\alpha}_1$ defined in (B.7). It holds that*

$$R_{0,t}^{\mathsf{H}} = \left\{\varepsilon_t^\circ + \breve{\alpha}_1\beta_1^{*\circ\prime}\Delta_{\hat{\rho}}X_{t-1}^* + (\rho_\circ - \hat{\rho})\,\breve{\tau}_\perp\beta_{\rho\perp}^{\circ\prime}\Delta_1 X_{t-1}\middle|\, U_{\hat{\rho},t-1}, 1\right\}.$$

Defining $R_{\varepsilon,t}^{\mathsf{H}} = (\varepsilon_t^\circ\,|U_{\hat{\rho},t-1}, 1)$ it follows that

$$R_{0,t}^{\mathsf{H}} = R_{\varepsilon,t}^{\mathsf{H}} + \breve{\alpha}_1\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}} + (\rho_\circ - \hat{\rho})\,\breve{\tau}_\perp\left(W_{t-1}^\circ|U_{\hat{\rho},t-1}, 1\right).$$

Proof of Lemma B.3. Note first the identities

$$\begin{aligned}
(1-\rho_\circ)\,\Delta_\rho X_{t-1}^* &= (1-\rho)\,\Delta_{\rho_\circ}X_{t-1}^* + (\rho - \rho_\circ)\begin{pmatrix}\Delta_1 X_{t-1} \\ 0\end{pmatrix}, && \text{(B.12)}\\
\Delta_1\Delta_\rho X_{t-j} &= \Delta_1\Delta_{\rho_\circ}X_{t-j} + (\rho_\circ - \rho)\,\Delta_1 X_{t-j-1} && \text{(B.13)}\\
\Delta_1 X_{t-j-1} &= \rho_\circ^{-j}\left(\Delta_1 X_{t-1} - \sum_{l=1}^{j}\rho_\circ^{l-1}\Delta_1\Delta_{\rho_\circ}X_{t-l}\right). && \text{(B.14)}
\end{aligned}$$

Now, rewrite the regressand $\Delta_1\Delta_{\hat{\rho}}X_t$ of $R_{0,t}^{\mathsf{H}}$ using (B.13) as

$$\Delta_1\Delta_{\hat{\rho}}X_t = \Delta_1\Delta_{\rho_\circ}X_t + (\rho_\circ - \hat{\rho})\Delta_1 X_{t-1}.$$

Then substitute $\Delta_1\Delta_{\rho_\circ}X_t$ using the equation (B.4) to arrive at

$$\begin{aligned}\Delta_1\Delta_{\hat\rho}X_t &= \varepsilon_t^\circ + \alpha_1^\circ\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1} + \alpha_\rho^\circ\beta_\rho^{\circ\prime}\Delta_1 X_{t-1} \\ &\quad + \sum_{j=1}^{k-2}\Phi_j^\circ\Delta_1\Delta_{\rho_\circ}X_{t-j} + (\rho_\circ - \hat\rho)\Delta_1 X_{t-1} + \mu^\circ. \qquad \text{(B.15)}\end{aligned}$$

Finally, substitute the terms $\Delta_{\rho_\circ}X_{t-1}$ and $\Delta_1\Delta_{\rho_\circ}X_{t-j}$ using (B.12)-(B.14) to get

$$\begin{aligned}\Delta_1\Delta_{\hat\rho}X_t &= \varepsilon_t^\circ + \frac{1-\rho_\circ}{1-\hat\rho}\alpha_1^\circ\beta_1^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^* + (\rho_\circ - \hat\rho)\,\breve\Psi_{\hat\rho}\Delta_1 X_{t-1} \\ &\quad + (\rho_\circ - \hat\rho)\sum_{j=1}^{k-2}\Phi_j^\circ\sum_{l=1}^{j}\hat\rho^{l-1}\Delta_1\Delta_{\hat\rho}X_{t-l} + \mu^\circ.\end{aligned}$$

The expression for $R_{0,t}^{\mathsf{H}}$ follows since the regression on $U_{\hat\rho,t-1}$ eliminates the constant and the $\Delta_1\Delta_{\hat\rho}X_{t-l}$ terms. ■

An analytic expression for the maximum likelihood estimator $\hat\rho$ in the restricted model H can now be found. This looks at bit tedious at present, but will be rather helpful for the asymptotic analysis.

Introduce the notation $\widehat{\mathsf{corr}}(x_t, y_t) = (\sum_{t=1}^T x_t^{\otimes 2})^{-1/2}\sum_{t=1}^T x_t y_t'(\sum_{t=1}^T y_t^{\otimes 2})^{-1/2}$ as well as $\widehat{\mathsf{m}}(x_t) = \sum_{t=1}^T x_t^{\otimes 2}$ and define

$$\begin{aligned}C_{\varepsilon_t^\circ, W_{t-1}^\circ} &= \widehat{\mathsf{corr}}\left(\varepsilon_t^\circ, W_{t-1}^\circ \,\middle|\, \hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right), \\ C_{\varepsilon_t^\circ, \hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}} &= \widehat{\mathsf{corr}}\left(\varepsilon_t^\circ, \hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}} \,\middle|\, \hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right), \\ C_{\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ} &= \widehat{\mathsf{corr}}\left(\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ \,\middle|\, \hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right),\end{aligned}$$

$$M_{W_{t-1}^\circ} = \widehat{\mathsf{m}}\left(W_{t-1}^\circ \,\middle|\, \hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right), \qquad M_{\varepsilon_t^\circ} = \widehat{\mathsf{m}}\left(\varepsilon_t^\circ \,\middle|\, \hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right)^{\otimes 2}.$$

Lemma B.4 *In the model* H *the maximum likelihood estimator for* ρ *satisfies*

$$\begin{aligned}&\left(\hat\tau_\perp^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\tau_\perp\right)(\hat\rho - \rho_\circ)\,M_{W_{t-1}^\circ}^{-1/2} = \hat\tau_\perp^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}M_{\varepsilon_t^\circ}{}^{1/2}C_{\varepsilon_t^\circ, W_{t-1}^\circ} \\ &\quad -\hat\tau_\perp^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\alpha_1\left(\hat\alpha_1^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\alpha_1\right)^{-1}\hat\alpha_1^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}M_{\varepsilon_t^\circ}^{1/2}C_{\varepsilon_t^\circ, \hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}}C_{\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ} \\ &\quad -\hat\tau_\perp^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\alpha_1\left(\hat\alpha_1^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\alpha_1\right)^{-1}\hat\alpha_1^{\mathsf{H}\prime}\hat\Omega_{\mathsf{H}}^{-1}\breve\tau_\perp(\rho_\circ - \hat\rho)\,M_{W_{t-1}^\circ}^{1/2}C'_{\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ}C_{\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ}.\end{aligned}$$

Proof of Lemma B.4. The likelihood equation for $\hat{\rho}$ in (B.11) is

$$0 = \hat{\tau}_{\perp}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^{\circ}.$$

Following Johansen (1995, p.182) the likelihood equation for β_1^* is

$$0 = \hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}\left(\Delta_{\hat{\rho}}X_{t-1}^{*}\right)'.$$

Post-multiplying this expression with $\hat{\beta}_{1\perp}^{*\mathsf{H}}$ and using $\hat{\varepsilon}_t^{\mathsf{H}} = (R_{0,t}^{\mathsf{H}}|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}})$ this implies

$$0 = \hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}\left(\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right)'.$$

Now, replace $R_{0,t}^{\mathsf{H}}$ occuring in $\hat{\varepsilon}_t^{\mathsf{H}}$ by the expression found in Lemma B.3 to get

$$\begin{aligned}
0 &= \hat{\tau}_{\perp}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\left\{\varepsilon_t^{\circ} + \breve{\alpha}_1\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}} + (\rho_{\circ}-\hat{\rho})\,\breve{\tau}_{\perp}W_{t-1}^{\circ}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right\}W_{t-1}^{\circ},\\
0 &= \hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\left\{\varepsilon_t^{\circ} + \breve{\alpha}_1\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}} + (\rho_{\circ}-\hat{\rho})\,\breve{\tau}_{\perp}W_{t-1}^{\circ}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right\}\left(\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right)'.
\end{aligned}$$

