DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

The cointegrated vector autoregressive model with general deterministic terms

Søren Johansen and Morten Ørregaard Nielsen

CREATES Research Paper 2016-22

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

The cointegrated vector autoregressive model with general deterministic terms

Søren Johansen*
University of Copenhagen
and CREATES

Morten Ørregaard Nielsen†
Queens University
and CREATES

July 24, 2016

Abstract

In the cointegrated vector autoregression (CVAR) literature, deterministic terms have until now been analyzed on a case-by-case, or as-needed basis. We give a comprehensive unified treatment of deterministic terms in the additive model $X_t = \gamma Z_t + Y_t$, where Z_t belongs to a large class of deterministic regressors and Y_t is a zero-mean CVAR. We suggest an extended model that can be estimated by reduced rank regression and give a condition for when the additive and extended models are asymptotically equivalent, as well as an algorithm for deriving the additive model parameters from the extended model parameters. We derive asymptotic properties of the maximum likelihood estimators and discuss tests for rank and tests on the deterministic terms. In particular, we give conditions under which the estimators are asymptotically (mixed) Gaussian, such that associated tests are χ^2-distributed.

Keywords: Additive formulation, cointegration, deterministic terms, extended model, likelihood inference, VAR model.

JEL Classification: C32.

1 Introduction

The cointegrated vector autoregressive (CVAR) model continues to be one of the most commonly applied model in many areas of empirical economics, as well as other disciplines. However, the formulation and modeling of deterministic terms in the CVAR model has until now been analyzed on a case-by-case basis because no general treatment exists. Moreover, the role of deterministic terms is not always intuitive and is often difficult to interpret. Indeed, Hendry and Juselius (2001, p. 95) note that "In general, parameter inference, policy

*Corresponding author. Søren Johansen, Department of Economics, University of Copenhagen, Øster Farimagsgade 5, building 26,DK-1353, Copenhagen K, Denmark. Telephone: +45 35323071. Email: Soren.Johansen@econ.ku.dk

†Department of Economics, Dunning Hall, Queen's University, Kingston, Ontario K7L 3N6, Canada. Email: mon@econ.queensu.ca

simulations, and forecasting are much more sensitive to the specification of the deterministic than the stochastic components of the VAR model."

In this paper we give a comprehensive unified treatment of the CVAR model for a large class of deterministic regressors and derive the relevant asymptotic theory. There are two ways of modeling deterministic terms in the CVAR model, and we call these the additive and innovative formulations. In the additive formulation the deterministic terms are added to the process and in the innovative formulation they are added to the equations.

1.1 The additive formulation

In this paper, we analyze the additive formulation. To fix ideas, let the p-dimensional time series X_t be given by the additive model,

$$\begin{aligned}\mathcal{H}_r^{add}: \quad & X_t = Y_t + \gamma Z_t, \quad t = 1-k, \ldots, -1, 0, \ldots, T, \\ & \Pi(L) Y_t = \varepsilon_t, \quad t = 1, \ldots, T,\end{aligned} \tag{1}$$

where Z_t is a multivariate deterministic regressor and

$$\Pi(z) = (1-z) I_p - \alpha\beta' z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i \tag{2}$$

is the lag-polynomial defining the cointegrated $I(1)$ process Y_t. Furthermore, ε_t is i.i.d. $(0, \Omega)$, $Y_0, \ldots, Y_{1-k}$ are fixed initial values, and $\lambda = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1}, \gamma)$ and Ω are freely varying parameters where α, β are $p \times r$ for some $r < p$.

The advantage of the formulation in (1) is that the role of the deterministic terms for the properties of the process is explicitly modeled, and the interpretation is relatively straightforward. One can, for example, focus on the mean of the stationary processes ΔX_t and $\beta' X_t$, for which we find from (1) that

$$E(\Delta X_t) = \gamma \Delta Z_t \text{ and } E(\beta' X_t) = \beta' \gamma Z_t. \tag{3}$$

Thus, γ can be interpreted as a "growth rate", and, moreover, $\beta'\gamma$ can be more accurately estimated than the rest of γ, because the information $\sum_{t=1}^T Z_t Z_t'$ in general is larger than $\sum_{t=1}^T \Delta Z_t \Delta Z_t'$. Note that if Z_t contains the constant with parameter $\gamma_1 \in \mathbb{R}^p$, then the corresponding entry in ΔZ_t is zero and does not contain information about γ_1, and we can therefore only determine $\beta'\gamma_1$.

When analyzing properties of the process, the following $I(1)$ conditions are important, see Johansen (1996, Theorem 4.2).

Assumption 1. *The roots of* $\det \Pi(z) = 0$ *are either greater than one in absolute value or equal to 1. The matrices* α *and* β *are* $p \times r$ *of rank* r*, and for* $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$*, we assume that* $\det \alpha_\perp' \Gamma \beta_\perp \neq 0$*, such that* Y_t *is an* $I(1)$ *process,* $\beta' Y_t$ *is a stationary* $I(0)$ *process, and* $C = \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp'$ *is well defined.*

It follows from Assumption 1, specifically $\det \alpha_\perp' \Gamma \beta_\perp \neq 0$, that

$$(\beta, \Gamma'\alpha_\perp)'(\bar{\beta}, \beta_\perp) = \begin{pmatrix} I_r & 0 \\ \alpha_\perp' \Gamma \bar{\beta} & \alpha_\perp' \Gamma \beta_\perp \end{pmatrix} \tag{4}$$

has full rank, so that also $(\beta, \Gamma'\alpha_\perp)$ has full rank. Here, and throughout, for any $p \times s$ matrix a of rank $s \leq p$, we define $\bar{a} = a(a'a)^{-1}$. This full rank result is used repeatedly, in particular in the proof of Lemma 1 below, which gives an algorithm for calculating the parameters in the additive model from the parameters in the extended model.

We assume throughout that the data generating process satisfies Assumption 1, but the parameters will be assumed to be freely varying in the statistical models. For example, α and β will be freely varying $p \times r$ matrices in the statistical model but of full rank r in the data generating process.

The solution of the equations for Y_t is given by the following version of Granger's Representation Theorem, which states that

$$Y_t = C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{t-1} C_i^* \varepsilon_{t-i} + A_t, \tag{5}$$

where A_t depends on initial values of Y_t and $\beta' A_t$ decreases to zero exponentially. The representation for X_t is therefore

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{t-1} C_i^* \varepsilon_{t-i} + \gamma Z_t + A_t, \tag{6}$$

which again illustrates the explicit role of the deterministic terms in the additive formulation.

The additive formulation has been analyzed by, e.g., Lütkepohl and Saikkonen (2000a,b,c), Nielsen (2004, 2007), and Trenkler, Saikkonen, and Lütkepohl (2007); each for specific choices of deterministic terms.

1.2 The innovative formulation

The most commonly applied method of modeling deterministic terms in the cointegrated VAR model is the innovative formulation, where the regression variables are added in the equation, i.e.,

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \tilde{\gamma} Z_t + \varepsilon_t, \tag{7}$$

and the deterministic terms are possibly restricted to lie in the cointegrating space; see Johansen (1996) for a detailed treatment of the case $Z_t = (t, 1)'$ or Rahbek and Mosconi (1999) for stochastic regressors, Z_t, in the innovative formulation. They point out that the asymptotic distribution for the test for rank contains nuisance parameters, and that they can be avoided by including the cumulated Z_t as a regressor with a coefficient proportional to α. We show below that starting with the additive formulation, the highest order regressor automatically appears with a coefficient proportional to α in the innovative formulation, and we find conditions for inference to be asymptotically free of nuisance parameters.

Under Assumption 1, the $I(1)$ solution for the process X_t in (7) is given, see (5), by

$$X_t = C \sum_{i=1}^{t} (\varepsilon_i + \tilde{\gamma} Z_i) + \sum_{i=0}^{t-1} C_i^* (\varepsilon_{t-i} + \tilde{\gamma} Z_{t-i}) + A_t. \tag{8}$$

A model like (7) is easy to estimate using reduced rank regression, but it follows from (8) that the deterministic terms are generated by the dynamics of the model. We see that the deterministic term in the process is a combination of the cumulated regressors in the first term and a weighted sum of lagged regressors. Thus, for instance, an outlier dummy in the equation (7) becomes a combination of a step dummy from the first term in the process (8) and an exponentially decreasing function from the second term in (8), giving a gradual shift from one level to another. A constant in the equation (7) becomes a linear function in the process (8), see for instance Johansen (1996, Chapter 5) for a discussion of some simple models and Johansen, Mosconi, and Nielsen (2000) for a discussion of a model with broken trends and impulse dummies to eliminate a few observations just after the break. Thus, one can use the innovative formulation to model the deterministic terms in the process by taking into account the dynamics of the model.

Applications including broken trends and several types of dummy variables are also given in, for example, Doornik, Hendry, and Nielsen (1998), Hendry and Juselius (2001), Juselius (2006, 2009), and Belke and Beckmann (2015). For an application using various dummies, including a "volcanic function" dummy variable for modeling volcanic eruptions, see Model V of Pretis (2015) and also Pretis et al. (2016) for the definition of the volcanic function.

The remainder of the paper is organized as follows. In the next section we discuss the structure of the regressors, derive the extended model, and consider identification and estimation. In Section 3 we derive the asymptotic theory for the parameter estimators in both the extended and additive models, and in Section 4 we derive and discuss tests on the cointegrating rank and on the coefficients to the regressors. Finally, we conclude and give some general recommendations in Section 5. The proofs of all results are given in the appendix.

2 The regressors and the additive and extended models

Going back to the additive formulation in (1), we eliminate Y_t to find the equations for X_t,

$$\Pi(L)X_t = \Pi(L)Y_t + \Pi(L)\gamma Z_t \tag{9}$$

or

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \gamma\Delta Z_t - \alpha\beta'\gamma Z_{t-1} - \sum_{i=1}^{k-1} \Gamma_i \gamma \Delta Z_{t-i} + \varepsilon_t. \tag{10}$$

From (10) it follows that maximum likelihood estimation and inference is not so straightforward as in the model with no deterministic terms, and this is the issue we want to address in the present paper.

In the model equation (10) for X_t, the coefficients $(\gamma, -\alpha\beta'\gamma, -\Gamma_1\gamma, \ldots, -\Gamma_{k-1}\gamma)$ involve γ. These depend nonlinearly on the model parameters, so the model becomes a nonlinear restriction in the usual linear CVAR model with k lags and an innovative formulation of the deterministic terms, $(\Delta Z_t, Z_{t-1}, \Delta Z_{t-1}, \ldots, \Delta Z_{t-k+1})$.

A general technique for handling such nonlinear models consists of finding a larger model where the estimation problem is easier to handle. As a simple special example of this principle, consider a linear regression with autoregressive errors, i.e. $X_t = Y_t + \gamma Z_t$, where $Y_t = \rho Y_{t-1} + \varepsilon_t$ and ε_t is i.i.d. $(0, \sigma^2)$. The equation for X_t is $X_t = \rho X_{t-1} + \gamma Z_t - \rho\gamma Z_{t-1} + \varepsilon_t$ and maximum likelihood leads to non-linear least squares estimation. We extend the model to $X_t = \rho X_{t-1} + \gamma Z_t + \gamma_1 Z_{t-1} + \varepsilon_t$ with $\rho, \gamma, \gamma_1, \sigma^2$ freely varying. This extended statistical model can be easily estimated by (linear) least squares, and asymptotic properties of the estimators are derived under the assumption that the original (non-linear) model is the data generating process. If we are interested in the original parameters, we can choose the estimators ρ, γ from the extended model. We can use these (consistent) estimators as starting values for an iteration to the maximum likelihood estimator.

Extending model (10) in a similar way to the simple example above, leads to the problem that the regressors Z_{t-1} and ΔZ_{t-i} for $i = 0, \ldots, k-1$ may be linearly dependent. As a simple example of this, consider $Z_{t-1} = (t-1, 1)'$ with $\Delta Z_{t-i} = (1, 0)'$ for $i \geq 0$, which are clearly linearly dependent. Such a linear dependence between the regressors has to be avoided before the parameters can be estimated and the properties of the estimators derived. We therefore first discuss a formulation of the regressors that allows an analysis of the additive model and its extension.

2.1 A formulation of a class of regressors

If $U_t \in \mathbb{R}$ has the property that it is linearly dependent on some of its differences, $\sum_{i=0}^{n} c_i \Delta^i U_t = 0$ for all t, say, then U_t is the solution to a linear difference equation. A basis for the solution of such an equation is of the form $a^t \sum_{i=0}^{p} a_i t^i$, where a is a root of multiplicity $p+1$ of $\sum_{i=0}^{n} c_i a^i = 0$, see Miller (1968). For $a = 1$ we therefore get a polynomial, for $a = -1$ and $p = 0$ we get a seasonal (semi-annual) dummy $(-1)^t$, and for $a = \pm \mathsf{i}, \mathsf{i} = \sqrt{-1}$, we can find quarterly dummies. We do not deal with exponential regressors $Z_t = a^t, |a| > 1$, because the asymptotic theory is different since the Central Limit Theorem does not apply to sums of the form $\sum_{t=1}^{T} \varepsilon_t a^t$ for $|a| > 1$.

Thus, in the following we consider all regressors that are linearly independent on their differences, but for regressors that are linearly dependent on their differences we only consider a polynomial and a seasonal dummy. We note specifically that for $U_t = (-1)^t$ we have $\Delta U_t = -2U_t = M_2 U_t$, say, and for the quarterly dummy $U_{1t} = \mathsf{i}^t + (-1)^t + \mathsf{i}^{-t}$ (also orthogonalized on the constant) we find for $U_t = (U_{1t}, U_{1,t-1}, U_{1,t-2})' \in \mathbb{R}^3$ that $\Delta U_t = M_4 U_t$, where

$$M_4 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}. \qquad (11)$$

This matrix can be diagonalized and has eigenvalues $\omega_j, j = 1, 2, 3$, which are such that $1 - \omega_j$ is a seasonal unit root; that is $(-1, \pm \mathsf{i})$.

For a general regressor we define its order as follows.

Definition 1. *For a regressor $U_t \in \mathbb{R}$ we define the information as $\sum_{t=1}^{T} U_t^2$. If the information of U_t diverges, we define the order of U_t as the smallest integer $i \geq -1$ for which the information of $\Delta^{i+1} U_t$ is bounded, i.e.*

$$m = \inf\{i \geq -1 : \sum_{t=1}^{T} (\Delta^{i+1} U_t)^2 \to c < \infty \text{ as } T \to \infty\}.$$

Thus, if the information of U_t is bounded, $\lim_{T\to\infty} \sum_{t=1}^{T} U_t^2 < \infty$, we define the order to be $m = -1$, and if the information of $\Delta^i U_t$ diverges for all i we define the order to be ∞.

