Econometric Theory, **26**, 2010, 882–915.
doi:10.1017/S0266466609990144

ANALYSIS OF COEXPLOSIVE PROCESSES

BENT NIELSEN
University of Oxford

A vector autoregressive model allowing for unit roots as well as an explosive characteristic root is developed. The Granger-Johansen representation shows that this results in processes with two common features: a random walk and an explosively growing process. Cointegrating and coexplosive vectors can be found that eliminate these common factors. The likelihood ratio test for a simple hypothesis on the coexplosive vectors is analyzed. The method is illustrated using data from the extreme Yugoslavian hyperinflation of the 1990s.

1. INTRODUCTION

A vector autoregressive model allowing for unit roots as well as an explosive characteristic root is developed. This is a situation that is of relevance, for instance, in the study of hyperinflations or asset bubbles. A Granger-Johansen representation is derived showing the structure of the common stochastic trends. It is demonstrated that there are cointegrating vectors eliminating random walk common trends and "coexplosive" vectors eliminating an explosive common trend. The statistical analysis of the model is discussed. It is shown that the cointegration aspects of the model can be analyzed using the standard cointegration techniques suggested by Johansen (1988, 1995). A new test for a simple hypothesis on the coexplosive vectors is derived.

The methods are illustrated using data from the extreme Yugoslavian hyperinflation of the 1990s, with a full analysis given in the companion paper, Nielsen (2008a). As pointed out by Juselius and Mladenović (2002), hyperinflationary data can very well be explosive in contrast to the established literature by Sargent and Wallace (1973), Sargent (1977), Evans (1978), Christiano (1987), and Taylor (1991), all of whom assume that logarithms of money and prices are I(2) variables. It is shown that the Yugoslavian hyperinflation data are better described as explosive, where real money is a coexplosive relation and hence I(1), whereas inflation measured as first differences of logarithm of prices is explosive.

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 in Rio de Janeiro in July 2002. I have also benefited from discussions with Martin Wagner and insightful comments from the referees. Computations were done using PcGive (Doornik and Hendry, 2001). Financial support from ESRC grant RES-000-27-0179 is gratefully acknowledged. Address correspondence to Bent Nielsen, Nuffield College, Oxford OX1 1NF, UK; e-mail: bent.nielsen@nuffield.ox.ac.uk.

 © Cambridge University Press 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

The presented methods have also been applied by Engsted (2006) to analyze U.S. stock prices and dividends for the possibility of speculative explosive bubbles.

Related approaches to analyzing coexplosive variables have been established recently by Phillips and Magdalinos (2008) and Magdalinos and Phillips (2009), who look at triangular systems. In the first paper, the regressors are purely explosive as here, and just as here the asymptotic theory and the finite sample distributions are strongly related. In the second paper, the regressors are moderately explosive, which permits mixed normal limit theory.

The outline of this paper is as follows: In Section 2 the model is developed and a Granger-Johansen representation theorem is presented. In Section 3 the proposed statistical analysis is presented. The asymptotic theory follows in Section 4. An empirical illustration is presented in Section 5. Mathematical proofs are given in the Appendix.

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

2. THE MODEL, THE RESTRICTIONS, AND INTERPRETATION

The starting point for the analysis is the vector autoregressive model. Then the restriction of cointegration is presented, followed by the restriction of coexplosiveness. Finally, a Granger-Johansen representation is given.

2.1. The Vector Autoregressive Model

Suppose a p-dimensional time series, $X_{1-k}, \ldots, X_0, \ldots, X_T$, is observed. The vector autoregressive model with a linear trend is given by the equation

$$\mathsf{M}: \quad X_t = \sum_{j=1}^{k} A_j X_{t-j} + \mu + \nu t + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2.1}$$

conditionally on the initial values $X_{1-k}, \ldots, X_0$. The innovations ε_t are assumed independent, identically $\mathsf{N}_p(0, \Omega)$-distributed, and the parameters vary freely so $A_j, \Omega \in \mathbb{R}^{p \times p}$, where Ω is positive definite, while $\mu, \nu \in \mathbb{R}^p$.

2.2. The Cointegrated Vector Autoregressive Model

The cointegrated model arises by restricting the vector autoregressive model as described by Johansen (1995). It is well known how the vector autoregression (2.1) is reparametrized in equilibrium correction form

$$\mathsf{M}: \quad \Delta_1 X_t = \Pi X_{t-1} + \Pi_l t + \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} + \mu + \varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2.2}$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

where Δ_1 is the difference operator, so $\Delta_1 X_t = X_t - X_{t-1}$. The usual cointegration hypothesis arises as the reduced rank restriction

$$\mathsf{H}_1(r): \quad \operatorname{rank}(\Pi, \Pi_l) \leq r.$$

This hypothesis remains of interest, although its interpretation will only become clear in connection with the Granger-Johansen representation in Theorem 1, in Section 2.4.

Imposing the reduced-rank hypothesis gives rise to the usual cointegrated vector autoregressive model

$$\mathsf{M}_1: \quad \Delta_1 X_t = \alpha\left(\beta_1' X_{t-1} + \delta_1' t\right) + \sum_{j=1}^{k-1} \Gamma_j \Delta_1 X_{t-j} + \mu + \varepsilon_t. \tag{2.3}$$

The parameters of this model vary freely, so $\Gamma_j, \Omega \in \mathbb{R}^{p\times p}$ where Ω is positive definite, $\alpha, \beta_1 \in \mathbb{R}^{p\times r}$, while $\mu \in \mathbb{R}^p$ and $\delta_1 \in \mathbb{R}^r$.

2.3. The Coexplosive Model

The situation of main interest here is when there are both unit roots and a single explosive root present. The applications would be to situations that for normal economies would be modeled by I(2) analysis, but where a single I(2) common trend has switched to an explosive mode as in hyperinflations or asset bubbles. The analysis is restricted to the situation with just one positive explosive root, $\rho > 1$ say. This covers a range of situations, where it would seem implausible that the economy is torn apart along several explosive common trends. It also avoids some inconsistency problems that can arise with multiple explosive roots. A further discussion is given in Section 2.5. The subsequent analysis would also apply if $\rho < -1$ were assumed.

For most macroeconomic data, two lags or more are typically needed, $k \geq 2$; see Juselius (2006, p. 72). It is maintained throughout the paper that $k \geq 2$.

The coexplosive model now arises as follows: The standard cointegration model (2.3) is parametrized in terms of first differences $\Delta_1 X_t$ so as to focus the attention on the matrix $\alpha\beta_1'$. This parametrization is, however, not readily informative about the explosive root ρ, so another parametrization is required. Introducing the filter $\Delta_\rho X_t = X_t - \rho X_{t-1}$ and defining $\Delta_\rho X_{t-1}^* = \{\Delta_\rho X_{t-1}', (1-\rho)t\}'$, the model equation (2.3), with $k \geq 2$, is reparametrized as

$$\mathsf{M}_1: \quad \Delta_1\Delta_\rho X_t = \alpha_1\beta_1^{*\prime}\Delta_\rho X_{t-1}^* + \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} + \sum_{j=1}^{k-2} \Phi_j \Delta_1\Delta_\rho X_{t-j} + \mu + \varepsilon_t, \tag{2.4}$$

where $\beta_1^* = (\beta_1', \delta_1')'$. A detailed derivation of the equivalence between (2.3) and (2.4) is given in Appendix A.1. The new parameters $\alpha_1, \alpha_\rho, \beta_\rho$, and Φ_j relate to

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

the original parameters by

$$\alpha_1 = \frac{\alpha}{1-\rho}, \qquad \alpha_\rho \beta_\rho' = -\rho \left(I_p + \frac{\alpha \beta_1'}{1-\rho} - \sum_{j=1}^{k-1} \rho^{-j} \Gamma_j \right), \qquad \Phi_j = \sum_{l=j+1}^{k-1} \rho^{j-l} \Gamma_l. \tag{2.5}$$

The characteristic polynomial, with argument z^{-1}, for a process of type (2.3) or (2.4) is given by the determinant of

$$\left(1-z^{-1}\right) I_p - z^{-1} \alpha \beta_1' - \sum_{j=1}^{k-1} z^{-j} \left(1-z^{-1}\right) \Gamma_j = \frac{z-1}{z} \left(I_p + \frac{\alpha \beta_1'}{1-z} - \sum_{j=1}^{k-1} z^{-j} \Gamma_j \right). \tag{2.6}$$

When $z = \rho$, the latter expression reduces to $(1-\rho^{-1}) \alpha_\rho \beta_\rho'$; see (2.5). This shows that the matrix $\alpha_\rho \beta_\rho'$ has rank $p-1$ precisely when there is one characteristic root at ρ. The new parameter space is therefore given in terms of freely varying parameters satisfying $\alpha_1, \beta_1 \in \mathbb{R}^{p \times r}$, $\alpha_\rho, \beta_\rho \in \mathbb{R}^{p \times (p-1)}$, $\mu \in \mathbb{R}^p$, $\delta_1' \in \mathbb{R}^r$, $\Phi_j, \Omega \in \mathbb{R}^{p \times p}$ so Ω is positive definite, and $\rho > 1$.

2.4. Interpretation and Hypothesis of Interest

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 (1995, Thm. 4.2).

Assumption 1. The parameters satisfy the following conditions:

(i) The matrices $\alpha_1, \beta_1 \in \mathbb{R}^{p \times r}$ and $\alpha_\rho, \beta_\rho \in \mathbb{R}^{p \times (p-1)}$ have full column rank.

(ii) The nonstationary characteristic roots of X_t are at 1 or at ρ, where $\rho > 1$.

(iii) The $\det\left(\alpha_{1\perp}' \Psi_1 \beta_{1\perp}\right) \neq 0$ and $\det\left(\alpha_{\rho\perp}' \Psi_\rho \beta_{\rho\perp}\right) \neq 0$, where

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

The representation theorem can now be formulated. This shows that processes satifying (2.4) have one explosive root and two common features in the form of a random walk component and an explosive component. The parameters β_1 and β_ρ have interpretation as cointegrating and coexplosive 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 any cointegrating-coexplosive linear relations given by $\operatorname{span}(\beta_1) \cap \operatorname{span}(\beta_\rho)$. Due to the dimensions of β_1 and β_ρ, this space is a linear subspace of $\mathbb{R}^p$, which is typically of dimension $r-1$, but it would have dimension r if $\beta_1 \subset \operatorname{span}(\beta_\rho)$. The proof of the representation theorem follows in Appendix A.1.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

THEOREM 1. *Consider a process of the form (2.4) satisfying Assumption 1. Recall the notation* $\Delta_\rho X_t = X_t - \rho X_{t-1}$ *and* $\Delta_\rho X^*_{t-1} = \{\Delta_\rho X'_{t-1}, (1-\rho)t\}'$. *Then*

$$U_t = \{(\Delta_\rho X^*_t)'\beta^*_1, (\Delta_1 X_t)'\beta_\rho, (\Delta_1\Delta_\rho X_t)', \ldots, (\Delta_1\Delta_\rho X_{t-k+3})'\}'$$

can be given a stationary initial distribution ensuring the representation

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

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* $\beta_c \in \operatorname{span}(\beta_1) \cap \operatorname{span}(\beta_\rho)$.

The linear slope coefficient, τ_l, *can be expressed as*

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

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

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

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 that 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 (see Lai and Wei, 1983), noting that the convergence holds even if ε_t is a martingale difference sequence.

Note that it is a consequence of the representation that under Assumption 1, the process has $p-r$ unit roots and one explosive root at ρ. The result could be generalized to a situation with $p-r_\rho$ roots at ρ; see Section 2.5. The case of one explosive root is emphasized here.

It is convenient to use the notations $\mathsf{I}(1,x)$ for variables with both explosive and random walk components and $\mathsf{I}(x)$ for variables with just explosive common trends.

The hypothesis of interest for the coexplosive vectors is that these are known:

$$\mathsf{H}_x: \qquad \beta_\rho = \beta^\circ_\rho. \tag{2.8}$$

In the empirical illustration reported in Section 5, this hypothesis is used to formulate a homogeneity restriction. The model M_1 restricted by H_x is denoted M_{1x}.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

2.5. Multiple Explosive Roots

It would be possible to generalize the Granger-Johansen representation in Theorem 1 to situations with multiple explosive roots. Three types of generalizations could be made: First, the matrix $\alpha_\rho\beta_\rho'$ could have some rank $0 \le r_\rho < p$ and $\det\left(\alpha_{\rho\perp}'\Psi_\rho\beta_{\rho\perp}\right) \neq 0$ so that the $p - r_\rho$ roots are all simple, generalizing the usual I(1) theory. Second, if $\det\left(\alpha_{\rho\perp}'\Psi_\rho\beta_{\rho\perp}\right) = 0$, an I(2)-type representation could be made. Third, if there are explosive roots at different locations, a seasonal integration-type representation arises.