In both equations, post-multiply $\beta_1^{*\circ\prime}$ by $I_{p+1} = \bar{\hat{\beta}}_{1\perp}^{*\mathsf{H}}\hat{\beta}_{1\perp}^{*\mathsf{H}\prime} + \bar{\hat{\beta}}_1^{*\mathsf{H}}\hat{\beta}_1^{*\mathsf{H}\prime}$, noting that the term $\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}$ disappears by regression. Solve the first equation for $(\rho_{\circ}-\hat{\rho})$ and the second for $\beta_1^{*\circ\prime}\bar{\hat{\beta}}_{1\perp}^{*\mathsf{H}}$ to get

$$\begin{aligned}
&\left(\hat{\tau}_{\perp}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\breve{\tau}_{\perp}\right)(\hat{\rho}-\rho_{\circ})\sum_{t=1}^{T}\left(W_{t-1}^{\circ}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right)^2\\
&= \hat{\tau}_{\perp}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\left\{\varepsilon_t^{\circ} + \breve{\alpha}_1\left(\beta_1^{*\circ\prime}\bar{\hat{\beta}}_{1\perp}^{*\mathsf{H}}\right)\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right\}W_{t-1}^{\circ},\\
&\left(\hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\breve{\alpha}_1\right)\left(-\beta_1^{*\circ\prime}\bar{\hat{\beta}}_{1\perp}^{*\mathsf{H}}\right)\sum_{t=1}^{T}\left\{\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right\}^{\otimes 2}\\
&= \hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^{T}\left\{\varepsilon_t^{\circ} + (\rho_{\circ}-\hat{\rho})\,\breve{\tau}_{\perp}W_{t-1}^{\circ}\left|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right.\right\}\left(\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right).
\end{aligned}$$

Insert the second expression in the first to get the desired expression. ∎

B.5 The restricted model: A reparametrisation

For the unrestricted model M the analytic expressions for the estimators can be given directly. For the restricted model H, however, the estimators are expressed in terms of the estimator for the explosive root ρ, for which no analytic expression is available. In the following a reparametrisation of the model is therefore suggested which will facilitate the asymptotic analysis of $\hat{\rho}$.

Lemma B.5 *Suppose β_ρ is identified by $\beta_\rho' \overline{\beta}_\rho^\circ = I_{p-1}$ while β_1^* satisfies* (3.2). *Let*

$$\Psi_{\rho_\circ} = I + \frac{\Pi_1}{1-\rho_\circ} - \sum_{j=1}^{k-2} \rho_\circ^{-j} \Phi_j,$$

and define $\theta = (\theta_\rho^U, \theta_1^U, \theta_{\rho,1}^U, \ldots, \theta_{\rho,k-2}^U, \theta^V, \theta^W, \theta^\mu)$ where

$$\begin{aligned}
\theta_\rho^U &= \alpha_\rho^\circ - \alpha_\rho + (\rho_\circ - \rho)\, \Psi_{\rho_\circ} \overline{\beta}_\rho^\circ, \\
\theta_1^U &= \alpha_1^\circ - \frac{1-\rho}{1-\rho_\circ} \alpha_1, \\
\theta_{\rho,j}^U &= \Phi_j^\circ - \frac{\rho}{\rho_\circ} \Phi_j + (\rho_\circ - \rho) \sum_{m=j+1}^{k-2} \Phi_m \rho_\circ^{j-1-m}, \\
\theta^V &= -\frac{1-\rho}{1-\rho_\circ} \Pi_1^* \overline{\beta}_{1\perp}^{*\circ}, \\
\theta^W &= -\alpha_\rho \beta_\rho' \overline{\beta}_{\rho\perp}^\circ + (\rho_\circ - \rho)\, \Psi_{\rho_\circ} \overline{\beta}_{\rho\perp}^\circ, \\
\theta^\mu &= \mu^\circ - \mu.
\end{aligned}$$

Then the error term satisfies $\varepsilon_{\vartheta,t} = \varepsilon_t^\circ + \theta S_{t-1}^\circ$. Note the identity

$$\sum_{j=1}^{k-2} \rho^{-j} \theta_{\rho,j}^U = \sum_{j=1}^{k-2} \rho^{-j} \Phi_j^\circ - \sum_{j=1}^{k-2} \rho_\circ^{-j} \Phi_j. \tag{B.16}$$

Proof of Lemma B.5. Consider the error equation (B.1). Replace $\Delta_1 \Delta_\rho X_t$, $\Delta_\rho X_{t-1}^*$, $\Delta_1 \Delta_\rho X_{t-j}$ by the expressions (B.15), (B.12) and (B.13)-(B.14) to get

$$\begin{aligned}
\varepsilon_{\vartheta,t} = {} & \varepsilon_t^\circ + \left(\Pi_\rho^\circ - \Pi_\rho\right) \Delta_1 X_{t-1} + \left(\Pi_1^{\circ *} - \frac{1-\rho}{1-\rho_\circ} \Pi_1^*\right) \Delta_{\rho_\circ} X_{t-1}^* \\
& + \sum_{j=1}^{k-2} \left(\Phi_j^\circ - \Phi_j + (\rho_\circ - \rho) \sum_{m=j}^{k-2} \Phi_m \rho_\circ^{j-1-m}\right) \Delta_1 \Delta_{\rho_\circ} X_{t-j} \\
& + (\mu^\circ - \mu) + (\rho_\circ - \rho) \left(I_p + \frac{\Pi_1}{1-\rho_\circ} - \sum_{j=1}^{k-2} \Phi_j \rho_\circ^{-j}\right) \Delta_1 X_{t-1}.
\end{aligned}$$

Now, pre-multiply $\Delta_1 X_{t-1}$ with the identity $I_p = \overline{\beta}^{\circ}_{\rho}\beta^{\circ\prime}_{\rho} + \overline{\beta}^{\circ}_{\rho\perp}\beta^{\circ\prime}_{\rho\perp}$ and $\Delta_{\rho_\circ} X^*_{t-1}$ with the identity $I_{p+1} = \overline{\beta}^{*\circ}_1 \beta^{*\circ\prime}_1 + \overline{\beta}^{*\circ}_{1\perp}\beta^{*\circ\prime}_{1\perp}$, and use the identification in (3.2).

The identity (B.16) follows by straight forward inspection. ■

The matrix $\breve{\Psi}_\rho$ introduced in (B.8) is non-zero in the following sense.

Lemma B.6 *Suppose Assumption 2.1 is satisfied. Let $k = 2$. It then holds*

$$(1-\rho)\breve{\Psi}_\rho \overline{\beta}^{\circ}_{\rho\perp} \neq 0 \qquad \text{for all } \rho > 1.$$

In particular it holds that a $c > 0$ exists so for all $\rho \geq 1$ then

$$\left|\breve{\Psi}_\rho \overline{\beta}^{\circ}_{\rho\perp}\right| \geq c$$

Proof. Consider the companion form $\mathbf{x}_t = A^\circ \mathbf{x}_{t-1} + \mathbf{e}^\circ_t$, where

$$\mathbf{x}_t = \begin{pmatrix} \beta^{\circ\prime}_\rho \Delta_1 X_t \\ \Delta_{\rho_\circ} X_t \\ \vdots \\ \Delta_{\rho_\circ} X_{t-k+2} \end{pmatrix}, \qquad \mathbf{e}^\circ_t = \begin{pmatrix} \beta^{\circ\prime}_\rho \\ I_p \\ 0 \end{pmatrix} \varepsilon^\circ_t,$$

and A° is the matrix

$$\left\{ \begin{array}{cccccc} \rho_\circ I + \beta^{\circ\prime}_\rho \alpha^\circ_\rho & \beta^{\circ\prime}_\rho (\Pi^\circ_1 + \Phi^\circ_1) & \beta^{\circ\prime}_\rho (\Phi^\circ_2 - \Phi^\circ_1) & \cdots & \beta^{\circ\prime}_\rho \left(\Phi^\circ_{k-2} - \Phi^\circ_{k-3}\right) & -\beta^{\circ\prime}_\rho \Phi^\circ_{k-2} \\ \alpha^\circ_\rho & I + \Pi^\circ_1 + \Phi^\circ_1 & \Phi^\circ_2 - \Phi^\circ_1 & \cdots & \Phi^\circ_{k-2} - \Phi^\circ_{k-3} & -\Phi^\circ_{k-2} \\ & I & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & I & \end{array} \right\}.$$