Example 1. For the impulse dummy $U_t = 1_{\{t=t_0\}}$, where $1_{\{A\}}$ denotes the indicator function for the event A, we find $\sum_{t=1}^T U_t^2 = 1$ so that $m = -1$. On the other hand, when the information for $\Delta^i U_t$ diverges, which is important for proving tightness of the coefficient of $\Delta^i U_t$, then $i \leq m$. For a polynomial we find that the order is the degree of the polynomial. More generally, for the power function t^a, with $a \in \mathbb{R}$ and $a > -1/2$, the order is $m = [a + 1/2]$, where $[x]$ denotes the integer part of x. For the broken linear trend $U_t = (t - t_0)^+$, with $x^+ = \max\{0, x\}$, we see that all differences are linearly independent, but because $\Delta U_t = 1_{\{t \geq t_0+1\}}$ satisfies $\sum_{t=1}^T (\Delta U_t)^2 \to \infty$ and $\sum_{t=1}^T (\Delta^2 U_t)^2 = \sum_{t=1}^T 1_{\{t=t_0+1\}} = 1$, the order of U_t in this case is $m = 1$. Finally, for a seasonal dummy variable like $U_t = (-1)^t$ it is seen that $\Delta^{i+1} U_t = (-2)^{i+1}(-1)^t$, so the information diverges for all $i \geq -1$, and the order is infinite. ♦

The regressors considered are conveniently expressed in differences (rather than lags) since these have natural interpretations in many cases. Furthermore, as the examples suggest, the sums of squares of differences of the regressors will typically have different orders of magnitude, and hence different normalizations. We therefore define the structure of regressors in terms of differences.

Definition 2. *Let $U_t = (U_{1t}, \ldots, U_{qt})' \in \mathbb{R}^q$ be a set of linearly independent regressors of orders $m_v < \infty$, $v = 1, \ldots, q$. Assume further that $\{\Delta^i U_{vt}, i \geq 0\}$ are either linearly independent or (for a polynomial) equal to zero for $i > m_v$. Let $U_{se,t} \in \mathbb{R}^{s-1}$ be an $(s-1)$-dimensional seasonal dummy variable orthogonalized to the constant term, which is such that $\Delta U_{se,t} = M_s U_{se,t}$, where M_s has eigenvalues $\{\omega_j, j = 1, \ldots, s-1\}$ such that $1 - \omega_j$ is a seasonal unit root. We consider the regressor defined as*

$$Z_t = (U_t', \Delta U_t', \ldots, \Delta^n U_t', U_{se,t}')',$$

which is of dimension $(n+1)q + s - 1$. We decompose γ correspondingly,

$$\gamma = (\gamma^0, \ldots, \gamma^n, \gamma^{se}), \ \gamma^i = (\gamma_1^i, \ldots, \gamma_q^i), \ i = 0, \ldots, n,$$

such that

$$\gamma Z_t = \sum_{i=0}^n \gamma^i \Delta^i U_t + \gamma^{se} U_{se,t} = \sum_{v=1}^q \sum_{i=0}^n \gamma_v^i \Delta^i U_{vt} + \gamma^{se} U_{se,t}.$$

It is important to note that some of the components of Z_t may be zero (if a polynomial is differenced too many times), or more generally have bounded information if the order of the component is less than n.

2.2 Some reparametrizations of the additive model

To express the deterministic term in the additive model in terms of differences of U_t, we expand $\Pi(z)$ around $z = 1$ and find the coefficients

$$\Pi(z) = \Phi_0 + \Phi_1(1-z) + \cdots + \Phi_k(1-z)^k, \quad \Phi_i = (-1)^i D_z^i \Pi(z)|_{z=1}/i!,$$

where Φ_i are functions of the parameters; in particular, see (1),

$$\Phi_0 = -\alpha\beta', \qquad \Phi_1 = -\alpha\beta' - (I_p - \sum_{i=1}^{k-1} \Gamma_i) = -\alpha\beta' - \Gamma. \tag{12}$$

We then find the deterministic term in the additive model equation, see (2),

$$\Pi(L)\gamma Z_t = \sum_{i=0}^{k} \Phi_i (\sum_{j=0}^{n} \gamma^j \Delta^{i+j} U_t + \gamma^{se} \Delta^i U_{se,t}) = \sum_{i=0}^{n+k} \Upsilon_i \Delta^i U_t + \Upsilon_{se} U_{se,t}, \tag{13}$$

where we have introduced the coefficient $\Upsilon = (\Upsilon_0, \dots, \Upsilon_{n+k}, \Upsilon_{se})$ given by

$$\Upsilon_i = \sum_{j=\max\{0,i-n\}}^{\min\{i,k\}} \Phi_j \gamma^{i-j},\ i = 0, \dots, n, \quad \Upsilon_{se} = \sum_{i=0}^{k} \Phi_i \gamma^{se} M_s^i. \tag{14}$$

It is clear from (14) that, for given values of the dynamic parameters $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1})$, the parameter Υ is a linear function of the parameter γ. In Lemma 1 we next give an algorithm for recovering the parameter γ as a linear function of the parameter Υ, also for given values of the dynamic parameters $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1})$.

Lemma 1. *Let Assumption 1 be satisfied and define Υ as in (14). Then, for $i = 0, \dots, n$,*

$$\bar{\alpha}'\Upsilon_i = -\beta'\gamma^i + \bar{\alpha}' \sum_{j=1}^{\min\{i,k\}} \Phi_j \gamma^{i-j} \text{ and } \alpha_\perp' \Upsilon_{i+1} = -\alpha_\perp' \Gamma \gamma^i + \alpha_\perp' \sum_{j=2}^{\min\{i+1,k\}} \Phi_j \gamma^{i+1-j}. \tag{15}$$

Thus, because the matrix $(\beta, \Gamma'\alpha_\perp)$ has full rank, see (4), the parameters $\gamma^0, \dots, \gamma^n$ can be recovered recursively as linear functions of $(\bar{\alpha}'\Upsilon_0, \Upsilon_1, \dots, \Upsilon_n, \alpha_\perp'\Upsilon_{n+1})$ for given values of $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1})$.

The coefficient γ^{se} is uniquely determined as a linear function of $\Upsilon_{se} = \sum_{i=0}^{k} \Phi_i \gamma^{se} M_s^i$,

$$\mathrm{vec}(\gamma^{se}) = (\sum_{i=0}^{k} M_s^i \otimes \Phi_i)^{-1} \mathrm{vec}(\Upsilon_{se}). \tag{16}$$

2.3 The extended model

We define the extended model based on the results in the previous subsection and the coefficients in (14). We note in particular that Υ_0 is proportional to α, and define the parameter $\rho' = \bar{\alpha}'\Upsilon_0 = \bar{\alpha}'\Phi_0\gamma^0 = -\beta'\gamma^0$, such that the extended model is, see also (13),

$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' U_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{n+k} \Upsilon_i \Delta^i U_t + \Upsilon_{se} U_{se,t} + \varepsilon_t, \tag{17}$$

where the parameters $\xi = (\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1}, \rho, \Upsilon_1, \dots, \Upsilon_{n+k}, \Upsilon_{se})$ and Ω are freely varying.

The additive model $\mathcal{H}_r^{add}$ in (10) is now expressed as the submodel of the extended model $\mathcal{H}_r^{ext}$ in (17), where the restrictions (14) and $\rho' = -\beta'\gamma^0$ give the extended model parameter, ξ, as a function of the additive model parameter, $\lambda = (\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1}, \gamma)$.

In general the additive model is a submodel of the extended model, but there is a special case where the two models are the same, as given in the next theorem. Define the polynomials $f_i(t) = t(t-1)\cdots(t-i+1)/i!$, which satisfy $\Delta f_i(t) = f_{i-1}(t)$. The regressor $Z_t = (t^m, \dots, 1)'$ is equivalent to the regressor $Z_t = (f_m(t), \dots, f_0(t))'$ in the sense that they span the same space. For $m = 0$ and $m = 1$ the models with these regressors were denoted $H_0^*(r)$ and $H_1^*(r)$, respectively, in Johansen (1996).

Theorem 1. *Let Assumption 1 be satisfied. Then the additive model for the regressor* $Z_t = (f_m(t), \ldots, f_0(t))'$ *is a reparametrization of the extended model.*

Note that the result in Theorem 1 also holds if Z_t is extended with a seasonal dummy like $(-1)^t$. In general, of course, the simple result in Theorem 1 does not hold, so that the additive model is not a reparametrization of the extended model. For the general case, we next discuss identification and estimation of the parameters in the situation where we allow a polynomial regressor U_{1t} of order m_1, say, in the additive model and have removed zero regressors.

2.4 Identification of the parameters in the extended and additive models

For identification of the parameters in the extended model (17), the zero regressors $\Delta^i U_{1t} = 0$, $i > m_1$, have been removed together with their coefficients, so that the remaining regressors are linearly independent (Definition 2). Then the coefficient ξ is identified, because if the likelihood functions for parameters ξ_1 and ξ_2 are the same, then $\xi_1 = \xi_2$, except for α and β, where only their product is identified. A convenient normalization to identify β, see Johansen (1996, p. 179), is to assume that $\beta'\bar{\beta}_0 = I_r$. This will be assumed throughout.

We next consider identification of the additive model (10) as a submodel of the extended model (17). This is a consequence of the following result, which is based on Lemma 1. The result is formulated for the additive model with a polynomial regressor, which may generate zero regressors.

Theorem 2. *Let Assumption 1 be satisfied. Let* $\lambda = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1}, \gamma)$ *be the parameters in the additive model (10), which contains a polynomial* $P_t = U_{1t}$*, say, of order* m_1*, and assume that the regressors* $\Delta^i U_{1t} = 0$*,* $i > m_1$*, have been removed together with their coefficients* γ_1^i*. Let* $\xi = \xi(\lambda) = (\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1}, \rho, \Upsilon_1, \ldots, \Upsilon_{n+k}, \Upsilon_{se})$*, where* $\Upsilon_0, \ldots, \Upsilon_{n+k}, \Upsilon_{se}$ *are defined by (14) and* $\rho' = \bar{\alpha}'\Upsilon_0$*, and assume the coefficients* $\Upsilon_{i1}, i > m_1$*, have been removed. Then, for any set of parameters* λ_0*,* λ_h*,* $h \to 0$*, we find*

$$\xi(\lambda_h) \to \xi(\lambda_0) \text{ as } h \to 0 \text{ implies } \lambda_h \to \lambda_0, \tag{18}$$

except if $n \geq m_1$*, where for the constant term with coefficient* $\gamma_{1,h}^{m_1}$*, we only find* $\beta'\gamma_{1,h}^{m_1} \to \beta'\gamma_{1,0}^{m_1}$*.*

Identification of the additive model as a submodel of the extended model follows from Theorem 2 because if $\xi(\lambda_1) = \xi(\lambda_0)$ then, choosing $\lambda_h = \lambda_1$, we find from (18) that $\lambda_1 = \lambda_0$. Thus, a special case of Theorem 2 implies identification of the parameters of the additive model in the usual sense. However, in anticipation of our proof of consistency, Theorem 2 proves the more general result that ξ depends continuously on the parameter λ, which one could call "continuous identification". The result in Theorem 2 shows continuous identification of γ, with the exception that, if $n \geq m_1$ (so that the constant term, $\Delta^{m_1} P_t = \Delta^{m_1} U_{1t}$, is included in the model), then the coefficient to the constant term is only identified in the β-directions.

2.5 Estimation of the parameters in the extended and additive models

For estimation of the extended model (17), we continue to assume that the zero regressors $\Delta^i U_{1t} = 0$, $i > m_1$, have been removed together with their coefficients, so that the remaining regressors are linearly independent. Then maximum likelihood estimation of the parameters of the extended model can be conducted by reduced rank regression of ΔX_t on $(X'_{t-1}, U'_t)'$ corrected for the non-zero regressors. See Anderson (1951) and Johansen (1996, Chapter 6).

Next, the additive model (10) can be estimated by maximum likelihood using an optimizing algorithm, as a submodel of the extended model subject to the restrictions (14). Starting values for the iterations in the numerical optimization of the likelihood function can be found, using Lemma 1, from parameter estimates of the extended model.

3 Asymptotic theory for parameter estimators

We first give some conditions on the regressors which are needed for the asymptotic analysis. We then discuss consistency of the parameter estimators and find their asymptotic distributions for both the extended and additive models.

3.1 Normalization and partition of regressors

We introduce the notation for product moments of sequences $U_t, V_t, W_t, t = 1, \ldots, T$,

$$\langle U, V \rangle_T = T^{-1} \sum_{t=1}^{T} U_t V_t',$$

and for residuals

$$(U_t | V_t) = U_t - \langle U, V \rangle_T \langle V, V \rangle_T^{-1} V_t$$

and conditional product moments

$$\langle U, V | W \rangle_T = \langle U, V \rangle_T - \langle U, W \rangle_T \langle W, W \rangle_T^{-1} \langle W, V \rangle_T .$$

When the limit of a product moment exists, we use the notation $\langle U, V \rangle_T \to \langle U, V \rangle$, for example, to denote the limit as $T \to \infty$.

For the asymptotic analysis, regressors with bounded information will not give consistent estimation of their associated coefficients. That is, for any deterministic term U_{vt} with order m_v, the coefficients to the regressors $\Delta^i U_{vt}, i > m_v$, cannot be consistently estimated because $\sum_{t=1}^{T} (\Delta^i U_{vt})^2$ is bounded, see Definition 1. However, as shown below, this has no influence on asymptotic inference for the remaining parameters.

To conduct asymptotic inference, we thus partition the regressors into those with divergent and those with bounded information, respectively, and for the former we also separate those that are proportional to α in the extended model (17). These regressors and their associated coefficients are defined next.