When it comes to inference, such generalizations would be difficult to implement, in that intriguing issues can arise with multiple explosive roots in vector autoregressions, as pointed out by Anderson (1959). Consider an explosive vector autoregression $W_t = \mathbf{W}W_{t-1} + e_{W,t}$ satisfying $|\text{eigen}(\mathbf{W})| > 1$. The Marcinkiewicz-Zygmund result, generalized by Lai and Wei (1985, Lem. 2), shows that $\mathbf{W}^{-t}W_t \xrightarrow{\text{a.s.}} W$ where $W = W_0 + \sum_{t=1}^{\infty}\mathbf{W}^{-t}e_{W,t}$ is a random vector with a continuous distribution if $e_{W,t}$ is a martingale difference satisfying mild conditions. Lai and Wei (1985, Cor. 2) shows that the denominator of the least squares estimator satisfies

$$\mathbf{W}^{-T}\sum_{t=1}^{T}W_{t-1}W_{t-1}'\left(\mathbf{W}^{-T}\right)' \xrightarrow{\text{a.s.}} \mathbf{F}_W = \sum_{t=1}^{\infty}\mathbf{W}^{-t}WW'\left(\mathbf{W}^{-t}\right)'.$$

The limiting matrix $\mathbf{F}_W$ is positive definite if and only if the explosive roots have geometric multiplicity of one; that is, for all $\rho \in \mathbb{C}$ so $|\rho| > 1$, the geometric multiplicity $\dim\mathbf{W} - \text{rank}(\mathbf{W} - \rho I_{\dim\mathbf{W}})$ must be unity. This is referred to as the regular case by Duflo, Senoussi, and Touati (1991) and covers univariate autoregressions. The singular case where the geometric multiplicity of a root is larger than one can only arise for vector autoregressions of higher dimension than one. Examples are

$$\mathbf{W}_1 = \begin{pmatrix}\rho & 1\\ 0 & \rho\end{pmatrix}, \qquad \mathbf{W}_2 = \begin{pmatrix}\rho & 0\\ 0 & \rho\end{pmatrix};$$

for both of these, the algebraic multiplicity of ρ is two, whereas the geometric multiplicity of ρ is one for $\mathbf{W}_1$ and two for $\mathbf{W}_2$. In a case like $\mathbf{W}_2$, the limiting matrix is $\mathbf{F}_W = WW'/(\rho^2 - 1)$, which has rank one. This singularity shows up in the first type of generalized Granger-Johansen representation, discussed above, in which $r_\rho = \text{rank}(\alpha_\rho\beta_\rho') < p$. Then, one explosive component is of order ρ^t and $p - r_\rho - 1$ explosive components obey the law of large numbers, and yet they are uncorrelated with any stationary components in the representation; see Nielsen (2008b) for a discussion.

Phillips and Magdalinos (2008) have recently analyzed a triangular systems framework for variables Y_t, W_t say, in which the equation for the variable W_t has no feedback from Y_t and is purely explosive with a diagonal autoregressive matrix. Even in that framework, correlation between errors of the Y_t and

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

W_t equations yields inconsistent least squares estimators. If the errors are uncorrelated, the triangular setup does, however, produce consistent estimators. Recently, Nielsen (2008b) has generalized the asymptotic analysis of Phillips and Magdalinos (2008) to include martingale difference errors and nondiagonal $\mathbf{W}$-matrices. This is done with a vector autoregressive analysis in mind. It is found that the results for lag-length determination by Nielsen (2006b) also apply in the singular case.[1]

3. STATISTICAL ANALYSIS

In this section, tests for the validity of the unrestricted model M are discussed. Analysis of the hypotheses H_1 and H_x follows. The asymptotic analysis is given in Section 4.

3.1. Misspecification Analysis of Unrestricted Model M

Most diagnostic tests for the vector autoregressive model M as given by (2.1) are derived under an assumption of stationarity, $|z| < 1$, or at most a random walk-like behavior, $|z| \leq 1$. It is worth noting that while such an assumption is mathematically convenient, it may not be necessary. Indeed, tests with asymptotic distributions that are invariant to the unknown parameters would be preferable.

Test for Lag Length. Lag length can be determined consistently using likelihood methods, regardless of the value of the parameters. This applies both for regular and singular processes; see Nielsen (2006b, 2008b) and note 1. Methods that are based on the Yule-Walker equations are, however, not generally applicable across the parameter space. This means that the Box-Pierce test is biased, and correlograms cannot be used as they appear in econometric software at the moment; see Nielsen (2006a).

Test for Normality. Standard tests based on the residuals are applicable across the parameter space, at least for regular processes, which is sufficient for the current analysis. This has been proved for QQ plots when an intercept is included in the model as here; see Engler and Nielsen (2009).

3.2. Analysis of Cointegration Rank Hypothesis H_1

The cointegrating rank is established by writing the autoregressive model M as given by (2.1) in equilibrium correction form (2.2) and testing the reduced rank hypothesis $\mathsf{H}_1(r)$: $\text{rank}(\Pi, \Pi_l) \leq r$. The likelihood ratio test statistic is the usual 'trace-test-statistic' of Johansen (1995, Chap. 6).

The maximum likelihood estimation of the restricted model M_1 is described by Johansen (1995, Chap. 6). The estimation of the parameters, ϑ say, is done in two steps. In a first step, the residuals $R_{0,t}$ and $R_{1,t}$, say, from regressing $\Delta_1 X_t$ and $(X'_{t-1}, t)'$ on $\Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}$ and a constant are found. In a second step,

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

the squared canonical correlations $1 \geq \hat{\lambda}_1^{\mathsf{M}} \geq \cdots \geq \hat{\lambda}_p^{\mathsf{M}} \geq 0$ of $R_{0,t}$ and $R_{1,t}$ are then found. The maximized log likelihood function of the model M_1 is then

$$\hat{\ell}^{\mathsf{M}} = \max_{\vartheta \text{ in } \mathsf{M}_1} \ell(\vartheta) = -\frac{T}{2}\left\{\log\det\left(S_{00}^{\mathsf{M}}\right) + \sum_{j=1}^{r}\log\left(1-\hat{\lambda}_j^{\mathsf{M}}\right)\right\},$$

where $S_{00}^{\mathsf{M}} = T^{-1}\sum_{t=1}^{T} R_{0,t}R_{0,t}'$.

The likelihood ratio test for the hypothesis $\mathsf{H}_1(r)$ is given by

$$LR(\mathsf{M}_1|\mathsf{M}) \stackrel{\text{def}}{=} LR\{\mathsf{H}_1(r)|\mathsf{M}\} = -T\sum_{j=r+1}^{p}\log\left(1-\hat{\lambda}_j^{\mathsf{M}}\right).$$

In Johansen's 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 not necessary. This will be shown in Theorem 2 in Section 4.1.

The coexplosive model given by (2.4) is a reparametrized version of M_1. The maximum likelihood estimators for this version of the model are found by exploiting the reparametrization. In particular, the explosive root ρ is estimated by the explosive root of the sample characteristic polynomial for (2.3). The other parameters of (2.4) are estimated using (2.5). Since $\hat{\alpha}_\rho\hat{\beta}_\rho'$ has reduced rank of $(p-1)$ with probability one, by construction, α_ρ and β_ρ can be estimated from eigenvectors of $\hat{\alpha}_\rho\hat{\beta}_\rho'$ associated with the nonzero eigenvalues.

In applications it will of interest to test whether an explosive root is significantly different from unity. The explosiveness can be investigated by imposing additional unit root restrictions on the model either as an I(1)-rank restriction or as an I(2) restriction. If the explosive root is then eliminated, this suggests that the explosive root is not significantly different from unity. Under the hypothesis of no explosiveness, the tests for these restrictions will have the conventional Dickey-Fuller-type distributions.

3.3. Estimation When the Coexplosive Vector Is Known

Under the hypothesis H_x that the coexplosive vector is known, the likelihood function has to be maximized by numerical methods.

A profile likelihood argument is convenient. For a given value of ρ, the remaining parameters, θ say, can be estimated by a standard reduced-rank regression as outlined in Section 3.2. In a first step, the residuals $R_{0,t}^{\mathsf{H}}(\rho)$ and $R_{1,t}^{\mathsf{H}}(\rho)$ from regressing $\Delta_1\Delta_\rho X_t$ and $\Delta_\rho X_{t-1}^*$ on $\beta_\rho^{\circ\prime}\Delta_1 X_{t-1}$, $\Delta_1\Delta_\rho X_{t-1}, \ldots, \Delta_1\Delta_\rho X_{t-k+2}$ and a constant are found. In a second step, the squared canonical correlations $1 \geq \hat{\lambda}_1^{\mathsf{H}}(\rho) \geq \cdots \geq \hat{\lambda}_p^{\mathsf{H}}(\rho) \geq 0$ of $R_{0,t}^{\mathsf{H}}(\rho)$ and $R_{1,t}^{\mathsf{H}}(\rho)$ are found. The profile log likelihood function for ρ is then

$$\ell^{\mathsf{H}}(\rho) = \max_{\theta \text{ in } \mathsf{H}_x} \ell^{\mathsf{H}}(\rho,\theta) = -\frac{T}{2}\left[\log\det\left\{S_{00}^{\mathsf{H}}(\rho)\right\} + \sum_{j=1}^{r}\log\left\{1-\hat{\lambda}_j^{\mathsf{H}}(\rho)\right\}\right],$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

where $S_{00}^{\mathsf{H}}(\rho) = T^{-1}\sum_{t=1}^{T}\{R_{0,t}^{\mathsf{H}}(\rho)\}\{R_{0,t}^{\mathsf{H}}(\rho)\}'$. This profile log likelihood is maximized by search over ρ for $\rho > 1$, giving the maximum

$$\hat{\ell}^{\mathsf{H}} = \max_{\rho>1}\ell^{\mathsf{H}}(\rho) = -\frac{T}{2}\left[\log\det\{S_{00}^{\mathsf{H}}(\hat{\rho}^{\mathsf{H}})\} + \sum_{j=1}^{r}\log\{1-\hat{\lambda}_j^{\mathsf{H}}(\hat{\rho}^{\mathsf{H}})\}\right].$$

The maximum likelihood estimators are shown to be consistent in Section 4.2. In finite samples, it is difficult to establish theoretically that the likelihood has a unique maximum. In practice, one will only embark on this analysis if there is a single explosive root in the models M and M_1. If no maximum or several maximums are found over the region $\rho > 1$ under the hypothesis H_x, the specification of the model and the hypothesis should be investigated. A local maximum at $\rho = 1$ is a possibility, but again, it happens with vanishing probability in increasing samples if the model is correct.

The likelihood ratio test statistic for the hypothesis H_x within the model M_1 is

$$LR(\mathsf{M}_{1x}|\mathsf{M}_1) \stackrel{\text{def}}{=} LR(\mathsf{H}_x|\mathsf{M}_1) = -2(\hat{\ell}^{\mathsf{H}} - \hat{\ell}^{\mathsf{M}}).$$

In Section 4.3, it is shown that the test statistic $LR(\mathsf{H}_x|\mathsf{M}_1)$ is χ^2.

4. ASYMPTOTIC RESULTS

In the following, asymptotic results are derived under the restrictions H_1 and H_x, so the coexplosive vector is known. First, the cointegration rank test statistic $LR(\mathsf{M}_1|\mathsf{M})$ is considered, then consistency is shown for the estimators in the model M_{1x}, and finally the test statistic $LR(\mathsf{H}_x|\mathsf{M}_1)$ is analyzed.

The asymptotic theory is derived using results of Lai and Wei (1985) and later adaptations by Nielsen (2005). While the analysis is based on the Gaussian likelihood derived from the model in Section 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, the main component of the explosive common trend, $\sum_{s=1}^{t}\rho^{-s}\varepsilon_s$, must converge. The Marcinkiewicz-Zygmund theorem (see Lai and Wei, 1983) shows this convergence almost surely. It suffices to assume that $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence for some filtration $\mathcal{F}_t$ satisfying the following condition:

Assumption 2. For some $\gamma > 0$, it holds $\sup_t \mathsf{E}\{(\varepsilon_t'\varepsilon_t)^{(2+\gamma)/2}|\mathcal{F}_{t-1}\} < \infty$ a.s.

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

Assumption 3. Suppose $\mathsf{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Omega$ a.s., where Ω is positive definite.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

4.1. Cointegration Rank Test

Johansen (1995, Chap. 6) analyzed the cointegration rank test in the situation where the process is I(1) so that it satisfies the following assumptions: (i) the number of unit roots is $p-r$, and (ii) the remaining characteristic roots are stationary. The latter assumption is not necessary. This was shown for the univariate case by Nielsen (2001). Here it is shown for the model M when one of the remaining characteristic roots is explosive, as stipulated in Assumption 1.

THEOREM 2. *Suppose model* M_1 *and Assumptions 1–3 hold. Then LR* $(\mathsf{M}_1|\mathsf{M})$ *has the same limiting distribution as in Johansen (1995, Chap. 6),*

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

where $F_u = (B_u' - \int_0^1 B_s' ds, u - 1/2)'$ *with* B_u *being a standard Brownian motion of dimension* $p-r$.

The proof follows in Appendix A.3 and resembles that of Johansen (1995, Chap. 11).

4.2. Consistency in the Coexplosive Model

The consistency result for the maximum likelihood estimators under the hypothesis H_x can now be formulated. The consistency is established when ρ varies in a closed set so $\rho > 1$. It is likely that consistency actually holds for $\rho \geq 1$, as it can also be shown that the likelihood function does not achieve its maximum at $\rho = 1$.

THEOREM 3. *Consider the maximum likelihood estimators in model* M_{1x}. *Suppose* $\rho \geq \varrho$ *for some* $\varrho > 1$ *and Assumptions 1–3 hold. Choose* $\beta_{1\perp}^*$ *as*

$$\beta_{1\perp}^* = \begin{pmatrix} \beta_{1\perp} & -\overline{\beta}_1 \delta' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_{p-r} & 0 \\ \varphi & 1 \end{pmatrix}, \tag{4.1}$$

where $\varphi = \{1 + \delta(\beta_1'\beta_1)^{-1}\delta'\}^{-1}\tau_l'\beta_{1\perp}$; *identify* $\hat{\beta}_1^*$ *and* $\hat{\beta}_{1\perp}^*$ *by*

$$\hat{\beta}_1^{*\prime}\overline{\beta}_1^* = I_r, \qquad \hat{\beta}_{1\perp}^* = (I_{p+1} - \overline{\beta}_1^*\hat{\beta}_1^{*\prime})\beta_{1\perp}^*, \tag{4.2}$$

so $\overline{\beta}_{1\perp}^{*\prime}\hat{\beta}_{1\perp}^* = I_{p+1-r}$; *and define the block-diagonal normalization factors*

$$N_V = \operatorname{diag}(T^{-1/2}I_{p-r}, T^{-1}) \quad \text{and} \quad N_W = T^{1/2}\rho^{-T}. \tag{4.3}$$

Then it holds that

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

(i) $\hat{\mu} = \mu + o_P(1)$,
(ii) $\hat{\rho} = \rho + o_P(N_W)$,
(iii) $(\hat{\alpha}_\rho, \hat{\alpha}_1, \hat{\Phi}_1, \ldots, \hat{\Phi}_{k-2}) = (\alpha_\rho, \alpha_1, \Phi_1, \ldots, \Phi_{k-2}) + o_P(1)$,
(iv) $(\hat{\beta}_1^* - \beta_1^*)'\overline{\beta}_{1\perp}^* N_V^{-1} = o_P(1)$,
(v) $\hat{\Omega} = \Omega + o_P(1)$.