Due to Theorem 2.2 the eigenvalues of the companion matrix A° are bounded by one in absolute value, $|\text{eigen}(A^\circ)| \leq 1$. This in turn implies that

$$|\text{eigen}(\rho I - A^\circ)| \geq \rho - 1 > 0 \quad \text{for } \rho > 1.$$

Partition the matrix $\rho I - A^\circ$ as a (2×2) block matrix with a $(2p-1)$-dimensional upper left block. The rule for determinants of partitioned matrices implies invertibility of

$$\left[\begin{array}{cc} (\rho - \rho_\circ) I - \beta^{\circ\prime}_\rho \alpha^\circ_\rho & -\beta^{\circ\prime}_\rho \left\{ \Pi^\circ_1 + \sum_{j=1}^{k-2} (\rho - 1)\rho^{-j} \Phi^\circ_1 \right\} \\ -\alpha^\circ_\rho & (\rho - 1) I - \Pi^\circ_1 - \sum_{j=1}^{k-2} (\rho - 1)\rho^{-j} \Phi^\circ_1 \end{array} \right]$$

Now, assume the first of the stated results does not hold. Then

$$(\rho - 1)\,\breve{\Psi}_\rho = \eta \beta_\rho^{\circ\prime} \qquad \text{for some } \eta \in \mathbf{R}^{p\times(p-1)}.$$

This in turn implies that the companion matrix satisfies

$$\rho I - A^\circ = \left\{ \begin{pmatrix} (\rho - \rho_\circ)\, I - \beta_\rho^{\circ\prime}\alpha_\rho^\circ \\ -\alpha_\rho^\circ \end{pmatrix}, \begin{pmatrix} \beta_\rho^{\circ\prime}\eta - (\rho - 1)\, I_{p-1} \\ \eta \end{pmatrix} \beta_\rho^{\circ\prime} \right\}$$

which is a reduced rank matrix *contradicting* that $\rho I - A^\circ$ is invertible.

The second result is proved by looking at four different special cases. (i) When ρ belongs to any compact set so $\rho > 1$ then both $\breve{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp}$ and $1 - \rho$ have outcomes in compact sets not including zero due to the first result. (ii) For large ρ then $\breve{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp}$ converges to $\overline{\beta}^\circ_{\rho\perp}$. (iii) For ρ approaching 1 and $\beta_1^{\circ\prime}\overline{\beta}^\circ_{\rho\perp} \neq 0$ then $\breve{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp}$ has an infinite asymptote. (iv) For $\beta_1^{\circ\prime}\overline{\beta}^\circ_{\rho\perp} = 0$ it suffices to argue that $\rho_{\min}$ can be chosen so close to one that $\breve{\Psi}_\rho \overline{\beta}^\circ_{\rho\perp}$ has no zero points for $1 \leq \rho \leq \rho_{\min}$. Now, the vector polynomial $(I - \sum_{j=1}^{k-2} z^j \Phi_j^\circ)\overline{\beta}^\circ_{\rho\perp}$ has finitely many zero points in a neighbourhood of 1. It suffices to show that such zero points cannot fall at one. Assumption 2.1(C) implies that $\Psi_1^\circ \beta_{1\perp}^\circ = \{I_p + \alpha_\rho^\circ \beta_\rho^{\circ\prime}/(\rho_\circ - 1) - \sum_{j=1}^{k-2} z^j \Phi_j^\circ\}\beta_{1\perp}^\circ$ is non-zero. Since $\beta_1^{\circ\prime}\overline{\beta}^\circ_{\rho\perp} = 0$ then $\beta^\circ_{\rho\perp} \in \mathrm{span}(\beta^\circ_{1\perp})$ this implies that

$$\Psi_1^\circ \beta^\circ_{\rho\perp} = \left(I_p + \frac{\alpha_\rho^\circ \beta_\rho^{\circ\prime}}{\rho_\circ - 1} - \sum_{j=1}^{k-2} z^j \Phi_j^\circ \right) \beta^\circ_{\rho\perp} = \left(I_p - \sum_{j=1}^{k-2} z^j \Phi_j^\circ \right) \beta^\circ_{\rho\perp}$$

is non-zero as desired. ■

C Asymptotic analysis

In this appendix the asymptotic properties of the likelihood function are explored. This done for a data generating process with parameters $(\upsilon^\circ, \Omega^\circ)$ and innovations ε_t°, see (B.4). The asymptotic theory is unusual in two ways. First, the explosive root is unknown and it therefore has to be estimated giving a non-linear estimation problem as that for the $\mathsf{I}(2)$ analysis considered by Johansen (1997). Secondly, because of the explosive characteristic roots the usual asymptotic theory for stationary and integrated time series has to be enhanced.

This section is organised as follows. In §C.1 some general asymptotic results for a vector autoregression are listed. These are mainly due to Lai and Wei (1985) and adaptations by Nielsen (2005). The asymptotic theory for the unrestricted variance estimator in model M then follows in §C.2. The consistency result for the restricted model H of Theorem 3.3 is given in §C.3 followed by an analysis of the restricted variance estimator in model H in §C.4. This in turn leads to an improved consistency rate for $\hat{\rho}$ in §C.5. Finally, in §C.6 the asymptotic distribution of the test statistic stated in Theorem 3.4 is derived.

C.1 *Asymptotic results for vector autoregressive processes*

In the following the data generating process is defined and asymptotic results of Lai and Wei (1985) as modified by Nielsen (2005) are reported.

It is convenient to introduce the notation.

$$U_{t-1}^{\circ} = \{(\Delta_1 X_{t-1})' \beta_{\rho}^{\circ}, (\Delta_{\rho_\circ} X_{t-1}^{*})' \beta_1^{*\circ}, (\Delta_1 \Delta_{\rho_\circ} X_{t-1})', \ldots, (\Delta_1 \Delta_{\rho_\circ} X_{t-k+2})'\}', \quad \text{(C.1)}$$

$$V_t^{\circ} = \beta_{1\perp}^{*\circ\prime} \Delta_{\rho_\circ} X_t^{*}, \qquad W_t^{\circ} = \beta_{\rho\perp}^{\circ\prime} \Delta_1 X_t, \quad \text{(C.2)}$$

where W_t° was also introduced in (B.5). According to the Granger-Johansen representation in theorem 2.2 the process U_t° can be given a stationary initial distribution while V_t° and W_t° are the common trends of random-walk and of explosive behaviour, respectively. The companion vectors for the process X_t can then be wirtten as

$$S_t^{\circ} = (U_t^{\circ\prime}, V_t^{\circ\prime}, W_t^{\circ\prime}, 1)', \quad \text{(C.3)}$$

which satisfies a first order autoregressive equation.

The following Lemma summarises the asymptotic behaviour of product moments of the processes $\varepsilon_t^{\circ}, S_{t-1}^{\circ}$. Recall the normalisations N_V, N_W given in (3.3) and define $N_S = \mathrm{diag}(I_{\dim U}, N_V, N_W, 1)$.