Definition 3. *The non-zero regressors in $\Delta^i U_{vt}, 1 \leq v \leq q, 0 \leq i \leq n + k$, are partitioned as*

$$Z_{0t} = (U'_{vt}, 0 \leq m_v)',$$
$$Z_{1t} = (\Delta^i U'_{vt}, 1 \leq i \leq \min\{n + k, m_v\}; U'_{se,t})',$$
$$Z_{2t} = (U'_{vt}, m_v < 0; \Delta^i U'_{vt}, m_v < i \leq n + k)',$$

with coefficients given by

$$\begin{aligned}\rho^0 &= (\rho_v, m_v \geq 0),\\ \Upsilon^1 &= (\Upsilon_{iv}, 1 \leq i \leq \min\{n+k, m_v\}; \Upsilon_{se}),\\ \Upsilon^2 &= (\alpha\rho_v, m_v < 0; \Upsilon_{iv}, m_v < i \leq n+k).\end{aligned}$$

Similarly to ρ^0 *we define* $\gamma^{0*} = (\gamma_v^0, m_v \geq 0)$ *such that* $\rho^{0\prime} = -\beta'\gamma^{0*}$.

According to Definition 3, we have partitioned the regressors such that Z_{0t} and Z_{1t} have divergent information and Z_{2t} has bounded information, see Definitions 1 and 2, and we note that Z_{jt} may be empty in which case the remainder of the paper is simplified accordingly. With the notation in Definition 3 we find that the deterministic terms in the extended model (17) can be reparametrized as

$$\alpha\rho' U_t + \sum_{i=1}^{n+k} \Upsilon_i \Delta^i U_t + \Upsilon_{se} U_{se,t} = \alpha\rho^{0\prime} Z_{0t} + \Upsilon^1 Z_{1t} + \Upsilon^2 Z_{2t}. \tag{19}$$

Finally, for the asymptotic analysis we need the following normalizations and a mild condition to rule out asymptotically multicollinear regressors.

Assumption 2. *For a regressor* $\Delta^i U_{vt}$, *for which* $m_v \geq 0$, *there exists normalizations* M_{Tiv} *for* $i = 0, \ldots, m_v$, *satisfying* $M_{Tiv} M_{T,i+1,v}^{-1} \to 0$ *and* $M_{Tiv} T^{-1/2} \to 0$, *and for which the normalized regressors* $\Delta^i U_{vTt} = M_{Tiv} \Delta^i U_{vt}$ *satisfy that*

$$\left\langle \Delta^i U_{vT}, \Delta^j U_{vT} \right\rangle_T = T^{-1} \sum_{i=1}^{T} (M_{Tiv} \Delta^i U_{vt})(M_{Tjv} \Delta^j U_{vt})'$$

is convergent.

Corresponding to Z_{0t} and Z_{1t}, we collect their normalizations in the diagonal matrices $N_{T0} = \mathrm{diag}(M_{T0v}, m_v \geq 0, 1 \leq v \leq q)$ and $N_{T1} = \mathrm{diag}(M_{Tiv}, 1 \leq i \leq \min\{n+k, m_v\}, 1 \leq v \leq q, \iota_{s-1}')$, where ι_{s-1} is an $(s-1)$-vector of ones. This defines the normalized regressors

$$Z_{0Tt} = N_{T0} Z_{0t} \text{ and } Z_{1Tt} = N_{T1} Z_{1t}. \tag{20}$$

Assumption 3. *The asymptotic information matrix for* $(Z_{0Tt}', Z_{1Tt}')'$ *is nonsingular, i.e. satisfies*

$$\langle (Z_{0T}', Z_{1T}')', (Z_{0T}', Z_{1T}')' \rangle_T \to \langle (Z_0', Z_1')', (Z_0', Z_1')' \rangle > 0.$$

Example 2. The nonsingularity condition in Assumption 3 rules out asymptotically multicollinear regressors, and is easily satisfied in practice. As an example of what is ruled out, consider the regressor $U_t = (1 + 1_{\{t=t_0+1\}}, 1 + 1_{\{t=t_0-1\}})'$, which satisfies Definition 2 with $m_1 = m_2 = 0$, but the information $\langle U, U \rangle_T \to \langle U, U \rangle = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is clearly singular in the limit and thus violates Assumption 3. In this case, one could apply instead $U_t = (1 + 1_{\{t=t_0+1\}}, 1_{\{t=t_0-1\}} - 1_{\{t=t_0+1\}})'$, which spans the same space, but where the information is $\langle U, U \rangle_T \to \langle U, U \rangle = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and the first element gives rise to consistent estimation with a non-singular asymptotic information matrix. Thus, for this example, we set $Z_{0t} = 1 + 1_{\{t=t_0+1\}}$ and $Z_{2t} = 1_{\{t=t_0-1\}} - 1_{\{t=t_0+1\}}$. ♦

Finally, for the asymptotic analysis we make the following high-level assumption, for which primitive sufficient conditions are well-known.

Assumption 4. *We assume that ε_t is i.i.d. $(0,\Omega)$, and for $S_t = \sum_{i=1}^{t}\varepsilon_i$ we have the weak limit $T^{-1/2}S_{[Tu]} \overset{D}{\to} W_\varepsilon$, where W_ε denotes Brownian motion generated by ε_t. Furthermore, the following limits exist and the convergences hold jointly,*

$$T^{1/2}\left\langle Z_{jT},\varepsilon\right\rangle_T = T^{-1/2}N_{Tj}\sum_{t=1}^{T} Z_{jt}\varepsilon_t' \overset{D}{\to} \left\langle Z_j,\varepsilon\right\rangle \text{ for } j=0,1,$$

$$T^{-1/2}\left\langle Z_{jT},S_{t-1}\right\rangle_T = T^{-3/2}N_{Tj}\sum_{t=1}^{T} Z_{jt}S_{t-1} \overset{D}{\to} \left\langle Z_j,W_\varepsilon\right\rangle \text{ for } j=0,1,$$

$$T^{-1/2}\left\langle S_{t-1},\varepsilon\right\rangle_T = T^{-3/2}\sum_{t=1}^{T} S_{t-1}\varepsilon_t' \overset{D}{\to} \int_0^1 W_\varepsilon(dW_\varepsilon)' = \left\langle W_\varepsilon,\varepsilon\right\rangle.$$

Again, we use $\left\langle Z_j,\varepsilon\right\rangle$, for example, as the notation for the limit of a product moment, because simple expressions in terms of stochastic integrals are not possible for all regressors. Examples of the conditions in Assumption 4 are given next.

Example 3. Let $U_t = (t,(t-[Tv_0])^+)'$ with $\Delta U_t = (1, 1_{\{t\geq [Tv_0]+1\}})'$. Then $M_{T0} = T^{-1}$ and $M_{T1} = 1$, and we note that $M_{T0}M_{T1}^{-1} \to 0$, reflecting that the order of the regressor in this case decreases when differenced. We define $u(v) = \lim_{T\to\infty} U_{T,[Tv]} = \lim_{T\to\infty} M_{T0}U_{[Tv]} = (v,(v-v_0)^+)'$ and $\dot{u}(v) = \lim_{T\to\infty}\Delta U_{T,[Tv]} = \lim_{T\to\infty} M_{T1}\Delta U_{[Tv]} = (1, 1_{\{v\geq v_0\}})'$. For this example we find the limits

$$\left\langle U_T,\Delta U_T\right\rangle_T = T^{-1}\sum_{t=1}^{T}(T^{-1}U_t)(\Delta U_t)' \to \int_0^1 u(v)\dot{u}(v)'dv = \left\langle U,\Delta U\right\rangle,$$

$$T^{1/2}\left\langle U_T,\varepsilon\right\rangle_T = T^{-1/2}\sum_{t=1}^{T} T^{-1}U_t\varepsilon_t' \overset{D}{\to} \int_0^1 u(v)dW_\varepsilon(v)' = \left\langle U,\varepsilon\right\rangle,$$

$$T^{-1/2}\left\langle U_T,S_{t-1}\right\rangle_T = T^{-3/2}\sum_{t=1}^{T} T^{-1}U_tS_{t-1}' \overset{D}{\to} \int_0^1 u(v)W_\varepsilon(v)'dv = \left\langle U,W_\varepsilon\right\rangle.$$

♦

The previous example illustrates a relatively simple regressor, which when appropriately normalized has a limit, $u(v)$, in L_2. In this case, the limit of the product moment $T^{1/2}\left\langle U_T,\varepsilon\right\rangle_T$, for example, can be expressed as a stochastic integral of $u(v)$ with respect to Brownian motion, W_ε. However, such simple limit expressions are not always possible, as the following example shows.

Example 4. Let $U_{se,t} = (-1)^t$ be a seasonal dummy variable. Then

$$T^{1/2}\left\langle U_{se},\varepsilon\right\rangle_T = T^{-1/2}\sum_{t=1}^{T} U_{se,t}\varepsilon_t' \overset{D}{\to} N(0,\Omega) = \left\langle U_{se},\varepsilon\right\rangle,$$

$$T^{-1/2}\left\langle S_{t-1},U_{se}\right\rangle_T = T^{-3/2}\sum_{t=1}^{T} S_{t-1}U_{se,t} = O_P(T^{-1}),$$

where we note that $\langle U_{se}, \varepsilon \rangle$ is not a stochastic integral involving a limit of $U_{se,t}$ because $U_{se,t}$ does not converge in L_2. ♦

3.2 Consistency of parameter estimators

To prove consistency, we use the fact that both the additive model (10) and the extended model (17) can be expressed as nonlinear submodels of a linear regression model, such that we can apply the following result from Johansen (2006).

Lemma 2. *Let $\varsigma = \varsigma(\tau)$ be a continuously identified parametrization of a submodel of the regression model $y_t = \varsigma' z_t + \varepsilon_t$, ε_t i.i.d. $(0, \Omega)$, with stochastic or deterministic regressors. Assume that the information diverges in probability for all components of z_t,*

$$P\left(\omega_{\min}(\sum_{t=1}^{T} z_t z_t') > A\right) \to 1 \text{ for all } A > 0 \text{ as } T \to \infty,$$

where $\omega_{\min}(\cdot)$ denotes the smallest eigenvalue of the argument. Then $\hat{\tau}$ exists with probability converging to one and is consistent, as $T \to \infty$.

Consistency of the continuously identified parameters in both the additive and extended models thus follows from Theorem 2 and Lemma 2 for those regressors that have divergent information. That is, in the extended model, we cannot obtain consistency for $\hat{\Upsilon}^2$, but the remaining parameters are consistently estimated, as formulated in the next result.

Theorem 3. *Suppose Assumptions 1–3 are satisfied. Then the maximum likelihood estimators exist in both the extended model and the additive model, with probability converging to one, and both are consistent for the regressors with divergent information.*

3.3 Asymptotic distribution of the parameters of the extended model

We apply the Gaussian likelihood function and let ξ denote all parameters in the conditional mean of the extended model, see (17), with true value ξ_0. The normalized (negative) log-likelihood function for model $\mathcal{H}_r^{ext}$ is

$$L(\xi, \Omega) = -T^{-1} \log L_T(\xi, \Omega) = \frac{1}{2} \log\det(\Omega) + \frac{1}{2} \operatorname{tr}\{\Omega^{-1} T^{-1} \sum_{t=1}^{T} \varepsilon_t(\xi) \varepsilon_t(\xi)'\}, \tag{21}$$

where

$$\varepsilon_t(\xi) = \Delta X_t - \alpha(\beta' X_{t-1} + \rho^{0\prime} Z_{0t}) - \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} - \Upsilon^1 Z_{1t} - \Upsilon^2 Z_{2t}. \tag{22}$$

We normalize $\beta' \bar{\beta}_0 = I_r$ and use the decomposition $\beta = \beta_0 + \beta_{0\perp} \bar{\beta}_{0\perp}' (\beta - \beta_0)$. We then stack $T^{-1/2} \beta_{0\perp}' Y_{t-1}$ and $Z_{0Tt} = N_{T0} Z_{0t}$, see (20), as

$$G_{Tt} = \begin{pmatrix} T^{-1/2} \beta_{0\perp}' Y_{t-1} \\ Z_{0Tt} \end{pmatrix},$$

and define the variance $\Sigma_{stat} = Var(Y_{t-1}' \beta_0, \Delta Y_{t-1}', \ldots \Delta Y_{t-k+1}')'$.

Theorem 4. *Suppose Assumptions 1–4 are satisfied. Then it holds jointly that*

$$T^{1/2}(\hat{\alpha}-\alpha_0,\hat{\Gamma}_1-\Gamma_{1,0},\dots,\hat{\Gamma}_{k-1}-\Gamma_{k-1,0}) \xrightarrow{D} N_{p\times r}\left(0,\Sigma_{stat}^{-1}\otimes\Omega_0\right) \tag{23}$$

$$T^{1/2}\begin{pmatrix} T^{1/2}\bar{\beta}_{0\perp}'(\hat{\beta}-\beta_0) \\ N_{T0}^{-1}(\hat{\rho}^0+\gamma_0^{0*\prime}\hat{\beta}) \end{pmatrix} \xrightarrow{D} -\left\langle G,G|Z_1\right\rangle^{-1}\left\langle G,\varepsilon_\alpha|Z_1\right\rangle, \tag{24}$$

$$T^{1/2}(\hat{\Upsilon}^1-\Upsilon_0^1)N_{T1}^{-1} \xrightarrow{D} -\left\langle \varepsilon,Z_1\right\rangle\left\langle Z_1,Z_1\right\rangle^{-1}+\alpha_0\left\langle \varepsilon_\alpha,G|Z_1\right\rangle\left\langle G,G|Z_1\right\rangle^{-1}\left\langle G,Z_1\right\rangle\left\langle Z_1,Z_1\right\rangle^{-1}, \tag{25}$$

where $\rho^0=-\gamma^{0*\prime}\hat{\beta}$ *and* $\varepsilon_{\alpha,t}=(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\varepsilon_t$*. Furthermore, the distribution (23) is asymptotically independent of the distributions (24) and (25).*

As discussed above, only the parameter Υ^1 appears in Theorem 4. The parameter Υ^2 corresponds to the regressors with bounded information, and cannot be consistently estimated (Theorem 3) and hence has no place in Theorem 4.

The distribution in (24) is mixed Gaussian (MG), and an important consequence is that asymptotic inference on β can be conducted using the χ^2-distribution. However, the distribution in (25) is not MG, although we can obtain Gaussianity for some linear combinations of $(\hat{\Upsilon}^1-\Upsilon_0^1)$ by pre-multiplication by $\alpha_{0\perp}'$.

3.4 Asymptotic distribution of the parameters of the additive model

In order to derive the simple result that the additive model and the extended model are asymptotically equivalent, because the only difference is in some regressors with bounded information, we make the next assumption. We show below that the result for polynomials in Theorem 1 is asymptotically satisfied for a general additive model, if enough regressors are included in the formulation of the additive model.