The proof follows in Appendix A.4. For fixed ρ, the proof would resemble the I(2) analysis of Johansen (1997), but the estimation of ρ adds an extra layer of complexity.

The choice of identification of $\hat{\beta}_1^*$ in (4.2) follows the approach of Johansen (1995, Sec. 13.2). Identification could be obtained alternatively 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 by an argument as that of Johansen (1995, Sec. 13.2).

4.3. Asymptotic Distribution of the Test Statistic for H_x

In the following, the asymptotic distribution of the likelihood ratio test statistic for H_x is described. At first, an asymptotic expansion is given that only exploits the martingale property of the residuals. To formulate the asymptotic result, introduce a parameter $\tau_\perp = \Psi_\rho \beta_{\rho\perp}$. This is a p-vector that is nonzero due to Assumption 1(iii), and its orthogonal complement can be chosen as $\tau = (I_p - \tau_\perp \overline{\tau}'_\perp)\alpha_\rho$; see Johansen (1995, Ex. 3.7).

THEOREM 4. *Suppose model* M_{1x} *with* $\rho \geq \varrho$ *for some* $\varrho > 1$ *and Assumptions 1–3 hold with* $\gamma > 2$. *Then the log likelihood ratio test statistic for* H_x *in* M_1 *satisfies*

$$LR(H_x|M_1) = -2(\hat{\ell}^H - \hat{\ell}^M) = H'\tau(\tau'\Omega\tau)^{-1}\tau' H\{1 + o_P(1)\} + o_P(1),$$

where

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

The proof of this result is unusual in that it requires an improvement of the consistency rate established in Theorem 3. This is done in Appendix A.5. The proof of the theorem then follows in Appendix A.6.

Since the statistic H is a moving average of $\varepsilon_1, \ldots, \varepsilon_T$, it does not obey the central limit theorem. If, however, the innovations are independent and identically distributed, it holds, as pointed out by Anderson (1959), that

$$\frac{\sum_{t=1}^T \rho^{t-T}\varepsilon_t}{\{\sum_{t=1}^T \rho^{2(t-T)}\}^{1/2}} \overset{D}{=} \frac{\sum_{t=1}^T \rho^{1-t}\varepsilon_t}{\{\sum_{t=1}^T \rho^{2(1-t)}\}^{1/2}}.$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

By exploiting the distribution of the innovations, the limiting distribution of H can then be found. The following result covers the normal case:

COROLLARY 1. *Suppose model* M_{1x} *with* $\rho \geq \varrho$ *for some* $\varrho > 1$ *and Assumptions 1–3 hold with* $\gamma > 2$, *and that* $\tau'\varepsilon_t$ *is independent* $\mathsf{N}\left(0, \tau'\Omega\tau\right)$*-distributed. Then* $LR(\mathsf{H}_x|\mathsf{M}_1) \xrightarrow{\mathsf{D}} \chi^2(p-1)$.

The asymptotic normality in Theorem 4 relies directly on the normality assumption. In practice, the best we can do is to check if the normality assumption looks reasonable and then use the above test. The normality may not hold exactly, so infences should only be made cautiously when test statistics are close to the chosen critical value. In this respect it is actually not that different from the inference in standard models, where normality arising from the central limit theorem should also be used with caution in applications.

5. EMPIRICAL ILLUSTRATION: THE EXTREME YUGOSLAVIAN HYPERINFLATION

The money demand in hyperinflations is studied. It is usually assumed that log prices, p_t, and log money, m_t, measured at a monthly frequency, are $\mathsf{I}(2)$ and $m_t - p_t$ is $\mathsf{I}(1)$ so that money demand is a cointegrating relation between $m_t - p_t$ and $\Delta_1 p_t$. Juselius and Mladenović (2002) have pointed out that data from the first Yugoslavian hyperinflation show explosive behavior, contrary to these assumptions. With the new econometric techniques presented here, this issue can be studied formally. Data from the second Yugoslavian hyperinflation are studied. It is found that m_t and p_t are $\mathsf{I}(1, x)$ with $m_t - p_t$ as a coexplosive relation, hence $m_t - p_t$ is $\mathsf{I}(1)$, in line with the literature. However, $\Delta_1 p_t$ is found to $\mathsf{I}(x)$, in contrast to the assumptions of the literature, so that the frequently made relation between $m_t - p_t$ and $\Delta_1 p_t$ is unbalanced. A different empirical analysis is then called for; see Nielsen (2008a).

The seminal paper on money demand in hyperinflations is by Cagan (1956). The assumptions that $p_t, m_t \sim \mathsf{I}(2)$ and $m_t - p_t \sim \mathsf{I}(1)$ so $\Delta_1 p_t \sim \mathsf{I}(1)$ were later introduced and used by Sargent and Wallace (1973), Sargent (1977), Evans (1978), and Christiano (1987). The idea that $m_t - p_t$ and $\Delta_1 p_t$ cointegrate was introduced by Taylor (1991). Taylor tested the assumptions for a series of hyperinflationary episodes using one-sided unit root tests against the stationary alternative, ignoring the possibility of explosiveness. Indeed, a number of his tests indicate explosiveness. The advantage of looking at the second Yugoslavian episode is that it is very long, very extreme, and increasing very smoothly.

The institutional background for the extreme Yugoslavian hyperinflation is described in Petrović and Vujošević (1996) and Petrović, Bogetić, and Vujošević (1999). In short, the Federal Republic of Yugoslavia was falling apart in 1991, the war then started, and the United Nations embargo was introduced in May 1992. The inflation started to increase, and monthly inflation reached the 50% level in

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

February 1992 and only ended on 24 January 1994 with currency reform after prices had risen by a factor of 1.6×10^{21} over 24 months.

The data are taken from Petrović and Mladenović (2000) and are available from the *Journal of Money, Credit and Banking* online data archive. Plots (a) and (d) of Figure 1 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. Following precedence in the empirical hyperinflation literature, the end of the episode is not analyzed; only the data until 1993:10 are considered, due to measurement problems for the price series. Plots (b) and (e) of Figure 1 show first differences of the series. Both in levels and in differences, the series shows exponential growth over time and hence an increasing growth in prices. This exponential growth can also be discerned from the plot of the second differences in Figure 1(c,f), although these plots are somewhat more volatile. Real money, measured as $m_t - p_t$, is shown in levels, in differences, and in second differences in Figure 1(g,h,i). It appears that real money, $m_t - p_t$, has a random walk behavior while the depreciation rate, $\Delta_1 p_t$, is explosive. Using the developed statistical model, this can now be analyzed formally.

A vector autoregression as (2.1) with a constant, trend, and three lags is fitted. Misspecification tests for lag length and QQ plots for testing normality are applicable, as discussed in Section 3.1; see Nielsen (2006b, 2008b) and Engler and

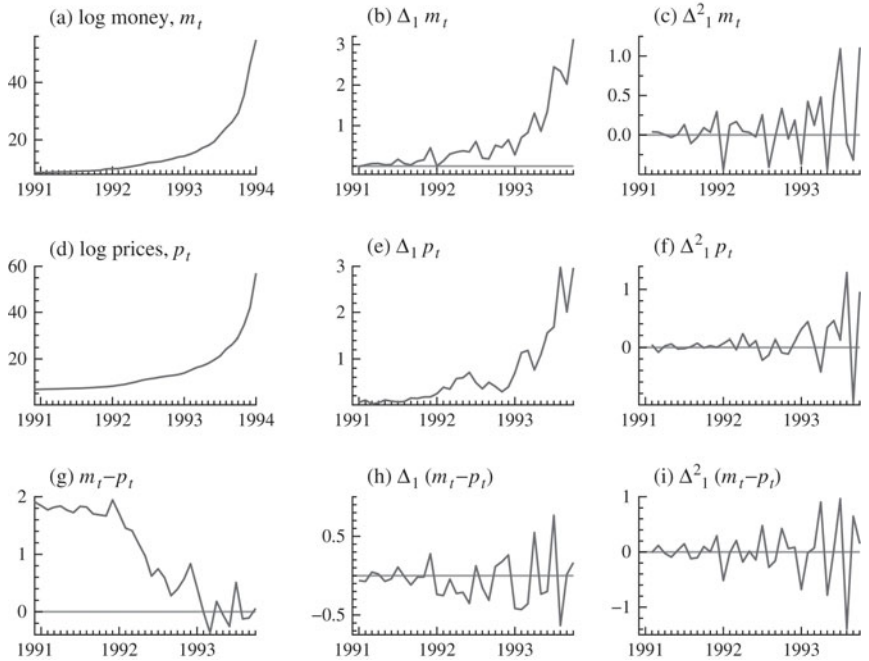


FIGURE 1. Data in levels, differences and second differences.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

TABLE 1. Misspecification tests for the vector autoregressive model for m, p

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]

Note: Single-equation as well as system tests are reported, and p-values are given in brackets.

Nielsen (2009). The cumulant-based tests for normality rather than QQ plots are reported, although their validity has not been established yet. The tests reported in Table 1 indicate that the model is well-specified. The lag-length test statistics are reported in an F-form as advocated by Doornik and Hendry (2001) in an attempt to deal with finite sample issues, although this aspect has not been explored formally. This model has a single explosive characteristic root of $\hat{\rho} = 1.216$. A nonexplosive I(2) description of the data was firmly excluded using the system analysis proposed by Rahbek, Kongsted, and Jørgensen (1999). This suggests that $p_t, m_t \sim \mathsf{I}(1, x)$ and $\Delta_1 p_t, \Delta_1 m_t \sim \mathsf{I}(x)$.

A cointegration analysis can be carried out in the usual way, as described in Section 3.2. The Bartlett-type correction of Johansen (2002) is not applicable, due to the explosive root. The results are given in Table 2, pointing toward a rank of one, $\widehat{r} = 1$. The model M given in (2.4) with $r = 1$, therefore appears to give a reasonable description of the data. When it is imposed that $r = 1$, the estimate of the explosive characteristic root changes slightly to $\hat{\rho} = 1.205$.

The hypothesis that β_ρ is known and given by the homogeneity condition H_x : $\beta'_\rho = (1, -1)$ can now be tested. This would imply that real money, $m_t - p_t$, is a coexplosive relation with random walk behavior but no explosive behavior, so $m_t - p_t \sim \mathsf{I}(1)$. Since β_ρ is completely specified, the model is estimated by reduced-rank regression for each value of ρ. This in turn results in a profile likelihood in ρ that is maximized by grid search. Searching in the region $\rho > 1$, the maximum of the likelihood function of 10.74 is found at $\hat{\rho} = 1.200$. The test statistic for H_x against M_1 is 0.06, which is small compared to the $\chi^2(1)$ distribution arising from Theorem 4.

The sample analyzed here is short compared to standard macroeconomic cointegration analyses. It is, however, one of the longest hyperinflationary episodes

TABLE 2. Test for cointegration rank; 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	

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

available, and its information content is rather large, with enormous changes in the variables. It does seem as if the $I(1,x)$ description of p_t, m_t gives a more reasonable description of the dependence structure than the usual $I(2)$ description. Asymptotic inferences drawn from an $I(1,x)$ description must therefore be more reliable than the usual analysis. The explosive behavior found here cannot last forever, and, indeed, the hyperinflation stops in January 1994. While this out-of-sample structural break is important for forecasts, it does not have much bearing on the in-sample inferences made here.

From an economic viewpoint, the most interesting conclusion is that $m_t - p_t$ is more stable than p_t, m_t. Extrapolating from this study and studies of standard economies, it seems as if m_t and p_t tend to have common stochastic features, either $I(2)$ or $I(1,x)$ or perhaps structural breaks, which disappear by taking the linear combination $m_t - p_t$.

In summary, the above analysis shows that the two variables m_t, p_t are $I(1,x)$. The series coexplodes, so $m_t - p_t$ is an $I(1)$ process. Extrapolating from this study and studies of standard economies, it seems as if m_t and p_t tend to have common stochastic features, either $I(2)$ or $I(1,x)$ or perhaps structural breaks, which disappear by taking the linear combination $m_t - p_t$. In addition, the differenced series $\Delta_1 m_t, \Delta_1 p_t$ are $I(x)$. This indicates that linking, for instance, $m_t - p_t$ with Δp_t, as in the model of Taylor (1991), will not give a balanced regression in this situation. A suggestion for getting around the issue of the unbalanced regression is given in Nielsen (2008a). The problem lies in the discretization of the continuous time money demand schedule considered by Cagan (1956). By measuring inflation as the cost of holding money, $c_t = \Delta_1 P_t / P_t = 1 - \exp(-\Delta_1 p_t)$ instead of $\Delta_1 p_t$, a standard $I(1)$ cointegration analysis can be carried out for $m_t - p_t$ and c_t. Many of the ideas in the hyperinflation literature seem to carry over to this new choice of variables.