Lemma C.1 *Consider the process X_t given by (B.4) and suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Let ξ, η be constants satisfying $\xi < \gamma/(1+\gamma)$ and $\eta > 0$, recalling the definition of γ in Assumption 3.1.*
Define sample variances $\widehat{\mathsf{var}}(x_t) = T^{-1} \sum_{t=1}^{T} x_t x_t'$. Then

(*i*) $\widehat{\mathsf{var}}(\varepsilon_t^{\circ}) \overset{a.s.}{=} \Omega_\circ + \mathrm{o}(T^{-\xi})$ *and* $\widehat{\mathsf{var}}(\varepsilon_t^{\circ}|1) \overset{a.s.}{=} \Omega_\circ + \mathrm{o}(T^{-\xi})$.
(*ii*) $\widehat{\mathsf{var}}(U_{t-1}|1) \overset{a.s.}{\to} \Sigma_U > 0$.
(*iii*) $\rho_\circ^{-2T} T \widehat{\mathsf{var}}(W_{t-1}|1) \overset{a.s.}{\to} \Sigma_W \overset{a.s.}{>} 0$.
(*iv*) $\widehat{\mathsf{var}}(N_V V_{t-1}|1) \overset{\mathsf{D}}{\to} \Sigma_V \overset{a.s.}{>} 0$.
(*v*) $\widehat{\mathsf{var}}\,(N_S S_{t-1}) \overset{\mathsf{D}}{\to} \Sigma_S \overset{a.s.}{>} 0$.

The matrices $\Sigma_W, \Sigma_V, \Sigma_S$ are stochastic, while $\Omega_\circ, \Sigma_U$ are deterministic. The sample variance of the joint process S also satisfies

(*vi*) $\max\mathrm{eigen}(\sum_{t=1}^{T} S_{t-1} S_{t-1})' \overset{a.s.}{=} \mathrm{O}(\rho_\circ^{2T})$.

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*

(vii) $\widehat{\mathsf{corr}}(S_{t-1}, \varepsilon_t) \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2})$.
$(viii)$ $\widehat{\mathsf{corr}}\{(U_{t-1}', V_{t-1}', 1)', \varepsilon_t\} \overset{a.s.}{=} \mathrm{o}(T^{\eta-1/2})$.
(ix) $\widehat{\mathsf{corr}}(W_{t-1}, 1) \overset{a.s.}{=} \mathrm{O}(T^{-1/2})$.
(x) $\widehat{\mathsf{corr}}(U_{t-1}, W_{t-1}|1) \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2})$.
(xi) $\widehat{\mathsf{corr}}(V_{t-1}, W_{t-1}|1) \overset{a.s.}{=} \mathrm{o}(1) = \mathrm{O}_{\mathsf{P}}(T^{-1/2})$.
(xii) $\widehat{\mathsf{corr}}(U_{t-1}, V_{t-1}|1) \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2}) = \mathrm{o}_{\mathsf{P}}(T^{\eta-1/2})$.

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

$(xiii)$ $T^{-1/2} \sum_{t=1}^T N_V (V_{t-1}^{\circ}|1) \varepsilon_t^{\circ\prime} \overset{\mathsf{D}}{\to} \Sigma_{V\varepsilon}$.
(xiv) $T^{-1/2} \sum_{t=1}^T N_V (V_{t-1}^{\circ}|1) U_{t-1}^{\circ\prime} \overset{\mathsf{D}}{\to} \Sigma_{VU}$.

Proof of Lemma C.1. Most of the results follow from Nielsen (2005), noting that the definition of S_t is slightly different from here.

(i): Corollary 2.6. (ii): Example 6.6. (iii): Corollary 7.2 and Theorem 9.1. (vi): Theorem 7.1. (vii),$(viii)$: Theorem 2.4 and Corollary 2.6. (ix): Theorem 9.1. (x): Table 2. (xi): Theorem 9.2, Remark 9.3 and (ix). (xii): Example 6.6 and Theorem 9.4. The results (iv), $(xiii)$ can be proved using the techniques of Chan and Wei (1988). For (xiv) see also Johansen (1995, §B). ■

C.2 Asymptotic theory for the unrestricted variance estimator

The expression for the unrestricted variance estimator Ω^M is analysed asymptotically. For the sake of discussing the log likelihood ratio test statistic LR it suffices to show consistency for the estimators $\widehat{\alpha}_1^M$, $\widehat{\beta}_1^{*M}$. These are reduced rank estimators, so consistency can be shown along the lines of Johansen (1995, §13).

Lemma C.2 *Consider the maximum likelihood estimators in the unrestricted model* M *identified by* (3.2)*. Consider processes* X_t *satisfying* (*B.4*) *and hence also* H *and suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then, for some positive definite matrix* $\Sigma_{\beta\beta}$ *and some random matrices* $\Sigma_{\beta\varepsilon}, \Sigma_{V\varepsilon}$*, it holds*

$$\hat{\Omega}_{\mathsf{M}} \overset{\mathsf{P}}{\to} \Omega_{\circ}, \qquad \begin{aligned} T^{1/2}\left(\hat{\alpha}_1^{\mathsf{M}} - \alpha_1^{\circ}\right) &= \Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1} + \mathrm{o}_{\mathsf{P}}(1), \\ T^{1/2} N_V^{-1} \overline{\beta}_{1\perp}^{*\circ\prime} \hat{\beta}_1^{*\mathsf{M}} &= \Sigma_V^{-1}\Sigma_{V\varepsilon} + \mathrm{o}_{\mathsf{P}}(1). \end{aligned}$$

First a result like Lemma 10.3 of Johansen (1995) is needed.

Lemma C.3 *Consider the process X_t given by (B.4) and suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then,*

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\beta_1^{*\circ\prime}R_{1,t}\\ R_{0,t}\end{pmatrix}^{\otimes 2} &\overset{\mathsf{P}}{\to} \begin{pmatrix}\Sigma_{\beta\beta} & \Sigma_{\beta\beta}\alpha_1^{\circ\prime}\\ \alpha_1^{\circ}\Sigma_{\beta\beta} & \Omega_{\circ}+\alpha_1^{\circ}\Sigma_{\beta\beta}\alpha_1^{\circ\prime}\end{pmatrix},\\
\frac{1}{T}\sum_{t=1}^{T}(N_V\beta_{1\perp}^{*\circ\prime}R_{1,t})^{\otimes 2} &\overset{\mathsf{D}}{\to} \Sigma_V,\\
\frac{1}{T^{1/2}}\sum_{t=1}^{T}(N_V\beta_{1\perp}^{*\circ\prime}R_{1,t})\,\varepsilon_t' &\overset{\mathsf{D}}{\to} \Sigma_{V\varepsilon},\\
\frac{1}{T^{1/2}}\sum_{t=1}^{T}(N_V\beta_{1\perp}^{*\circ\prime}R_{1,t})\,R_{1,t}'\beta_1^{*\circ} &= \mathrm{O}_{\mathsf{P}}(1).
\end{aligned}
$$

Proof of Lemma C.3. It is first argued, with ξ defined in Lemma C.1, that

$$
M^{\circ}=\frac{1}{T}\sum_{t=1}^{T}\left\{\left.\begin{matrix}\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^{*}\\ U_{\rho,t-1}^{\circ}\end{pmatrix}\\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^{*}\\ \varepsilon_t^{\circ}\\ N_W W_{t-1}^{\circ}\end{matrix}\right|1\right\}^{\otimes 2}= \tag{C.4}
$$

$$
\begin{bmatrix}\Sigma_U+\mathrm{o}_{\mathsf{P}}(1) & \mathrm{O}_{\mathsf{P}}(T^{-1/2}) & \mathrm{o}_{\mathsf{P}}(T^{-\xi/2}) & \mathrm{o}_{\mathsf{P}}(1)\\
 & \Sigma_V+\mathrm{o}_{\mathsf{P}}(1) & T^{-1/2}\{\Sigma_{V\varepsilon}+\mathrm{o}_{\mathsf{P}}(1)\} & \mathrm{O}_{\mathsf{P}}(T^{-1/2})\\
 & & \Omega_{\circ}+\mathrm{o}_{\mathsf{P}}(1) & T^{-1/2}\{\Sigma_{W\varepsilon}+\mathrm{o}_{\mathsf{P}}(1)\}\\
 & & & \Sigma_W+\mathrm{o}_{\mathsf{P}}(1)\end{bmatrix},
$$

where $\Sigma_{W\varepsilon}=\rho_{\circ}^{-T}\sum_{t=1}^{T}\varepsilon_t^{\circ}W_{t-1}^{\circ}$. This result arises using the following matrix of items from Lemma C.1

$$
\begin{Bmatrix}(ii) & (xiv) & (vii) & (x)\\ & (iv) & (xiii) & (xi)\\ & & (i) & (ix)\\ & & & (iii,ix)\end{Bmatrix}
$$