Assumption 5. *The number, n, of differences $\Delta^i U_t$ in Z_t, see Definition 2, satisfies $n\geq m_v$ for $v=1,\dots,q$.*

For any parameter θ, let the maximum likelihood estimator in the additive model be denoted by $\breve{\theta}$.

Theorem 5. *Suppose Assumptions 1–5 are satisfied. Then it holds jointly that*

$$T^{1/2}(\breve{\alpha}-\alpha_0,\breve{\Gamma}_1-\Gamma_{1,0},\dots,\breve{\Gamma}_{k-1}-\Gamma_{k-1,0}) \xrightarrow{D} N_{p\times r}\left(0,\Sigma_{stat}^{-1}\otimes\Omega_0\right), \tag{26}$$

$$T^{1/2}\begin{pmatrix} T^{1/2}\bar{\beta}_{0\perp}'(\breve{\beta}-\beta_0) \\ -N_{T0}^{-1}(\breve{\gamma}^{0*}-\gamma_0^{0*})'\breve{\beta} \end{pmatrix} \xrightarrow{D} -\left\langle G,G|Z_1\right\rangle^{-1}\left\langle G,\varepsilon_\alpha|Z_1\right\rangle, \tag{27}$$

where $\varepsilon_\alpha=(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\varepsilon_t$*, see (23) and (24) in Theorem 4. For $v=1,\dots,q$ we express the asymptotic distributions of $\breve{\gamma}_v^i$ in terms of the maximum likelihood estimators of the parameters in the extended model,*

$$T^{1/2}\beta_0'(\breve{\gamma}_v^i-\gamma_{v0}^i)M_{Tiv}^{-1}=-T^{1/2}\bar{\alpha}_0'(\hat{\Upsilon}_{iv}-\Upsilon_{iv0})M_{Tiv}^{-1}+o_P(1),\ 1\leq i\leq m_v, \tag{28}$$

$$T^{1/2}\alpha_{0\perp}'\Gamma_0(\breve{\gamma}_v^i-\gamma_{v0}^i)M_{T,i+1,v}^{-1}=-T^{1/2}\alpha_{0\perp}'(\hat{\Upsilon}_{i+1,v}-\Upsilon_{i+1,v,0})M_{T,i+1,v}^{-1}+o_P(1),\ 1\leq i<m_v, \tag{29}$$

$$T^{1/2}\operatorname{vec}(\breve{\gamma}^{se}-\gamma_0^{se})=\Big(\sum_{i=0}^{k}M_s^{i\prime}\otimes\Phi_i\Big)^{-1}T^{1/2}\operatorname{vec}(\hat{\Upsilon}_{se}-\Upsilon_{se,0}). \tag{30}$$

Because all $\hat{\Upsilon}_{iv}$ and $\hat{\Upsilon}_{se}$ coefficients on the right-hand sides of (28)–(30) are included in $\hat{\Upsilon}^1$, their asymptotic distributions are given in (25) in Theorem 4. Finally, the distribution in (26) is asymptotically independent of the distributions (27)–(30).

The main result in Theorem 5 is that the simple condition in Assumption 5 implies that the additive and extended models are (asymptotically) equivalent. The consequence is that, under Assumption 5, asymptotic inference is identical in the two models.

3.5 Asymptotic distributions when $m > n$

If the condition that $\max_{1\leq v\leq q} m_v \leq n$ in Assumption 5 is violated, inference becomes much more involved. To simplify, we consider the additive model for a single regressor $U_t \in \mathbb{R}$ with order m. In particular, we give the proof for the case $m = 1, n = 0, k = 1$, and use the notation γ instead of γ^{0*}; that is, we consider the models

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha(\beta' X_{t-1} - \beta'\gamma U_{t-1}) + \gamma\Delta U_t + \varepsilon_t, \tag{31}$$

$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} - \rho U_{t-1}) + \upsilon\Delta U_t + \varepsilon_t. \tag{32}$$

The general case follows similarly, but with more complicated notation.

As an illustration, consider the following example.

Example 5. Consider the model $X_t = Y_t + \gamma(t - t_0)^+$ and $\Delta Y_t = \alpha\beta' Y_{t-1} + \varepsilon_t$, where the innovative formulation of the additive model and the associated extended model are, compare also (31) and (32),

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha(\beta' X_{t-1} - \beta'\gamma(t - t_0 - 1)^+) + \gamma 1_{\{t\geq t_0+1\}} + \varepsilon_t,$$
$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho(t - t_0 - 1)^+) + \upsilon 1_{\{t\geq t_0+1\}} + \varepsilon_t.$$

Thus, the extended model has two parameters in the deterministic term, ρ and υ, both of which can be consistently estimated, whereas the additive model has only one parameter, γ. Obviously the two models are not asymptotically equivalent, and this is an example of a case where $n = 0$, but $m = m_1 = 1$, and hence Assumption 5 is not satisfied. ♦

We now consider the asymptotic distributions when $m > n$, i.e., when Assumption 5 is violated. The asymptotic theory for the extended model in Theorem 4 covers case of $m > n$, but the theory for the additive model in Theorem 5 does not. In the next theorem, we compare inference in the two models.

Theorem 6. *Suppose Assumptions 1–4 are satisfied, but Assumption 5 is violated. For the additive model, the asymptotic distribution for $\breve{\alpha}, \breve{\Gamma}_1, \ldots, \breve{\Gamma}_{k-1}$ in (26) continues to hold, but the asymptotic distribution of*

$$\breve{\zeta} = T^{1/2} \begin{pmatrix} T^{1/2}\bar{\beta}'_{0\perp}(\breve{\beta} - \beta_0) \\ -M_{T0}^{-1}(\breve{\gamma} - \gamma_0)'\breve{\beta} \end{pmatrix},$$

or any linear combination of it, is not mixed Gaussian. Furthermore, the asymptotic distribution $\alpha'_{0\perp}(\breve{\gamma} - \gamma_0)M_{T1}^{-1}$ is neither asymptotically Gaussian nor mixed Gaussian and the same holds for any linear combination of it.

Finally, the asymptotic information matrix for ζ in the extended model is larger than the asymptotic information matrix for ζ in the additive model, in the sense that the difference is positive definite.

Note that when $n < m$, inference for $\breve{\alpha}, \breve{\Gamma}_1, \ldots, \breve{\Gamma}_{k-1}$ in the additive model is asymptotically the same as for $n \geq m$. This can be explained by the block-diagonality of the information matrix for the parameters $(\alpha, \Gamma_1, \ldots, \Gamma_{k-1})$ and the remaining parameters, such that inference on $(\alpha, \Gamma_1, \ldots, \Gamma_{k-1})$ can be conducted as if the remaining parameters were known.

In order to explain what happens with the regression parameters in the additive model, we decompose γ into $\beta'\gamma$ and $\alpha'_\perp\gamma$. The first parameter is estimated as the coefficient to U_{t-1}, and the contribution to $\beta'\gamma$ from the coefficient to ΔU_t is asymptotically negligible, whereas the parameter $\alpha'_\perp\gamma$ is estimated from the coefficient to ΔU_t. Thus the information in $\beta'\gamma\Delta U_t$ is not used in the additive model.

By extending the model, we replace the coefficient to ΔU_t by a freely varying parameter, and can then exploit all the information in the data. This simplifies inference with a loss of efficiency as measured by the ratio of the information matrices. More precisely, the limiting asymptotic conditional variance of the mixed Gaussian distribution of $\hat{\zeta}$ in the extended model is larger than the corresponding expression for the additive model, but the interpretation of the limit distribution is entirely different in the two models.

The difficult inference problems in the additive model could possibly be solved by an application of the bootstrap along the lines of Cavaliere, Rahbek, and Taylor (2012) and Cavaliere, Nielsen, and Rahbek (2016). However, enlarging the model to have $n \geq m$ is a simple device to achieve simple inference. The latter possibility is illustrated as follows.

Example 6. Continuation of Example 5. Note that Assumption 5 would be satisfied by including a step dummy, $1_{\{t \geq t_0+1\}} = \Delta(t - t_0)^+$, in the additive model formulation such that $X_t = Y_t + \gamma^1(t - t_0)^+ + \gamma^2 1_{\{t \geq t_0+1\}}$ giving

$$\mathcal{H}_r^{add} : \Delta X_t = \alpha(\beta' X_{t-1} - \beta'\gamma^1(t - t_0 - 1)^+ - \beta'\gamma^2 1_{\{t \geq t_0+1\}}) + \gamma^1 1_{\{t \geq t_0+1\}} + \gamma^2 1_{\{t = t_0+1\}} + \varepsilon_t,$$
$$\mathcal{H}_r^{ext} : \Delta X_t = \alpha(\beta' X_{t-1} + \rho(t - t_0 - 1)^+) + \upsilon_1 1_{\{t \geq t_0+1\}} + \upsilon_2 1_{\{t = t_0+1\}} + \varepsilon_t.$$

With this slightly larger additive model we have $n = 1$ and $m = 1$ such that Assumption 5 is satisfied. It is seen that the two models are not reparametrizations as for polynomials, see Theorem 1, but the coefficient υ_2 is associated with a regressor with information $\sum_{t=1}^T 1^2_{\{t=t_0+1\}} = 1$, and hence does not contribute to the asymptotic analysis. That is, by including the missing step dummy, $1_{\{t \geq t_0+1\}}$, in the additive model, and hence allowing the broken trend to have a discontinuity at the breakpoint, t_0, Assumption 5 is now satisfied and the two models are asymptotically equivalent. In this case the asymptotic analysis is relatively simple as shown in Theorem 5. ♦

4 Hypothesis testing

We first give the asymptotic distribution of the test for cointegration rank and then discuss tests on coefficients of deterministic terms.

4.1 Test of cointegration rank

We consider the extended model (17) for $r = p$,

$$\mathcal{H}_p^{ext} : \Delta X_t = \Pi(X_{t-1} - \gamma^{0*} Z_{0t}) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon^1 Z_{1t} + \Upsilon^2 Z_{2t} + \varepsilon_t. \tag{33}$$

The likelihood ratio test for rank r or $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices, is denoted $LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext})$. For the general class of models and deterministic terms considered, we can provide a unified result for the asymptotic distribution of the test of cointegration rank, and this is given next.

Theorem 7. *Under Assumptions 1–5, the asymptotic distribution of the test of cointegrating rank in either the extended model (17) or in the additive model (1) is given by*

$$-2\log LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext}) \xrightarrow{D} \mathrm{tr}\{\langle G, \varepsilon_{\alpha_\perp}|Z_1\rangle' \langle G, G|Z_1\rangle^{-1} \langle G, \varepsilon_{\alpha_\perp}|Z_1\rangle\},$$

where $\varepsilon_{\alpha_\perp,t} = (\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1/2}\alpha_{0\perp}'\varepsilon_t$ *is i.i.d.* $N(0, I_{p-r})$.

Note that the limit distribution of the rank test depends on the type of regressors and needs to be simulated for the various cases. However, it does not depend on the values of the regression parameters, i.e. the rank test is asymptotically similar with respect to the regression parameters, see Nielsen and Rahbek (2000). This is a consequence of starting from the additive formulation with $n \geq \max_{1\leq v\leq q} m_v$, and deriving the extended model from the additive model. In the innovative formulation (7), this is not the case, see for example the analysis of the model with an unrestricted constant term in Johansen (1996).

4.2 Tests of hypotheses on deterministic terms

We consider inference on the coefficients $\gamma_v^0, m_v \geq 0$, and $\gamma_v^i, i = 1, \ldots, m_v \leq n$, in the additive model and denote by $\breve{\gamma}_v^i$ the maximum likelihood estimator in the additive model. It follows from Theorems 4 and 5 that the limit distribution of $\breve{\gamma}_v^i - \gamma_{0v}^i$ naturally decomposes in two parts, and we therefore split the hypothesis $\gamma_v^i = 0$ into a test that $\beta'\gamma_v^i = 0$ and a test that $\alpha_\perp'\Gamma\gamma_v^i = 0$.

Theorem 8. *Let Assumptions 1–5 be satisfied. When* $m_v \geq 0$*, the likelihood ratio test for the hypothesis* $\beta'\gamma_v^0 = 0$ *is asymptotically* χ^2*-distributed, and when* $m_v \geq 1$*, the likelihood ratio test for the hypothesis* $\alpha_\perp'\Gamma\gamma_v^0 = 0$ *is asymptotically* χ^2*-distributed.*

We apply the results in Theorem 8 as follows. It appears natural first to investigate if γ_v^0, the coefficient of U_{vt}, is zero. If we cannot reject that it is zero, then we can proceed to test that the coefficient of ΔU_{vt} is zero; that is, test the hypothesis $\gamma_v^1 = 0$, assuming $\gamma_v^0 = 0$.

Under Assumption 5 the additive and extended models are asymptotically equal. Under the hypothesis $\beta'\gamma_v^0 = 0$, we find $\Upsilon_v^0 = 0$, so we estimate the other parameters by reduced rank regression leaving out the regressor U_{vt} in the extended model. If also $\alpha_\perp'\Gamma\gamma_v^0 = 0$, then by (4) we have $\gamma_v^0 = 0$, so that U_{vt} is no longer a regressor in the additive model and ΔU_{vt} becomes the highest order term. By reformulating the model to take this into account, the coefficient γ_v^1 is split into $\beta'\gamma_v^1$ and $\alpha_\perp'\Gamma\gamma_v^1$. The first is in the cointegrating space where $\breve{\beta}'(\breve{\gamma}_v^1 - \gamma_{v0}^1)$ has an asymptotic mixed Gaussian distribution. The second is found from the new $\alpha_\perp'\Phi_1$, and the asymptotic distribution of $T^{1/2}(\breve{\alpha}_\perp'\hat{\Gamma}\breve{\gamma}_v^1 - \alpha_{0\perp}'\Gamma_0\gamma_{v0}^1)M_{T1v}^{-1}$ is Gaussian.

Thus we can apply the asymptotic distributions in Theorem 4 to test recursively that $\gamma_v^i = 0$, provided we assume that $\gamma_v^j = 0, 0 \leq j < i$.

5 Conclusions

We define the CVAR model with additive deterministic terms and derive the corresponding innovative formulation which is nonlinear in the parameters. This additive model is extended to a model which is linear in the coefficients of the deterministic terms and hence allows estimation by reduced rank regression. A general class of regressors is defined and for each regressor its order. This setup allows a discussion of the relation between the innovative formulation of the additive model and its extension.