NOTE

1. The possibility that $\mathbf{F}_W$ can be singular was overlooked by Lai and Wei (1985). The generalization of their results to models with deterministic terms by Nielsen (2005) therefore only applies in the regular case. Likewise, the proofs for the results for lag-length determination by Nielsen (2006b) only apply in the regular case, although the results hold more generally as shown in Nielsen (2008b).

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 M. Friedman (ed.), *Studies in the Quantity Theory of Money*, pp. 25–117. University of Chicago Press.

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

Christiano, L.J. (1987) Cagan's model of hyperinflation under rational expectations. *International Economic Review* 28, 33–49.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

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

Duflo, M., R. Senoussi, & R. Touati (1991) Propriétés asymptotiques presque sûre de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 27, 1–25.

Engler, E. & B. Nielsen (2009) The empirical process of autoregressive residuals. *Econometrics Journal* 12, 367–381.

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

Evans, P. (1978) Time series analysis of the German hyperinflation. *International Economic Review* 19, 195–209.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

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

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

Johansen, S. (2002) A small sample correction for the test of cointegration rank in the vector autoregressive model. *Econometrica* 70, 1929–1962.

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

Juselius, K. (2006) *The Cointegrated VAR Model*. Oxford University Press.

Juselius, K. & Z. Mladenović (2002) High inflation, hyper inflation and explosive roots. The case of Jugoslavia. Mimeo, University of Copenhagen. http://www.econ.ku.dk/okokj/papers/hyper.pdf.

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

Magdalinos, T. & P.C.B. Phillips (2009) Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* 25, 482–526.

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

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

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

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

Nielsen, B. (2008a) On the explosive nature of hyper-inflation data. *Economics: The Open-Access, Open-Assessment E-Journal* 2, 2008-21.

Nielsen, B. (2008b) Singular Vector Autoregressions with Deterministic Terms: Strong Consistency and Lag Order Determination. Discussion paper, Nuffield College.

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

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

Petrović, P. & Z. Vujošević (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.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Phillips, P.C.B. & T. Magdalinos (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865–887.
Rahbek, A., Kongsted, H.C. & C. Jørgensen (1999) Trend stationarity in the I(2) cointegration model. *Journal of Econometrics* 90, 265–289.
Sargent, T. (1977) The demand for money during hyperinflations under rational expectations: I. *International Economic Review* 18, 59–82.
Sargent, T. & N. Wallace (1973) Rational expectations and the dynamics of hyperinflation. *International Economic Review* 14, 328–350.
Taylor, M.P. (1991) The hyper-inflation model of money demand revisited. *Journal of Money, Credit, and Banking* 23, 327–351.

APPENDIX: Proofs of Results

A.1. Proof of Granger-Johansen Representation.

Proof of Equation (2.4). Note first the identities

$$(1-\rho)X_{t-1} = \Delta_\rho X_{t-1} - \rho\Delta_1 X_{t-1}, \tag{A.1}$$

$$\rho^j \Delta_1 X_{t-j-1} = \Delta_1 X_{t-1} - \sum_{l=1}^{j} \rho^{l-1}\Delta_1\Delta_\rho X_{t-l}. \tag{A.2}$$

Now, turn to the model equation (2.3). Subtract $\rho\Delta_1 X_{t-1}$ on both sides and apply (A.1) to the regressor X_{t-1} to get

$$\begin{aligned}\Delta_1 X_t - \rho\Delta_1 X_{t-1} = {} & \frac{\alpha}{1-\rho}\left\{\beta_1'\left(\Delta_\rho X_{t-1} - \rho\Delta_1 X_{t-1}\right) + \delta_1' t(1-\rho)\right\} \\ & + \sum_{j=1}^{k-1}\Gamma_j\Delta_1 X_{t-j} + \mu + \varepsilon_t - \rho\Delta_1 X_{t-1}.\end{aligned} \tag{A.3}$$

From relation (A.2), it follows that

$$\begin{aligned}\sum_{j=1}^{k-1}\Gamma_j\Delta_1 X_{t-j} &= \Gamma_1\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Gamma_{j+1}\rho^{-j}\rho^j\Delta_1 X_{t-j-1} \\ &= \Gamma_1\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Gamma_{j+1}\rho^{-j}\Delta_1 X_{t-1} - \sum_{j=1}^{k-2}\Gamma_{j+1}\rho^{-j} \\ &\quad \times \sum_{l=1}^{j}\rho^{l-1}\Delta_1\Delta_\rho X_{t-l}.\end{aligned}$$

Collecting the first two terms and changing the summation index in the third term gives

$$\sum_{j=1}^{k-1}\Gamma_j\Delta_1 X_{t-j} = \sum_{j=1}^{k-1}\Gamma_j\rho^{1-j}\Delta_1 X_{t-1} - \sum_{l=1}^{k-2}\left(\sum_{j=l+1}^{k-1}\Gamma_j\rho^{l-j}\right)\Delta_1\Delta_\rho X_{t-l}.$$

Substitute this expression into (A.3) and gather the $\Delta_1 X_{t-1}$ terms to get (2.4). ■

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Proof of Theorem 2.1. When μ, δ_1 are both zero, the result follows as in Johansen and Schaumburg (1998, Thm. 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. This gives the solution $\tilde{X}_t$ to the homogenous equation (2.7).

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 (2.4), giving

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

Collecting the coefficients 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, premultiply the third equation with $\alpha_{1\perp}'$ and postmultiply Ψ_1 with $I_p = \overline{\beta}_1 \beta_1' + \beta_{1\perp} \overline{\beta}_{1\perp}'$. Finally, to get $\beta_1' \tau_c$, premultiply the third equation with $\overline{\alpha}_1'$. ■

A.2. The Data Generating Process and Preliminary Asymptotic Results. The data generating process is described, then some asymptotic results from the literature are listed.

It is convenient to start by introducing some additional notation: 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 for $t = 1, \ldots, T$, and the symbol $S_{XY.Z}$ is used for the partial sample covariance of X_t and Y_t given Z_t, that is $T^{-1} \sum_{t=1}^{T} X_t (Y_t | Z_t)'$.

In the probabilistic analysis, the properties of the likelihood function will be analyzed for each parameter $(\vartheta^\circ, \Omega_\circ)$ satisfying the model M_1. This defines the data generating process. Since β_ρ is known for this process, it will also satisfy H_x. From the model equation (2.4), define

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

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

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

$$U_{t-1}^\circ = \{(\beta_1^{*\circ\prime} \Delta_{\rho_\circ} X_{t-1}^*)', U_{\rho,t-1}^{\circ\prime}\}', \quad \textbf{(A.5)}$$

$$V_t^\circ = \beta_{1\perp}^{*\circ\prime} \Delta_{\rho_\circ} X_{t-1}^*, \qquad W_{t-1}^\circ = \beta_{\rho\perp}^{\circ\prime} \Delta_1 X_{t-1}. \quad \textbf{(A.6)}$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Note that $U^{\circ}_{\rho,t-1}$ is a subset of U°_{t-1}. According to the Granger-Johansen representation in Theorem 1 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 behavior, respectively. The companion vectors for the process X_t can then be written as

$$S^{\circ}_t = (U^{\circ\prime}_t, V^{\circ\prime}_t, W^{\circ\prime}_t, 1)', \tag{A.7}$$

which satisfies a first-order autoregressive equation.

Finally, the analysis of the model M_1 is based on the residuals $R_{0,t}$ and $R_{1,t}$ introduced in Section 3.2. With the above definitions, they satisfy

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

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

$$R_{\varepsilon,t} = \left(\varepsilon^{\circ}_t \,\middle|\, U^{\circ}_{\rho,t-1}, W^{\circ}_{t-1}, 1\right) = R_{0,t} - \alpha^{\circ}_1 \beta^{*\circ\prime}_1 R_{1,t}. \tag{A.8}$$

Asymptotic results from Lai and Wei (1985) and Nielsen (2005) for the product moments of the processes $\varepsilon^{\circ}_t, S^{\circ}_{t-1}$ are summarized. These results apply for regular explosive processes. The process of interest is indeed regular, since it has only one explosive root. Results for singular processes can be found in Nielsen (2008b). Recall the normalizations N_V, N_W given in (4.3) and define $N_S = \operatorname{diag}(I_{\dim U}, N_V, N_W, 1)$.

LEMMA A.1. *Consider the process X_t satisfying M_1 as given by (A.4) and suppose Assumptions 1–3 are satisfied. Let ξ be a constant satisfying $\xi < \gamma/(2+\gamma)$, recalling the definition of γ in Assumption 2.*

Define sample variances $\widehat{\mathsf{var}}(x_t) = T^{-1}\sum_{t=1}^{T} x_t x_t'$. Then stochastic matrices Σ_W, Σ_V, Σ_S, and a deterministic matrix Σ_U exist so

(i) $\widehat{\mathsf{var}}(\varepsilon^{\circ}_t) \overset{\text{a.s.}}{=} \Omega_\circ + \mathrm{o}(T^{-\xi})$ *and* $\widehat{\mathsf{var}}(\varepsilon^{\circ}_t|1) \overset{\text{a.s.}}{=} \Omega_\circ + \mathrm{o}(T^{-\xi})$,

(ii) $\widehat{\mathsf{var}}\left(U^{\circ}_{t-1}|1\right) \overset{\text{a.s.}}{\to} \Sigma_U > 0$,

(iii) $\rho_\circ^{-2T}\, T\, \widehat{\mathsf{var}}\left(W^{\circ}_{t-1}|1\right) \overset{\text{a.s.}}{\to} \Sigma_W \overset{\text{a.s.}}{>} 0$,

(iv) $\widehat{\mathsf{var}}\left(N_V V^{\circ}_{t-1}|1\right) \overset{\mathsf{D}}{\to} \Sigma_V \overset{\text{a.s.}}{>} 0$,

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

The sample variance of the joint process S also satisfies

(vi) $\max \operatorname{eigen}\left(\sum_{t=1}^{T} S_{t-1} S'_{t-1}\right) \overset{\text{a.s.}}{=} \mathrm{O}\left(\rho_\circ^{2T}\right)$.

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

(vii) $\widehat{\mathsf{corr}}(S^{\circ}_{t-1}, \varepsilon^{\circ}_t) \overset{\text{a.s.}}{=} \mathrm{o}(T^{-\xi/2})$,

(viii) $\widehat{\mathsf{corr}}\{(U^{\circ\prime}_{t-1}, V^{\circ\prime}_{t-1}, 1)', \varepsilon^{\circ}_t\} = \mathrm{O}_{\mathsf{P}}(T^{-1/2})$,

(ix) $\widehat{\mathsf{corr}}(W^{\circ}_{t-1}, 1) \overset{\text{a.s.}}{=} \mathrm{O}(T^{-1/2})$,

(x) $\widehat{\mathsf{corr}}(U^{\circ}_{t-1}, W^{\circ}_{t-1}|1) \overset{\text{a.s.}}{=} \mathrm{o}(T^{-\xi/2})$,

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

(xi) $\widehat{\mathrm{corr}}(V_{t-1}^{\circ}, W_{t-1}^{\circ}|1) = \mathrm{O_P}(T^{-1/2})$,

(xii) $\widehat{\mathrm{corr}}(U_{t-1}^{\circ}, V_{t-1}^{\circ}|1) \overset{\text{a.s.}}{=} \mathrm{o}(T^{-\xi/2})$.

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

(xiii) $T^{-1/2}\sum_{t=1}^{T}(U_{t-1}^{\circ}|1)\varepsilon_t^{\circ\prime} \xrightarrow{\mathsf{D}} \Sigma_{U\varepsilon}$,

(xiv) $T^{-1/2}\sum_{t=1}^{T} N_V(V_{t-1}^{\circ}|1)\varepsilon_t^{\circ\prime} \xrightarrow{\mathsf{D}} \Sigma_{V\varepsilon}$,

(xv) $T^{-1/2}\sum_{t=1}^{T} N_V(V_{t-1}^{\circ}|1)U_{t-1}^{\circ\prime} \xrightarrow{\mathsf{D}} \Sigma_{VU}$.

Proof of Lemma A.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): Theorem 2.4. (ix): Theorem 9.1. (x): Table 2. (xi): Remark 9.3 and (ix). (xii): Example 6.6 and Theorem 9.4. The results (iv), (viii), and (xiv) can be proved using the techniques of Chan and Wei (1988). For (xiii), and (xv), see also Johansen (1995, App. B). (v): Combine (ii)–(iv) and (x)–(xii). ■

A.3. Asymptotic Theory for Cointegration Analysis under $\mathsf{M_1}$**.** Theorem 2, concerning the cointegration rank test statistic $LR(\mathsf{M_1}|\mathsf{M})$, is proved.

At first, a profile likelihood function for the model $\mathsf{M_1}$ is derived. Two equivalent expressions for the model equation were given in (2.3) and (2.4):

$$\varepsilon_t = \Delta_1 X_t - \alpha_1\beta_1^{*\prime} X_{t-1}^{*} - \sum_{j=1}^{k-1}\Gamma_j \Delta_1 X_{t-j} - \mu, \tag{A.9}$$

$$= \Delta_1\Delta_\rho X_t - \alpha_1\beta_1^{*\prime}\Delta_\rho X_{t-1}^{*} - \alpha_\rho\beta_\rho'\Delta_1 X_{t-1} - \sum_{j=1}^{k-2}\Phi_j \Delta_1\Delta_\rho X_{t-j} - \mu. \tag{A.10}$$

The Gaussian log likelihood function is

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

where ϑ represents the parameters $\Pi, \Pi_l, \Gamma_1, \ldots, \Gamma_{k-1}, \mu$. 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

$$-2\log L(\vartheta) = -2\max_{\Omega}\log L(\vartheta, \Omega) = T\log\det\{\hat{\Omega}(\vartheta)\}. \tag{A.11}$$

The asymptotic theory for the rank test statistic is derived along the lines of Johansen (1995, Chaps. 10, 11). In addition, the asymptotic theory for the estimators in $\mathsf{M_1}$ is established following Johansen (1995, Chap. 13).