Partialling out $U_{\rho,t-1}^{\circ}, W_{t-1}^{\circ}$ shows that

$$
\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\beta_1^{*\circ\prime}R_{1,t}\\ N_V\beta_{1\perp}^{*\circ\prime}R_{1,t}\\ R_{\varepsilon,t}\end{pmatrix}^{\otimes 2}=\begin{bmatrix}\Sigma_{\beta\beta}+\mathrm{o}_{\mathsf{P}}(1) & \mathrm{O}_{\mathsf{P}}(T^{-1/2}) & \mathrm{o}_{\mathsf{P}}(T^{-\xi/2})\\ & \Sigma_V+\mathrm{o}_{\mathsf{P}}(1) & T^{-1/2}\{\Sigma_{V\varepsilon}+\mathrm{o}_{\mathsf{P}}(1)\}\\ & & \Omega_{\circ}+\mathrm{o}_{\mathsf{P}}(1)\end{bmatrix}.
$$

The desired result then follows by noting that $R_{0,t}=\alpha_1^{\circ}\beta_1^{*\circ\prime}R_{1,t}+R_{\varepsilon,t}$, see (B.6). ■

The asymptotic theory for cointegration analysis can now be derived exactly as in §10-13 of Johansen (1995), replacing his Lemma 10.3 by the above Lemma C.3.

Proof of Lemma C.2. Use the arguments in the proof of Lemmas 13.1 and 13.2 of Johansen (1995). ■

Proof of Theorem 3.5. Use the arguments in the proof of Lemma 11.1 of Johansen (1995). ■

C.3 Consistency in the restricted model

The consistency of the maximum likelihood estimators in the restricted model is now argued. The argument is given in three steps exploring the profile likelihood for ϑ derived in §B.1 above. First, consistency is argued for $\hat{\Omega}$ as well as the auxillary parameter θ defined in §B.5. Secondly, the consistency of $\hat{\rho}$ as stated in Theorem 3.3 is proved. Finally, the consistency of the remaining parameters is proved.

Lemma C.4 *Consider the model given by the regression equation* $\varepsilon_{\theta,t} = \varepsilon_t^\circ + \theta S_{t-1}^\circ$ *as outlined in Lemma B.5, where* θ, Ω *vary freely, so* $\theta \in \mathbf{R}^{p \times \dim S}$ *and* Ω *is positive definite. Then the maximum likelihood estimator for* θ, Ω *exists with probability tending to one, and satisfies* $\hat{\theta} N_S^{-1} \xrightarrow{\mathsf{P}} 0$.

Note that this proves Theorem 3.3(i), as $\mu = \theta^\mu$.

Proof of Lemma C.4. As outlined in §B.1 the level curves of the likelihood function are given by $\det\{\hat{\Omega}(\theta)\}$ where $T\hat{\Omega}(\theta) = \sum_{t=1}^T \varepsilon_{\theta,t}^{\otimes 2}$. Partial regression then implies that

$$T\hat{\Omega}(\theta) = \sum_{t=1}^T \left(\varepsilon_{\theta,t} | S_{t-1}^\circ\right)^{\otimes 2} + \sum_{t=1}^T \varepsilon_{\theta,t} S_{t-1}^{\circ\prime} \left\{\sum_{t=1}^T \left(S_{t-1}^\circ\right)^{\otimes 2}\right\}^{-1} \sum_{t=1}^T S_{t-1}^\circ \varepsilon_{\theta,t}'.$$

According to Lemma C.1,i,v,vii it holds

$$\begin{aligned} T^{-1} \sum_{t=1}^T \left(\varepsilon_{\theta,t} | S_{t-1}^\circ\right)^{\otimes 2} &\xrightarrow{\mathsf{P}} \Omega_\circ, \\ T^{-1/2} \sum_{t=1}^T \varepsilon_t^\circ S_{t-1}^\circ \left\{\sum_{t=1}^T \left(S_{t-1}^\circ\right)^{\otimes 2}\right\}^{-1/2} &\xrightarrow{\mathsf{P}} 0, \\ T^{-1} \sum_{t=1}^T \left(N_S S_{t-1}^\circ\right)^{\otimes 2} &\xrightarrow{\mathsf{D}} \Sigma_S \overset{a.s.}{>} 0. \end{aligned}$$

The level curves then satisfy

$$T\hat{\Omega}(\theta) = \Omega_\circ + \mathsf{o}_\mathsf{P}(1) + \left\{\mathsf{o}_\mathsf{P}(1) + \theta N_S^{-1}\Sigma_S^{1/2}\right\}\left\{\mathsf{o}_\mathsf{P}(1) + \theta N_S^{-1}\Sigma_S^{1/2}\right\}'.$$

Thus for any parameter value θ and any constant $\zeta > 0$ so $\left\|\theta N_S^{-1}\right\| \geq \zeta$ then $R \geq \zeta^2\Sigma_S + \mathsf{o}_\mathsf{P}(1)$, which is non-zero, positive semi-definite, and not depending on θ as desired. Since ζ is arbitrary the consistency result follows. ■

The consistency result of $\hat{\rho}$ can now be established.

Proof of Theorem 3.3(ii). First, the assumption of non-convergence of $\hat{\rho}$ is combined with the convergence for $\hat{\theta}$. On the one hand the assumption to $\hat{\rho}$ implies that there exists $\kappa, \delta > 0$ so $\mathsf{P}(N_W^{-1}|\hat{\rho} - \rho_\circ| > \kappa) > \delta$ infinitely often. On the other hand Lemma C.4 shows that $\hat{\theta}N_S^{-1} \xrightarrow{\mathsf{P}} 0$. Thus, for all $\zeta > 0$ exists a sub-sequence (T') so for all T' it holds $\mathsf{P}(\Theta_{T'}) > \delta$ where

$$\Theta_{T'} = (N_W^{-1}|\hat{\rho} - \rho_\circ| > \kappa \text{ and } |\hat{\theta}N_S^{-1}| < \zeta).$$

Secondly, consider outcomes in $\Theta_{T'}$. Then it holds that

$$\left|\frac{\hat{\theta}^W N_W^{-1}}{(\hat{\rho} - \rho_\circ)N_W^{-1}}\right| = \left|\frac{\hat{\theta}^W}{\hat{\rho} - \rho_\circ}\right| < \frac{\zeta}{\kappa}. \tag{C.5}$$

Since $\beta_\rho = \beta_\rho^\circ$ under the hypothesis H then

$$\frac{\hat{\theta}^W}{\hat{\rho} - \rho_\circ} = \Psi_{\rho_\circ}\overline{\beta}_{\rho\perp}^\circ = \breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ + \left(\Psi_{\rho_\circ} - \breve{\Psi}_\rho\right)\overline{\beta}_{\rho\perp}^\circ.$$

Recall that ρ belongs to a closed subset of the parameter space, hence, $\rho \geq \rho_{\min}$ for some $\rho_{\min} > 1$. Due to Lemma B.6 the absolute value of the first term first term is therefore bounded from below by some $c > 0$ not depending on $\rho_{\min}$. The second term can be rewritten using the identity (B.16) as

$$\left(\Psi_{\rho_\circ} - \breve{\Psi}_\rho\right)\overline{\beta}_{\rho\perp}^\circ = \frac{1}{1-\rho}\left(\theta_1^U\beta_\rho^{\circ\prime} + \theta^V\beta_{\rho\perp}^{\circ\prime}\right) - \sum_{j=1}^{k-2}\rho^{-j}\theta_{\rho,j}^U.$$

Due to the consistency of the estimator for θ established in Lemma C.4 and the inequalities $|\rho|^{-j} \leq |\rho_{\min}|^{-j}$ and $|1-\rho|^{-1} \leq |1-\rho_{\min}|^{-1}$ then for all η the norm of the second term is bounded by η for large T'. Since ζ, η are arbitrary they can be chosen to *contradict* the inequality (C.5). ■

The consistency of the remaining parameters can now be argued.