A simple condition for when the additive and the extended model are (asymptotically) identical is given. The condition, given as Assumption 5, is that for each regressor in the additive model one should also include its differences, as long as they have diverging information. If this recommendation is not followed, asymptotic inference is considerably more complicated. For example, when the regressor is a polynomial or power function, say t^a for some $a > -1/2$, the recommendation is to include (at least) $m = [a + 1/2]$ differences of t^a, which seems like a natural thing to do. Indeed, not doing so seems very strange. On the other hand, for the broken trend function, $(t - t_0)^+$, it may in fact be reasonable to exclude the first difference, $1_{\{t \geq t_0+1\}}$, when insisting on continuity of the trend function as in Example 5. However, the recommendation is to include the first difference anyway, even if it may be zero, because including it leads to simple inference.

The asymptotic distribution of the parameter estimates is found to be a mixture of a Gaussian distribution and a mixed Gaussian distribution, and we show how it can be applied to test that the regression coefficients are zero. Finally, we derive the asymptotic distribution of the rank test and show that it is similar with respect to the regression parameters.

A Appendix: proofs of results

A.1 Proof of Lemma 1

The result (15) follows from (14) when multiplying by $\bar{\alpha}'$ and $\alpha'_\perp$ using $\Phi_0 = -\alpha\beta'$ and $\Phi_1 = -\alpha\beta' - \Gamma$. The relations (15) can be solved for γ^i because $(\beta, \Gamma'\alpha_\perp)$ has full rank under Assumption 1, see (4), and therefore γ^i is determined recursively as a linear function of $\Upsilon_0, \ldots, \Upsilon_i, \alpha'_\perp \Upsilon_{i+1}$.

To solve for γ^{se}, we let $(\omega_j, v_j), j = 1, \ldots, s-1$, be the eigenvalues and eigenvectors of M_s. It is clear from (14) that Υ_{se} is a linear function of γ^{se}, and we want to show that this function is non-singular, that is, that $\Upsilon_{se} = \sum_{i=0}^k \Phi_i \gamma^{se} M_s^i = 0$ implies $\gamma^{se} = 0$. To see this, post-multiply by v_j and use $M_s^i v_j = \omega_j^i v_j$, such that

$$0 = \Upsilon_{se} v_j = \sum_{i=0}^k \Phi_i \gamma^{se} M_s^i v_j = \sum_{i=0}^k \Phi_i \omega_j^i \gamma^{se} v_j = \Pi(1 - \omega_j)\gamma^{se} v_j. \qquad (34)$$

Now $1 - \omega_j$ is a seasonal root, see Definition 2 and (11), and by Assumption 1 this implies that $\Pi(1 - \omega_j)$ has full rank, such that $\gamma^{se} v_j = 0$ for all j and hence $\gamma^{se} = 0$. The definition of Υ_{se} is therefore

$$\operatorname{vec}(\Upsilon_{se}) = (\sum_{i=0}^k M_s^i \otimes \Phi_i) \operatorname{vec}(\gamma^{se}),$$

where $\sum_{i=0}^k M_s^i \otimes \Phi_i$ is of full rank. The solution is then given by (16).

A.2 Proof of Theorem 1

The additive formulation of model (1) with $Z_t = (f_m(t), \dots, f_0(t))'$ has deterministic term

$$\Pi(L)\gamma Z_t = \sum_{i=0}^{k} \Phi_i \Delta^i \sum_{j=0}^{m} \gamma^j f_{m-j}(t) = \sum_{i=0}^{k}\sum_{j=0}^{m} \Phi_i \gamma^j f_{m-j-i}(t) = \sum_{s=0}^{m} \Upsilon_s f_{m-s}(t).$$

Lemma 1 shows that $\gamma^0, \dots, \gamma^{m-1}, \beta'\gamma^m$ can be determined from $\rho' = \bar{\alpha}'\Upsilon_0, \Upsilon_1, \dots, \Upsilon_m$ for given values of $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1})$. Thus, for this choice of regressors, the additive model parametrized by $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1}, \gamma^0, \dots, \gamma^{m-1}, \beta'\gamma^m)$ is the same as the extended model parametrized by $(\alpha, \beta, \Gamma_1, \dots, \Gamma_{k-1}, \rho, \Upsilon_1, \dots, \Upsilon_m)$.

A.3 Proof of Theorem 2

The proof follows from Lemma 1 because $\xi(\lambda)$ determines λ as a linear, and hence continuous, function except for $\alpha'_\perp \Gamma \gamma_1^{m_1}$ (in the case $n \geq m_1$).

A.4 Proof of Theorem 3

We can express both the additive model (10) and the extended model (17) as nonlinear submodels of a linear regression model as follows. Because we have normalized β on $\beta'\bar{\beta}_0 = I_r$, we can define $\theta = \bar{\beta}'_{0\perp}(\beta - \beta_0)$ such that $\beta = \beta_0 + \beta_{0\perp}\theta$. Then the extended model (17) is

$$\Delta X_t = \alpha\beta_0' X_{t-1} + \alpha\theta'\beta_{0\perp}' X_{t-1} + \alpha\rho^{0\prime} Z_{0t} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon^1 Z_{1t} + \Upsilon^2 Z_{2t} + \varepsilon_t,$$

which is a submodel of the linear regression model

$$\Delta X_t = \alpha(\beta_0' X_{t-1}) + \phi(\beta_{0\perp}' X_{t-1}) + \psi Z_{0t} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon^1 Z_{1t} + \Upsilon^2 Z_{2t} + \varepsilon_t \qquad (35)$$

defined by the restrictions $\phi = \alpha\theta', \psi = \alpha\rho^{0\prime}$, and the remaining parameters being the same in the two models. From Theorem 2 it follows that, because $\alpha_0\rho_h^{0\prime} \to \alpha_0\rho_0^{0\prime}$ and $\alpha_0\theta_h' \to \alpha_0\theta_0'$ implies $\theta_h \to \theta_0$ and $\rho_h^0 \to \rho_0^0$, the extended model is continuously identified in the larger linear regression model (35). Similarly, the additive model is continuously identified in the extended model and hence in the larger linear regression model. The result now follows immediately from Theorem 2 and Lemma 2.

A.5 Proof of Theorem 4

Let $\Pi_0(L)$ be the characteristic polynomial with the true values inserted. The data is generated by $\Pi_0(L)(X_t - \gamma_0 Z_t) = \varepsilon_t$ and we define

$$\begin{aligned}\varepsilon_t(\xi) &= \Pi(L)(X_t - \gamma Z_t) = \Pi(L)Y_t - \Pi(L)(\gamma - \gamma_0)Z_t \\ &= (\Pi(L) - \Pi_0(L))Y_t - \Pi(L)(\gamma - \gamma_0)Z_t + \varepsilon_t,\end{aligned}$$

where, see (2), (13), (19), and Definition 3,

$$-\Pi(L)(\gamma - \gamma_0)Z_t = \alpha\beta'(\gamma^{0*} - \gamma_0^{0*})Z_{0t} - (\Upsilon^1 - \Upsilon_0^1)Z_{1t} - (\Upsilon^2 - \Upsilon_0^2)Z_{2t}, \qquad (36)$$

$$(\Pi(L) - \Pi_0(L))Y_t = -(\alpha - \alpha_0)\beta_0' Y_{t-1} - \alpha(\beta - \beta_0)'\bar{\beta}_{0\perp}'\beta_{0\perp}' Y_{t-1} - \sum_{i=1}^{k-1}(\Gamma_i - \Gamma_{i,0})\Delta Y_{t-i}. \qquad (37)$$

We can simplify the notation by redefining the parameters to account for the different orders of magnitude of the regressors. We therefore use

$$G_{Tt} = \begin{pmatrix} T^{-1/2}\beta_{0\perp}' Y_{t-1} \\ Z_{0Tt} \end{pmatrix}, \ \zeta = \begin{pmatrix} T^{1/2}\bar{\beta}_{0\perp}'(\beta - \beta_0) \\ -N_{T0}^{-1}(\gamma^{0*} - \gamma_0^{0*})'\beta \end{pmatrix}, \text{ and } \upsilon = (\Upsilon^1 - \Upsilon_0^1)N_{T1}^{-1},$$

and also define

$$\alpha^* = (\alpha - \alpha_0, \Gamma_1 - \Gamma_{1,0}, \ldots, \Gamma_{k-1} - \Gamma_{k-1,0}) \text{ and } Y_t^* = (Y_{t-1}'\beta_0, \Delta Y_{t-1}', \ldots, \Delta Y_{t-k+1}')'. \quad (38)$$

Then we find the expression

$$\varepsilon_t(\xi) = -\alpha^* Y_t^* - \alpha\zeta' G_{Tt} - \upsilon Z_{1Tt} - (\Upsilon^2 - \Upsilon_0^2)Z_{2t} + \varepsilon_t. \quad (39)$$

Elimination of Z_{2t}: At the true values ($\alpha_0^* = 0, \zeta_0 = 0, \upsilon_0 = 0$) we find the derivatives

$$\begin{aligned} D_{\alpha^*}\varepsilon_t(\xi_0; d\alpha^*) &= -(d\alpha^*)Y_t^*, \\ D_\zeta\varepsilon_t(\xi_0; d\zeta) &= -\alpha_0(d\zeta)'G_{Tt}, \\ D_\upsilon\varepsilon_t(\xi_0; d\upsilon) &= -(d\upsilon)Z_{1Tt}, \\ D_{\Upsilon^2}\varepsilon_t(\xi_0; d\Upsilon^2) &= -(d\Upsilon^2)Z_{2t}. \end{aligned}$$

Because the scores for Y_t^*, G_{Tt}, and Z_{1Tt} all need to be normalized by $T^{-1/2}$ by definition of N_{T1} and G_{Tt}, while the score for Z_{2t} need not be normalized, showing that the information is asymptotically block diagonal with respect to Υ^2 entails showing that $T^{1/2}\langle Z_{0T}, Z_2\rangle_T \to 0$, $T^{1/2}\langle Z_{1T}, Z_2\rangle_T \to 0$, $T^{1/2}\langle G_T, Z_2\rangle_T \overset{P}{\to} 0$, and $T^{1/2}\langle Y^*, Z_2\rangle_T \overset{P}{\to} 0$. The proofs are almost identical, so we prove only that $T^{1/2}\langle Z_{1T}, Z_2\rangle_T \to 0$. We use that Z_{2t} has bounded information while Z_{1t} has diverging information, and for $T_1 < T$ we decompose as

$$T^{1/2}\langle Z_{1T}, Z_2\rangle_T = T^{-1/2}\sum_{t=1}^T Z_{1Tt}Z_{2t}' = T^{-1/2}N_{T1}\sum_{t=1}^{T_1} Z_{1t}Z_{2t}' + T^{-1/2}\sum_{t=T_1+1}^T Z_{1Tt}Z_{2t}'.$$

Here the first term tends to zero for any fixed T_1 because $T^{-1/2}N_{T1} \to 0$. The second term is bounded in norm by the Cauchy-Schwarz inequality,

$$\begin{aligned} \sum_{t=T_1+1}^T T^{-1/2}||Z_{1Tt}||||Z_{2t}|| &\le (\sum_{t=T_1+1}^T T^{-1}||Z_{1Tt}||^2)^{1/2}(\sum_{t=T_1+1}^T ||Z_{2t}||^2)^{1/2} \\ &\le \operatorname{tr}\{\langle Z_{1T}, Z_{1T}\rangle_T\}^{1/2}(\sum_{t=T_1+1}^\infty ||Z_{2t}||^2)^{1/2} \to 0 \end{aligned}$$

as $T_1 \to \infty$ because $\sum_{t=T_1+1}^\infty ||Z_{2t}||^2 \to 0$ as $T_1 \to \infty$. Thus, we can conduct inference separately on $(\alpha^*, \zeta, \upsilon)$ and Υ^2, and we continue fixing $\Upsilon^2 = \Upsilon_0^2$.

Score and information: We denote the normalized score function with respect to α^*, for example, in the direction $d\alpha^*$ as $S_{\alpha^*} = T^{1/2}D_{\alpha^*}L(\xi, \Omega; d\alpha^*)|_{\xi=\xi_0}$. The information with respect to α^*, ζ, for example, is similarly denoted by $I_{\alpha^*\zeta} = D^2_{\alpha^*\zeta}L(\xi,\Omega; d\alpha^*, d\zeta)|_{\xi=\xi_0}$. It follows that the scores are

$$T^{-1/2}S_{\alpha^*} = -\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)T^{1/2}\langle Y^*,\varepsilon\rangle_T\} \overset{\mathsf{D}}{\to} -\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*,\varepsilon\rangle\}, \tag{40}$$

$$T^{-1/2}S_{\zeta} = -\operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'T^{1/2}\langle G_T,\varepsilon\rangle_T\} \overset{\mathsf{D}}{\to} -\operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G,\varepsilon\rangle\}, \tag{41}$$

$$T^{-1/2}S_{\upsilon} = -\operatorname{tr}\{\Omega_0^{-1}(d\upsilon)T^{1/2}\langle Z_{1T},\varepsilon\rangle_T\} \overset{\mathsf{D}}{\to} -\operatorname{tr}\{\Omega_0^{-1}(d\upsilon)\langle Z_1,\varepsilon\rangle\}, \tag{42}$$

and the diagonal elements of the information are

$$T^{-1}I_{\alpha^*\alpha^*} = \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*,Y^*\rangle_T(d\alpha^*)'\} + o_P(1) \overset{P}{\to} \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\Sigma_{stat}(d\alpha^*)'\}, \tag{43}$$

$$T^{-1}I_{\zeta\zeta} = \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G_T,G_T\rangle_T(d\zeta)\alpha_0'\} + o_P(1) \overset{\mathsf{D}}{\to} \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G,G\rangle(d\zeta)\alpha_0'\}, \tag{44}$$

$$T^{-1}I_{\upsilon\upsilon} = \operatorname{tr}\{\Omega_0^{-1}(d\upsilon)\langle Z_{1T},Z_{1T}\rangle_T(d\upsilon)'\} + o_P(1) \overset{\mathsf{D}}{\to} \operatorname{tr}\{\Omega_0^{-1}(d\upsilon)\langle Z_1,Z_1\rangle(d\upsilon)'\}, \tag{45}$$

where $\Sigma_{stat} = Var(Y_t^*)$. There is one non-zero off-diagonal block,

$$T^{-1}I_{\zeta\upsilon} = \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G_T,Z_{1T}\rangle_T(d\upsilon)'\} + o_P(1) \overset{\mathsf{D}}{\to} \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G,Z_1\rangle(d\upsilon)'\}, \tag{46}$$

and the following are asymptotically negligible,

$$T^{-1}I_{\alpha^*\zeta} = \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*,G_T\rangle_T(d\zeta)'\alpha_0'\} + o_P(1) \overset{P}{\to} 0, \tag{47}$$

$$T^{-1}I_{\alpha^*\upsilon} = \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*,Z_{1T}\rangle_T(d\upsilon)'\} + o_P(1) \overset{P}{\to} 0. \tag{48}$$

Because the information is asymptotically block diagonal, $\hat{\alpha}^*$ and $(\hat{\zeta},\hat{\upsilon})$ are asymptotically independent, and we consider inference separately for α^* and (ζ,υ), in both cases fixing $\Omega = \Omega_0$ and $\Upsilon^2 = \Upsilon_0^2$.