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

LEMMA A.2. *Consider the process* X_t *satisfying* $\mathsf{M_1}$ *as given by* (*A*.4), *and suppose Assumptions 1–3 are satisfied. Then, for some positive definite matrix* $\Sigma_{\beta\beta}$ *and some*

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

random matrices $\Sigma_{\beta\varepsilon}, \Sigma_{U\varepsilon}, \Sigma_{V\varepsilon}, \Sigma_V$, *it holds that*

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\beta_1^{*\circ\prime}R_{1,t}\\ N_V\beta_{1\perp}^{*\circ\prime}R_{1,t}\\ R_{\varepsilon,t}\end{pmatrix}^{\otimes 2}=\begin{bmatrix}\Sigma_{\beta\beta}+\mathrm{o_P}(1) & \mathrm{O_P}\left(T^{-1/2}\right) & T^{-1/2}\left\{\Sigma_{\beta\varepsilon}+\mathrm{o_P}(1)\right\}\\ * & \Sigma_V+\mathrm{o_P}(1) & T^{-1/2}\left\{\Sigma_{V\varepsilon}+\mathrm{o_P}(1)\right\}\\ * & * & \Omega_\circ+\mathrm{o_P}(1)\end{bmatrix},$$

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\beta_1^{*\circ\prime}R_{1,t}\\ R_{0,t}\end{pmatrix}^{\otimes 2}\xrightarrow{\mathsf{P}}\begin{pmatrix}\Sigma_{\beta\beta} & \Sigma_{\beta\beta}\alpha_1^{\circ\prime}\\ \alpha_1^{\circ}\Sigma_{\beta\beta} & \Omega_\circ+\alpha_1^{\circ}\Sigma_{\beta\beta}\alpha_1^{\circ\prime}\end{pmatrix}.$$

Proof of Lemma A.2. It is first argued that

$$M^\circ=\frac{1}{T}\sum_{t=1}^{T}\left\{\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^*\\ U_{\rho,t-1}^\circ\\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^*\\ \varepsilon_t^\circ\\ N_WW_{t-1}^\circ\end{pmatrix}\middle|\,1\right\}^{\otimes 2} \tag{A.12}$$

$$=\begin{bmatrix}\Sigma_U+\mathrm{o_P}(1) & \mathrm{O_P}\left(T^{-1/2}\right) & T^{-1/2}\left\{\Sigma_{U\varepsilon}+\mathrm{o_P}(1)\right\} & \mathrm{o_P}\left(T^{-\xi/2}\right)\\ * & \Sigma_V+\mathrm{o_P}(1) & T^{-1/2}\left\{\Sigma_{V\varepsilon}+\mathrm{o_P}(1)\right\} & \mathrm{O_P}\left(T^{-1/2}\right)\\ * & * & \Omega_\circ+\mathrm{o_P}(1) & T^{-1/2}S_{\varepsilon W}\\ * & * & * & \Sigma_W+\mathrm{o_P}(1)\end{bmatrix},$$

where $S_{W\varepsilon}=\rho_\circ^{-T}\sum_{t=1}^{T}\varepsilon_t^\circ W_{t-1}^\circ=\mathrm{o_P}(T^{(1-\xi)/2})$. This result arises using the following matrix of items from Lemma A.1:

$$\begin{Bmatrix}\text{(ii)} & \text{(xv)} & \text{(viii)} & \text{(x)}\\ * & \text{(iv)} & \text{(xiv)} & \text{(xi)}\\ * & * & \text{(i)} & \text{(iii), (vii), (ix)}\\ * & * & * & \text{(iii,ix)}\end{Bmatrix}.$$

Partialing out $U_{\rho,t-1}^\circ, W_{t-1}^\circ$ shows the first result. The second result then follows by noting that $R_{0,t}=\alpha_1^\circ\beta_1^{*\circ\prime}R_{1,t}+R_{\varepsilon,t}$; see (A.8). ■

The asymptotic distribution of the rank test statistic then follows.

Proof of Theorem 2. Use the arguments in the proof of Lemma 11.1 of Johansen (1995). ■

Finally, the asymptotic distribution for the maximum likelihood estimators in M_1 is established. It is convenient to mark these estimators by an index M.

LEMMA A.3. *Consider the maximum likelihood estimators in the unrestricted model* M_1 *identified by (4.2). Consider processes* X_t *satisfying* M_1 *as given by (A.4). Suppose*

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Assumptions 1–3 hold. Then, for $\Sigma_{\beta\beta}$, $\Sigma_{\beta\varepsilon}$, Σ_V, *and* $\Sigma_{V\varepsilon}$ *defined in Lemmas A.1 and A.2, it holds that*

$$\hat{\Omega}_{\mathsf{M}} \xrightarrow{\mathsf{P}} \Omega_\circ, \qquad T^{1/2}\left(\hat{\alpha}_1^{\mathsf{M}} - \alpha_1^\circ\right) = \Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1} + o_{\mathsf{P}}(1),$$

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

Proof of Lemma A.3. Use the arguments in the proof of Lemmas 13.1 and 13.2 of Johansen (1995). ■

A.4. Consistency under $\boldsymbol{M_{1x}}$**.** For fixed ρ, the proof of the consistency result in Theorem 3 would resemble the I(2) analysis of Johansen (1997). As ρ is unknown, an additional layer of complexity is added. The following steps are taken: First, some new parameters are introduced and a full rank property shown for these parameters. Second, an auxiliary parametrization is introduced as in the I(2) analysis, and consistency is shown for the auxiliary parameters. Third, the consistency of the original parametrization is shown.

First some notation. Recall the parameter $\tau_\perp = \Psi_\rho\overline{\beta}_{\rho\perp}^\circ$ 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:

$$\hat{\tau}_\perp = \hat{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ, \qquad \breve{\tau}_\perp = \breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ, \qquad \breve{\alpha}_1 = \frac{1-\rho_\circ}{1-\hat{\rho}}\alpha_1^\circ, \tag{A.13}$$

defined in terms of the quantities

$$\hat{\Psi}_\rho = I + \frac{\hat{\Pi}_1}{1-\hat{\rho}} - \sum_{j=1}^{k-2}\hat{\rho}^{-j}\hat{\Phi}_j, \qquad \breve{\Psi}_\rho = I + \frac{\Pi_1^\circ}{1-\rho} - \sum_{j=1}^{k-2}\rho^{-j}\Phi_j^\circ. \tag{A.14}$$

The matrix $\breve{\tau}_\perp$ depends on ρ. It is shown to be nonzero.

LEMMA A.4. *Suppose Assumption 1 is satisfied. Then*

(i) $(1-\rho)\breve{\tau}_\perp \neq 0$ *for all* $\rho > 1$;
(ii) a constant $c > 0$ *exists so, for all* $\rho \geq 1$, *then* $\breve{\tau}_\perp'\breve{\tau}_\perp \geq c$.

Proof of Lemma A.4. (i) Consider the companion form $\mathbf{x}_t = A^\circ\mathbf{x}_{t-1} + \mathbf{e}_t^\circ$, where

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

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

and A° is the matrix

$$\begin{Bmatrix} \rho_\circ I+\beta_\rho^{\circ\prime}\alpha_\rho^\circ & \beta_\rho^{\circ\prime}(\Pi_1^\circ+\Phi_1^\circ) & \beta_\rho^{\circ\prime}(\Phi_2^\circ-\Phi_1^\circ) & \cdots & \beta_\rho^{\circ\prime}(\Phi_{k-2}^\circ-\Phi_{k-3}^\circ) & -\beta_\rho^{\circ\prime}\Phi_{k-2}^\circ \\ \alpha_\rho^\circ & I+\Pi_1^\circ+\Phi_1^\circ & \Phi_2^\circ-\Phi_1^\circ & \cdots & \Phi_{k-2}^\circ-\Phi_{k-3}^\circ & -\Phi_{k-2}^\circ \\ & I & & & & \\ & & \ddots & & & \\ & & & & I & \end{Bmatrix}.$$

Due to Theorem 1, 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

$$\begin{bmatrix} (\rho-\rho_\circ)I-\beta_\rho^{\circ\prime}\alpha_\rho^\circ & -\beta_\rho^{\circ\prime}\{\Pi_1^\circ+\Sigma_{j=1}^{k-2}(\rho-1)\rho^{-j}\Phi_1^\circ\} \\ -\alpha_\rho^\circ & (\rho-1)I-\Pi_1^\circ-\Sigma_{j=1}^{k-2}(\rho-1)\rho^{-j}\Phi_1^\circ \end{bmatrix}.$$

Now, assume (i) does not hold. Then $(\rho-1)\breve{\Psi}_\rho = \eta\beta_\rho^{\circ\prime}$ 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.

(ii) Consider four different special cases: (a) When ρ belongs to any compact set so $\rho > 1$, then both $\breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ$ and $1-\rho$ have outcomes in compact sets not including 0 due to the first result. (b) For large ρ then, $\breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ$ converges to $\overline{\beta}_{\rho\perp}^\circ$. (c) For ρ approaching 1 and $\beta_1^{\circ\prime}\overline{\beta}_{\rho\perp}^\circ \neq 0$ then, $\breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ$ has an infinite asymptote. (d) For $\beta_1^{\circ\prime}\overline{\beta}_{\rho\perp}^\circ = 0$ then, $\breve{\tau}_\perp$ is continuous in ρ on $1 \leq \rho \leq \varrho$ so it suffices to argue that ϱ can be chosen so close to 1 that $\breve{\Psi}_\rho\overline{\beta}_{\rho\perp}^\circ$ has no 0 points for $1 \leq \rho \leq \varrho$. Now the vector polynomial $\big(I - \Sigma_{j=1}^{k-2} z^j\Phi_j^\circ\big)\overline{\beta}_{\rho\perp}^\circ$ has finitely many 0 points in a neighborhood of 1. It suffices to show that such 0 points cannot fall at 1. Assumption 1(iii) implies that $\Psi_1^\circ\beta_{1\perp}^\circ = \big\{I_p + \alpha_\rho^\circ\beta_\rho^{\circ\prime}/(\rho_\circ - 1) - \Sigma_{j=1}^{k-2} z^j\Phi_j^\circ\big\}\beta_{1\perp}^\circ$ is nonzero. Since $\beta_1^{\circ\prime}\overline{\beta}_{\rho\perp}^\circ = 0$, then $\beta_{\rho\perp}^\circ \in \text{span}(\beta_{1\perp}^\circ)$; this implies that

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

is nonzero as desired. ■

Second, an auxiliary parametrization is introduced as in Johansen (1997).

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

LEMMA A.5. *Suppose β_ρ is identified by $\beta_\rho'\overline{\beta}_\rho^\circ = I_{p-1}$, while β_1^* satisfies (4.2). Let*

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

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

$$\theta_\rho^U = \alpha_\rho^\circ - \alpha_\rho + (\rho_\circ - \rho)\Psi_{\rho_\circ}\overline{\beta}_\rho^\circ, \qquad \theta_1^U = \alpha_1^\circ - \frac{1-\rho}{1-\rho_\circ}\alpha_1,$$

$$\theta_{\rho,j}^U = \Phi_j^\circ - \frac{\rho}{\rho_\circ}\Phi_j + (\rho_\circ - \rho)\sum_{m=j+1}^{k-2}\Phi_m\rho_\circ^{j-1-m}, \qquad \theta^V = -\frac{1-\rho}{1-\rho_\circ}\Pi_1^*\overline{\beta}_{1\perp}^{*\circ},$$

$$\theta^W = -\alpha_\rho\beta_\rho'\overline{\beta}_{\rho\perp}^\circ + (\rho_\circ - \rho)\Psi_{\rho_\circ}\overline{\beta}_{\rho\perp}^\circ, \qquad \theta^\mu = \mu^\circ - \mu.$$

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

$$\sum_{j=1}^{k-2}\rho^{-j}\theta_{\rho,j}^U = \sum_{j=1}^{k-2}\rho^{-j}\Phi_j^\circ - \sum_{j=1}^{k-2}\rho_\circ^{-j}\Phi_j. \tag{A.15}$$

Proof of Lemma A.5. Note first the identities

$$(1-\rho_\circ)\Delta_\rho X_{t-1}^* = (1-\rho)\Delta_{\rho_\circ}X_{t-1}^* + (\rho-\rho_\circ)\begin{pmatrix}\Delta_1 X_{t-1}\\ 0\end{pmatrix}, \tag{A.16}$$

$$\Delta_1\Delta_\rho X_{t-j} = \Delta_1\Delta_{\rho_\circ}X_{t-j} + (\rho_\circ - \rho)\Delta_1 X_{t-j-1}, \tag{A.17}$$

$$\Delta_1 X_{t-j-1} = \rho_\circ^{-j}\left(\Delta_1 X_{t-1} - \sum_{l=1}^{j}\rho_\circ^{l-1}\Delta_1\Delta_{\rho_\circ}X_{t-l}\right). \tag{A.18}$$

Use (A.17) to write $\Delta_1\Delta_\rho X_t = \Delta_1\Delta_{\rho_\circ}X_t - (\rho-\rho_\circ)\Delta_1 X_{t-1}$. Then substitute $\Delta_1\Delta_{\rho_\circ}X_t$ using the equation (A.4) to arrive at

$$\Delta_1\Delta_\rho X_t = \varepsilon_t^\circ + \alpha_1^\circ\beta_1^{*\circ\prime}\Delta_{\rho_\circ}X_{t-1}^* + \alpha_\rho^\circ\beta_\rho^{\circ\prime}\Delta_1 X_{t-1} + \sum_{j=1}^{k-2}\Phi_j^\circ\Delta_1\Delta_{\rho_\circ}X_{t-j} - (\rho-\rho_\circ)\Delta_1 X_{t-1} + \mu^\circ. \tag{A.19}$$

Insert these expressions in the error equation (A.9) to get

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

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Now, premultiply $\Delta_1 X_{t-1}$ 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 (4.2).