Proof of Theorem 3.3(iii, iv)**.** Once $\hat{\rho}$ is known the remaining estimators can be derived from the auxillary estimator $\hat{\theta}$. Due to the consistency of $\hat{\rho}$ and of $\hat{\theta}$ established in Theorem 3.3(ii) and Lemma C.4 the consistency can now be established.

Starting with $\hat{\alpha}_1$ note that $\hat{\alpha}_1 = (\alpha_1^\circ - \hat{\theta}_1^U)(1-\rho_\circ)/(1-\hat{\rho})$. Since $\hat{\theta}_1^U$ and $\hat{\rho}$ are $o_{\mathsf{P}}(1)$ then $\hat{\alpha}_1 = \alpha_1^\circ + o_{\mathsf{P}}(1)$.

Similarly the consistency of $\hat{\Phi}_{k-2}, \ldots, \hat{\Phi}_1, \hat{\Pi}_1^* \overline{\beta}_{1\perp}^{*\circ}, \hat{\alpha}_\rho$ can be established one by one from $\hat{\theta}_{\rho,k-2}^U, \ldots, \hat{\theta}_{\rho,1}^U, \hat{\theta}^V, \hat{\theta}_\rho^U$. ■

C.4 Asymptotic theory for the restricted variance estimator

The expression for the unrestricted variance estimator Ω^{H} is analysed asymptotically. Here, the consistency for the reduced rank estimators $\widehat{\alpha}_1^{\mathsf{H}}$, $\widehat{\beta}_1^{*\mathsf{H}}$ is shown. Once again this is done along the lines of Johansen (1995, §13), see also Lemma C.2.

Lemma C.5 *Consider the restricted model* H *identified by* (3.2) *and processes* X_t *satisfying* (B.4)*. Suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then,*

$$\hat{\Omega}_{\mathsf{H}} \xrightarrow{\mathsf{P}} \Omega_\circ, \qquad \begin{aligned} T^{1/2}\left(\hat{\alpha}_1^{\mathsf{H}} - \alpha_1^\circ\right) &= \Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1} + o_{\mathsf{P}}(1), \\ T^{1/2} N_V^{-1} \overline{\beta}_{1\perp}^{*\circ\prime} \hat{\beta}_1^{*\mathsf{H}} &= \Sigma_{VV}^{-1}\Sigma_{V\varepsilon} + o_{\mathsf{P}}(1). \end{aligned}$$

First a result like Lemma 10.3 of Johansen (1995) and Lemma C.3 is needed.

Lemma C.6 *Consider the process* X_t *given by* (B.4) *with* $\rho \geq \rho_{\min}$ *and suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then,*

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T \begin{pmatrix} \beta_1^{*\circ\prime} R_{1,t}^{\mathsf{H}} \\ R_{0,t}^{\mathsf{H}} \end{pmatrix}^{\otimes 2} &\xrightarrow{\mathsf{P}} \begin{pmatrix} \Sigma_{\beta\beta} & \Sigma_{\beta\beta}\alpha_1^{\circ\prime} \\ \alpha_1^\circ \Sigma_{\beta\beta} & \Omega_\circ + \alpha_1^\circ \Sigma_{\beta\beta}\alpha_1^{\circ\prime} \end{pmatrix}, \\
\frac{1}{T}\sum_{t=1}^T \left(N_V \beta_{1\perp}^{*\circ\prime} R_{1,t}^{\mathsf{H}}\right)^{\otimes 2} &\xrightarrow{\mathsf{D}} \Sigma_V, \\
\frac{1}{T^{1/2}}\sum_{t=1}^T \left(N_V \beta_{1\perp}^{*\circ\prime} R_{1,t}^{\mathsf{H}}\right)\varepsilon_t' &\xrightarrow{\mathsf{D}} \Sigma_{V\varepsilon}, \\
\frac{1}{T}\sum_{t=1}^T \left(N_V \beta_{1\perp}^{*\circ\prime} R_{1,t}^{\mathsf{H}}\right) R_{1,t}^{\mathsf{H}\prime}\beta_1^{*\circ} &= O_{\mathsf{P}}(1).
\end{aligned}$$

Proof of Lemma C.6. As in the proof of Lemma C.3 the idea is to argue that

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\begin{array}{c|c}\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^{*}\\ U_{\hat\rho,t-1}\end{pmatrix} & \\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^{*} & 1\\ \varepsilon_t^{\circ} & \\ N_W W_{t-1}^{\circ} & \end{array}\right\}^{\otimes 2}=M^{\circ},$$

where M° is defined in (C.4). To see this, note that the relations (B.12)-(B.14) imply

$$\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^{*}\\ U_{\hat\rho,t-1}\\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^{*}\end{pmatrix}=\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^{*}\\ U_{\rho,t-1}^{\circ}\\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^{*}\end{pmatrix}+(\hat\rho-\rho_\circ)\begin{pmatrix}Q_{U,1}\\ Q_{U,\rho}\\ N_VQ_V\end{pmatrix}S_{t-1},$$

for some deterministic matrices Q.

When taking sums of squares of this expression the last term can be omitted since $T^{-1}(\hat\rho-\rho_\circ)^2\sum_{t=1}^{T}S_{t-1}^{\otimes 2}=\mathsf{o}_{\mathsf{P}}(1)$ due Theorem 3.3(ii) and Lemma C.1,vi, applied with a similar argument for the cross terms.

When considering the cross product with ε_t° the last term can likewise be ignored since $T^{-1}(\hat\rho-\rho_\circ)\sum_{t=1}^{T}S_{t-1}\varepsilon_t^{\circ\prime}=\mathsf{o}_{\mathsf{P}}(T^{-\xi/2})$ using Theorem 3.3(ii) and Lemma C.1,vii.

When considering the cross product with W_{t-1}° the last term can also be ignored since $T^{-1}(\hat\rho-\rho_\circ)\sum_{t=1}^{T}S_{t-1}W_{t-1}^{\circ}=\mathsf{o}_{\mathsf{P}}(N_W^{-1})$ by Theorem 3.3(ii) and Lemma C.1,vi.

Now, partial out $U_{\hat\rho,t-1}$ to see that, with ξ defined in Lemma C.1, it holds

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\begin{array}{c}\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}}\\ N_V\beta_{1\perp}^{*\circ\prime}R_{1,t}^{\mathsf{H}}\\ R_{\varepsilon,t}^{\mathsf{H}}\\ N_W\left(W_{t-1}^{\circ}|U_{\hat\rho,t-1}\right)\end{array}\right\}^{\otimes 2}$$
$$=\left[\begin{array}{cccc}\Sigma_{\beta\beta}+\mathsf{o}_{\mathsf{P}}(1) & \mathsf{O}_{\mathsf{P}}\left(T^{-1/2}\right) & \mathsf{o}_{\mathsf{P}}\left(T^{-\xi/2}\right) & \mathsf{o}_{\mathsf{P}}(1)\\ & \Sigma_V+\mathsf{o}_{\mathsf{P}}(1) & T^{-1/2}\{\Sigma_{V\varepsilon}+\mathsf{o}_{\mathsf{P}}(1)\} & \mathsf{O}_{\mathsf{P}}\left(T^{-1/2}\right)\\ & & \Omega_\circ+\mathsf{o}_{\mathsf{P}}(1) & \mathsf{o}_{\mathsf{P}}\left(T^{-\xi/2}\right)\\ & & & \Sigma_W+\mathsf{o}_{\mathsf{P}}(1)\end{array}\right].$$

The result now follows by noting the link between $R_{0,t}^{\mathsf{H}}$, $R_{1,t}^{\mathsf{H}}$, $R_{\varepsilon,t}^{\mathsf{H}}$ established in Lemma B.3 and using the consistency in Theorem 3.3. ■

The asymptotic theory for cointegration analysis can now be derived exactly as in §10-13 of Johansen (1995), replacing his Lemma 10.3 by the above Lemma C.6.