The asymptotic distribution of $T^{1/2}\hat{\alpha}^*$: By the usual Taylor expansion of the likelihood equations, we find that the equations for the asymptotic distribution of $T^{1/2}\hat{\alpha}^*$ are given by

$$\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*,Y^*\rangle_T(T^{1/2}\hat{\alpha}^*)'\} = -\operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)T^{1/2}\langle Y^*,\varepsilon\rangle_T\} + o_P(1) \text{ for all } d\alpha^*,$$

and hence

$$\Sigma_{stat}T^{1/2}\hat{\alpha}^{*\prime} = -T^{1/2}\langle Y^*,\varepsilon\rangle_T + o_P(1),$$

which by the Central Limit Theorem gives the result in (23).

The asymptotic distribution of $T^{1/2}(\hat{\zeta},\hat{\upsilon})$: Similarly, we find that the equations for determining the limit distribution of $(\hat{\zeta},\hat{\upsilon})$ are

$$\langle G_T,G_T\rangle_T(T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G_T,Z_{1T}\rangle_T(T^{1/2}\hat{\upsilon})'\Omega_0^{-1}\alpha_0 = -T^{1/2}\langle G_T,\varepsilon\rangle_T\Omega_0^{-1}\alpha_0 + o_P(1), \tag{49}$$

$$\langle Z_{1T},G_T\rangle_T(T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1} + \langle Z_{1T},Z_{1T}\rangle_T(T^{1/2}\hat{\upsilon})'\Omega_0^{-1} = -T^{1/2}\langle Z_{1T},\varepsilon\rangle_T\Omega_0^{-1} + o_P(1). \tag{50}$$

Pre-multiplying (50) by $\langle G_T,Z_{1T}\rangle_T\langle Z_{1T},Z_{1T}\rangle_T^{-1}$, post-multiplying it by α_0, and subtracting the result from (49) we find

$$\langle G_T,G_T|Z_{1T}\rangle_T(T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 = -\langle G_T,\varepsilon|Z_{1T}\rangle_T\Omega_0^{-1}\alpha_0,$$

which implies (24). Finally, inserting (24) into (50), we find (25).

A.6 Proof of Theorem 5

The additive model is parametrized by the dynamic parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$, the parameters $\beta'\gamma_v^0, \gamma^{se}$, if $m_v = 0$, and $\gamma_v^0, \ldots, \gamma_v^{m_v-1}, \beta'\gamma_v^{m_v}, \gamma^{se}$, if $m_v \geq 1$, as well as the parameters $\alpha_\perp'\Gamma\gamma_v^{m_v}, \gamma_v^{m_v+1}, \ldots, \gamma_v^n$, where the latter are coefficients to regressors with bounded information, see Definitions 1 and 2. Thus, for the asymptotic analysis, the last set of parameters is irrelevant. The extended model is similarly parametrized by the dynamic parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$, $\rho_v' = \bar{\alpha}'\Upsilon_{0v} = -\beta'\gamma_v^0$ for $m_v \geq 0$, Υ_{iv} for $1 \leq i \leq m_v$, and Υ_{se}, as well as the parameters $\Upsilon_{iv}, m_v < i \leq n+k$. Again, the latter are coefficients to regressors with bounded information and therefore not relevant.

Lemma 1 thus shows that there is a simple one-to-one relation between the relevant parameters of the additive model and the relevant parameters in the extended model. The likelihoods for the two models are therefore asymptotically the same. Consequently, the results (26) and (27) follow directly from (23) and (24).

We next find that $\breve{\gamma}_v^0, \ldots, \breve{\gamma}_v^{m_v-1}, \breve{\beta}'\breve{\gamma}_v^{m_v}$ can be expressed in terms of $\hat{\rho}_v, \hat{\Upsilon}_{1v}, \ldots, \hat{\Upsilon}_{m_v,v}$, using that the maximum likelihood estimators of the dynamic parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$ in both models are consistent (Theorem 3). That is, by Lemma 1,

$$\begin{aligned}
-T^{1/2}\beta_0'(\breve{\gamma}_v^i - \gamma_{v0}^i)M_{Tiv}^{-1} &= -T^{1/2}\bar{\alpha}_0'(\hat{\Upsilon}_{iv} - \Upsilon_{iv0})M_{Tiv}^{-1} \\
&\quad + T^{1/2}\bar{\alpha}_0' \sum_{j=1}^{\min\{i,k,m_v\}} \Phi_{j0}(\hat{\gamma}_v^{i-j} - \gamma_{v0}^{i-j})M_{Tiv}^{-1} \\
&= -T^{1/2}\bar{\alpha}_0'(\hat{\Upsilon}_{iv} - \Upsilon_{iv0})M_{Tiv}^{-1} + o_P(1)
\end{aligned}$$

and

$$\begin{aligned}
T^{1/2}\alpha_{0\perp}'\Gamma_0(\breve{\gamma}_v^i - \gamma_{v0}^i)M_{T,i+1,v}^{-1} &= -T^{1/2}\alpha_{0\perp}'(\hat{\Upsilon}_{i+1,v} - \Upsilon_{i+1,v,0})M_{T,i+1,v}^{-1} \\
&\quad + T^{1/2}\alpha_{0\perp}' \sum_{j=2}^{\min\{i+1,k,m_v\}} \Phi_{j0}(\hat{\gamma}_v^{i+1-j} - \gamma_{v0}^{i+1-j})M_{T,i+1,v}^{-1} \\
&= -T^{1/2}\alpha_{0\perp}'(\hat{\Upsilon}_{i+1,v} - \Upsilon_{i+1,v,0})M_{T,i+1,v}^{-1} + o_P(1)
\end{aligned}$$

because $M_{T,i-j,v}M_{Tiv}^{-1} \to 0$ for $j = 1, \ldots, \min\{i, k, m_v\}$ and $M_{T,i-j+1,v}M_{T,i+1,v}^{-1} \to 0$ for $j = 2, \ldots, \min\{i+1, k, m_v\}$. This proves (28) and (29). Note that $T^{1/2}\bar{\alpha}_0'(\hat{\Upsilon}_{iv} - \Upsilon_{iv0})M_{Tiv}^{-1}$ is neither asymptotically Gaussian nor mixed Gaussian.

Finally, (30) follows directly from vectorization of (14) and noting that $\sum_{i=0}^k M_s^{i\prime} \otimes \Phi_i$ is invertible by the proof of Theorem 2.

A.7 Proof of Theorem 6

The extended model: From (49) and (50) we find the equations to determine the limit distribution of the maximum likelihood estimator $T^{1/2}(\hat{\zeta}, \hat{\upsilon})$,

$$\langle G, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle (T^{1/2}\hat{\upsilon})\Omega_0^{-1}\alpha_0 \xrightarrow{\mathsf{D}} - \langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0, \tag{51}$$

$$\langle Z_1, G\rangle (T^{1/2}\hat{\zeta})\alpha_0'\Omega_0^{-1} + \langle Z_1, Z_1\rangle (T^{1/2}\hat{\upsilon})\Omega_0^{-1} \xrightarrow{\mathsf{D}} - \langle Z_1, \varepsilon\rangle \Omega_0^{-1}. \tag{52}$$

Eliminating $T^{1/2}\hat{\upsilon}$ from the equations, the right-hand side becomes

$$- \langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle \langle Z_1, Z_1\rangle^{-1} \langle Z_1, \varepsilon\rangle \Omega_0^{-1}\alpha_0.$$

Conditional on G, this expression has mean zero because $\alpha_0'\Omega_0^{-1}W_\varepsilon$ is independent of $\alpha_{0\perp}'W_\varepsilon$. This implies that the limit distribution of $T^{1/2}\hat{\zeta}$ is mixed Gaussian as given in Theorem 4, and subsequently that $\alpha_{0\perp}'T^{1/2}\hat{\upsilon}$ is Gaussian.

The additive model: We decompose $\gamma - \gamma_0$ as

$$\gamma - \gamma_0 = \alpha(\beta'\alpha)^{-1}\beta'(\gamma - \gamma_0) + \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(\gamma - \gamma_0).$$

We note that $\zeta_2' = -\beta'(\gamma - \gamma_0)M_{T0}^{-1}$ and define $\phi' = T^{1/2}(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(\gamma - \gamma_0)M_{T1}^{-1}$ as well as $\alpha^* = \alpha - \alpha_0, Y_t^* = \beta_0'Y_{t-1}$. The corresponding likelihood is then based on

$$\varepsilon_t(\alpha^*, \zeta, \phi) = -\alpha^* Y_t^* - \alpha\zeta' G_{Tt} - \alpha(\beta'\alpha)^{-1}\zeta_2' M_{T0}M_{T1}^{-1}Z_{1Tt} - \beta_\perp\phi' Z_{1Tt} + \varepsilon_t,$$

see (39). We note that ζ_2 appears in two places, but $M_{T0}M_{T1}^{-1} \to 0$, and therefore the term with ζ_2 and Z_{1Tt} disappears in the asymptotic analysis of the score and information, because it is dominated by the term $\alpha\zeta' G_{Tt}$. Note also that $\alpha = \alpha(\alpha^*)$ and $\beta = \beta(\zeta)$ when we calculate derivatives below.

Mimicking the analysis of the extended model, we find at the true values, $\alpha_0^* = 0, \zeta_0 = 0, \phi_0 = 0$, the derivatives

$$\begin{aligned} D_{\alpha^*}\varepsilon_t(\alpha^*, \zeta, \phi; d\zeta) &= -(d\alpha^*)Y_t^*, \\ D_\zeta\varepsilon_t(\alpha^*, \zeta, \phi; d\zeta) &= -\alpha_0(d\zeta)' G_{Tt} - \alpha_0(\beta_0'\alpha_0)^{-1}(d\zeta_2)' M_{T0}M_{T1}^{-1}Z_{1Tt} = -\alpha_0(d\zeta)'G_{Tt} + o(1), \\ D_\phi\varepsilon_t(\alpha^*, \zeta, \phi; d\phi) &= -\beta_{0\perp}(d\phi)' Z_{1Tt}. \end{aligned}$$

The scores for α^* and ζ are given in (40) and (41), and for ϕ we find

$$T^{-1/2}S_\phi = -\operatorname{tr}\{\Omega_0^{-1}\beta_{0\perp}(d\phi)'T^{1/2}\langle Z_{1T}, \varepsilon\rangle_T\} \xrightarrow{\mathsf{D}} -\operatorname{tr}\{\Omega_0^{-1}\beta_{0\perp}(d\phi)'\langle Z_1, \varepsilon\rangle\}.$$

The information matrix blocks $I_{\alpha^*\alpha^*}$, $I_{\zeta\zeta}$, and $I_{a^*\zeta}$ are given in (43), (44), and (47), respectively, and for ϕ we find

$$\begin{aligned} T^{-1}I_{\phi\phi} &= \operatorname{tr}\{\Omega_0^{-1}\beta_{0\perp}(d\phi)'\langle Z_{1T}, Z_{1T}\rangle_T (d\phi)\beta_{0\perp}'\} + o_P(1) \xrightarrow{\mathsf{D}} \operatorname{tr}\{\Omega_0^{-1}\beta_{0\perp}(d\phi)'\langle Z_1, Z_1\rangle (d\phi)\beta_{0\perp}'\}, \\ T^{-1}I_{\alpha^*\phi} &= \operatorname{tr}\{\Omega_0^{-1}(d\alpha^*)\langle Y^*, Z_{1T}\rangle_T (d\phi)\beta_{0\perp}'\} + o_P(1) \xrightarrow{P} 0, \\ T^{-1}I_{\zeta\phi} &= \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G_T, Z_{1T}\rangle_T (d\phi)\beta_{0\perp}'\} + o_P(1) \xrightarrow{\mathsf{D}} \operatorname{tr}\{\Omega_0^{-1}\alpha_0(d\zeta)'\langle G, Z_1\rangle (d\phi)\beta_{0\perp}'\}. \end{aligned}$$

Thus, the only difference compared with the extended model is the factor $\beta_{0\perp}$, which comes from only estimating $\alpha_\perp'\gamma$ from the coefficient to ΔU_t. It is seen that the limit information is block-diagonal corresponding to α^* and (ζ, ϕ), such that the asymptotic distribution of $T^{1/2}\breve{\alpha}^* = T^{1/2}(\breve{\alpha} - \alpha_0)$ is as given in (26) in Theorem 5.

We find the equations for determining the limit distribution of the maximum likelihood estimator $T^{1/2}(\breve{\zeta}, \breve{\phi})$, compare (49) and (50),

$$\langle G, G\rangle (T^{1/2}\breve{\zeta})\alpha_0'\Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle (T^{1/2}\breve{\phi})\beta_{0\perp}'\Omega_0^{-1}\alpha_0 \xrightarrow{\mathsf{D}} -\langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0, \tag{53}$$

$$\langle Z_1, G\rangle (T^{1/2}\breve{\zeta})\alpha_0'\Omega_0^{-1}\beta_{0\perp} + \langle Z_1, Z_1\rangle (T^{1/2}\breve{\phi})\beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp} \xrightarrow{\mathsf{D}} -\langle Z_1, \varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}. \tag{54}$$

Eliminating $T^{1/2}\breve{\phi}$ from (53), we find the right-hand side

$$-\langle G, \varepsilon\rangle \Omega_0^{-1}\alpha_0 + \langle G, Z_1\rangle \langle Z_1, Z_1\rangle^{-1}\langle Z_1, \varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}(\beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp})^{-1}\beta_{0\perp}'\Omega_0^{-1}\alpha_0.$$

If we condition on G, or equivalently on $\alpha_{0\perp}' W_\varepsilon$, the right-hand side is Gaussian with mean proportional to

$$E(\langle Z_1,\varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}|\alpha_{0\perp}' W_\varepsilon) = \langle Z_1,\varepsilon\rangle \alpha_{0\perp}(\alpha_{0\perp}'\Omega_0\alpha_{0\perp})^{-1}\alpha_{0\perp}'\beta_{0\perp} \neq 0. \tag{55}$$

Thus, the limit distribution of $T^{1/2}\breve{\zeta}$ is not mixed Gaussian, and the same holds for any linear combination of $T^{1/2}\breve{\zeta}$.