The identity (A.15) follows by straightforward inspection. ■

Consistency for the estimator of θ, Ω is shown as in Johansen (1997).

LEMMA A.6. *Consider the auxiliary model in which* $\varepsilon_t = \varepsilon^{\circ}_t + \theta S^{\circ}_{t-1}$ *as outlined in Lemma A.5, where* θ, Ω *vary freely, so* $\theta \in \mathbf{R}^{p\times \dim S}$ *and* Ω *is positive definite. Suppose Assumptions 1, 2, and 3 are satisfied. Then the maximum likelihood estimator for* θ, Ω *exists with probability tending to 1 and satisfies* $\hat{\theta} N_S^{-1} \xrightarrow{\mathsf{P}} 0$.

Proof of Lemma A.6. Corresponding to the argument in Section A.3, the level curves of the likelihood function under M_{1x} are given in terms of $\det\{\hat{\Omega}(\theta)\}$, where $T\hat{\Omega}(\theta) = \sum_{t=1}^{T} \varepsilon_t^{\otimes 2}$. Partial regression implies

$$T\hat{\Omega}(\theta) = \sum_{t=1}^{T} \left(\varepsilon_t | S^{\circ}_{t-1}\right)^{\otimes 2} + \sum_{t=1}^{T} \varepsilon_t S^{\circ\prime}_{t-1} \left\{ \sum_{t=1}^{T} \left(S^{\circ}_{t-1}\right)^{\otimes 2} \right\}^{-1} \sum_{t=1}^{T} S^{\circ}_{t-1} \varepsilon'_t.$$

According to Lemma A.1(i, v, vii), it holds that

$$T^{-1} \sum_{t=1}^{T} \left(\varepsilon_t | S^{\circ}_{t-1}\right)^{\otimes 2} = T^{-1} \sum_{t=1}^{T} \left(\varepsilon^{\circ}_t | S^{\circ}_{t-1}\right)^{\otimes 2} \xrightarrow{\mathsf{P}} \Omega_{\circ},$$

$$T^{-1/2} \sum_{t=1}^{T} \varepsilon^{\circ}_t S^{\circ}_{t-1} \left\{ \sum_{t=1}^{T} \left(S^{\circ}_{t-1}\right)^{\otimes 2} \right\}^{-1/2} \xrightarrow{\mathsf{P}} 0,$$

$$T^{-1} \sum_{t=1}^{T} (N_S S^{\circ}_{t-1})^{\otimes 2} \xrightarrow{\mathsf{D}} \Sigma_S \overset{\text{a.s.}}{>} 0.$$

The level curves then satisfy $T\hat{\Omega}(\theta) = \Omega_{\circ} + R$, where

$$R\hat{\Omega}(\theta) = \mathrm{o}_{\mathsf{P}}(1) + \left\{ \mathrm{o}_{\mathsf{P}}(1) + \theta N_S^{-1} \Sigma_S^{1/2} \right\} \left\{ \mathrm{o}_{\mathsf{P}}(1) + \theta N_S^{-1} \Sigma_S^{1/2} \right\}'.$$

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

Proof of Theorem 3. Theorem 3 (i) concerns $\hat{\mu}$. This follows directly, since $\mu = \theta^{\mu}$.

Theorem 3(ii) concerns $\hat{\rho}$ (proof by contradiction). First, the assumption of nonconvergence 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 A.6 shows that $\hat{\theta} N_S^{-1} \to 0$ in probability. Thus, for all $\zeta > 0$, there exists a subsequence (T'), so for all T' it holds that $\mathsf{P}(\Theta_{T'}) > \delta$ where $\Theta_{T'} = \left(N_W^{-1}|\hat{\rho} - \rho_{\circ}| > \kappa\right.$ and $\left.|\hat{\theta} N_S^{-1}| < \zeta\right)$.

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

$$\left| \frac{\hat{\theta}^W N_W^{-1}}{\left(\hat{\rho} - \rho_{\circ}\right) N_W^{-1}} \right| = \left| \frac{\hat{\theta}^W}{\hat{\rho} - \rho_{\circ}} \right| < \frac{\zeta}{\kappa}. \tag{A.20}$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

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

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

Recall that ρ belongs to a closed subset of the parameter space, hence, $\rho \geq \varrho$ for some $\varrho > 1$. Due to Lemma A.4, the absolute value of the first term is therefore bounded from below by some $c > 0$ not depending on ϱ. The second term can be rewritten using the identity (A.15) as

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

Due to the consistency of the estimator for θ established in Lemma A.6 and the inequalities $|\rho|^{-j} \leq |\varrho|^{-j}$ and $|1-\rho|^{-1} \leq |1-\varrho|^{-1}$, then for all $\eta > 0$ 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 (A.20).

For Theorem 3(iii, iv), once $\hat\rho$ is known, the remaining estimators can be derived from the auxiliary estimator $\hat\theta$. Due to the consistency of $\hat\rho$ and $\hat\theta$ established in Theorem 3(ii) and Lemma A.6, 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 $\mathrm{o_P}(1)$, then $\hat\alpha_1 = \alpha_1^\circ + \mathrm{o_P}(1)$. 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 $\hat\theta_{\rho,k-2}^U, \ldots, \hat\theta_{\rho,1}^U, \hat\theta^V, \hat\theta_\rho^U$.

For Theorem 3(v), the consistency of $\hat\Omega$ is stated directly in Lemma A.6. ■

A.5. Improving the Rate of Consistency under H_x. The consistency rate of the estimators under H_x is now improved through a detailed analysis of the likelihood equations. Denoting the estimators by an index H, the result in Theorem 3 shows, for instance, that $\hat\rho_\mathsf{H} - \rho_\circ = \mathrm{o_P}(T^{1/2}\rho_\circ^{-T})$. This rate of consistency is now improved by a factor of $T^{-\xi/2}$, where ξ arose in Lemma A.1. When $\gamma > 2$ in Assumption 2, which amounts to assuming fourth moments, then the consistency rate is improved to $\hat\rho_\mathsf{H} - \rho_\circ = \mathrm{o_P}(T^{1/4}\rho_\circ^{-T})$. The following steps are taken: First, the relevant residuals under H_x are introduced and studied. Then likelihood equations for $\hat\rho_\mathsf{H}$ and $\hat\beta_1^{*\mathsf{H}}$ are derived and solved. The solution is used to improve the consistency rate.

Introduce the vector $U_{\hat\rho,t-1} = \{(\beta_\rho^{\circ\prime}\Delta_1 X_{t-1})', (\Delta_1\Delta_{\hat\rho}X_{t-1})', \ldots, (\Delta_1\Delta_{\hat\rho}X_{t-k+2})'\}'$, and define the residuals $R_{0,t}^{\mathsf{H},\rho} = R_{0,t}^\mathsf{H}(\rho) = \left(\Delta_1\Delta_\rho X_t \middle| U_{\rho,t-1}, 1\right)$ and $R_{1,t}^{\mathsf{H},\rho} = R_{1,t}^\mathsf{H}(\rho) = \left(\Delta_\rho X_{t-1}^* \middle| U_{\rho,t-1}, 1\right)$.

The profile likelihood for $\rho, \alpha_1, \beta_1^*$ is then given in terms of

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

Profile estimates for α_1, β_1^* given ρ are found by reduced-rank regression of $R_{0,t}^{\mathsf{H},\rho}$ on $R_{1,t}^{\mathsf{H},\rho}$, giving estimators $\hat\alpha_1^{\mathsf{H},\rho}, \hat\beta_1^{*\mathsf{H},\rho}$, and

$$\hat\Omega_\mathsf{H}(\rho) = T^{-1}\sum_{t=1}^T \left(\hat\varepsilon_t^{\mathsf{H},\rho}\right)^{\otimes 2}, \quad \text{where } \hat\varepsilon_t^{\mathsf{H},\rho} = R_{0,t}^{\mathsf{H},\rho} - \hat\alpha_1^{\mathsf{H},\rho}\hat\beta_1^{*\mathsf{H},\rho\prime}R_{1,t}^{\mathsf{H},\rho}.$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Minimizing the determinant of $\hat{\Omega}_{\mathsf{H}}(\rho)$ with respect to ρ gives maximum likelihood estimator $\hat{\rho}$, and hence $\hat{\alpha}_1^{\mathsf{H}} = \hat{\alpha}_1^{\mathsf{H},\hat{\rho}}$, $\hat{\beta}_1^{*\mathsf{H}} = \hat{\beta}_1^{*\mathsf{H},\hat{\rho}}$, and the residuals

$$R_{0,t}^{\mathsf{H}} = R_{0,t}^{\mathsf{H}}(\hat{\rho}_{\mathsf{H}}) = \left(\Delta_1 \Delta_{\hat{\rho}} X_t \,\middle|\, U_{\hat{\rho},t-1}, 1\right),$$

$$R_{1,t}^{\mathsf{H}} = R_{1,t}^{\mathsf{H}}(\hat{\rho}_{\mathsf{H}}) = \left(\Delta_{\hat{\rho}} X_{t-1}^* \,\middle|\, U_{\hat{\rho},t-1}, 1\right).$$

The maximum likelihood estimator for the variance is $\hat{\Omega}_{\mathsf{H}} = T^{-1}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t^{\mathsf{H}}\right)^{\otimes 2}$, where

$$\hat{\varepsilon}_t^{\mathsf{H}} = R_{0,t}^{\mathsf{H}} - \hat{\alpha}_1^{\mathsf{H}}\hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}} = (R_{0,t}^{\mathsf{H}} | \hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}}) = (\Delta_1 \Delta_{\hat{\rho}} X_t | \hat{\beta}_1^{*\mathsf{H}\prime} \Delta_{\hat{\rho}} X_{t-1}^*, U_{\hat{\rho},t-1}, 1). \tag{A.21}$$

Define also the residuals $R_{\varepsilon,t}^{\mathsf{H}} = \left(\varepsilon_t^{\circ} | U_{\hat{\rho},t-1}^{\mathsf{H}}, 1\right)$ and $R_{W,t}^{\mathsf{H}} = \left(W_{t-1}^{\circ} | U_{\hat{\rho},t-1}^{\mathsf{H}}, 1\right)$.

First, an expression linking the residuals akin to (A.8) is derived.

LEMMA A.7. *Recall $\breve{\tau}_\perp$, $\breve{\alpha}_1$ defined in (A.13). Then*

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

Proof of Lemma A.7. Insert (A.16)–(A.18) in (A.19) to get

$$\Delta_1 \Delta_{\hat{\rho}} X_t = \varepsilon_t^{\circ} + \frac{1-\rho_{\circ}}{1-\hat{\rho}_{\mathsf{H}}} \alpha_1^{\circ} \beta_1^{*\circ\prime} \Delta_{\hat{\rho}} X_{t-1}^* + \alpha_\rho^{\circ} \beta_\rho^{\circ\prime} \Delta_1 X_{t-1} - \left(\hat{\rho}_{\mathsf{H}} - \rho_{\circ}\right) \breve{\Psi}_{\hat{\rho}}^{\mathsf{H}} \Delta_1 X_{t-1}$$
$$+ \sum_{j=1}^{k-2} \Phi_j^{\circ} \Delta_1 \Delta_{\hat{\rho}} X_{t-j} - \left(\hat{\rho}_{\mathsf{H}} - \rho_{\circ}\right) \sum_{j=1}^{k-2} \Phi_j^{\circ} \sum_{l=1}^{j} \hat{\rho}_{\mathsf{H}}^{l-1} \Delta_1 \Delta_{\hat{\rho}} X_{t-l} + \mu^{\circ}.$$

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

Second, the likelihood equations for ρ and β_1^* can be found.

LEMMA A.8. *The likelihood equations for ρ and β_1^* under H_x are*

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

Proof of Lemma A.8. Following the approach of Johansen (1995, p. 182), a first version the likelihood equations for ρ and β_1^* is

$$0 = \operatorname{tr}\left(\hat{\Omega}^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} Y_{t-1}'\right), \qquad 0 = \hat{\alpha}_1^{\mathsf{H}\prime} \hat{\Omega}_{\mathsf{H}}^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} \left(\Delta_{\hat{\rho}} X_{t-1}^*\right)', \tag{A.22}$$

where $Y_{t-1} = \Delta_1 X_{t-1} - \hat{\Pi}_1^* \left(X_{t-2}', t\right)' - \sum_{j=1}^{k-2} \hat{\Gamma}_j \Delta_1 X_{t-j-1}$.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Equation for ρ. Since $\hat{\varepsilon}_t^{\mathsf{H}}$ involves regresson on $\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*, U_{\hat{\rho},t-1}, 1$, it is first argued that

$$\left(Y_{t-1}|\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*, U_{\hat{\rho},t-1}, 1\right) = \hat{\tau}_\perp\left(W_{t-1}^\circ|\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*, U_{\hat{\rho},t-1}, 1\right). \tag{A.23}$$

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 that

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

Second, by the identity $\left(1-\hat{\rho}\right)\left(X_{t-2}', t\right)' = \Delta_{\hat{\rho}}X_{t-1}^* - \left(\Delta_1 X_{t-1}', 0\right)'$, it holds that

$$\left\{\hat{\Pi}_1^*(X_{t-2}', t)'|\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*, U_{\hat{\rho},t-1}, 1\right\} = \frac{\hat{\Pi}_1\overline{\beta}_{\rho\perp}^\circ}{1-\hat{\rho}}\left(W_{t-1}^\circ|\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*, U_{\hat{\rho},t-1}, 1\right).$$

Third, use (A.2) for a similar substitution in the third term. Insert (A.23) in (A.22) to get $0 = \mathrm{tr}\{\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^T\hat{\varepsilon}_t^{\mathsf{H}}\left(\hat{\tau}_\perp W_{t-1}^\circ\right)'\} = \hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\sum_{t=1}^T\hat{\varepsilon}_t^{\mathsf{H}}W_{t-1}^\circ$, as desired.