Proof of Lemma C.5. Use the arguments in the proof of Lemmas 13.1 and 13.2 of Johansen (1995). ■

C.5 Improving the rate of consistency in the restricted model

Theorem 3.3, which was proved above, shows that parameters are consistent. In particular, it holds that $\hat{\rho} - \rho_\circ = \mathrm{o_P}(N_W^{-1})$. This rate of consistency can be improved by a factor of $T^{-\xi/2}$, where ξ arose in Lemma C.1.

Lemma C.7 *Consider the model* H *with* $\rho \geq \rho_{\min}$ *and suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then, for all* $\xi < \gamma/(1+\gamma)$ *it holds*

$$\hat{\rho}_{\mathsf{H}} = \rho_\circ + \mathrm{o_P}(T^{-\xi/2}N_W^{-1}).$$

Proof of Lemma C.7. Recall the expression for $\hat{\rho}$ in Lemma B.4. The desired result follows from the consistency of the estimators established in Theorem 3.3 and then applying following orders of magnitude

$$M_{W_{t-1}^\circ} = \mathrm{O_P}\left(\rho_\circ^{2T}\right), \qquad M_{\varepsilon_t^\circ} = \mathrm{O_P}\left(T\right),$$

$$C_{\varepsilon_t^\circ, W_{t-1}^\circ} = \mathrm{o_P}\left(T^{-\xi/2}\right), \qquad C_{\varepsilon_t^\circ, \hat{\beta}_{1\perp}^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}}} = \mathrm{o_P}\left(T^{-\xi/2}\right), \qquad C_{\hat{\beta}_{1\perp}^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}}, W_{t-1}^\circ} = \mathrm{O_P}\left(T^{-1/2}\right),$$

following from the proof of Lemma C.6. ■

As a corollary the asymptotic behaviour of the leading term in the statistic H in Theorem 3.4 can be established.

Lemma C.8 *Consider the restricted model* H *with* $\rho \geq \rho_{\min}$ *and a process* X_t *given by* (*B.4*). *Suppose Assumptions 2.1, 3.1, 3.2 are satisfied. Then,*

$$\rho_\circ^{-T}\check{\tau}^{\mathsf{H}\prime}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^\circ = \rho_\circ^{-T}\tau'\sum_{t=1}^{T}\varepsilon_t^\circ W_{t-1}^\circ + \mathrm{o_P}\left(T^{1/2-\xi}\right).$$

Proof of Lemma C.8. Using the identity for $R_{0,t}^{\mathsf{H}}$ established in Lemma B.3 and the projection identity $I_{p+1} = \overline{\hat{\beta}}_1^*\hat{\beta}_1^{*\prime} + \overline{\hat{\beta}}_{1\perp}^*\hat{\beta}_{1\perp}^{*\prime}$ then $\hat{\varepsilon}_t^{\mathsf{H}} = (R_{0,t}^{\mathsf{H}}|\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}})$ satisfies

$$\hat{\varepsilon}_t^{\mathsf{H}} = \left\{R_{\varepsilon,t}^{\mathsf{H}} + \frac{1-\rho_\circ}{1-\hat{\rho}}\alpha_1^\circ\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^*\hat{\beta}_{1\perp}^{*\prime}R_{1,t}^{\mathsf{H}} + (\rho_\circ - \hat{\rho})\,\check{\tau}_\perp^{\mathsf{H}}W_{t-1}^\circ\,\middle|\,\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}}\right\}.$$

Due to this usage $\hat{\beta}_{1\perp}^*$ and the identification (3.2) the estimator $\hat{\beta}_{1\perp}^*$ can be chosen as $\hat{\beta}_{1\perp}^{*\mathsf{H}} = (I_{p+1} - \overline{\beta}_1^{*\circ}\hat{\beta}_1^{*\mathsf{H}\prime})\beta_{1\perp}^{*\circ}$, see Johansen (1995, p.191). It then follows from Lemma C.5 that $N_V^{-1}\overline{\beta}_{1\perp}^{*\circ\prime}\hat{\beta}_1^{*\mathsf{H}}$ and $N_V^{-1}\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\beta_1^{*\circ}$ are both $\mathrm{O_P}(T^{-1/2})$. Using in addition Theorem 3.3 it holds

$$\check{\tau}^{\mathsf{H}\prime}\hat{\varepsilon}_t^{\mathsf{H}} = \check{\tau}^{\mathsf{H}\prime}\left\{R_{\varepsilon,t}^{\mathsf{H}} + \mathrm{O_P}(T^{-1/2})N_V\hat{\beta}_{1\perp}^{*\prime}R_{1,t}^{\mathsf{H}}\,\middle|\,\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}}\right\}.$$

Moreover, it follows that

$$\left(\begin{array}{c}\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}}\\ N_V\hat{\beta}_{1\perp}^{*\prime}R_{1,t}^{\mathsf{H}}\end{array}\right)=\left\{\begin{array}{cc}I & \mathrm{O}_{\mathsf{P}}\left(T^{-1/2}\right)\\ \mathrm{O}_{\mathsf{P}}\left(T^{-1/2}\right) & I\end{array}\right\}\left(\begin{array}{c}\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}}\\ N_V\beta_{1\perp}^{*\circ\prime}R_{1,t}^{\mathsf{H}}\end{array}\right).$$

A result as that in the proof of Lemma C.6 therefore holds when $\beta_1^{*\circ}$ is estimated. Moreover, due to the improved consistency rate for $\hat{\rho}$ it follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\beta_1^{*\circ\prime}R_{1,t}^{\mathsf{H}}\left(W_{t-1}^{\circ}\left|U_{\hat{\rho},t-1}\right.\right) &= \mathrm{o}_{\mathsf{P}}\left(T^{-\xi/2}\right),\\ \frac{1}{T^{1/2}}\sum_{t=1}^{T}R_{\varepsilon,t}^{\mathsf{H}}N_W\left(W_{t-1}^{\circ}\left|U_{\hat{\rho},t-1}\right.\right) &= \rho_{\circ}^{-T}\sum_{t=1}^{T}\varepsilon_t^{\circ}W_{t-1}^{\circ}+\mathrm{o}_{\mathsf{P}}\left(T^{1/2-\xi}\right),\end{aligned}$$

rather than just $\mathrm{o}_{\mathsf{P}}(1)$ and $\mathrm{o}_{\mathsf{P}}\left(T^{1/2-\xi/2}\right)$, respectively. Therefore, when partialling out $U_{\hat{\rho},t-1}$ as well as $\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}}$ it holds

$$\frac{N_W}{T}\sum_{t=1}^{T}\left\{\begin{array}{c}N_V\hat{\beta}_{1\perp}^{*\prime}R_{1,t}^{\mathsf{H}}\\ R_{\varepsilon,t}^{\mathsf{H}}\end{array}\left|\hat{\beta}_1^{*\prime}R_{1,t}^{\mathsf{H}}\right.\right\}W_{t-1}^{\circ}=\left[\begin{array}{c}\mathrm{O}_{\mathsf{P}}\left(T^{-1/2}\right)\\ T^{-1/2}\left\{\Sigma_{W\varepsilon}+\mathrm{o}_{\mathsf{P}}\left(T^{1/2-\xi}\right)\right\}\end{array}\right].$$

This in turn gives the desired result. ■

C.6 Combining the preliminary results

The asymptotic result in Theorem 3.4 can now be proved by combining the previous results for $\hat{\Omega}_{\mathsf{M}}$ and $\hat{\Omega}_{\mathsf{H}}$. Due to the following auxillary result some terms cancel out.