If we eliminate $T^{1/2}\breve{\zeta}$ from the equations (53) and (54), we find that the right-hand side becomes

$$\langle Z_1,G\rangle \langle G,G\rangle^{-1} \langle G,\varepsilon\rangle \Omega_0^{-1}\alpha_0(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\alpha_0'\Omega_0^{-1}\beta_{0\perp} - \langle Z_1,\varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}.$$

Conditional on G the distribution has mean $-E(\langle Z_1,\varepsilon\rangle \Omega_0^{-1}\beta_{0\perp}|\alpha_{0\perp}' W_\varepsilon)$, see (55), and the limit distribution of $T^{1/2}\breve{\phi}$ is neither Gaussian nor mixed Gaussian, and the same holds for any linear combination.

Comparison of information matrices: The limit of the information matrix for (ζ,υ) in the extended model is, see (51) and (52),

$$\begin{pmatrix} \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,G\rangle & \Omega_0^{-1}\alpha_0 \otimes \langle G,Z_1\rangle \\ \alpha_0'\Omega_0^{-1} \otimes \langle Z_1,G\rangle & \Omega_0^{-1} \otimes \langle Z_1,Z_1\rangle \end{pmatrix} = \begin{pmatrix} I_{\zeta\zeta}^{ext} & I_{\zeta\upsilon}^{ext} \\ I_{\upsilon\zeta}^{ext} & I_{\upsilon\upsilon}^{ext} \end{pmatrix},$$

say. In the additive model the limit of the information matrix for (ζ,ϕ) is, see (53) and (54),

$$\begin{pmatrix} \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,G\rangle & \alpha_0'\Omega_0^{-1}\beta_{0\perp} \otimes \langle G,Z_1\rangle \\ \beta_{0\perp}'\Omega_0^{-1}\alpha_0 \otimes \langle Z_1,G\rangle & \beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp} \otimes \langle Z_1,Z_1\rangle \end{pmatrix} = \begin{pmatrix} I_{\zeta\zeta}^{add} & I_{\zeta\phi}^{add} \\ I_{\phi\zeta}^{add} & I_{\phi\phi}^{add} \end{pmatrix}.$$

We note that the left factors in the information matrix for (ζ,ϕ) satisfy the relation

$$\alpha_0'\Omega_0^{-1}\alpha_0 - \alpha_0'\Omega_0^{-1}\beta_{0\perp}(\beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp})^{-1}\beta_{0\perp}'\Omega_0^{-1}\alpha_0 = \alpha_0'\beta_0(\beta_0'\Omega_0^{-1}\beta_0)^{-1}\beta_0'\alpha_0 > 0.$$

This has the consequence that

$$\begin{aligned} &I_{\zeta\zeta}^{add} - I_{\zeta\phi}^{add}(I_{\phi\phi}^{add})^{-1}I_{\phi\zeta}^{add} \\ &\quad = \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,G\rangle - \alpha_0'\Omega_0^{-1}\beta_{0\perp}(\beta_{0\perp}'\Omega_0^{-1}\beta_{0\perp})^{-1}\beta_{0\perp}'\Omega_0^{-1}\alpha_0 \otimes \langle G,Z_1\rangle \langle Z_1,Z_1\rangle^{-1} \langle Z_1,G\rangle \\ &\quad > \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,G\rangle - \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,Z_1\rangle \langle Z_1,Z_1\rangle^{-1} \langle Z_1,G\rangle \\ &\quad = \alpha_0'\Omega_0^{-1}\alpha_0 \otimes \langle G,G|\Delta Z\rangle = I_{\zeta\zeta}^{ext} - I_{\zeta\upsilon}^{ext}(I_{\upsilon\upsilon}^{ext})^{-1}I_{\upsilon\zeta}^{ext}. \end{aligned}$$

A.8 Proof of Theorem 7

Because the regressors with bounded information, Z_{2t}, do not contribute in the asymptotic analysis, see the proof of Theorem 4, we continue setting them equal to zero.

Normalization of parameters and an auxiliary model: We introduce the $p \times r$ matrix β_0 of rank r and decompose Π as

$$\Pi = \Pi\bar{\beta}_0\beta_0' + \Pi\bar{\beta}_{\perp 0}\beta_{\perp 0}',$$

and define the auxiliary hypothesis

$$\mathcal{H} = \{\Pi\bar{\beta}_{\perp 0} = 0 \text{ and } \beta_0'(\gamma^{0*} - \gamma_0^{0*}) = 0\}.$$

We note that under the assumptions in $\mathcal{H}$, $\Pi = \alpha\beta_0'$ for $\alpha = \Pi\bar{\beta}_0$, such that

$$\mathcal{H} = \{\Pi = \alpha\beta_0' \text{ and } \beta_0'(\gamma^{0*} - \gamma_0^{0*}) = 0\}.$$

Thus, if $\beta^* = (\beta', \beta'\gamma^{0*})'$ then

$$\mathcal{H}_p^{ext} \cap \mathcal{H} = \mathcal{H}_r^{ext} \cap \{\beta^* = \beta_0^*\}.$$

To facilitate the analysis of the test for rank, we introduce the extra hypothesis $\mathcal{H}$ in models $\mathcal{H}_p^{ext}$ and $\mathcal{H}_r^{ext}$, see Lawley (1956) for an early application of this idea or Johansen (2002, p. 1947) and Johansen and Nielsen (2012) for applications to the (fractional) CVAR model. We then find

$$LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext}) = \frac{\max_{\mathcal{H}_r^{ext}} L_T(\xi,\Omega)}{\max_{\mathcal{H}_p^{ext}} L_T(\xi,\Omega)} = \frac{\max_{\mathcal{H}_p^{ext}\cap\mathcal{H}} L_T(\xi,\Omega)}{\max_{\mathcal{H}_p^{ext}} L_T(\xi,\Omega)} \Big/ \frac{\max_{\mathcal{H}_r^{ext}\cap\{\beta^*=\beta_0^*\}} L_T(\xi,\Omega)}{\max_{\mathcal{H}_r^{ext}} L_T(\xi,\Omega)}.$$

That is, instead of the rank test statistic, we analyze the ratio of two test statistics,

$$LR(\mathcal{H}_r^{ext}|\mathcal{H}_p^{ext}) = \frac{LR(\mathcal{H}|\mathcal{H}_p^{ext})}{LR(\beta^* = \beta_0^*|\mathcal{H}_r^{ext})}. \tag{56}$$

Analysis of $LR(\mathcal{H}|\mathcal{H}_p^{ext})$: We apply the formulas (36) and (37) for $\mathcal{H}_p^{ext}$, using $\alpha - \alpha_0 = (\Pi - \Pi_0)\bar{\beta}_0$,

$$\begin{aligned} -\Pi(L)(\gamma - \gamma_0)Z_t &= \Pi(\gamma^{0*} - \gamma_0^{0*})N_{T0}^{-1}Z_{0Tt} - (\Upsilon^1 - \Upsilon_0^1)N_{T1}^{-1}Z_{1Tt}, \\ (\Pi(L) - \Pi_0(L))Y_t &= -\alpha^* Y_t^* - T^{1/2}\Pi\bar{\beta}_{0\perp}'T^{-1/2}\beta_{0\perp}'Y_{t-1}, \end{aligned}$$

where α^* and Y_t^* are given in (38), such that for

$$\zeta' = (T^{1/2}\Pi\bar{\beta}_{0\perp}', \Pi(\gamma^{0*} - \gamma_0^{0*})N_{T0}^{-1}) \text{ and } G_{Tt} = \begin{pmatrix} T^{-1/2}\beta_{0\perp}'Y_{t-1} \\ Z_{0Tt} \end{pmatrix}, \tag{57}$$

we find the residuals

$$\varepsilon_t(\xi) = -\alpha^* Y_t^* - \zeta' G_{Tt} - (\Upsilon^1 - \Upsilon_0^1)N_{T1}^{-1}Z_{1Tt} + \varepsilon_t.$$

This shows that the likelihood for $\mathcal{H}_p^{ext}$ is maximized by regression of ε_t on (G_{Tt}, Y_t^*, Z_{1Tt}), and the maximized likelihood function becomes

$$-2T^{-1}\log L_{\max}(\mathcal{H}_p^{ext}) = \log\det\langle\varepsilon,\varepsilon|G_T, Y^*, Z_{1T}\rangle_T = \log\det\langle\varepsilon,\varepsilon|G_T, Z_{1T}\rangle_T + o_P(1)$$

because $\langle\varepsilon, Y^*\rangle_T = O_P(T^{-1/2})$. The hypothesis $\mathcal{H}$ is just $\zeta = 0$, and we find similarly

$$-2T^{-1}\log L_{\max}(\mathcal{H}_p^{ext} \cap \mathcal{H}) = \log\det\langle\varepsilon,\varepsilon|Y^*, Z_{1T}\rangle_T = \log\det\langle\varepsilon,\varepsilon|Z_{1T}\rangle_T + o_P(1).$$

It follows from

$$\langle\varepsilon,\varepsilon|G_T, Z_{1T}\rangle_T = \langle\varepsilon,\varepsilon|Z_{1T}\rangle_T - \langle\varepsilon, G_T|Z_{1T}\rangle_T \langle G_T, G_T|Z_{1T}\rangle_T^{-1} \langle G_T, \varepsilon|Z_{1T}\rangle_T$$

and $\langle \varepsilon, \varepsilon | Z_{1T} \rangle_T \xrightarrow{P} \Omega_0$ that

$$-2 \log LR(\mathcal{H}|\mathcal{H}_p^{ext}) = \mathrm{tr}\{\Omega_0^{-1} T^{1/2} \langle \varepsilon, G_T | Z_{1T} \rangle_T \langle G_T, G_T | Z_{1T} \rangle_T^{-1} T^{1/2} \langle G_T, \varepsilon | Z_{1T} \rangle_T\} + o_P(1)$$
$$\xrightarrow{\mathsf{D}} \mathrm{tr}\{\Omega_0^{-1} \langle \varepsilon, G | Z_1 \rangle \langle G, G | Z_1 \rangle^{-1} \langle G, \varepsilon | Z_1 \rangle\}. \tag{58}$$

Analysis of $LR(\beta^* = \beta_0^* | \mathcal{H}_r^{ext})$: The hypothesis we want to test here involves only β and γ^{0*}, and because inference on (α^*, Ω) is asymptotically independent of inference on (β, γ^{0*}), we can assume that $\alpha^* = \alpha_0^* = 0$ and $\Omega = \Omega_0$ for the asymptotic analysis of this statistic. We now find for $\beta = \beta_0 + \beta_{\perp 0} \bar{\beta}_{\perp 0}'(\beta - \beta_0)$ that

$$-\Pi(L)|_{\alpha^* = \alpha_0^*}(\gamma - \gamma_0) Z_t = \alpha_0 \beta'(\gamma^{0*} - \gamma_0^{0*}) N_{T0}^{-1} Z_{0Tt} - (\Upsilon^1 - \Upsilon_0^1) N_{T1}^{-1} Z_{1Tt},$$
$$(\Pi(L) - \Pi_0(L))|_{\alpha^* = \alpha_0^*} Y_t = -\alpha_0 (\beta - \beta_0)' \bar{\beta}_{\perp 0} \beta_{0\perp}' Y_{t-1}.$$

We define G_{Tt} as above, see (57), and define

$$\zeta' = (T^{1/2} \bar{\beta}_{\perp 0}'(\beta - \beta_0), \beta'(\gamma^{0*} - \gamma_0^{0*}) N_{T0}^{-1}),$$

and note that the hypothesis $\beta = \beta_0$ and $\beta' \gamma^{0*} = \beta_0' \gamma_0^{0*}$ is again $\zeta = 0$. We then find

$$\mathcal{H}_r^{ext} : \varepsilon_t(\xi) = -\alpha_0 \zeta' G_{Tt} - (\Upsilon^1 - \Upsilon_0^1) N_{T1}^{-1} Z_{1Tt} + \varepsilon_t.$$

We split the residuals by multiplying by $\alpha_{\Omega_0}' = (\alpha_0' \Omega_0^{-1} \alpha_0)' \alpha_0' \Omega_0^{-1}$ and $\alpha_{0\perp}'$ into

$$\alpha_{\Omega_0}' \varepsilon_t(\xi) = -\zeta' G_{Tt} - \alpha_{\Omega_0}'(\Upsilon^1 - \Upsilon_0^1) N_{T1}^{-1} Z_{1Tt} + \alpha_{\Omega_0}' \varepsilon_t,$$
$$\alpha_{0\perp}' \varepsilon_t(\xi) = -\alpha_{0\perp}'(\Upsilon^1 - \Upsilon_0^1) N_{T1}^{-1} Z_{1Tt} + \alpha_{0\perp}' \varepsilon_t.$$

The errors $\alpha_{\Omega_0}' \varepsilon_t$ and $\alpha_{0\perp}' \varepsilon_t$ are independent and both sets of residuals are analyzed by regression, and we find

$$-2T^{-1} \log L_{\max}(\mathcal{H}_r^{ext}) = \log\det \left\langle \alpha_{\Omega_0}' \varepsilon, \alpha_{\Omega_0}' \varepsilon | G_T, Z_{1T} \right\rangle_T + \log\det \langle \alpha_{0\perp}' \varepsilon, \alpha_{0\perp}' \varepsilon | Z_{1T} \rangle_T$$

and

$$-2T^{-1} \log L_{\max}(\mathcal{H}_r^{ext} \cap \{\zeta = 0\}) = \log\det \left\langle \alpha_{\Omega_0}' \varepsilon, \alpha_{\Omega_0}' \varepsilon | Z_{1T} \right\rangle_T + \log\det \langle \alpha_{0\perp}' \varepsilon, \alpha_{0\perp}' \varepsilon | Z_{1T} \rangle_T .$$