Equation for β_1^*. Postmultiply the equation with $\hat{\beta}_{1\perp}^{*\mathsf{H}}$. ■

The third argument is to solve the likelihood equations in Lemma A.8 for $\hat{\rho}$ and $\hat{\beta}_1^*$.

LEMMA A.9. *In the model* M_{1x}, *the maximum likelihood estimators for* ρ, β_1^* *satisfy*

$$\begin{aligned}
&\hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\left(\hat{\rho}_{\mathsf{H}} - \rho_\circ\right)\left(\check{\tau}_\perp\hat{\mathcal{M}}_{WW} - \mathcal{P}_\alpha\check{\tau}_\perp\hat{\mathcal{M}}_{W\beta_\perp}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}^{-1}\hat{\mathcal{M}}_{\beta_\perp W}\right)\\
&\quad = \hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\left(\hat{\mathcal{M}}_{\varepsilon W} - \mathcal{P}_\alpha\hat{\mathcal{M}}_{\varepsilon\beta_\perp}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}^{-1}\hat{\mathcal{M}}_{\beta_\perp W}\right),\\
&\hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\left(\check{\alpha}_1\beta_1^{*\circ\prime}\overline{\beta}_{1\perp}^{*\mathsf{H}}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp} - \mathcal{P}_{\tau_\perp}\check{\alpha}_1\beta_1^{*\circ\prime}\overline{\beta}_{1\perp}^{*\mathsf{H}}\hat{\mathcal{M}}_{\beta_\perp W}\hat{\mathcal{M}}_{WW}^{-1}\hat{\mathcal{M}}_{W\beta_\perp}\right)\\
&\quad = -\hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\left(\hat{\mathcal{M}}_{\varepsilon\beta_\perp} - \mathcal{P}_{\tau_\perp}\hat{\mathcal{M}}_{\varepsilon W}\hat{\mathcal{M}}_{WW}^{-1}\hat{\mathcal{M}}_{W\beta_\perp}\right),
\end{aligned}$$

where $\mathcal{P}_\alpha = \check{\alpha}_1\left(\hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\check{\alpha}_1\right)^{-1}\hat{\alpha}_1^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}$, $\mathcal{P}_{\tau_\perp} = \check{\tau}_\perp\left(\hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}\check{\tau}_\perp\right)^{-1}\hat{\tau}_\perp^{\mathsf{H}\prime}\hat{\Omega}_{\mathsf{H}}^{-1}$ *and*

$$\hat{\mathcal{M}} = \begin{pmatrix} \hat{\mathcal{M}}_{\beta_\perp\beta_\perp} & \hat{\mathcal{M}}_{\beta_\perp\varepsilon} & \hat{\mathcal{M}}_{\beta_\perp W}\\ \hat{\mathcal{M}}_{\varepsilon\beta_\perp} & \hat{\mathcal{M}}_{\varepsilon\varepsilon} & \hat{\mathcal{M}}_{\varepsilon W}\\ \hat{\mathcal{M}}_{W\beta_\perp} & \hat{\mathcal{M}}_{W\varepsilon} & \hat{\mathcal{M}}_{WW} \end{pmatrix} = \sum_{t=1}^T\left(\begin{array}{c|c} \hat{\beta}_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}} & \\ R_{\varepsilon,t}^{\mathsf{H}} & \hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\\ R_{W,t}^{\mathsf{H}} & \end{array}\right)^{\otimes 2}. \tag{A.24}$$

Proof of Lemma A.9. Recall from (A.21) that $\hat{\varepsilon}_t^{\mathsf{H}} = \left(R_{0,t}^{\mathsf{H}}|\hat{\beta}_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right)$. Insert the expression for $R_{0,t}^{\mathsf{H}}$ using Lemma A.7 to get

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

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Postmultiply $\beta_1^{*\circ\prime}$ by the identity $I_{p+1} = \overline{\hat\beta}_{1\perp}^{*\mathsf{H}}\hat\beta_{1\perp}^{*\mathsf{H}} + \overline{\hat\beta}_1^{*\mathsf{H}}\hat\beta_1^{*\mathsf{H}}$. Since $\hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}$ disappears by regression this results in

$$\hat\varepsilon_t^{\mathsf{H}} = \{R_{\varepsilon,t}^{\mathsf{H}} + \breve\alpha_1^{\mathsf{H}}\beta_1^{*\circ\prime}\overline{\hat\beta}_{1\perp}^{*\mathsf{H}}\hat\beta_{1\perp}^{*\mathsf{H}}R_{1,t}^{\mathsf{H}} - (\hat\rho_{\mathsf{H}} - \rho_\circ)\breve\tau_\perp^{\mathsf{H}}R_{W,t}^{\mathsf{H}}|\hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\}. \tag{A.25}$$

Insert this in the likelihood equations in Lemma A.8.

Solve the first equation for $\hat\rho_{\mathsf{H}} - \rho_\circ$, insert this in the second equation, and solve for $\beta_{1\perp}^{*\circ\prime}\overline{\hat\beta}_{1\perp}^{*\mathsf{H}}$. Next, solve the second equation for $\beta_{1\perp}^{*\circ\prime}\overline{\hat\beta}_{1\perp}^{*\mathsf{H}}$, insert this in the first equation, and solve for $\hat\rho_{\mathsf{H}} - \rho_\circ$. ■

Next the asymptotic behavior of the matrix $\hat{\mathcal{M}}$ in (A.24) is derived.

LEMMA A.10. *Consider the model* M_1 *with* $\rho \geq \varrho$ *for some* $\varrho > 1$, *the process* X_t *given by* (A.4), *and suppose Assumptions 1–3 are satisfied. Suppose* $T^{-1/2}\rho_\circ^T(\hat\rho_{\mathsf{H}} - \rho_\circ) = o_P(T^{-\zeta})$ *and* $\hat\beta_1^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}N_V^{-1} = o_P(T^{-\zeta})$ *for some* $0 \leq \zeta < 1/2$. *Then*

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\\ N_V\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\\ R_{\varepsilon,t}^{\mathsf{H}}\\ N_W R_{W,t}^{\mathsf{H}}\end{pmatrix}R_{1,t}^{\mathsf{H}\prime}\hat\beta_1^{*\mathsf{H}} = \begin{pmatrix}\Sigma_{\beta\beta} + o_P(1)\\ O_P(T^{-1/2})\\ T^{-1/2}\tilde\Sigma_{\varepsilon\beta}\\ o_P(1)\end{pmatrix},$$

where $\tilde\Sigma_{\varepsilon\beta} = \Sigma_{\varepsilon\beta} + o_P(1) + o_P(T^{1/2-\zeta-\xi/2})$, *and*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\begin{array}{c}N_V\hat\beta_{1\perp}^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\\ \varepsilon_t^\circ\\ N_W W_{t-1}^\circ\end{array}\middle|\,\hat\beta_1^{*\mathsf{H}\prime}R_{1,t}^{\mathsf{H}}\right)^{\otimes 2} = \begin{pmatrix}\Sigma_V + o_P(1) & T^{-1/2}\tilde\Sigma_{V\varepsilon} & o_P(T^{-\zeta})\\ * & \Omega_\circ + o_P(1) & T^{-1/2}\tilde S_{\varepsilon W}\\ * & * & \Sigma_W + o_P(1)\end{pmatrix},$$

where $\tilde\Sigma_{V\varepsilon} = \Sigma_{V\varepsilon} + o_P(1) + o_P(T^{1/2-\zeta-\xi/2})$ *and* $\tilde S_{\varepsilon W} = S_{\varepsilon W} + o_P(1) + o_P(T^{1/2-\zeta-\xi/2})$ *with* $S_{\varepsilon W} = \rho_\circ^{-T}\sum_{t=1}^{T}\varepsilon_t^\circ W_{t-1}^\circ$.

Proof of Lemma A.10. First, recall M° defined in (A.12). Argue that

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\left(\begin{array}{c}\beta_1^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^*\\ U_{\hat\rho,t-1}\end{array}\right)\left(\begin{array}{c}\beta_1^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^*\\ U_{\hat\rho,t-1}\end{array}\right)'\otimes\begin{pmatrix}N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\hat\rho}X_{t-1}^*\\ \varepsilon_t^\circ\\ N_W W_{t-1}^\circ\end{pmatrix}\middle|\,1\right\}^{\otimes 2} = M^\circ + o_P(T^{-\zeta})\begin{pmatrix}1 & 1 & T^{-\xi/2} & 1\\ 1 & 1 & T^{-\xi/2} & 1\\ T^{-\xi/2} & T^{-\xi/2} & 0 & 0\\ 1 & 1 & 0 & 0\end{pmatrix}.$$

To see this, note that the relations (A.16)–(A.18) imply

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

for some deterministic matrices Q.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

When taking sums of squares of this expression, the last term can be omitted since $T^{-1}(\hat{\rho}_{\mathsf{H}}-\rho_\circ)^2\sum_{t=1}^T(S_{t-1}^\circ)^{\otimes 2}=\mathrm{o}_{\mathsf{P}}(T^{-2\zeta})$ due to Theorem 3(ii) and Lemma A.1(vi), whereas the cross-terms are $\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$ by the Cauchy-Schwarz inequality.

When considering the cross-product with ε_t°, use Theorem 3(ii) and Lemma A.1(vii) to see $T^{-1}(\hat{\rho}_{\mathsf{H}}-\rho_\circ)\sum_{t=1}^T S_{t-1}^\circ\varepsilon_t^{\circ\prime}=\mathrm{o}_{\mathsf{P}}(T^{-\xi/2-\zeta})$.

When considering the cross-product with W_{t-1}°, use Theorem 3(ii) and Lemma A.1(vi) to see $T^{-1}(\hat{\rho}_{\mathsf{H}}-\rho_\circ)\sum_{t=1}^T S_{t-1}^\circ W_{t-1}^\circ N_W=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$.

Second, argue that

$$\frac{1}{T}\sum_{t=1}^T\left\{\begin{array}{c|c}\begin{pmatrix}\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*\\ U_{\hat{\rho},t-1}\end{pmatrix} & \\ N_V\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^* & 1\\ \varepsilon_t^\circ & \\ N_W W_{t-1}^\circ & \end{array}\right\}^{\otimes 2}=M^\circ+\mathrm{o}_{\mathsf{P}}(T^{-\zeta})\begin{pmatrix}1&1&T^{-\xi/2}&1\\1&1&T^{-\xi/2}&1\\T^{-\xi/2}&T^{-\xi/2}&0&0\\1&1&0&0\end{pmatrix}.$$

By exploiting the identity $I_{p+1}=\overline{\beta}_1^{*\circ}\beta_1^{*\circ\prime}+\overline{\beta}_{1\perp}^{*\circ}\beta_{1\perp}^{*\circ\prime}$, it is seen that

$$\begin{pmatrix}\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*\\ N_V\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*\end{pmatrix}=\begin{pmatrix}\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_1^{*\circ}&\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}N_V^{-1}\\ N_V^2N_V^{-1}\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\overline{\beta}_1^{*\circ}&N_V\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}N_V^{-1}\end{pmatrix}\begin{pmatrix}\beta_1^{*\circ\prime}\Delta_{\hat{\rho}}X_{t-1}^*\\ N_V\beta_{1\perp}^{*\circ\prime}\Delta_{\hat{\rho}}X_{t-1}^*\end{pmatrix}.$$

Due to the identification (4.2), $\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_1^{*\circ}$ and $\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}$ are identity matrices. By assumption $\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}N_V^{-1}=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$ and the identification (4.2) also implies that $N_V^{-1}\hat{\beta}_{1\perp}^{*\mathsf{H}\prime}\overline{\beta}_1^{*\circ}=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$.

Third, partial out $\hat{\beta}_1^{*\mathsf{H}\prime}\Delta_{\hat{\rho}}X_{t-1}^*$ and $U_{\hat{\rho},t-1}$ to get the desired result. ■

The convergence rates of $\hat{\rho}_{\mathsf{H}}$ and $\beta_{1\perp}^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*\mathsf{H}}$ are now improved.

LEMMA A.11. *Consider model* M_{1x}*, identified by (4.2) with* $\rho\geq\varrho$ *for some* $\varrho>1$*. Suppose Assumptions 1–3 are satisfied. Then, for all* $\xi<\gamma/(2+\gamma)$*, it holds that*

$$T^{-1/2}\rho_\circ^T(\hat{\rho}_{\mathsf{H}}-\rho_\circ),\ \beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*\mathsf{H}}N_V^{-1}=\mathrm{o}_{\mathsf{P}}(T^{-\xi/2}).$$

Proof of Lemma A.11. Recall the likelihood equations in Lemma A.9. Theorem 3 implies $T^{-1/2}\rho_\circ^T(\hat{\rho}_{\mathsf{H}}-\rho_\circ)=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$ and $\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}N_V^{-1}=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$ with $\zeta=0$. Due to the identification (4.2) and the consistency of $\hat{\beta}_1^{*\mathsf{H}}$, then

$$\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*\mathsf{H}}=-\hat{\beta}_1^{*\mathsf{H}\prime}\overline{\beta}_{1\perp}^{*\circ}\{1+\mathrm{o}_{\mathsf{P}}(1)\},\tag{A.26}$$

so $\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*\mathsf{H}}N_V^{-1}=\mathrm{o}_{\mathsf{P}}(T^{-\zeta})$. Lemma A.10 then implies that

$$N_W\hat{\mathcal{M}}_{W\beta_\perp}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}^{-1}\hat{\mathcal{M}}_{\beta_\perp W}N_W,\ N_V\hat{\mathcal{M}}_{\beta_\perp W}\hat{\mathcal{M}}_{WW}^{-1}\hat{\mathcal{M}}_{W\beta_\perp}N_V=\mathrm{o}_{\mathsf{P}}\big(T^{1-2\zeta}\big),$$

$$\hat{\mathcal{M}}_{\varepsilon\beta_\perp}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}^{-1}\hat{\mathcal{M}}_{\beta_\perp W}N_W,\ \hat{\mathcal{M}}_{\varepsilon W}\hat{\mathcal{M}}_{WW}^{-1}\hat{\mathcal{M}}_{W\beta_\perp}N_V=\mathrm{o}_{\mathsf{P}}\big(T^{1-\zeta-\xi/2}\big),$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

$T^{-1}N_W\hat{\mathcal{M}}_{WW}N_W$ and $T^{-1}N_V\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}N_V$ are convergent, while $\hat{\mathcal{M}}_{\varepsilon W}N_W$ and $\hat{\mathcal{M}}_{\varepsilon\beta_\perp}N_V$ are $o_P(T^{1-\xi/2})$, then the expressions in Lemma A.9 imply that $T^{-1/2}\rho_\circ^T(\hat{\rho}_H - \rho_\circ)$ and $\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*H}N_V^{-1}$ are $o_P(T^{-\xi/2})$. ■

The improved convergence rate can be used to find the limiting expression for $\sum_{t=1}^T \hat{\varepsilon}_t^H W_{t-1}^\circ$ and $\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*H}$.