Lemma C.9 *Suppose $\hat{\alpha}_1$ and $\hat{\beta}_1^*$ are estimators identified by $\hat{\beta}_1^* = \beta_1^{*\circ}+\beta_{1\perp}^{*\circ}\overline{\beta}_{1\perp}^{*\circ\prime}\hat{\beta}_1^*$ and satisfying*

$$T^{1/2}\left(\hat{\alpha}_1-\alpha_1^{\circ}\right)=\Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1}+\mathrm{o}_{\mathsf{P}}(1),\qquad T^{1/2}N_V^{-1}\overline{\beta}_{1\perp}^{*\circ\prime}\hat{\beta}_1^*=\Sigma_V^{-1}\Sigma_{V\varepsilon}+\mathrm{o}_{\mathsf{P}}(1).$$

Consider processes X_t satisfying (B.4) and hence also H, and suppose Assumptions 2.1,3.1,3.2 are satisfied. Then

$$\sum_{t=1}^{T}\left\{R_{\varepsilon,t}+\left(\alpha_1^{\circ}\beta_1^{*\circ\prime}-\hat{\alpha}_1\hat{\beta}_1^{*\prime}\right)R_{1,t}\right\}^{\otimes 2}=Q^{\circ}+\mathrm{o}_{\mathsf{P}}(1),$$

where Q° is some function of $\varepsilon_t^{\circ},S_{t-1}^{\circ},\theta^{\circ},\Omega_{\circ}$ not depending on $\hat{\alpha}_1$,$\hat{\beta}_1^*$.

Proof of Lemma C.9. First note, that by the identification it holds

$$\alpha_1^\circ \beta_1^{*\circ\prime} - \hat{\alpha}_1 \hat{\beta}_1^{*\prime} = -\left(\hat{\alpha}_1 - \alpha_1^\circ\right) \beta_1^{*\circ\prime} - \hat{\alpha}_1 \hat{\beta}_1^{*\prime} \overline{\beta}_{1\perp}^{*\circ} \beta_{1\perp}^{*\circ\prime}.$$

The asymptotic properties for $\hat{\alpha}_1$ and $\hat{\beta}_1^*$ then imply

$$T^{1/2}\left(\alpha_1^\circ \beta_1^{*\circ\prime} - \hat{\alpha}_1 \hat{\beta}_1^{*\prime}\right) = -\left(\Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1}\beta_1^{*\circ\prime} + \alpha_1^\circ \Sigma_{\varepsilon V}\Sigma_V^{-1} N_V \beta_{1\perp}^{*\circ\prime}\right)\left\{1 + o_{\mathsf{P}}(1)\right\}.$$

The expression of interest then equals, by

$$\sum_{t=1}^{T} R_{\varepsilon,t}^{\otimes 2} - \Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\varepsilon} - \alpha_1^\circ \Sigma_{\varepsilon V}\Sigma_V^{-1}\Sigma_{V\varepsilon} - \Sigma_{\varepsilon V}\Sigma_V^{-1}\Sigma_{V\varepsilon}\alpha_1^{\circ\prime} + \alpha_1^\circ \Sigma_{\varepsilon V}\Sigma_V^{-1}\Sigma_{V\varepsilon}\alpha_1^{\circ\prime} + o_{\mathsf{P}}(1),$$

showing that the desired result is satisfied. ■

Proof of Theorem 3.4. Due to the regression nature of the problem the log likelihood ratio test statistic satisfies

$$LR(\mathsf{H}|\mathsf{M}) = -T\log\det(\hat{\Omega}_{\mathsf{M}}\hat{\Omega}_{\mathsf{H}}^{-1}) = -T\log\det\left\{I_p - \hat{\Omega}_{\mathsf{H}}^{-1}\left(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}}\right)\right\}.$$

Lemma C.5 immediately shows that $\hat{\Omega}_{\mathsf{H}} \overset{\mathsf{P}}{\to} \Omega_\circ$. Thus, if it is argued that $T(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}})$ converges in distribution then a Taylor expansion shows

$$LR(\mathsf{H}|\mathsf{M}) = \mathrm{tr}\left\{\Omega_\circ^{-1} T\left(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}}\right)\right\} + o_{\mathsf{P}}(1).$$

Now, combining Lemmas C.2,C.5 shows

$$\begin{aligned} T\left(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}}\right) &= \sum_{t=1}^{T}\left\{R_{\varepsilon,t} + \left(\alpha_1^\circ\beta_1^{*\circ\prime} - \hat{\alpha}_1^{\mathsf{H}}\hat{\beta}_1^{*\mathsf{H}\prime}\right) R_{1,t}\right\}^{\otimes 2} \\ &\quad - \sum_{t=1}^{T}\left\{R_{\varepsilon,t} + \left(\alpha_1^\circ\beta_1^{*\circ\prime} - \hat{\alpha}_1^{\mathsf{M}}\hat{\beta}_1^{*\mathsf{M}\prime}\right) R_{1,t}\right\}^{\otimes 2} \\ &\quad + \left[\hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}}(\breve{\tau}^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}\hat{\tau}^{\mathsf{H}})^{-1}\breve{\tau}^{\mathsf{H}\prime}\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^\circ\left\{\sum_{t=1}^{T}\left(W_{t-1}^\circ\right)^2\right\}^{-1/2}\right]^{\otimes 2}. \end{aligned}$$

The first two terms cancel asymptotically by Lemma C.9. As for the last term note that $\hat{\Omega}_{\mathsf{H}}$ and $\hat{\tau}^{\mathsf{H}}$ are consistent due to Theorem 3.3. By the assumption that $\gamma > 1$ Lemma C.8 shows that

$$\sum_{t=1}^{T}\hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^\circ\left\{\sum_{t=1}^{T}\left(W_{t-1}^\circ\right)^2\right\}^{-1/2} = \sum_{t=1}^{T}\varepsilon_t^\circ W_{t-1}^\circ\left\{\sum_{t=1}^{T}\left(W_{t-1}^\circ\right)^2\right\}^{-1/2} + o_{\mathsf{P}}(1).$$

The desired result follows by recognising the leading term H . ■

D References

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

Cagan, P. (1956) The monetary dynamics of hyper-inflation. In Friedman, M. (ed.), *Studies in the quantity theory of money*, pp. 25-117. Chicago: University of Chicago Press.

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

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

Johansen, S. (1996) *Likelihood-based inference in cointegrated vector autoregressive models.* 2nd printing. Oxford: Oxford University Press.

Johansen, S. (1997) Likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24, 433-462.

Johansen, S. and Schaumburg, E. (1999) Likelihood analysis of seasonal cointegration. *Journal of Econometrics* 88, 301-339.

Juselius, K. and Mladenović, Z. (2002) High inflation, hyper inflation and explosive roots. The case of Jugoslavia. Mimeo. http://eco.uninsubria.it/ESF-EMM/hyper.pdf.

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

Lai, T.L. and Wei, C.Z. (1985) Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah, ed., *Multivariate Analysis VI*, Elsevier Science Publishers, 375-393.

Nielsen, B. (2000) The asymptotic distribution of likelihood ratio test statistics for cointegration in unstable vector autoregressive processes. Discussion paper Nuffield College. See authors webpage.

Nielsen, B. (2001) The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica,* 69, 211-219.

Nielsen, B. (2001a) Weak consistency of criterions for order determination in a general vector autoregression. Discussion paper 2001-W10, Nuffield College. See authors webpage.

Nielsen, B. (2003) Correlograms for non-stationary autoregressions. Discussion paper 2003-W11, Nuffield College.

Nielsen, B. (2004) Money demand in the Yugoslavian hyper-inflation 1991-1994. Discussion paper 2004-W31, Nuffield College.

Nielsen, B. (2005) Strong consistency results for least squares statistics in general vector autoregressive models. Discussion papers 2001-W9 & 2003-W23, Nuffield College. To appear in *Econometric Theory.*

Petrović, P. and Vujošević, Z. (1996) The monetary dynamics in the Yugoslav hyper-inflation of 1991-1993: The Cagan money demand. *European Journal of Political Economy* 12, 467-483. Erratum (1997) in vol. 13, 385-387.

Petrović, P., Bogetić, Ž. and Vujošević, Z. (1999) The Yugoslav hyper-inflation of 1992-1994: Causes, dynamics, and money supply process. *Journal of Comparative Economics* 27, 335-353

Petrović, P. and Mladenović, Z. (2000) Money demand and exchange rate determination under hyper-inflation: Conceptual issues and evidence from Yugoslavia. *Journal of Money, Credit, and Banking* 32, 785-806.

Taylor, M.P. (1991) The hyper-inflation model of money demand revisited. *Journal of Money, Credit, and Banking* 23, 327-351.