The test of $\beta^* = \beta_0^*$ in $\mathcal{H}_r^{ext}$, using $\left\langle \alpha_{\Omega_0}' \varepsilon, \alpha_{\Omega_0}' \varepsilon \right\rangle_T^{-1} \xrightarrow{P} \alpha_{\Omega_0}' \Omega_0 \alpha_{\Omega_0} = (\alpha_0' \Omega_0^{-1} \alpha_0)^{-1}$, is

$$-2 \log LR(\beta^* = \beta_0^* | \mathcal{H}_r^{ext})$$
$$= \mathrm{tr}\{\alpha_0' \Omega_0^{-1} \alpha_0 T^{1/2} \left\langle \alpha_{\Omega_0}' \varepsilon, G_T | Z_{1T} \right\rangle_T \langle G_T, G_T | Z_{1T} \rangle_T^{-1} T^{1/2} \left\langle G_T, \alpha_{\Omega_0}' \varepsilon | Z_{1T} \right\rangle_T\} + o_P(1)$$
$$\xrightarrow{\mathsf{D}} \mathrm{tr}\{\Omega_0^{-1} \alpha_0 (\alpha_0' \Omega_0^{-1} \alpha_0)^{-1} \alpha_0' \Omega_0^{-1} \langle \varepsilon, G | Z_1 \rangle \langle G, G | Z_1 \rangle^{-1} \langle G, \varepsilon | Z_1 \rangle\}. \tag{59}$$

Analysis of $LR(\mathcal{H}_r^{ext} | \mathcal{H}_p^{ext})$: By (56), the test for rank r converges to the difference between (58) and (59), i.e.,

$$-2 \log LR(H_r^{ext} | \mathcal{H}_p^{ext}) \xrightarrow{\mathsf{D}} \mathrm{tr}\{(\Omega_0^{-1} - \Omega_0^{-1} \alpha_0 (\alpha_0' \Omega_0^{-1} \alpha_0)^{-1} \alpha_0' \Omega_0^{-1}) \langle \varepsilon, G | Z_1 \rangle \langle G, G | Z_1 \rangle^{-1} \langle G, \varepsilon | Z_1 \rangle\}.$$

Using the identity $\Omega_0^{-1} - \Omega_0^{-1} \alpha_0 (\alpha_0' \Omega_0^{-1} \alpha_0)^{-1} \alpha_0' \Omega_0^{-1} = \alpha_{0\perp} (\alpha_{0\perp}' \Omega_0 \alpha_{0\perp})^{-1} \alpha_{0\perp}'$ and defining $\varepsilon_{\alpha_\perp, t} = (\alpha_{0\perp}' \Omega_0 \alpha_{0\perp})^{-1/2} \alpha_{0\perp}' \varepsilon_t$, we obtain the result.

A.9 Proof of Theorem 8

The test that $\beta'\gamma_v^0 = 0$: We find from Theorem 5 and defining $\theta = T^{1/2}\bar{\beta}'_{0\perp}(\beta - \beta_0)$, that

$$\zeta'_{Tv} = T^{1/2}(\breve{\theta}', -N_{T0}^{-1}\breve{\beta}'(\breve{\gamma}_v^0 - \gamma_{v0}^0)) \xrightarrow{\mathsf{D}} - \langle \varepsilon_\alpha, G, |Z_1\rangle \langle G, G|Z_1\rangle^{-1} A_{1v} = \zeta_v, \tag{60}$$

say, for a suitable selection matrix A_{1v}. However, we cannot use (60) directly to test the hypothesis that $\beta'\gamma_v^0 = 0$, because the result is not given in terms of $\breve{\beta}'\breve{\gamma}_v^0 - \beta_0'\gamma_{v0}^0$. Instead, we use the expansion of $T^{1/2}(\breve{\beta}'\breve{\gamma}_v^0 - \beta_0'\gamma_{v0}^0)M_{T0v}^{-1}$ as

$$-T^{1/2}(\breve{\theta}'\beta'_{0\perp}\gamma_{v0}^0 T^{-1/2}M_{T0v}^{-1} - \breve{\beta}'(\breve{\gamma}_v^0 - \gamma_{v0}^0)M_{T0v}^{-1}) = \zeta'_{Tv}\begin{pmatrix} -\beta'_{0\perp}\gamma_{v0}^0 T^{-1/2}M_{T0v}^{-1} \\ 1 \end{pmatrix}, \tag{61}$$

and thus the asymptotic distribution depends on the relative orders of magnitude of the two terms on the right-hand side; that is, on the relation between the normalizations $T^{1/2}$ and M_{T0v}, and on the value of the parameter $\beta'_{0\perp}\gamma_{v0}^0$. Haldrup (1996) encounters a similar problem of different limit behaviour of estimators in the context of a Dickey-Fuller regression with a slope coefficient. Here we have to consider two cases.

If $\beta'_{0\perp}\gamma_{v0}^0 = 0$ or $T^{-1/2}M_{T0v}^{-1} \to c < \infty$, then

$$T^{1/2}(\breve{\beta}'\breve{\gamma}_v^0 - \beta_0'\gamma_{v0}^0)M_{T0v}^{-1} = \zeta'_{Tv}\begin{pmatrix} -\beta'_{0\perp}\gamma_{v0}^0 T^{-1/2}M_{T0v}^{-1} \\ 1 \end{pmatrix} \xrightarrow{\mathsf{D}} \zeta'_v \begin{pmatrix} -\beta'_{0\perp}\gamma_{v0}^0 c \\ 1 \end{pmatrix}.$$

On the other hand, if $\beta'_{0\perp}\gamma_{v0}^0 \neq 0$ and $T^{-1/2}M_{T0v}^{-1} \to \infty$, the estimator has to be renormalized and we find

$$T(\breve{\beta}'\breve{\gamma}_v^0 - \beta_0'\gamma_{v0}^0) = \zeta'_{Tv}\begin{pmatrix} -\beta'_{0\perp}\gamma_{v0}^0 \\ T^{-1/2}M_{T0v}^{-1} \end{pmatrix} \xrightarrow{\mathsf{D}} \zeta'_v \begin{pmatrix} -\beta'_{0\perp}\gamma_{v0}^0 \\ 0 \end{pmatrix}.$$

Thus, in any case the asymptotic distribution of the suitably normalized estimator is mixed Gaussian because ζ'_v is mixed Gaussian, see Theorem 5, and hence inference can be conducted using likelihood ratio tests and the χ^2-distribution.

The test that $\alpha'_\perp\Gamma\gamma_v^0 = 0$: Similarly to (61) we expand

$$T^{1/2}(\breve{\alpha}'_\perp\breve{\Gamma}\breve{\gamma}_v^0 - \alpha'_{0\perp}\Gamma_0\gamma_{v0}^0)M_{T1v}^{-1} = T^{1/2}\alpha'_{0\perp}\Gamma_0(\breve{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v}^{-1} + T^{1/2}(\breve{\alpha}'_\perp\breve{\Gamma} - \alpha'_{0\perp}\Gamma_0)\breve{\gamma}_v^0 M_{T1v}^{-1} \tag{62}$$

and define

$$\eta'_{Tv} = (T^{1/2}\alpha'_{0\perp}\Gamma_0(\breve{\gamma}_v^0 - \gamma_{v0}^0)M_{T1v}^{-1}, T^{1/2}(\breve{\alpha}'_\perp\breve{\Gamma} - \alpha'_{0\perp}\Gamma_0)) \xrightarrow{\mathsf{D}} \eta'_v,$$

which by (26) and (29) is Gaussian. Then

$$T^{1/2}(\breve{\alpha}'_\perp\breve{\Gamma}\breve{\gamma}_v^0 - \alpha'_{0\perp}\Gamma_0\gamma_{v0}^0)M_{T1v}^{-1} = \eta'_{Tv}\begin{pmatrix} 1 \\ \breve{\gamma}_v^0 M_{T1v}^{-1} \end{pmatrix},$$

and again there are two cases.

If $M_{T1v}^{-1} \to c < \infty$, then the asymptotic distribution of (62) is given by the asymptotic distribution of

$$\eta'_v \begin{pmatrix} 1 \\ \gamma_{v0}^0 c \end{pmatrix},$$

which is Gaussian. On the other hand, if $M_{T1v} \to 0$, we have to renormalize and find

$$T^{1/2}(\breve{\alpha}'_\perp\breve{\Gamma}\breve{\gamma}_v^0 - \alpha'_{0\perp}\Gamma_0\gamma_{v0}^0) = \eta'_{Tv}\begin{pmatrix} M_{T1v} \\ \breve{\gamma}_v^0 \end{pmatrix} \xrightarrow{\mathsf{D}} \eta'_v \begin{pmatrix} 0 \\ \gamma_{v0}^0 \end{pmatrix},$$

which is also Gaussian, so that asymptotic inference can again be conducted using likelihood ratio tests and the χ^2-distribution.

Acknowledgements

We are grateful to seminar participants at various universities and conferences for comments, and to Bent Nielsen for help. We would like to thank the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES - DNRF78, funded by the Danish National Research Foundation) for research support.

References

1. Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* **22**, pp. 327–351.
2. Belke, A., and Beckmann, J. (2015). Monetary policy and stock prices – cross-country evidence from cointegrated VAR models. *Journal of Banking and Finance* **54**, pp. 254–265.
3. Cavaliere, G., Nielsen, H.B., and Rahbek, A. (2016). Bootstrap testing of hypotheses on cointegration relations in VAR models. *Econometrica* **83**, pp. 813–831.
4. Cavaliere, G., Rahbek, A., and Taylor, A.M.R. (2012). Bootstrap determination of the cointegration rank in VAR models. *Econometrica* **80**, pp. 1721–1740.
5. Doornik, J.A., Hendry, D.F., and Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys* **12**, pp. 533–572.
6. Haldrup, N. (1996). Mirror image distributions and the Dickey-Fuller regression with a maintained trend. *Journal of Econometrics* **72**, pp. 301–312.
7. Hendry, D.F., and Juselius, K. (2001). Explaining cointegration analysis: part II. *Energy Journal* **22**, pp. 75–120.
8. Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (2.ed.), Oxford University Press, Oxford.
9. Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica* **70**, pp. 1929–1961.
10. Johansen, S. (2006). Statistical analysis of hypotheses on the cointegrating relations in the $I(2)$ model. *Journal of Econometrics* **132**, pp. 81–115.
11. Johansen, S., Mosconi, R., and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* **3**, pp. 216–249.
12. Johansen, S. and Nielsen, M.Ø. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica* **80**, pp. 2667–2732.
13. Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*, Oxford University Press, Oxford.
14. Juselius, K. (2009). The long swings puzzle: what the data tell when allowed to speak freely. In T. Mills and K. Patterson (eds.) *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, Palgrave-MacMillan, New York, pp. 349–384.
15. Lütkepohl, H., and Saikkonen, P. (2000a). Testing for the cointegrating rank of a VAR process with a time trend. *Journal of Econometrics* **95**, pp. 177–198.
16. Lütkepohl, H., and Saikkonen, P. (2000b). Testing for the cointegrating rank of a VAR process with an intercept. *Econometric Theory* **16**, pp. 373–406.
17. Lütkepohl, H., and Saikkonen, P. (2000c). Testing for the cointegrating rank of a

VAR process with structural shifts. *Journal of Business & Economic Statistics* **18**, pp. 451–464.

18. Miller, K.S. (1968). *Linear Difference Equations*, Benjamin, New York.
19. Nielsen, H.B. (2004). Cointegration analysis in the presence of outliers. *Econometrics Journal* **7**, pp. 249–271.
20. Nielsen, H.B. (2007). UK money demand 1873-2011: a long time series analysis and event study. *Cliometrics* **1**, pp. 45–61.
21. Nielsen, B., and Rahbek, A. (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics* **62**, pp. 5–22.
22. Pretis, F. (2015). Econometric models of climate systems: the equivalence of two-component energy balance models and cointegrated VARs. Department of Economics Discussion Paper Series 750, University of Oxford.
23. Pretis, F., Schneider, L., Smerdon, J.E., and Hendry, D.F. (2016). Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. *Journal of Economic Surveys* **30**, pp. 403–429.
24. Rahbek, A., and Mosconi, R. (1999). Cointegration rank inference with stationary regressors in VAR models. *Econometrics Journal* **2**, pp. 76–91.
25. Trenkler, C., Saikkonen, P., and Lütkepohl, H. (2007). Testing for the cointegrating rank of a VAR process with a level shift and a trend break. *Journal of Time Series Analysis* **29**, pp. 331–358.

Research Papers 2016

CREATES
Center for Research in Econometric Analysis of Time Series

2016-05:	Yunus Emre Ergemen: Generalized Efficient Inference on Factor Models with Long-Range Dependence
2016-06:	Girum D. Abate and Luc Anselin: House price fluctuations and the business cycle dynamics
2016-07:	Gustavo Fruet Dias, Cristina M. Scherrer and Fotis Papailias: Volatility Discovery
2016-08	N. Haldrup, O. Knapik and T. Proietti: A generalized exponential time series regression model for electricity prices
2016-09:	Ole E. Barndorff-Nielsen: Assessing Gamma kernels and BSS/LSS processes
2016-10:	Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Modeling and Forecasting (Un)Reliable Realized Covariances for More Reliable Financial Decisions
2016-11:	Tom Engsted and Thomas Q. Pedersen: The predictive power of dividend yields for future inflation: Money illusion or rational causes?
2016-12:	Federico A. Bugni, Mehmet Caner, Anders Bredahl Kock and Soumendra Lahiri: Inference in partially identified models with many moment inequalities using Lasso
2016-13:	Mikko S. Pakkanen and Jani Lukkarinen: Arbitrage without borrowing or short selling?
2016-14:	Andrew J.G. Cairns, Malene Kallestrup-Lamb, Carsten P.T. Rosenskjold, David Blake and Kevin Dowd: Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index
2016-15:	Mikkel Bennedsen, Ulrich Hounyo, Asger Lunde and Mikko S. Pakkanen: The Local Fractional Bootstrap
2016-16:	Martin M. Andreasen and Kasper Jørgensen: Explaining Asset Prices with Low Risk Aversion and Low Intertemporal Substitution
2016-17:	Robinson Kruse, Christian Leschinski and Michael Will: Comparing Predictive Accuracy under Long Memory - With an Application to Volatility Forecasting
2016-18:	Søren Johansen and Bent Nielsen: Tightness of M-estimators for multiple linear regression in time series
2016-19:	Tim Bollerslev, Jia Li and Yuan Xue: Volume, Volatility and Public News Announcements
2016-20:	Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: Retrieving Risk-Neutral Densities Embedded in VIX Options: a Non-Structural Approach
2016-21:	Mikkel Bennedsen: The Local Fractional Bootstrap
2016-22:	Søren Johansen and Morten Ørregaard Nielsen: The cointegrated vector autoregressive model with general deterministic terms