LEMMA A.12. *Consider model M_{1x}, identified by (4.2) with $\rho \geq \varrho$ for some $\varrho > 1$ and a process X_t given by (A.4). Suppose Assumptions 1–3 are satisfied. Then, for all $\xi < \gamma/(2+\gamma)$, it holds that*

(i) $T^{1/2}(\hat{\alpha}_1^H - \alpha) = \Sigma_{\varepsilon\beta}\Sigma_{\beta\beta}^{-1} + o_P(T^{1/2-\xi})$,

(ii) $(\hat{\beta}_1^{*H} - \beta_1^{*\circ})'\overline{\hat{\beta}}_{1\perp}^{*H}N_V^{-1}T^{1/2} = (\alpha_1^{\circ\prime}\Omega_\circ^{-1}\alpha_1^\circ)^{-1}\alpha_1^{\circ\prime}\Omega_\circ^{-1}\Sigma_{\varepsilon V}\Sigma_V^{-1} + o_P(T^{1/2-\xi})$,

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

Proof of Lemma A.12. (i) Given $\hat{\beta}_1^{*H}$, then α_1 is estimated by least squares regression so that

$$\hat{\alpha}_1^H - \alpha = \sum_{t=1}^T \left(R_{0,t}^H - \alpha\hat{\beta}_1^{*H\prime}R_{1,t}^H\right)R_{1,t}^{H\prime}\hat{\beta}_1^{*H}\left(\sum_{t=1}^T \hat{\beta}_1^{*H\prime}R_{1,t}^H R_{1,t}^{H\prime}\hat{\beta}_1^{*H}\right)^{-1}.$$

Insert $R_{0,t}^H$ from Lemma A.7. Postmultiply $\beta_1^{*\circ\prime}$ by $I_{p+1} = \overline{\hat{\beta}}_1^{*H}\hat{\beta}_1^{*H\prime} + \overline{\hat{\beta}}_{1\perp}^{*H}\hat{\beta}_{1\perp}^{*H\prime}$ to get

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

Apply that $\check{\alpha}_1^H$ is consistent at the same rate as $\hat{\rho}_H$, the consistency established in Lemma A.11, and combine with Lemma A.10.

(ii) Applying the proof of Lemma A.11, now with $\zeta = \xi/2$ due to the improved consistency in Lemma A.11, gives the likelihood equation

$$\hat{\alpha}_1^{H\prime}\hat{\Omega}_H^{-1}\check{\alpha}_1\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*H}\hat{\mathcal{M}}_{\beta_\perp\beta_\perp}\{1 + o_P(T^{-\xi})\} = -\hat{\alpha}_1^{H\prime}\hat{\Omega}_H^{-1}\{\hat{\mathcal{M}}_{\varepsilon\beta_\perp} + o_P(T^{-\xi})\}.$$

Insert the limiting expressions in Lemma A.10 to get the desired result.

(iii) Recall the expression for $\hat{\varepsilon}_t^H$ in (A.25). Manipulate, noting $\check{\tau}^{H\prime}\check{\tau}_\perp^H = 0$, to get

$$\rho_\circ^{-T}\check{\tau}^{H\prime}\sum_{t=1}^T \hat{\varepsilon}_t^H W_{t-1}^\circ = \rho_\circ^{-T}\check{\tau}^{H\prime}\sum_{t=1}^T \left\{R_{\varepsilon,t}^H + \check{\alpha}_1^H\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*H}\hat{\beta}_{1\perp}^{*H}R_{1,t}^H | \hat{\beta}_1^{*H\prime}R_{1,t}^H\right\}W_{t-1}^\circ.$$

By Lemmas A.10 and A.11 then, $\rho_\circ^{-T}\beta_1^{*\circ\prime}\overline{\hat{\beta}}_{1\perp}^{*H}\hat{\beta}_{1\perp}^{*H}\sum_{t=1}^T (R_{1,t}^H | \hat{\beta}_1^{*H\prime}R_{1,t}^H)W_{t-1}^\circ = o_P(T^{1/2-\xi})$ and $\rho_\circ^{-T}\check{\tau}^{H\prime}\sum_{t=1}^T R_{\varepsilon,t}^H W_{t-1}^\circ = \rho_\circ^{-T}\tau'\sum_{t=1}^T \varepsilon_t^\circ W_{t-1}^\circ + o_P(T^{1/2-\xi})$. ■

A.6. The Test on the Coexplosive Vector. The asymptotic result in Theorem 4 is proved. The following steps are taken: First, the variance estimators, $\hat{\Omega}_M$ and $\hat{\Omega}_H$ say, in the unrestricted model M_1 and under the hypothesis H_x are expanded. Second, two terms of the same appearance in these expressions are shown to have the same limit.

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

LEMMA A.13. *Consider the model* M_1 *and a process satisfying* M_1 *as given by (A.4). Then the maximum likelihood estimator for* Ω *satisfies*

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

Proof of Lemma A.13. The model equation (A.9) 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,

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

where, following the notation of Johansen (1995, Chap. 6), $\tilde{R}_{0,t}$ and $\tilde{R}_{1,t}$ are the residuals

$$\tilde{R}_{0,t} = \left(\Delta_1 X_t \,\middle|\, \Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}, 1 \right),$$

$$\tilde{R}_{1,t} = \left\{ (1-\rho_\circ) X_{t-1} \,\middle|\, \Delta_1 X_{t-1}, \ldots, \Delta_1 X_{t-k+1}, 1 \right\}.$$

Using the identity (A.1), 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 maximized by reduced-rank regression, giving unique maximum likelihood estimators $\hat{\alpha}_1^{\mathsf{M}}, \hat{\beta}_1^{\mathsf{M}*}$ and

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

Inserting $R_{0,t} = R_{\varepsilon,t} + \alpha_1 \beta_1^{*\prime} R_{1,t}$ as defined in (A.8) then gives the desired result. ■

LEMMA A.14. *Consider the model* M_{1x} *and a process satisfying (A.4). Then the maximum likelihood estimator for* Ω *satisfies*

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

Proof of Lemma A.14. Using partial regression on W_{t-1}°, then

$$T\hat{\Omega}_{\mathsf{H}} = \sum_{t=1}^{T} \left(\hat{\varepsilon}_t^{\mathsf{H}} | W_{t-1}^{\circ} \right)^{\otimes 2} + \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ} \left\{ \sum_{t=1}^{T} (W_{t-1}^{\circ})^2 \right\}^{-1} \sum_{t=1}^{T} W_{t-1}^{\circ} \hat{\varepsilon}_t^{\mathsf{H}\prime}. \tag{A.27}$$

For the first term in (A.27), recall that $\hat{\varepsilon}_t^{\mathsf{H}} = R_{0,t}^{\mathsf{H}} - \hat{\alpha}_1^{\mathsf{H}} \hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}^{\mathsf{H}}$ by (A.21). The residuals in the models M_1 and M_{1x} are related through partial regression. Since $U_{\hat{\rho},t-1}, W_{t-1}^{\circ}$ and $U_{\rho,t-1}^{\circ}, W_{t-1}^{\circ}$ span the same space, it holds that

$$R_{0,t} = \left(R_{0,t}^{\mathsf{H}} \,\middle|\, W_{t-1}^{\circ} \right), \qquad R_{1,t} = \left(R_{1,t}^{\mathsf{H}} \,\middle|\, W_{t-1}^{\circ} \right). \tag{A.28}$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

Therefore, $(\hat{\varepsilon}_t^{\mathsf{H}} | W_{t-1}^{\circ}) = R_{0,t} - \hat{\alpha}_1^{\mathsf{H}} \hat{\beta}_1^{*\mathsf{H}\prime} R_{1,t}$. As in the proof of Lemma A.13, apply the identity (A.8) to get

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

For the second term in (A.27), multiply $\hat{\varepsilon}_t^{\mathsf{H}}$ with the identity $I_p = \Omega \hat{\tau} (\check{\tau}' \Omega \hat{\tau})^{-1} \check{\tau}' + \check{\tau}_{\perp} (\hat{\tau}_{\perp}' \Omega^{-1} \check{\tau}_{\perp})^{-1} \hat{\tau}_{\perp}' \Omega^{-1}$. By Lemma A.8 then $\hat{\tau}_{\perp}' \Omega^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ} = 0$, so it holds, as desired, that $\sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ} = \hat{\Omega}_{\mathsf{H}} \hat{\tau}^{\mathsf{H}} (\check{\tau}^{\mathsf{H}\prime} \hat{\Omega}_{\mathsf{H}} \hat{\tau}^{\mathsf{H}})^{-1} \check{\tau}^{\mathsf{H}\prime} \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ}$. ■

The expressions for $\hat{\Omega}_{\mathsf{M}}$ and $\hat{\Omega}_{\mathsf{H}}$ in Lemmas A.13 and A.14 terms of similar appearances. These are shown to have the same limit.

LEMMA A.15. *Consider model* M_{1x}*, identified by (4.2) with* $\rho \geq \varrho$ *for some* $\varrho > 1$ *and a process* X_t *given by* (A.4)*. Suppose Assumptions 1–3 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$*, it holds that*

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

$$= \mathrm{o}_{\mathsf{P}}(T^{1/2-\xi}).$$

Proof of Lemma A.15. First note, that by the identification it holds that

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

The asymptotic properties for $\hat{\alpha}_1$ and $\hat{\beta}_1^{*}$ established in Lemmas A.3 and A.12 along with (A.26) then imply that

$$T^{1/2} \mathcal{S} = -\left(\Sigma_{\varepsilon\beta} \Sigma_{\beta\beta}^{-1} \beta_1^{*\circ\prime} + \mathcal{P}_{\alpha}^{\circ} \Sigma_{\varepsilon V} \Sigma_V^{-1} N_V \beta_{1\perp}^{*\circ\prime} \right) \left\{ 1 + \mathrm{o}_{\mathsf{P}} \left(T^{1/2-\xi} \right) \right\},$$

where $\mathcal{P}_{\alpha}^{\circ} = \alpha_1^{\circ} (\alpha_1^{\circ\prime} \Omega_{\circ} \alpha_1^{\circ})^{-1} \alpha_1^{\circ\prime}$. The expression of interest then follows by Lemma A.1. ■

Proof of Theorem 4. Due to the regression nature of the problem, the log likelihood ratio test statistic satisfies

$$LR(\mathsf{H}_x | \mathsf{M}_1) = -T \log \det \left(\hat{\Omega}_{\mathsf{M}} \hat{\Omega}_{\mathsf{H}}^{-1} \right) = -T \log \det \left\{ I_p - \hat{\Omega}_{\mathsf{H}}^{-1} \left(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}} \right) \right\}.$$

Theorem 3 shows that $\hat{\Omega}_{\mathsf{H}} \xrightarrow{\mathsf{P}} \Omega_{\circ}$. Thus, if it is argued that $T(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}})$ converges in distribution, then a Taylor expansion shows

$$LR(\mathsf{H}_x | \mathsf{M}_1) = \mathrm{tr} \left\{ \Omega_{\circ}^{-1} T \left(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}} \right) \right\} + \mathrm{o}_{\mathsf{P}}(1).$$

Now, combining Lemmas A.13 and A.14 shows

$$T(\hat{\Omega}_{\mathsf{H}} - \hat{\Omega}_{\mathsf{M}}) = \left[\hat{\Omega}_{\mathsf{H}} \hat{\tau}^{\mathsf{H}} \left(\check{\tau}^{\mathsf{H}\prime} \hat{\Omega}_{\mathsf{H}} \hat{\tau}^{\mathsf{H}} \right)^{-1} \check{\tau}^{\mathsf{H}\prime} \sum_{t=1}^{T} \hat{\varepsilon}_t^{\mathsf{H}} W_{t-1}^{\circ} \left\{ \sum_{t=1}^{T} \left(W_{t-1}^{\circ} \right)^2 \right\}^{-1/2} \right]^{\otimes 2}$$

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press

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

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

The last two terms are $o_{\mathsf{P}}(1)$ by Lemma A.15 assuming $\gamma > 2$. As for the first term, note that $\hat{\Omega}_{\mathsf{H}}$ and $\hat{\tau}^{\mathsf{H}}$ are consistent due to Theorem 3. By the assumption that $\gamma > 2$, Lemma A.12 shows that

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

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

https://doi.org/10.1017/S0266466609990144 Published online by Cambridge University Press